CONSTRUCTION AND SCREENING OF

A CHICKEN GENOMIC DNA LIBRARY

FOR MYOSIN HEAVY CHAIN GENE SEQUENCES

by

Ann Curtis B.Sc.

Thesis submitted for the
Degree of Doctor of Philosophy

Institute of Animal Genetics
University of Edinburgh
Kings Buildings
West Mains Road
Edinburgh

July 1983

I declare that the work presented in this Thesis is the result of my own independent investigation unless otherwise stated.

The work has not been, and is not being concurrently submitted in candidature for any other degree.

———————————————————

To My Parents and My Husband.

CONTENTS

ACKNOWLEDGEMENTS

ABBREVIATIONS

aa	Amino acid
ALD	Anterior Latissimus Dorsi
amp^r	Ampicillin resistance
AMV	Avian myeloblastosis virus
ATP	Adenosine 5' triphosphate
ATPase	Adenosine 5' triphosphatase
bis	N,N'-Methylene-bis-acrylamide
bp	Base pair
BSA	Bovine serum albumin
CTP	Cytidine 5' triphosphate
cm	Centimetre
°C	Degrees centigrade
cpm	Counts per minute
CsCl	Caesium chloride
DAP	Diaminopimelic acid
DEP	Diethylpyrocarbonate
DMSO	Dimethylsulphoxide
DNA	Deoxyribonucleic acid
cDNA	Complementary deoxyribonucleic acid
DNase	Deoxyribonuclease
dpm	Disintegrations per minute
dATP	Deoxyadenosine 5'-triphosphate
dCTP	Deoxycytidine 5'-triphosphate
dGTP	Deoxyguanosine 5'-triphosphate
dTTP	Deoxythymidine 5'-triphosphate
EDTA	Ethylenediaminetetracetic acid
EGTA	Ethylene glycol-bis(2-aminoethylether)N,N'-tetraacetic acid
g	Gravitation force
GTP	Guanosine 5'-triphosphate
Hepes	N-2-Hydroxyethylpiperazine-N'-2-ethanesulphonic acid
hnRNA	Heterogeneous nuclear RNA

kb	Kilobase
l	Litre
LC_1	Light chain 1
LC_2	Light chain 2
LC_3	Light chain 3
M	Molar
mA	milliamp
$MgSO_4$	Magnesium sulphate
MHC	Myosin heavy chain
min	Minute
ml	Millilitre
mM	millimolar
mm	millimetre
mRNA	Messenger ribonucleic acid
mRNP	Messenger ribonucleoprotein
nt	Nucleotide
OD	Optical density
pfu	Plaque forming unit
PLD	Posterior Latissimus Dorsi
POMC	Proopiomelanocortin
POPOP	1,4-di[2-(5-phenyloxazolyl)]benzene
PPO	2,5 diphenyloxazole
RNA	Ribonucleic acid
RNase	Ribonuclease
rRNA	Ribosomal ribonucleic acid
rpm	Revolutions per minute
secs	Seconds
SDS	Sodium dodecyl sulphate
SSC	Standard saline citrate
sq. in.	Square inch
TCA	Trichloroacetic acid
tcRNA	Translational control RNA
TEMED	N,N,N^1,N^1-Tetramethyl ethylenediamine
Tris	Tris (hydroxymethyl) aminomethane

w/v	Weight per volume
μl	Microlitre
μg	Microgram
λ^+	Wild type bacteriophage lambda

LIST OF FIGURES AND TABLES

ABSTRACT OF THESIS

The expression of muscle function genes during myogenesis is a highly controversial subject, and, therefore, it was recognised that highly specific nucleic acid probes for genes which are specifically expressed in muscle tissue must be developed to further research. Myosin heavy chain (MHC) makes up 30% of the total protein in differentiated muscle and, therefore, it was decided to attempt to isolate the MHC gene as a genomic sequence rather than as a cDNA, with a view to obtaining flanking and intervening sequences also. It was intended that such a gene would provide a highly specific probe to be used in further studies on the control of muscle gene expression during myogenesis and during the further development of different types of muscle.

A chicken genomic library was constructed by cloning 20 kilo base lengths of randomly cleaved chicken liver DNA in the bacteriophage λ vector, charon 4A. The vector was suitably manipulated to accept the DNA fragments and the recombinant molecules were packaged into viable phage particles *in vitro*. This chicken gene library was screened with a MHC probe which was also prepared as part of the project from 14-day chick embryo skeletal leg muscle heavy polysomes.

A fraction of the heavy polysomes, collected from the bottom of a sucrose gradient was shown by *in vitro* translation to predominantly direct the synthesis of a single high molecular weight (200,000 daltons) polypeptide which co-electrophoresed with MHC marker. This was confirmed by *in vitro* translation of the messenger RNA extracted from this polysome fraction in two heterologous systems. The mRNA was reverse transcribed and the ^{32}P-labelled cDNA was employed as a specific probe for screening the library by *in situ* plaque hybridisation. The cDNA was shown to represent a 10% copy of MHC mRNA (500 nucleotides), to

consist of at least two MHC sequences present in equal proportions, and to be contaminated with ribosomal RNA sequences.

A proportion of the library was screened with a nick-translated ribosomal DNA probe and it was estimated that 0.19% of the sequences in the library represented ribosomal RNA suggesting that the sequence representation of the library was complete and that the chances of detecting a much less abundant sequence such as MHC were high. By cross-screening a proportion of the library with both rDNA and MHC cDNA probes it was estimated that the cDNA was contaminated to an extent of 20% with ribosomal sequences.

A total of 3.5×10^5 plaques (7 genomes of chicken DNA) were screened with the MHC cDNA. 52 areas of positive hybridisation were detected indicating that this MHC probe recognises sequences equivalent to 0.0015% of the chicken genome. This suggests several possibilities about MHC gene structure and arrangement: Either the gene sequences represented in this 'embryonic' probe are present at 7-8 copies per genome, or MHC exists as a gene family with whose members the probe is able to cross-hybridise, or the probe may be cross reacting with other types of muscle myosin such as cardiac or smooth muscle.

Ann Curtis
110 Montgomery Street
Edinburgh EH7 5HE

July 1983

CHAPTER I

Introduction

CHAPTER I:

1.1 General Aim of the Project

The study of muscle function genes has been continuing in this laboratory for several years (John, 1974; Patrinou-Georgoulas and John, 1977; John, Patrinou-Georgoulas and Jones, 1977; Patrinou-Georgoulas and John, 1980; John 1980; John and Lawson, 1980), and the aim of this project is to further those investigations at the level of the nucleic acids.

The skeletal muscle myosin heavy chain gene was chosen for the study because the large protein of 200,000 molecular weight for which it codes makes up 30% of the total differentiated muscle protein and hence is very important in the overall structure of the muscle. The correspondingly large messenger RNA (mRNA) is convenient for isolation in a relatively pure form because it forms a heavy polyribosomal complex which can be purified using a variety of fractionation techniques.

In order that the myosin heavy chain gene could be studied at the genomic level, rather than as a mRNA or complementary DNA, as previously (Patrinou-Georgoulas and John, 1977), the necessity was recognised to isolate the DNA corresponding to the myosin heavy chain gene from actual tissue. For technical reasons, the embryonic chick was chosen as the source of material, and after consideration of various methods of obtaining complete copies of the myosin heavy chain gene, it was decided to construct a chicken gene library using techniques developed over the past four years. From this, the aim was to isolate the required gene with its flanking sequences and possible intervening sequences using purified probes which would also be prepared as part of the project.

1.2 Development of Muscle

The study of muscle development, especially regarding myosin is a particularly controversial subject, and it is a study of the gene at the most fundamental level which will help to resolve many of the existing controversies. These include:

a)   the number of different types of myosin;
b)   the number of genes which control their synthesis;
c)   the timing of gene expression during muscle development.

Muscle gene expression is influenced by many intrinsic and extrinsic factors such as nerve impulses (Sréter and Gergley, 1973; Weeds et al., 1974; Brevet et al., 1976; Salmons and Sréter, 1976), nervous trophic influences (Harris et al., 1971; Cohen and Fischbach, 1973; Walker and Wilson, 1975), hormones (Mandel and Pearson, 1974), collagen (Hauschka and Konigsberg, 1966; Elsdale and Bard, 1972; De La Haba, Kamali and Tiede, 1975; Duance et al., 1977; Mayne et al., 1977; John and Lawson, 1980), creatine (Konigsberg, 1971; Ingwall et al., 1972; Ingwall and Wildenthal, 1974) and cyclic AMP (Wahrman et al., 1973; Zalin and Montague, 1974), but the level of their action – whether for instance they affect transcription of certain genes or translation of their mRNAs – and the actual mechanisms are little understood.

Such fundamental studies may also help in the understanding of certain inherited muscle diseases. Although, to date, such investigations have not revealed defects in the structural proteins, this does not eliminate the possibility that a controlling factor associated with a structural gene could be primarily responsible.

1.2.1 Types of Muscle

The muscle of vertebrates is classified as several types:

a) *Smooth muscle* consists of uninucleate flame-shaped single cells, grouped together into muscle tissue. Innervation is from the autonomic nervous system. Smooth muscle is present in structures such as the lining of the intestine, blood vessels, ducts and walls of many organs.

b) *Cardiac muscle* is the least abundant type of muscle in the body, but is the most active. The myocardial cells are cylindrical in shape, arranged in columns which branch and anastomase. They show cross striations, but unlike skeletal muscle, do not form a syncytium, and are served by the autonomic nervous system.

c) *Skeletal muscle* is innervated by the voluntary nervous system, and its structure is detailed below in Section 1.2.2.

1.2.2 Structure of Skeletal Muscle (Wilkie, 1968; Bendall, 1969)

Mature skeletal muscle consists of groups of muscle fibres many centimetres long. Regular striations extend across the fibre and divide it into sarcomers. Within the fibres lie the actual contractile structures – the myofibrils which consist of rods of contractile protein running from one end of the fibril to the other. The fibril is divided by thin partitions called Z lines which run from fibril to fibril across the fibre and so form the basis of the sarcomers as shown in Figure 1.1. Each myofibril is made up of protein filaments. The thick filaments are myosin, containing two moles of heavy chains and 4 moles of light chains, which have projections to either side. The thin filaments are actin, made up of two strings of globular proteins wound together, on which are arranged the other

major muscle proteins, tropomyosin and three components of troponin (see Figure 1.1). The filaments interdigitate with each other, forming cross bridges which develop the force of contraction.

1.2.3 *General Structure of Myosin* (Wilkie, 1968; Bendall, 1969)

Myosin is the focal point of the contraction process, having both the binding sites for complexing with actin and the enzymatic sites which allow the hydrolysis of ATP - each being essential for muscle contraction to occur, as shown diagramatically in Figure 1.1.

The molecule consists of double α-helical rod with two globular heads at one end, which form the enzymatic site and the actin binding site. Brief trypsin digestion yields two fragments, having cleaved the tail approximately two-thirds the way along the length. Heavy meromyosin consists of the globular heads attached to a shortened tail, and light meromyosin consists of the remainder of the tail. The polypeptide chains in this structure consist of two similar heavy chains of which there are two or more types (Starr and Offer, 1973), each of molecular weight 200,000 daltons which make up the double helical rod and head structure. Fitted around the head are the light chains of which there are 5 types of molecular weights, 27,000, 25,000, 21,000, 18,000 and 16,500 daltons, and which are present in a combination which reflects the muscle type (Dow and Stracher, 1971). Two alkali, or non-phosphorylatable light chains (LC_1) and two regulatory, or phosphorylatable light chains (LC_2) are present per myosin molecule, and in fast skeletal muscle an additional alkali chain, LC_3, is present.

Figure 1. 1 →

FIGURE 1.1:   The Structure of Skeletal Muscle.

(a)   *The muscle fibre:*

The muscle fibre is a cylindrical structure which may be many centimetres long.   It is divided by regular striations into sarcomeres.   Within the fibre are many cylindrical sub-units, the fibrils which are the basic contractile elements of the muscle.

(b)   *The sarcomere:*

The individual sarcomeric units consist of rods of contractile protein which are fine filaments arranged longitudinally.   The thin filaments consist of an α-helix of globular actin molecules, and the thick filaments are myosin.   The myosin molecules have projections which extend laterally and make contact with the actin sub-units.   It is the sliding of the thick and thin filaments which form the molecular basis of muscle contraction, the actual length of the filaments remaining unchanged.

(c)   *Transverse section through a sarcomere*

This is a diagrammatic representation of the mechanism of ATPase activity, which is a particular feature of the actin – myosin interaction in muscle contraction, and depends on two other muscle specific proteins, troponin and tropomyosin. Calcium ions regulate contraction by releasing the inhibition of the actin – myosin interaction through a conformational change in the troponin-tropomyosin complex as shown.

Z-line▷

A-band⌐

H-zone

I-band⌐

◁actin

◁myosin

Sarcomere

a

b

actin

▷   ◁

ATP

troponin–
tropomyosin
complex

ATP

myosin

actin

▷   ◁

Ca²⁺

myosin

c
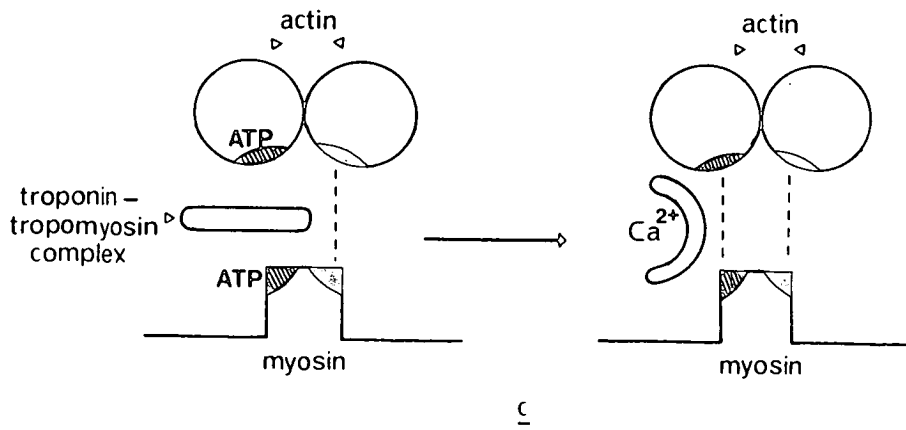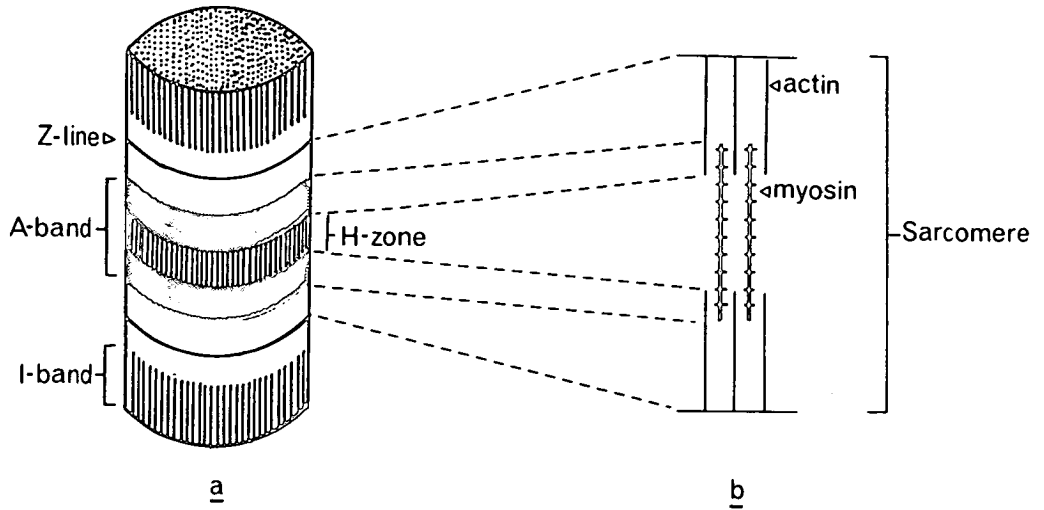
1.2.4 *Types of Myosin*

It has been shown by many workers that different isoforms of myosin exist, each being specific to a particular tissue type. The myosins of smooth muscle, cardiac muscle and skeletal muscle are antigenically different and these are different again to myosin found in non-myogenic tissue such as brain (Berl, Puszkin and Nicklas, 1973), fibroblasts (Adelstein *et al.*, 1972) and platelets (Bettex-Galland and Luescher, 1965; Adelstein, Pollard and Knehl, 1971). The non-myogenic myosin contributes to intracellular movement, mitosis and cellular adhesiveness. It only accounts for 0.5 - 3.0% of the total cell protein compared with 30% for muscle myosin. The two types of myosin differ in their solubility properties (myogenic myosin precipitates at high ionic strength whereas fibroblast myosin precipitates at low ionic strength (Ostlund and Pastan, 1974)), but otherwise many properties are similar (Adelstein *et al.*, 1972; Ostlund and Pastan, 1974). For example, ATPase activity is stimulated by EDTA in the presence of calcium and potassium and suppressed by magnesium in 0.6M potassium chloride, both bind to actin, have a heavy chain molecular weight of 200,000 daltons and there is evidence to suggest that non-myogenic myosin has two light chains of 20,000 and 16,000 daltons (Chi *et al.*, 1975).

Adult skeletal muscle is again subdivided according to muscle type which arises as a result of innervation by different sets of motor-neurons. Fast twitch muscle is innervated by a single motorneurone and is fast contracting; and slow tonic muscle is multiply innervated and slow contracting. Fast muscle myosin has three light chains of 25,000, 18,000 and 16,500 daltons associated with two heavy chains per molecule, and slow muscle myosin has two light chains of 27,700 and 21,000 daltons, with

two heavy chains, per molecule. Slow myosin also tends to have a lower ATPase activity than fast myosin. That innervation is important in the development of these two muscle types was demonstrated by Gordon *et al*. (1977) who replaced a slow muscle, the anterior latissimus dorsi (ALD), of the chicken with a minced fast muscle, the posterior latissimus dorsi (PLD), which, therefore, received the ALD nerve supply. The minced muscle regenerated as ALD, and in a similar experiment where minced ALD replaced PLD, the minced muscle regenerated as PLD, and it was shown that the relative myosins also changed as determined by their light chain composition. This was only achieved with regenerating muscle, though Weeds *et al*. (1974) by cross re-innervation of the slow and fast twitch muscle of the cat, and Sréter *et al*. (1975) performing cross re-innervation experiments in the rabbit, showed that the adult muscle phenotype could be changed to the opposite type and hence concluded that myosin phenotype is controlled through the nervous system.

That the adult forms of fast and slow muscle express different isoforms of myosin in a tissue specific manner is now well established, but the development of either form is highly controversial. Some investigators claim that all embryonic muscle fibres, whether destined to become fast or slow in the adult, initially synthesise both fast and slow myosins (Sréter *et al* ., 1972) and only after proper innervation has become established do they select the continued synthesis of one type and repress that of the other. Others propose that embryonic muscle consists of fast myosin only - slow myosin not being synthesised until proper innervation is established (Rubinstein and Kelly, 1978) and yet others claim that there is a distinct myosin unique to the embryo which is gradually replaced by one or other adult form (Sréter *et al*.,

1975; Whalen *et al.*, 1979; Whalen *et al.*, 1981). A most convincing demonstration of the expression of a distinct embryonic MHC gene has recently been provided by Jakowlew *et al.* (1982) who showed that the levels of a distinct MHC mRNA present in embryonic chick heart decrease during development. Several immunohistochemical studies (Hoh *et al.*, 1978; Whalen and Sell, 1980; Lompre *et al.*, 1981) on rat cardiac muscle have demonstrated three isoforms of cardiac MHC named V_1, V_2 and V_3 on the basis of their increasing electrophoretic mobility. These studies have shown that V_1 appears at birth and persists throughout life whereas V_2 and V_3 are foetal forms which also re-appear in the second month of post-natal life. The distribution of the three types throughout development suggests that different MHC genes are expressed at specific developmental stages and under different physiological and pathological conditions.

Using a fluorescent antibody staining technique with antibodies prepared specifically against slow, fast and cardiac myosin heavy chain, Masaki and Kinoshita (1974) and Masaki and Yoshizaki (1974) showed that all three antibodies stained the same myotubes and myofibrils in embryonic chick skeletal muscle at various stages of development from 4 days onwards. By the same technique, they also showed that the ability of embryonic fast skeletal muscle to synthesise cardiac and slow muscle myosin gradually decreases until they are almost undetectable in the pectoralis muscles of 17-day embryos. Similar results were obtained with early developing rat muscle, using an immunocytochemical approach (Gauthier *et al.*, 1978), when fast and slow isoenzymes of myosin were shown to co-exist in the same fibre. In this same study, the developing rat diaphragm, destined to become a fast muscle, retained ATPase activity after both acid and alkali treatment, but during post-natal development, only the alkali

stable form persisted. This may be an indication of the presence of both fast and slow myosin in the embryonic muscle, since, adult rat myosin ATPase is acid stable in slow fibres and alkali stable in fast fibres, unless there is an embryonic form which retains its activity in both acid and alkali conditions. In addition, it was shown by Syrový (1977) that the slow, semi-tendonous muscle of rabbits contains fast and slow muscle light chains in the newborn animal but only slow forms in the adult. The newborn also had a higher ATPase activity. Such a pattern of change in myosin isoenzymes is consistent with physiological changes which occur in the skeletal muscles of small mammals. Up until 7 days after birth contraction is slow, and until then, slow and fast myosin have been shown to exist together. By 21 days, however, when the speed of contraction in fast muscles has increased, the level of staining for slow myosin has decreased. In addition to these examples, a further 'superfast' molecule has been identified (MHC_{SF}) as the heavy chain in the myosin of the rapid maxillary muscle of the cat (Rowlerson *et al.*, 1981).

However, it must be noted that many studies have failed to identify slow myosin in embryonic muscle. Using sodium dodecylsulphate poly-acrylamide gel electrophoretic techniques to identify the various light chains, and antibody techniques to differentiate between fast and slow heavy chains, Rubinstein *et al.* (1977) claim that pectoralis (fast) and ALD (slow) muscles of the embryonic chick only contain fast myosin heavy chain and the 25,000 and 18,000 dalton light chains, and only after the establishment of innervation does the ALD begin to synthesise slow heavy chain and by 20 days of development the light chains which are characteristic of slow muscle become predominant. These results confirmed the early work of Sréter *et al.* (1972), who showed that the

ATPase activities of various embryonic myosins, whether destined to become fast or slow in the adult were very close to those of fast adult muscle. Rubinstein and Kelly (1978), working on the rabbit, also postulated a developmental progression from a fast type muscle to the adult form. The light chains of ELD (fast) isolated at various stages of development always migrated with the adult form. The light chains of the soleus (slow), however, also migrated with the adult fast light chains on polyacrylamide gels, up until 20 days of foetal development. By 2 days post partum, the soleus showed 5 light chains with mobilities identical to those of fast and slow, and by 10 days post partum, slow light chains predominated. The adult soleus shows no sign of the fast light chains.

The question of gene expression of the fast and slow isoforms of myosin heavy chain in developing muscle is complicated by the fact that some developing muscle contains a mixture of fast and slow fibres, and the study of individual fibres is difficult. In addition, potential cross-reactivity between the various antibodies to adult and embryonic fast and slow isoforms can present problems as shown by Whalen *et al*. (1981) and Gauthier *et al*. (1982). Billeter *et al*. (1981) have analysed individual fibres biochemically and shown co-expression of slow and fast light chains though it is unknown whether the genes are activated in the same nucleus. For this reason, further studies on cloned myoblasts were carried out (Keller and Emerson, 1980; Stockdale *et al*., 1981) and were found to express both slow and fast myosin light chain genes irrespective of the original source of the clone, though expression of the myosin HC under such conditions is unknown at present.

Perry and his co-workers (Trayer and Perry, 1966; Holland and Perry, 1969) and others (Obinata, 1969) stressed the differences between

'embryonic' and 'mature' muscle myosins to the extent that they suggested that the embryo has its own form of myosin with a lower ATPase activity which increases gradually during development as the myosin becomes replaced by the adult form. Evidence for this hypothesis has come from work on both the heavy and light chains. Perry claims that the heavy chain of the embryonic myosin is different to that of adult fast myosin, since, in recombination experiments, only the two largest adult light chains can combine with the embryonic heavy chain - that is, it appears to have no affinity for the smallest adult light chain, which has been shown to be absent, or present in very small amounts in embryonic tissue (Dow and Stracher, 1971; John, 1974; Sréter *et al.*, 1975; Dabrowska *et al.*, 1977). More recently (Whalen *et al.*, 1979), partial proteolytic cleavage and immunological analysis followed by two dimensional electrophoresis have demonstrated three types of skeletal muscle myosin heavy chain - a fast, a slow and an embryonic type. The fast and slow forms were identified in diaphragm and soleus muscles respectively, and the embryonic form was unique to embryonic muscle and some types of myogenic cell culture. The methods used are claimed to expose antigenic sites which have previously not been involved in immunological studies, and have resulted in an estimate of at least nine heavy chain genes.

Evidence for an embryonic light chain (LC_{1E}) similar to the largest light chain of fast muscle (LC_{1F}) was obtained (Whalen *et al.*, 1978) after two-dimensional electrophoresis of myosin from cultured muscle and cell lines. It was also identified in embryonic and newborn rats, but not the adult, so was not dismissed as a tissue culture phenomenon, and was shown to have similar properties to LC_{1F} by its co-purification properties and its association with myosin subfragment 1 (the globular part of the molecule) from which LC_{2F} is lost. However, this evidence

has been dismissed in another more recent study (Rubinstein and Holtzer, 1979) which has only been able to identify fast myosin heavy chain in muscle, prior to innervation, whether it is destined to become fast or slow. The largest light chain was identified as a doublet as previously reported, but when RNA was extracted from cultures of such muscle and translated *in vitro*, this second band was not apparent and was thus dismissed as an artifact which presented no real evidence for a distinct embryonic form. The amino acid sequence data on some of the myosin heavy chain proteins is very incomplete and thus it is not possible to conclude anything about the number of different forms of myosin heavy chain in this way. However, peptide mapping and immunological data show that the various forms have considerable sequence homology. The recent development of monoclonal antibodies (Masaki *et al*., 1982) against various regions of the MHC molecule should permit more accurate studies on the regions of homology between the various myosin heavy chains.

The fact that fast and slow muscle displays different properties suggests that several structural genes are required to produce a given functional thick filament. Obviously genetic decisions must be taken during myogenesis to specify which heavy and light chains are transcribed and translated in a particular fibre.

Most of the literature discourages the idea that prior to innervation and active contraction, all muscles translate messengers corresponding to the heavy and light chains of a less specialised 'embryonic' type which is later replaced by one or other adult form. All the literature is united in the idea that innervation is important in muscle development, and consequently in the genetic expression of the muscle but it is the precise nature of this influence which is controversial. One suggestion

is that the synthesis of fast myosin is endogenously programmed and the synthesis of slow myosin is dependent on innervation. This was demonstrated by the de-nervation of rabbits at birth (Rubinstein and Kelly, 1978) in which both types of muscle developed rapid contractions very similar to those of adult fast twitch muscle, suggesting that the speed of contraction of slow fibres depends on innervation. Other theories suggest that both types of muscle are synthesised without innervation, but a specific type of nervous input either suppresses the synthesis of a certain type of myosin or maintains the synthesis of the other type beyond a particular stage in development. Certainly, the speed of contraction of embryonic muscle, which is generally slow, does not correlate with the idea that it only contains fast myosin. Slow myosin seems to be most abundant when the speed of contraction is slow and fibres are multiply innervated, and its synthesis seems to be switched off when contraction speed increases and the fibres become innervated by single neurones. Recently, Kelly and Rubinstein (1980) have explained this apparent contradiction by suggesting that skeletal muscles develop from a small population of primary generation cells, and sub-sequently, second generation cells develop along the walls of these primary cells, eventually separating and becoming independent units of contraction. Using anti-slow and anti-fast myosin, Kelly and Rubinstein (1980) demonstrated that the primary generation cells react with anti-slow and the secondary generation cells react with anti-fast myosin, and that it is the influence of these primary cells which cause the slow contractions early in development.

However, the evidence for any of these theories is based on the type of myosin present in embryonic muscle, and if gene expression could be investigated directly at the DNA level, then such controversies may be resolved.

A partial cDNA, prepared from an embryonic chick skeletal muscle myosin heavy chain mRNA has been prepared in this laboratory by Patrinou-Georgoulas and John (1977) and using the technique of *in situ* hybridisation has enabled the detection of myosin mRNA in individual cells (John, Patrinou-Georgoulas and Jones, 1977). However, although the cDNA was a partial copy, it did appear to be specific for embryonic skeletal muscle myosin heavy chain mRNA as it did not cross react with fibroblast myosin heavy chain mRNA. The degree to which it may cross react with other muscle myosins remains to be determined. Any form of cross reactivity will make this probe a powerful tool for the isolation of complete genomic copies of the various myosin heavy chain genes which can then in turn be used as further probes for the detection of different myosin heavy chain messengers in the various tissues.

1.2.5 *Myogenesis*

The development of embryonic skeletal muscle is regarded by many workers as a model system for development studies on differentiating tissues.

Briefly, the muscle cells, or myoblasts, present early in development are single cells which become determined on one course of development. Eventually, cell fusion occurs quite rapidly to form multinuclear, highly specialised syncitial myotubes (Strehler *et al.*, 1963; Vertel and Fischman, 1976). This is the terminal differentiated state which is also marked by specific protein synthesis, including synthesis of the structural proteins, actin and myosin, muscle specific enzymes such as creatine phosphokinase and the appearance of acetylcholine receptors on the cell surface.

The fusion process is asynchronous *in vivo* but conditions have been developed *in vitro* where the transition from the single cell to the

myotube state is very sharp; 70-80% of all cells will fuse in one generation time (O'Neil and Stockdale, 1972) and, in fact, all the later stages of myogenesis can be reproduced in tissue culture (Buckingham, 1977). This led to many studies of cell fusion and various theories about muscle development. Several workers (Okazaki and Holtzer, 1966; Bischoff and Holtzer, 1969; O'Neil and Stockdale, 1972; Wahman et al., 1976; Moss and Strohman, 1976) have investigated cell fusion with respect to the cell cycle, and this prompted Holtzer (Bischoff and Holtzer, 1969) to put forward his theory of a critical mitosis which yields daughter cells with a pattern of synthetic activities which are very different from those of the parent cell. Fusion, it is agreed, is limited to cells in the G1 period of the cell cycle, and it has been reported (Okazaki and Holtzer, 1966; Bischoff and Holtzer, 1969) that nuclei radioactively labelled with tritiated thymidine at the end of the previous S period do not appear in myotubes sooner than 8 hours later which brings them into the final third of the G1 period. Under conditions which normally shorten the G1 period, it was found that this '8-hour rule' still applied, but the cells did not move on through the cell cycle. It was therefore concluded that cells withdraw from the cell cycle prior to fusion or enter a post-mitotic period, the previous mitosis being qualitatively different from that of other cell cycles.

In fresh myogenic cultures, it was observed (Bischoff and Holtzer, 1969) that 24% of the mononucleated cells fused without synthesising DNA, though they had ample opportunity to do so in vitro since fusion was not evident until the end of the first day in culture. It was implied, therefore, that these cells had withdrawn from the cell cycle prior to the setting up of the cultures and were awaiting suitable conditioning of the medium (O'Neil and Stockdale, 1972) before becoming competent to fuse.

This hypothesis suggests that myoblasts fall into two populations - dividing stem cells and post-mitotic cells, which have passed through the critical division and one, or both, of the daughter cells have repressed DNA synthesis and thus activated that part of the genome necessary for fusion. O'Neil and Stockdale (1972), however, have found no evidence for the existence of these two populations of cells and, therefore, discount the critical mitosis theory. They observed cell fusion occurring just 3 hours after the previous period of DNA synthesis - that is, at the beginning of G1, and if cells failed to make contact during the G1 period, they appeared to undergo another round of DNA synthesis before fusion. This means that the ability to fuse arises in a cyclic fashion in synchrony with the cell cycle.

It is possible that these differing observations could be accounted for by the vastly different culture conditions, which were shown to promote large differences in fusion rate, and various manipulations of the cell environment.

In support of his 'critical mitosis' theory, Holtzer has observed myoblasts which show striated myofibrils, i.e. contain actin and myosin, which later fuse into myotubes (Okazaki and Holtzer, 1966). The myoblasts do not undergo cell division and due to the presence of striated myofibrils, must be committed to myogenesis.

It appears, therefore, that the study of myogenesis in relation to the cell cycle can vary according to the *in vitro* conditions employed and more reliable factors which demonstrate terminal differentiation, such as the expression of specific muscle genes, for example myosin, will give a more meaningful indication of events. However, recent studies have focused on the expression of such genes at the protein or mRNA levels and various disagreements have arisen which reinforce the proposals that

developmental studies should be made at the level of the genome, and centre around the control of the expression of genomic DNA.

It has been proposed that myosin synthesis is not a reliable indicator of terminal differentiation since early investigations could not detect myosin in culture until after fusion (Coleman and Coleman, 1968; Yaffe and Dym, 1972; Pzybyla and Strohman, 1974) and yet others report detectable amounts in presumptive myoblasts (Chi *et al*., 1975). Emerson and Beckner (1975) showed that myosin synthesis increases 50-100 times during fusion over levels synthesised in the dividing myoblast, from a rate of 300 molecules/nucleus/min. before fusion to 20,000 molecules/nucleus/min in cultures which are 85% fused. In addition, when cells were inhibited from fusing by EGTA treatment to chelate all calcium ions, and prevented from dividing and synthesising DNA by lowering the serum and embryo extract concentrations in the medium, myosin synthesis paralleled those levels found in fusing cultures. However, other studies (Paterson and Strohman, 1972; Pzybyla and Strohman, 1974) involving inhibition of fusion could not detect myosin synthesis, though the conditions used did promote cell division. Using fluorescent antimyosin, Moss and Strohman (1976) detected myosin in most cells of their fusion blocked, post-mitotic myogenic chick cultures, and found that they could accumulate the protein at the same rate as fused cells, and it was also shown (Vertel and Fischman, 1976) that myosin was organised into fibrils in fusion blocked cells. The conflicting evidence arising from these early studies is probably due to differing levels of sensitivity of methods of detection of myosin heavy chain which therefore made it very difficult to establish the exact time of onset of of synthesis. More recent work (Devlin and Emerson, 1978; Shani *et al*., 1981a, b) has shown, using two dimensional gel electrophoretic

techniques that large numbers of contractile proteins are synthesised around the time of cell fusion, though the exact time of onset is still undetermined, but is probably a few hours before fusion (Shani *et al.*, 1981a,b). Therefore, it appears that the activation of myosin synthesis does not require the cells to be in their terminally differentiated state – that is, post-fusion, but is promoted in cells where cell division and DNA synthesis have ceased. These experiments are consistent with the view that cells withdraw from the cell cycle after a particular cell division and from then, the cell is determined on one course of development with the co-ordinated activation of specific genes.

Although some workers (Yaffe and Dym, 1972; Pzbyla and Strohman, 1974) could not detect myosin in myogenic cells prior to fusion, they do agree with others that fusion is not the primary event in gene activation during myogenesis, since at the time fusion occurs, some muscle specific gene transcription has taken place. This was implied by an experiment in which actinomycin D, applied to cultures at the time of fusion to block RNA synthesis, did not prevent the initial rise in myosin heavy chain synthesis, suggesting that myosin heavy chain mRNA is present prior to fusion. This led to the hypothesis (Yaffe and Dym, 1972; Morris *et al.*, 1972; Buckingham *et al.*, 1976) that the myosin heavy chain mRNA accumulates in advance of protein synthesis and is retained in the cell in an inactive form for a significant period of time. However, the actinomycin D experiments have also been interpreted (Strohman *et al.*, 1977) in terms of the existence of a very small amount of mRNA which may be present prior to fusion in order that the low level of myosin synthesis detected by Emerson and Beckner (1975) can be explained, and thus confuses interpretation of experiments on the accumulation of mRNA.

Several workers have isolated myosin heavy chain mRNA in pre-fusion myoblasts in the form of a messenger ribonucleoprotein particle (mRNP) sedimenting between 80S and 120S. However, the methods used for positively identifying the myosin messenger are open to dispute. Attempts at *in vitro* cell free translation have generally failed, but may be due to the messenger being of low translational efficiency or modified in some way when associated with proteins. Rubinstein and Holtzer (1974) isolated a 26S polyA containing RNA from myoblasts and called it the putative myosin message, though there was no other evidence (apart from the sedimentation value) that this actually was myosin message, except that it migrated at 32S on polyacrylamid gels, a value which has been determined for myosin heavy chain mRNA. It was shown that during the logarithmic growth of myoblasts, but not fibroblasts, this 26S species increased in amount. Its synthesis then decreased during the lag phase prior to fusion, though immediately before fusion, synthesis increased again. These changes also correlated with a change in half-life of the message from 10 hours before fusion to 56 hours after fusion. It was confirmed that these are the same RNA species by hybridisation to cDNA.

Other studies (Robbins and Heywood, 1976; Dym, Kennedy and Heywood, 1979) have identified the same mRNA, mainly in mRNP particles prior to fusion, and mainly in polysomes after fusion, by hybridisation with cDNA copies of 26S polyA$^+$ RNA. Using the same technique to quantify the levels of myosin heavy chain mRNA, it was shown by Robbins and Heywood (1976), that the total number of myosin transcripts decreases after cell fusion when the synthesis of myosin is increasing. This could be related to the increase in stability of the messenger found previously by Buckingham *et al.* (1974) and suggests

an increase in the efficiency of translation of mRNA during differen-
tiation when there is a general shift in the polysome profile from one
with many monosomes and small polysomes (with which the mRNP's
sediment) to one with many heavy polysomes. Thus, the detection
of mRNP particles prior to fusion which synthesises myosin in a cell
free system (Heywood *et al*., 1975; Bag and Sarkar, 1976) gives
credibility to the theory that terminal differentiation takes place when
the genome is already programmed for the synthesis of specific proteins
at the transcriptional level, but phenotypic changes occur later and
are controlled at the level of translation. When fusion occurs, the
messengers present may be stabilised, though this is not directly
associated with the appearance of heavy polysomes, which occurs some
16 hours later.

A model for the maintenance of mRNA in an inactive form has
been proposed by Bester *et al*. (1975), involving translational control
RNA (tcRNA) as shown in diagrammatic form in Figure 1.2. Evidence
for the existence of this is that small RNAs can be isolated from both
polysomes and mRNP fractions (Bester *et al*., 1975) which inhibit the
translation of the mRNA present in the same fraction but not in the other.
The tcRNA isolated from mRNPs contains a long polyU tract not found in
the tcRNA isolated from polysomes and therefore it was proposed that
this hybridises with the polyA tract of the messenger RNA as part of
the inactivation process. Furthermore, a specific nucleotide region at
the 3' end of the molecule hybridises with the 5' end of the message,
thus circularising and further inactivating it. On activation, the
polyU tract is hydrolysed, so making the RNA available to the ribosomes
for translation. Heywood and Kennedy (1976) showed that a small
poly(U) rich RNA interacts with myosin mRNA on a 1:1 basis, rendering

FIGURE 1.2:   The Mechanism of Action of Translational Control
RNA (Bester *et al.*, 1975).

Translational control RNA (tcRNA) is cleaved from the
heterogeneous nuclear RNA (HuRNA) in the nucleus and it
passes into cytoplasm as part of the mRNA molecule. Its
long poly(u) tract is able to hybridise with the poly(A)
tract at the 3' end of the mRNA, thus circularising and
preventing the messenger entering a polysomal form. On
hydrolysis of the polyU residues, the mRNA is released into
an active form.
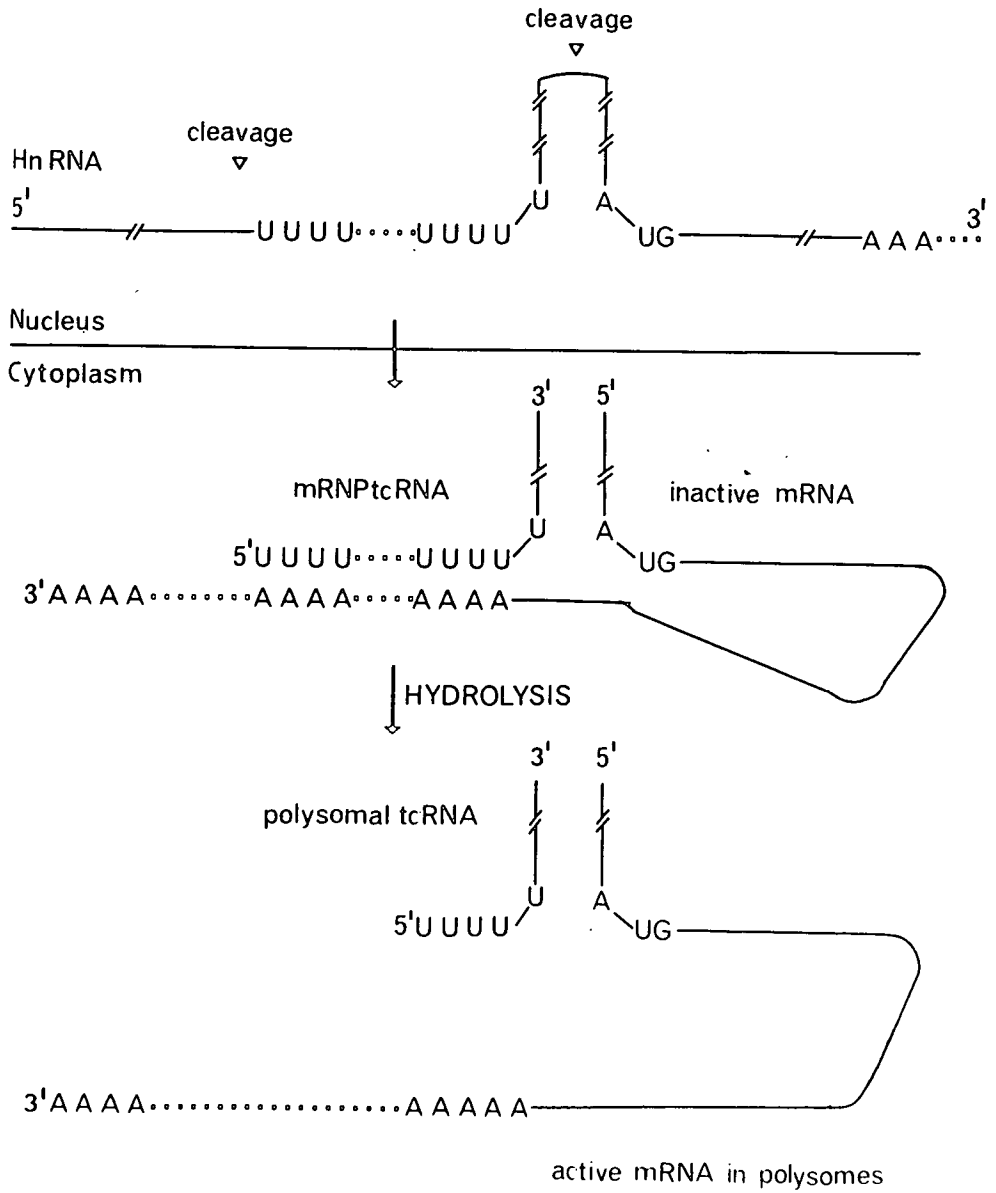
it resistant to T_1 and T_2 nuclease and completely inhibited from translation. They concluded that a translational control RNA does exist, controlling the utilization of messengers in the embryonic system.

Evidence against this model is provided by reports (Pzybyla and Strohman, 1974; Benoff and Nadal-Ginard, 1979) which fail to detect myosin heavy chain mRNA binding to oligo dT-cellulose, and therefore imply that the polyA stretch of the mRNA is non-existent or very short. Benoff and Nadal-Ginard (1979) suggest a series of 13 polyA residues only, although it is possible that, as the myosin heavy chain mRNA 'ages', its polyA tail is shortened, as has been observed for the globin message (Perry and Kelly, 1973). However, Dym, Kennedy and Heywood (1979) did not detect hybridisation of myosin heavy chain cDNA to the unbound fraction from their oligo dT-cellulose column and so could not confirm these results. Most investigators have used the polyA$^+$ fraction eluted by oligo dT-cellulose fractionation of total RNA as an initial step in the purification of myosin mRNA and later work (Benoff and Nadal-Ginard, 1980) has shown that myosin HC mRNA with a short polyA tail translates very poorly *in vitro* relative to myosin HC mRNA with a longer polyA tail.

The existence of mRNP particles implies a translational control mechanism for the synthesis of myosin. However, several workers favour a transcriptional control mechanism following studies on the levels of myosin HC mRNA in early (pre-fusion) and late (post-fusion) cultures. Equal amounts of total RNA from each of these cultures was translated *in vitro* and the relative amounts of protein synthesised by each preparation was determined (Strohman *et al*., 1977). Fused cultures gave a 30-fold increase in the myosin heavy chain mRNA activity over unfused cultures and it was demonstrated by mixing the two RNA preparations

that no inhibitory factor was acting on the translation of the RNA from the unfused cultures. The conclusion drawn from these experiments was that transcriptional control must be important because the increased synthesis in protein before and after fusion was reflecting direct changes in the mRNA levels of the cell. However, no actual measurements of RNA levels were made and therefore the experiment was simply measuring the translatable activity of the messengers present before and after fusion. It is possible, then, that myosin heavy chain mRNA is present before fusion in an inactive form, possibly an mRNP particle, which is not translatable in an *in vitro* cell free system, in which case these results imply the importance of translational control.

Two studies have estimated the quantitative nature of mRNA populations during development. Paterson and Bishop (1977) worked on the changes in mRNA populations during myogenesis and Axel, Fiegelson and Schultz (1976) analysed the messenger sequence complexity of the chicken liver and oviduct. Both studies concluded that transcriptional control is important during differentiation and when taken with studies on haemoglobin (Ross *et al.*, 1974), fibroin (Suzuki and Suzuki, 1974), and chorion proteins (Gelinas and Kafatos, 1972), there is a definite trend towards this view.

mRNA from pre-fusion cultures and fused cultures was hybridised to homologous cDNA to estimate the number of sequences expressed in each cell type and their frequency distribution, and cross hybridisation was carried out to determine the extent of homology between the two populations (Paterson and Bishop, 1977). This showed that the messenger population in the myoblast was ten times more complex than that in the myofibril. The myoblast expresses 17,000 different mRNA sequences present in three abundance groups, but the myofibril only

expresses 2400 sequences, 20% of which is only 6 different sequences present in 15,000 copies/cell. It was shown by cell free translation that the change in abundance of messengers as fusion proceeds correlates with a rapid increase in the synthesis of several muscle specific, predominantly structural, proteins, including myosin, actin and tropomyosin, providing evidence that myogenesis is controlled at the level of transcription.

Using a similar hybridisation technique, Axel, Fiegelson and Schultz (1976) showed that 70% of the oviduct mRNA represents only 10 different species, present as 14,000 copies/cell, and in the liver 43% of the messengers account for just 7 different types. 90% of the mRNA of the posterior silk gland of the silk moth is fibroin mRNA (Lizardi, 1976).

Thus, there is conflicting evidence, both for and against the operation of transcriptional and translational control mechanisms during myogenesis – particularly regarding the expression of the myosin gene. However, even if translational control does play an important role, then transcriptional control must operate at some stage since there are definite changes in the concentrations of various mRNA species at some developmental stages, and differences between cell types (Ross *et al.*, 1974; Suzuki and Suzuki, 1974; Axel, Fiegelson and Schultz, 1976; Lizardi, 1976; Gelinas and Kafatos, 1977; Paterson and Bishop, 1977). As well as the operation of some control mechanism over the onset of myosin heavy chain synthesis, there must also be co-ordinated switches between foetal and adult programs of myosin heavy chain gene expression These may be between embryonic and adult isoforms, or perhaps, the repression of slow isoforms in preference for the synthesis of fast forms, or visa versa. One way of obtaining a better understanding of gene

expression is to investigate the gene at the genomic level where the mechanisms of control may be found and understood.

1.3 Cloning of Muscle Function Genes

1.3.1 Introduction

The increasing number of controversies which are arising out of current research on muscle development indicate that a study of the genes at their most fundamental level is necessary to be able to resolve many of the problems. Such a study will reveal the number of myosin heavy chain genes, for example, and possibly lead to a conclusion about their expression during development, and in various physiological muscle types if suitable probes become available. From structural work on several contractile proteins (that is, actin, myosin, tropomyosin and troponim) it has become evident that they exist as different isoforms which have different primary sequences, and therefore are coded for by multi-gene families. The myosin heavy chain genes, in particular, appear to be expressed in a tissue specific manner as already discussed, and the advent of recombinant DNA technology is now making possible studies of such genes at the level of the genome.

Several attempts have been made to purify single genes from genomic DNA. Brenner *et al*. (1970) isolated tRNA from *E. coli* and by repeated hybridisation to total sheared, denatured, *E. coli* DNA and subsequent fractionation on a hydroxylapatite column to separate the hybrids from single stranded DNA, isolated the corresponding genomic DNA. Shih and Martin (1973) improved this basic system by chemically linking RNA to cellulose in a column maintained at 37°C with 50% form-amide to encourage hybridisation of the RNA to complementary sequences of sheared total genomic DNA which passed through the column. The

unbound DNA was then heated to 90 °C to denature any DNA/DNA

hybrids which had formed, and was then re-passed over the immobilised

RNA. If a specific RNA is used, it can be completely saturated with

complementary DNA fragments which are later eluted. The system was

tested by binding SV 40 RNA to the cellulose and circulating SV 40 DNA

mixed with *E. coli* DNA. The SV 40 DNA was specifically retained even

when the *E. coli* DNA was in a 10^4-fold excess.

In an effort to purify the ovalbumin gene sequence (Anderson and

Schimke, 1976; Woo *et al.*, 1977) which makes up 10^{-6} of the total

chicken genome, a cDNA copy of the specific mRNA was synthesised and

this was bound to phosphocellulose. 5 kilobase lengths of chicken DNA

were then circulated over this and it was found that the ovalbumin

sequence was specifically retained and could be enriched 15,000-fold.

Due to slight contamination of the original mRNA preparation with

messengers of the other egg white proteins, conalbumin and ovomucoid,

a lengthening of the hybridisation time allowed isolation of these gene

sequences also (Anderson and Schimke, 1976).

Many genes have now been isolated by complementary DNA (cDNA)

cloning in plasmids, a technique which has become quite routine in

many laboratories. This involves the synthesis of cDNA from either

a purified mRNA species, or a population of mRNA molecules, from one

particular tissue, for example. The cDNA is subsequently chemically

linked into a bacterial plasmid (a closed circular DNA molecule which can

be maintained in a bacterial cell as an extra-chromosomal element), as

shown diagrammatically in Figure 1.3 and used to transform a suitable

host bacterium which is then able to produce many copies of the particular

piece of DNA. Advantages of recombinant DNA technology over the

earlier methods are that gene sequences can be synthesised in quantity

FIGURE 1.3:   Strategy of Cloning in Bacterial Plasmids.

To clone a specific gene in a plasmid, a cDNA copy is
synthesised from a previously purified specific messenger
RNA.   This is made into a double-stranded from which is
then 'tailed' at the 3' ends of each strand with a short
stretch of cytidine residues.   The plasmid is cleaved at
a suitable site with a restriction enzyme and the free 3'
ends are extended with guanosine residues.   Mixing allows
base pairing between the polyG and polyC tails, and treat-
ment with DNA ligase to seal the 'nicks' gives a recombinant
molecule which can be used to transform a suitable bacterial
host.   This figure shows specifically the strategy involved
in using pBR322 as the bacterial plasmid which confers
resistance to both of the antibiotics, tetracycline and
ampicillin.   The plasmid is cut with the restriction enzyme
PstI which recognises a single site situated within the amp^r
gene and thus recombinants will lose their ability to confer
ampicillin resistance on transformed bacteria.

CTGCAG
GACGTC

double
stranded
plasmid
pBR322

cut with endo-
nuclease PstI

G G
G
G
G$_n$

'G' tail using ^{d}GTP
and terminal trans-
ferase

G^{G}G^n

G G
G
G
G$_n$

Anneal and
transform a
suitable host

– – – – – – – – purified mRNA

↓ reverse transcription

single stranded cDNA

↓ DNA polymerase I, Klenow fragment

double stranded cDNA

↓ S1 nuclease

removal of hairpin loop

↓

$_n$CCC ═══════ CCG$_n$ 'C' tail using
dCTP and terminal
transferase

anneal with plasmid by pairing
complementary single stranded
tails

$(^{Gn}_{Cn})$ ═══ $(^{nC}_{nG})$

and in a very pure form providing probes for subsequent study, such as actual sequencing and analysis of control of gene expression. In some cases, expression of the cloned eukaryotic gene can be achieved in the bacterial cell such that proteins, hormones, or other biological molecules important for medicine or industry, such as insulin, growth hormone, and interferon can be isolated in large quantity, at low cost, in a pure form.

However, the sequence which is cloned from cDNA is not necessarily the complete copy of the gene. On a technical level, the mRNA, if it is very long may not be completely copied by reverse transcriptase, due to inefficiencies in the methodology, producing a shortened cDNA. More significantly, the genomic sequence of a gene in many cases, has been shown to be much longer than the mature mRNA. For example, the β-globin gene has been extremely well studied, and it has been shown to contain two regions, called introns, which are not represented in the mature mRNA in the rabbit (Jeffreys and Favell, 1977; Van Den Berg *et al.*, 1978), mouse (Tilghman *et al.*, 1978a,b; Konkel, Tilghman and Leder, 1978) and human (Flavell *et al.*, 1978; Lawn *et al.*, 1978). Similarly, the ovalbumin gene contains six introns in the chicken genomic sequence (Dugaiczyk *et al.*, 1978; Gannon *et al.*, 1979). There is much speculation about the function of these intervening sequences, but as yet there is no identifiable consistent pattern emerging by way of nucleotide sequence or number of introns within the structural genes isolated so far. Indeed, the introns seem to be the most divergent part of some gene sequences which are very well conserved across the evolutionary tree, such as the globins and actins.

Within the past 3-4 years, several developments have yielded techniques which, in combination, make a direct approach to the actual

genomic sequence very feasible. These are cloning methods which

allow the whole genome of an organism to be maintained as a 'library'

of random DNA fragments incorporated into the genome of a modified

form of bacteriophage λ, and are illustrated diagrammatically in

Figure 4.1.

The DNA of organisms with small genomes, such as Drosophila

(Wensink et al., 1974; Glover et al., 1975) and Yeast (Carbon

1977), have been cloned as libraries in plasmids. More recently, the

cloning of the DNA of those organisms with larger genomes including

mouse (Blattner et al., 1978), rat (Sargent et al., 1979), human

(Lawn et al., 1978), rabbit (Maniatis et al., 1978) and silk moth

(Maniatis et al., 1978) has been possible due to the development of an

efficient screening technique by Benton and Davis (1977). This allows

10^6 recombinant phage to be screened daily, potentially allowing the

recovery of any single copy gene, providing that a suitable probe is

available.

The cloning of the genomes of higher organisms, particularly

mammals, has been made possible by the development of EK2 certified

cloning vectors - certain safe forms of bacteriophage lambda, which can

only survive under very special conditions which do not arise in the

natural environment, and due to the development of in vitro packaging

systems, where the recombinant DNA of a lambda phage can be incor-

porated into a phage protein coat, to produce a viable phage particle, in

the test tube.

There are some advantages in utilizing cDNA clones in preference

to genomic clones. For instance, some RNA sequences have no DNA

equivalent, such as RNA viruses, so cDNA cloning is the only possible

method. Also, a cDNA clone bank is generally simpler to screen than

a genomic library because it will contain fewer clones, being representative of the mRNA of a particular tissue only. If the aim of a particular cloning experiment is to produce a particular eukaryotic protein, then cDNA cloning in bacteria is more likely to result in expression of the cloned sequence, since, there is no evidence to indicate that prokaryotic genes contain introns, and therefore seems unlikely that bacteria will have the necessary splicing enzymes to remove the intervening sequence that interrupt the transcripts of eukaryotic genomic DNA.

However, the sequences represented in a cDNA clone bank will not be representative of the whole genome of the organism. Some sequences will be completely absent, and those which are present will be so in a proportion which reflects the abundance of the particular mRNA in the original tissue. This may be desirable for some studies, but here it was more important to obtain as equal a representation as possible of the whole genome of the chicken to enable the library to be used for the isolation of several myosin heavy chain genes which are probably expressed at different times during development and in different tissues.

Thus, it is the aim of this project to construct a DNA library of the chicken genome and to make a specific hybridisation probe for the myosin heavy chain which is expressed in embryonic skeletal tissue. The probe will be used to screen the library and hence to isolate one, or more (if there is cross reaction between the different types) of the various myosin heavy chain genes. Further advantages in isolating genes from a genomic library rather than constructing cDNA clones in plasmids are that sequences, which may function in controlling expression of the gene, on either side of the coding region are isolated too, as well as the previously mentioned fact that the gene can be obtained as a complete copy, including sequences within coding regions which themselves are not represented in the mature message.

1.3.2 *Structure of Bacteriophage Lambda (λ)*

The mature bacteriophage λ particle consists of an icosahedral protein head containing a single DNA molecule, and a protein tail which makes contact with the bacterial cell wall during infection of a susceptible host, and through which the phage DNA is able to pass into the bacterial cell, leaving the protein coat behind. The structure is shown diagrammatically in Figure 1.4.

The wild type DNA consists of a single linkage group (Jacob and Wollman, 1954) of 49,000 base pairs (30.8 x 10^6 daltons) (Schroeder and Blattner, 1978) which exists in a linear or closed circular form depending on conditions. The circular form arises due to annealing of the cohesive ends of the molecule (Hershey *et al.*, 1963). These are single-stranded ends of the same length which are complementary (Kaiser and Wu, 1965; Wu and Taylor, 1971) and thus can be annealed and sealed with ligase to form circular molecules.

Only λDNA molecules with cohesive ends are infective. Kaiser and Wu (1965) used DNA polymerase I to fill in the single-stranded ends at each end of the molecule, but these completely double-stranded molecules were not infective, until exonuclease III was used to remove (from the 3' end) the newly synthesised bases. The cohesive ends of separate molecules can also anneal, a left end to a right end (Hershey and Burgi, 1965), to form long concatamers of λDNA units. It is DNA in this state which is particularly important in the final developmental stage of a mature λ particle (Feiss and Widner, 1982). Weigel *et al.* (1973) showed the cos site to consist of a 22 base pair sequence with strong rotational symmetry as shown in Figure 1.5.

41 genes have so far been identified in the λ genome as shown in Figure 1.6. 21 code for the head, tail and assembly proteins, 12 are responsible
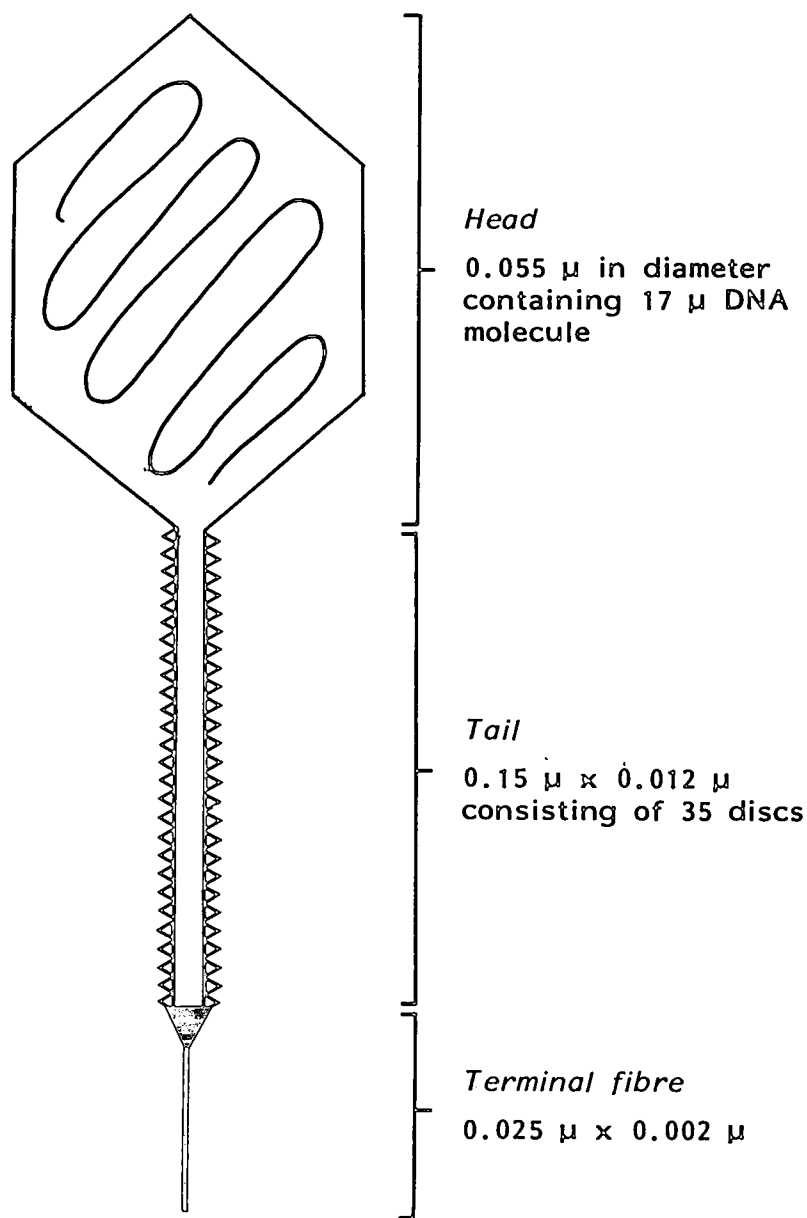
FIGURE 1.4: Diagrammatic Representation of a Bacteriophage λ Particle.

The mature bacteriophage λ particle consists of an icosahedral protein head, which contains the single DNA molecule, and a long tail piece which is assembled from 35 protein discs. During infection, the terminal fibre makes contact with the host bacterial cell wall and the DNA molecule is injected into the host cell following its passage through the tail.

OH—C C C G C C G C T G G A———— — — — —————————————————————A 5'

5' G—————————————————————— — — — ———G G G C G G C G A C C T—OH

Left hand end   Right hand end
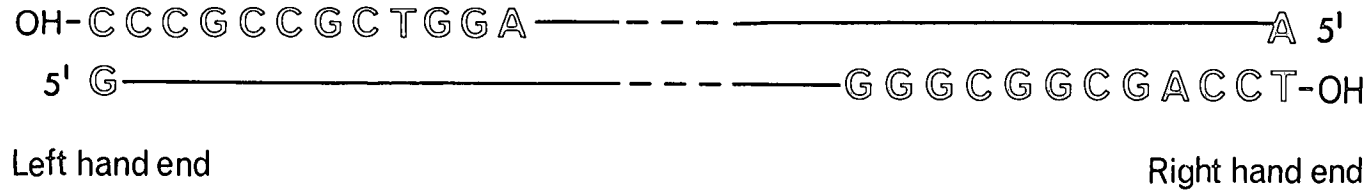
FIGURE 1.5:   Nucleotide Sequence of the Bacteriophage λ Cohesive Ends.

Before DNA replication can proceed, the DNA molecule which constitutes the bacteriophage
λ genome must circularise.   This is made possible by virtue of the single stranded protruding
ends which have a 12-nucleotide complementary sequence, as shown.

for DNA replication, recombination and lysis of the host cell, and 8 are regulatory genes. 27 of these are essential functions. The remaining 14 non-essential genes can be removed from the genome, and so long as the total length of the DNA remains greater than 75% of the wild type genome, then it will still be viable.

Phage λ is a temperate phage – that is, it can exist in a lytic or lysogenic state. The lytic state involves infection of a host bacterium by a single phage which then undergoes DNA replication and phage protein production, assembly of mature phage particles and finally results in the lysis of the host cell to release progeny phage. The lysogenic state, however, does not immediately destroy the host cell. The phage DNA molecule is incorporated into the host chromosome and is called a prophage. This renders the bacterium immune to further infection by a phage of the same type, but upon induction the prophage may excise from the bacterial chromosome and enter the lytic cycle. During lysogeny, integration into the bacterial chromosome is always at a specific point. This site on the phage chromosome is between the genes int and J. The maintenance of lysogeny depends initially on the synthesis of a repressor protein which is transcribed from the cI gene and which binds to the operators on either side of the cI gene, preventing transcription from these points, as shown in Figure 1.7a. This, in turn, prevents transcription from other points on the genome because all transcription is initially dependant on these two operators. In the absence of the repressor, RNA polymerase can reach the promotors as shown in Figure 1.7b, and move along to the initiation sites. Genes N and cro are transcribed initially, transcription ending at points beyond these genes. The cro gene product prevents repressor protein synthesis and this establishes the beginning of the lytic cycle, and so the balance

FIGURE 1.6: Functions of the Known Bacteriophage λ Genes.

This figure shows the functions of the 41 bacteriophage λ genes which have thus far been identified. The b_2 region, whose function is unknown, lies between genes J and att on the λ chromosome.

DNA packaging	Nu1	
	A	
head components and assembly	W	
	B	
	C	
	Nu3	
	D	second major head protein – 23%
	E	major head protein – 75%
DNA packaging	FI	
	FII	
tail components and assembly	Z	
	U	
	V	
	G	
	T	
	H	
	M	
	L	
	K	
	I	
	J	specifically controls host range by synthesis of the tail antigen
integration, excision and recombination	att	⎤
	int	⎬ integration of prophage
	xis	⎦
	exo	
	red A	⎤ vegetative recombination
	red B	⎦
	gam	recombination and replication
	kil	
immunity and regulation	cIII	establishment of lysogeny
	ral	
	N	control of recombination and replication genes
	rex	
	cI	repressor protein synthesis
	cro	controls repressor protein synthesis
	cII	establishment of lysogeny
DNA synthesis	O	
	P	
host lysis	Q	late gene regulation
	S	controls host lysis
	R	synthesis of endolysin

between lysogenic or lytic development depends on a direct antagonism between the cro and cI gene products. The N gene product then allows transcription of the delayed early genes (by a read through at the termination codons t_L and t_{R1}), O and P, which control replication and recombination of the phage genome, Q, which regulates the transcription of the late genes A-S, int which controls integration of the prophage and cII and cIII which are necessary for the transcription of cI. The transcription of int, cII and cIII leads to lysogeny, and the transcription of ⊚,P and Q leads to lytic development. Which of these occurs depends on the relative times and efficiencies of the action of the repressor protein and cro gene product.

Following transcription of the delayed early genes, the λ DNA, in a closed circular form (Ogawa and Tomizawa, 1967), due to annealing of the cohesive ends, undergoes replication to produce various forms of the molecule (Young and Sinsheimer, 1967, 1968). The first predominant form is a monomeric closed circle which replicates semi-conservatively and bidirectionally, and after 10 min of infection, both open and circles can be isolated. After 30 min, a heterogeneous population of molecules can be isolated which appear to be precursors of the mature linear form, being much larger than the single genome, and have been identified as concatamers. These are generated by a rolling circle method of replication where the circle is a unit genome but the tail extends for several genome lengths. Only DNA in this concatameric state can be used as a precursor for packaging into mature phage particles (Stahl *et al.*, 1972; Syvanan, 1974; Drabinka *et al.*, 1976; Hohn *et al.*, 1976). The specificity is for an uncut cohesive end site (cos site) since packaging is initiated by the A protein which complexes with the Nul gene product and the DNA at these specific sites. The complex formed between the

FIGURE 1.7: Mechanism of Repression and Induction of Bacterio-
phage λ Gene Expression.

(a)   shows the initial stages in the establishment of lysogeny.
A repressor protein, transcribed from gene cI binds to the
operators and promoters on either side of the cI gene, thus
preventing further transcription and, therefore, replication.

(b)   shows that in the absence of the repressor, the lytic cycle
can begin since RNA polymerase is able to bind at the
operators and transcription therefore occurs. Once the cro
gene product is available, to prevent synthesis of the
repressor, the lytic cycle becomes firmly established. The
N gene product prevents termination at t_l, t_{r_1} and t_{r_2} and
the Q gene positively regulates late transcription from P'_r
which establishes and maintains lysogeny.

a. Repression

repressor
mRNA

$O_l P_l$
$O_r P_r$
CI
N
t_l
cro
t_{r1}
t_{r2}
Q
P'_r
red
int
J
S
R
cos
Z F A

b. Induction

RNA polymerase

$O_l P_l$
$O_r P_r$
CI
12S mRNA
7S mRNA
N
t_l
cro
t_{r1}
t_{r2}
Q
P'_r
red
int
J
S
R
cos
Z F A

A and Nul gene products is called λ terminase (Hohn, 1975). The λ DNA is packaged directly into preformed precursor structures, called petite λ, made up of the phage coded protein E. ATP is necessary for this process, but before it is completed the D protein is added to cover the whole of the icosohedral shell. The λ terminase then cuts the concatameric DNA at the cos site to liberate two cohesive ends of a mature λ DNA molecule. By an analogy with type II restriction enzymes, it is the symmetry of the cos site sequence which is believed to be important in the terminase binding and nicking functions. The packaging of λ DNA into its protein coat is illustrated diagrammatically in Figure 1.8.

Syvanan and Yin (1978) using electron microscope studies have observed petite λ structures appearing like 'beads on a string'. It has been suggested that the primary icosohedral structure has two openings through which the DNA threads, becoming attached by a cos site at one opening while the remainder of the molecule is packaged through the other.

Thus, during one infectious cycle, λ DNA, either from an infecting particle or from a prophage, becomes circularised, undergoes several cycles of replication producing concatameric molecules of several units of DNA, and is packaged into precursor particles when the DNA is cleaved into monomers with cohesive ends. The final step in the assembly of a mature, viable, phage particle is the attachment of tail fibres to the DNA filled head from which point the DNA is then free to leave the head once more on initiation of another infective cycle.

FIGURE 1.8: Diagrammatic Scheme of the Assembly of a Bacterio-
             phage λ Particle.

This figure shows the stages involved in the assembly of a
mature bacteriophage λ particle and the gene products
involved. DNA is packaged into the prehead before the
head is completely assembled and before the attachment of
the tail. There is a physical restriction on the amount of
DNA which can be accommodated within the head particle,
and when this limit is achieved the DNA within the head is
cleaved from the rest of the concatamer by an enzyme which
recognises the cos site. The head expands by 20% during
this process. When the head is filled, the D protein locks
the head in place around the DNA. Finally, the head is
associated with preformed tail units which are associated by
a separate pathway.

	Structural proteins in the particle	Genes required for assembly

Prehead I consisting of protein core

E,B,C,Nu3

Empty prehead II

E,B,X1,X2

GroE

DNA packaging begins

E,B,X1,X2

A,Nu1

Fully expanded head. D protein is added around the outside

E,B,X1,X2,D

A,FI

Full head is stabilised for tail attachment. It contains a unit length of λ DNA

E,B,X1,X2,D FII,W

Tail components are assembled prior to attachment to the head.

Genes I,L,K,G,H,M,V, Z and U are required.

mature particle

1.3.3 *Manipulation of λ Bacteriophages to Produce Cloning Vectors*

As genetic engineering techniques have developed, several vector systems have been described due to the discovery of bacterial restriction enzymes which have made possible the physical manipulation of DNA. These endonucleases, which occur naturally, prevent the infection of a bacterial cell by foreign DNA by cleaving the invading DNA into fragments. The host DNA is protected from cleavage by methylation. There are three classes of restriction enzymes, and it is type II which is important in DNA manipulation. Each enzyme recognises a specific 4 or 6 base pair sequence and cleaves at certain phosphdiester bonds which may, or may not, lie opposite on each DNA strand of the double helix. When they are opposite, double stranded, blunt ends are formed and when they are not, single stranded, asymmetrical, complementary ends arise. Restriction enzymes are described in more detail in Chapter IV, Section 4.5.

Initially, eukaryotic genes were cloned in plasmids (Cohen *et al.*, 1973), extra-chromosomal DNA elements of bacteria, as briefly described in Section 1.3.1. A large number of plasmids are now available as cloning vectors, especially for use with *E. coli* as host (reviewed by Morrow, 1979) which possess the following necessary features:

1. capable of autonomous replication within the host cell;

2. contain a single restriction enzyme site for as many restriction enzymes as possible;

3. confer a selective phenotype on the transformed host cell since only a small number are transformed.

Plasmids often confer drug resistance on the host cell, which fulfils criterion 3 mentioned above, a feature which is useful when screening

for transformed bacteria. For instance, if a DNA fragment is inserted into the restriction enzyme sites of the endonucleases HindIII, SalI, or Bam HI, of plasmid pBR322, the result for the host *E. coli* cell will be loss of tetracylin resistance but the presence of ampicillin resistance. Therefore, successful transformants are selected by their ability to grow in the presence of the drug ampicillin but not in the presence of tetracyclin.

Plasmids have been used successfully for the construction of gene libraries from Xenopus (Smith *et al.*, 1979), Drosophila (Wensink *et al.*, 1974; Glover *et al.*, 1975), Sea Urchin (Kedes *et al.*, 1975) and Yeast (Carbon *et al.*, 1977), though the efficiency of transformation is relatively low and therefore plasmids are more useful for cloning cDNA which represents a small number of different sequences. The size of insert which is most efficiently transformed and detected is in the region of 5-10 kilobases, which is rather small if the object is to isolate a particular genomic DNA sequence with its flanking nucleotides.

Bacteriophage λ is extremely suitable for use in the construction of DNA recombinants containing large inserts of foreign DNA because a third of the λ genome, making up a single continuous block in the centre region, can be replaced without losing the ability to grow lytically. The current extensive knowledge of both the structure and genetics of bacteriophage λ has enabled workers to manipulate it in various ways so that it fulfils all the requirements of a receptor vector molecule as follows:

1. small and easily prepared in pure form;

2. able to replicate autonomously in a bacterial host;

3. only contains one, or a very limited number, of targets for a particular restriction enzyme;

4. carries some identifiable characteristic such as drug resistance, immunity, plaque formation or ability to complement an auxotrophic strain;

5. insertion of a foreign DNA does not destroy an essential gene;

6. recombinant vectors are easily selected;

7. able to clone a variety of sizes of DNA fragments;

8. able to clone with minimal manipulations.

The wild type form of the λ genome (λ^+) cannot be used as a vector directly for several reasons. The DNA has too many target sites for restriction enzymes which means that many fragments would have to be linked together during the formation of a recombinant molecule. There are also physical limitations on the actual size of the DNA molecule which can be packaged into a λ capsid. This ensures that viable particles are only formed when the DNA length is between 75% and 106% of the λ^+ genome, that is between 34 and 48 kilobases (Fiess *et al.*, 1977), indicating that the cloning vector must be of a reduced size to enable the insertion of foreign piece of DNA and its subsequent packaging. The maximum size of a DNA insert which can be introduced into the λ^+ genome is only about 5 kilobases. The recommendations of the Genetic Manipulation Advisory Group (GMAG) also make the use of λ^+ impossible since it does not meet the required containment standards.

Therefore, the λ^+ genome has been manipulated by several workers to enable it to fulfil these conditions. Deletion mutants provide a suitable starting point in the construction of vector molecules since they may remove unwanted restriction targets and provide physical space for the insertion of foreign pieces of DNA. Suitable genetic crosses

followed by selection of the required mutant may also remove unwanted restriction sites. There are now several derivatives of the phage λ genome which are suitable for cloning (reviewed by Morrow, 1979), each with their useful restriction enzyme target sites. For instance, λ 1059 has a single Bam HI site (Karn *et al.*, 1980), NM762 has a useful HindIII site (Murray *et al.*, 1977) and phages λgt WES (Tiemeier. *et al.*, 1976), NM641 (Murray *et al.*, 1977) and a range of Charon phages (Blattner *et al.*, 1977; De Wet *et al.*, 1980) all possess EcoR1 cloning sites. EcoR1 is a restriction enzyme which recognises a specific hexanucleotide sequence (Green *et al.*, 1974) and cuts it to give staggered ends. The $λ^+$ genome has 5 such sites (Allet *et al.*, 1973) but these have been reduced to one in some cloning vector forms. Therefore, it is possible to insert foreign pieces of DNA bearing the cut EcoR1 sequence at each end into the cut λ EcoR1 site by virtue of the complementary nature of the cleaved EcoR1 site.

Thomas *et al.* (1974) constructed a λ cloning vector, later modified by Tiemeier *et al.* (1976) and Leder *et al.* (1977) called λgtWES. The genome has been manipulated to contain two EcoR1 sites which allows the internal fragment, carrying only non-essential genes, to be removed leaving a molecule of 73% of the $λ^+$ genome which thus cannot be packaged alone. There must, therefore, be an insert of a certain minimum size, from 1-14 kilobases, inserted in place of the removed fragment before viable phage can be produced, which provides a positive selection method for recombinant phage particles. In order that this vector satisfies biological containment criteria, some specialised recombination genes have been substituted by an inserted fragment and three amber mutations have been introduced into very essential genes - W (the protein necessary for joining head and tail compoments), E (the major

capsid protein) and S (the host lysis gene). Therefore, λgt WES phages are only able to grow on hosts which suppress the W and E gene mutations and are only able to lyse hosts which suppress the S gene mutation.

The vector to be used in this study is a derivative of λ$^+$ called charon 4A, and is one of a series of 21 charon vectors constructed by Blattner and co-workers (Blattner *et al.*, 1977; Williams and Blattner, 1979; De Wet *et al.*, 1980). The charon phages contain mutations which make them easy to use and increase their biological safety. They are able to incorporate various lengths of foreign DNA up to 25 kilobases. The structure of charon 4A is compared with that of wild type λ in Figure 1.9.

Throughout the development of recombinant DNA research techniques there has been concern over the potential risk of transmission of foreign genes, via the cloning vector, into the natural environment. The guide lines set out by the National Institute of Health on this matter state that "for EK2 vector systems, when the vector is a phage, no more than 1 in 10^8 phage particles should be able to perpetuate itself and/or a cloned fragment under non-permissive conditions designed to represent the natural environment, either as a prophage in the laboratory or by transferring itself to another host in the natural environment".

Charon cloning vectors are designed with inherent safety mechanisms which reduce their survival chances outside specific hosts. In particular, they have two or more amber mutations introduced into essential genes which means that very specific hosts carrying the necessary suppressor tRNA's are essential for their propagation, as described for λgt WES. charon 4A is certified for use in EK2 vector systems because of the introduction of amber mutations into the capsid genes A and B, its

Figure 1.9 →

FIGURE 1.9:   Comparison of the Wild Type Bacteriophage λ Genome
              with that of Charon 4A.

The wild type bacteriophage λ genome (a) is compared with that
of charon 4A (b). The regions of the wild type genome which
have been replaced or deleted during its manipulation to produce
an EK2 certified cloning vector are shown. Certain recombination,
DNA replication and host cell lysis genes of the wild type have
each been substituted to limit the chance of propagation of charon
4A in the natural environment. Some deletions have also been
introduced to allow the recombination of foreign DNA fragments,
and thus the establishment of charon 4A as an efficient cloning
vector. The positions of the EcoR1 endonuclease recognition
sites are shown. Positions of some important genes are shown in
(a), but for a more comprehensive list of phage λ gene functions
see Figure 1.6. The verticle lines in (a) indicate distances of
1 kilobase along the genome.

cosL A W B E Z H J att int red N cI O P Q S R cos R

◁ - -b₂ region - - - - ▷

Non-essential and replaceable regions

a. Wild-type bacteriophage λ

cos L Lac 5 Bio 256 KH54 NIN5 cos R

Substitutions from the E.coli genome Deletions QSR substituted from Q 80
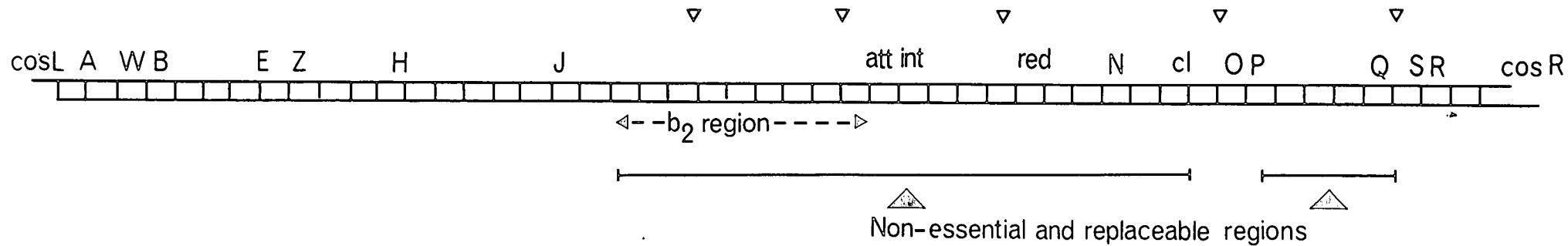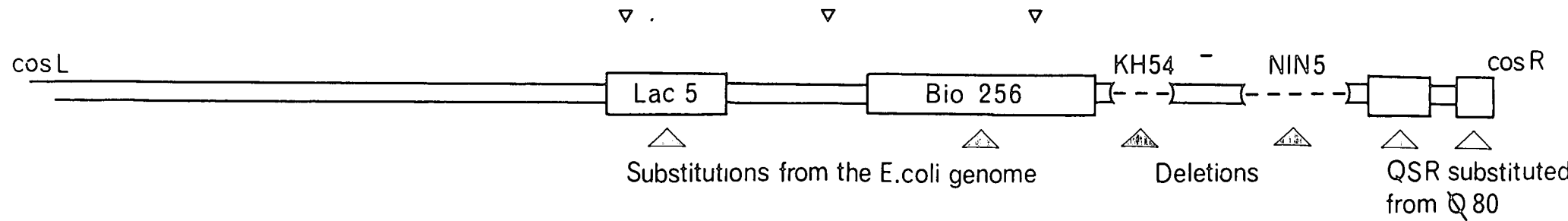
b. Charon 4A

▽ EcoRI sites

45.

inability to form a lysogen due to the deletion of K54 in the immunity
region, the elimination of the genes 'int' and 'att' which are necessary
for insertion into a bacterial chromosome, and the presence of nin5
deletion which enhances the lytic function.

λ phages in general have a poor survival rate in the natural
environment anyway. It was shown by Robeson *et al*. (1980) that less
than 1 in 10^{10} phage particles can survive pH 3.0 (stomach conditions),
less than 1 in 10^7 survive 30 mins in detergent, less than 2 in 10^4
survive 6 hours on a drying bench, and less than 3 in 10^6 will survive
20 hours in sewage. Therefore, there are initial barriers to escape
even before special precautions are introduced genetically.

The genomes of the charon phages have been further manipulated
by substitution, deletion and introduction of point mutations which alter
the distribution of restriction enzyme sites, especially those in the
essential regions of the genome. In the case of charon 4A, this has
resulted in a phage with three EcoR1 sites (as shown in Chapter IV,
Figure 4.6), which gives rise to two large external and two small
internal fragments. The latter are in non-essential regions and are,
therefore, removed during the cloning procedure to provide space for
insertion of foreign DNA fragments, and also leaving the genome smaller
than 75% of the λ^+ genome. As for λgtWES, this means that charon 4A
imposes a lower limit on the size of the DNA fragment which can be cloned
and thus provides its own positive selection system in that the ability
to form plaques is an indicator of successful cloning. Charon phages
10 and 11 do not impose this minimum size limit and therefore can be used
for cloning very small DNA fragments. In these cases, the cloning site
is in a non-essential gene whose function is monitored. Gene inactiva-
tion, therefore, indicates successful cloning. Charon 4A was considered

a better vector than λgtWES for this study because it was necessary to clone large fragments (approximately 20 kilobases) to increase the possibilities of eventually isolating a complete gene sequence for the myosin heavy chain and to reduce the amount of screening required to a minimum. A further advantage is that charon phages possess an ability to grow to high yields (up to 10^{10}/ml) in small volumes when given the correct conditions.

Many charon phages, including 4A, contain the lac5 region from *E. coli* which codes for β-galactosidase which acts as a second selection system. An EcoR1 site is situated in the lacZ promotor which means that when plated on the colourless chromogenic substrate 5-bromo-4-chloro-3 -indolyl-β-D-galactosidase (XG indicator plates), they produce blue plaques due to the synthesis of β-galactosidase which cleaves the chromogenic substrate and releases a blue indolyl derivative. In recombinant phage, the central EcoR1 fragment containing the lacZ promotor is substituted, and, therefore, β-galactosidase is not synthesised, giving rise to colourless or faint blue plaques.

More recently, cosmid cloning vectors have been developed (Collins and Bruning, 1978; Collins and Hohn, 1979). Cosmids are vectors derived primarily from plasmids but containing the phage λ cos site. Since only the cos site is necessary for recognition by the phage λ packaging system, cosmids can be packaged *in vitro* and transduced into *E. coli* at high efficiency. Once inside the bacterium, the cosmid circularises and behaves as a bacterial plasmid, conferring resistance to antibiotics, for instance. Cosmids are very small molecules, often less than 5 kb in size, and therefore the lower limit of DNA fragments which can be cloned is 30 kb and the upper limit approximately 45 kb - much larger than those which can be inserted into plasmids or λ vectors.

Therefore, the cosmid system is useful for cloning very large pieces of DNA, but so far the efficiency of cloning has not equalled that of λ vectors.

Until now, phage λ vectors have been the most widely used cloning vehicles, especially for the production of gene libraries because they are relatively easily handled, efficient and easily screened.

1.3.4 *In Vitro Packaging of Recombinant Bacteriophage λ DNA*

Bacteriophage λ DNA is packaged *in vivo* in the presence of the products of genes A, Nu1, D and F1, which are important in the packaging process itself (Hohn and Hohn, 1974; Hohn, 1975), the head precursor genes E, B, C and Nu3, and the genes W and FII which are responsible for the assembly of mature particles (Casjens *et al.*, 1972) and phage tails.

However, phage λ DNA, whether wild type or recombinant, can also be packaged *in vitro* into infectious virus particles (Casjens *et al.*, 1972; Hohn and Hohn, 1974; Hohn 1975; Hohn *et al.*, 1976). This occurs without propagation through a bacterial host and thus offers a degree of biological containment in the case of recombinant DNA. Furthermore, it eliminates the possibility of a selection mechanism acting on certain recombinant molecules by the bacterial host at this stage, as is thought to occur with transfection mechanisms (Cameron *et al.*, 1975). Transfection in this context involves treatment of bacteria with calcium chloride in order that they will take up recombinant bacteriophage DNA molecules and package them into viable phage *in vivo*.

Details of the methodology of the *in vitro* packaging system are presented in Chapter IV, Section 4.8, but the principles of the reaction will be described here. *In vitro* packaging is based on the use of two

E. coli bacterial strains, each carrying a λ prophage which is defective
in the synthesis of a different protein essential to the packaging process.
These lysogens, when induced by heating to 42°C, which inactivates
the repressor protein, begin to accumulate phage proteins, but due to
each having a genetic block in one essential gene, mature phage do not
arise and in effect, the bacteria are "bags of phage spare parts".
Upon lysis and mixing of the two induced strains, the two genetic blocks
complement each other such that all the necessary gene products for
the formation of mature phage λ are present. When exogenous λ DNA,
ATP and spermidine are added, packaging occurs spontaneously *in vitro*.

Sternberg *et al.* (1977) developed two lysogenic bacterial strains
suitable for *in vitro* packaging, one carrying a mutation in gene A and
one with a mutation in gene E. In addition, each has an amber mutation
in the host lysis gene, S, which means that lysogens fail to lyse and
as a result three times the normal amount of phage components accumulate.
To prevent endogenous phage DNA from being packaged, each prophage
also carries the λb2 mutation which results in a damaged attachment site
and thus prevents excision of the prophage from the host chromosome.
Also, both the hosts' and the prophage recombination systems are
inactivated by mutations in genes recA and red3 respectively, thus
preventing any recombination between exogenous and endogenous DNA.

Hohn and Murray (1977) have also developed two lysogenic strains
of *E. coli* suitable for use in *in vitro* packaging reactions based on the
same principal. One has a genetic block in gene E and one in gene D.
However, there is no genetic mutation to prevent packaging of endogenous
DNA, so this must be damaged by the application of ultra violet irradiation prior to the addition of the exogenous packaging substrate.

The final stage in the process of *in vitro* packaging involves a chlorform treatment which kills any existing bacteria and thus provides a means of chemical containment.

The efficiency of the *in vitro* packaging procedure is at least one hundred times greater than that of transfection to produce mature recombinant phage particles. Vector molecules carrying inserted DNA fragments are usually larger than the wild type molecule and for this reason have a lower transfection efficiency. Since the ligation of vector and insert DNA is usually inefficient, the number of packaged recombinants arising from a transfection experiment is, therefore, very low. Although phage λ will only package DNA molecules of a certain size limit, and there is some prejudice shown towards individual molecules which lie within the limits, such that wild type molecules are packaged more efficiently than recombinants, if there is no bacterial cell wall to traverse prior to packaging, then the *in vitro* method would be expected to be the more efficient. Estimates have indicated (Hohn and Murray, 1977) packaging *in vitro* of 1 in 10^3 wild type DNA molecules compared with only 1 in 10^5 appearing in mature phage particles during transfection. Although these figures are 10-100 times lower when recombinant DNA is used, their relative orders remain the same.

Transfection offers the opportunity for some recombinant DNA molecules to be selected for or against by the host, and for many to disappear altogether (Cameron *et al.*, 1975). Packaging does not allow the same selection pressures to arise since DNA is merely a substrate and only the cos sites have to be specific. Therefore, the characteristics of the DNA being packaged are theoretically of little consequence in the resulting plaque forming units and should reflect the abundance of the various DNA molecules present.

However, Sternberg showed that his packaging strains do show a preference towards the packaging of phage genomes which have a size closest to that of wild type λ DNA and, in fact, he found a direct relationship between the size of the DNA up to 100% of λ$^+$ and the packaging efficiency. Thus, there seems to be a selection mechanism operating on a certain size class of recombinant molecules. However, further selection pressures of a genetic nature will inevitably arise later during the replication and growth phase within host cells when the numbers of recombinant phage are amplified.

The strains developed by Hohn and Murray (1977), however, have been shown not to possess this selective mechanism based on size and all molecules within the size range 75% to 106% of the wild type λ genome are packaged equally well. It is these strains which were used in this work due to a generous gift by Professor K. Murray.

1.3.5 Selection and Screening of Recombinant Bacteriophage λ Clones

The final step in any cloning procedure is the screening of the clones to detect those containing the foreign DNA sequence of interest. In general, screening procedures have tended to fall into two classes: screening by phenotype and directly searching for the nucleotide sequence of interest.

The first class is applicable particularly to plasmid libraries containing prokaryotic DNA inserts when a particular gene can be expressed in *E. coli* and will confer upon the host a readily detectable phenotype. A number of genes from *E. coli* (Clarke and Carbon, 1975, 1976), fungi (Ratzkin and Carbon, 1977) and yeast (Struhl *et al.*, 1976) have been isolated in this way. It has also been possible to detect particular recombinant bacteriophage λ particles using functional expression by growing in a host defective in a particular function

(Cameron *et al.*, 1975; Moir and Brammer, 1976; Borck *et al.*, 1976; Hopkins *et al.*, 1976). Increased expression of a gene can be achieved by delaying cell lysis using an S$^-$ strain, and can be further exploited if the insert is made in genes N or Q when late proteins will not be synthesised, allowing large quantities of DNA to be available for transcription. If the inserted sequence does not have its own promotor, however, then to ensure expression it must come under the control of P$_L$. Hence, the exploitation of well studied λ genetics allows the useful amplification of a variety of genes and their products to be achieved with λ vectors even when a specific promotor is not cloned, such as restriction enzymes, DNA ligase and polymerase.

When the DNA insert is eukaryotic, the detection of clones by expression of the cloned sequence is not usually applicable due to the different modes of expression of eukaryotic and prokaryotic genes, and different biochemical pathways. Many eukaryotic genes contain intervening sequences and neither *E. coli*, nor bacteriophage λ is capable of synthesising the necessary splicing enzymes which remove the non-translated sequences from the primary transcript during the biosynthesis of the mature mRNA molecule. In addition, many eukaryotic gene products will not complement or suppress *E. coli* mutations, and therefore will not be detected. Occasionally, these problems have been overcome when cDNA has been cloned rather than genomic DNA, and a low level of gene expression has been observed, for example, the mouse dihydro-folate reductase gene was detected in *E. coli* clones because it conferred trimethoprim resistance (Chang *et al.*, 1978). Other gene products have been identified by immunological or biological assays.

The most widely used method of screening is the second class mentioned above – that is, a direct approach to the sequence of interest

by nucleic acid hybridisation using a gene specific radioactive DNA or RNA probe. *In situ* hybridisation methods, more commonly employed in cytological analyses (John *et al.*, 1969), were shown to be applicable using cRNA (Jones and Murray, 1975) as a probe. Sea urchin DNA was digested with EcoR1 restriction endonuclease and electrophoresed on agarose gels. The fragments were eluted and those containing histone genes were located by hybridisation to histone mRNA. These fragments were inserted into a λ vector and of the 200 recombinant plaques which were screened using cRNA, three were found to be positive.

A colony screening procedure for detecting recombinant plasmids was developed by Grunstein and Hogness (1975) and modified by Hanahan and Meselson (1980). A technique based on these procedures was developed by Benton and Davis (1977) and others (Kramer *et al.*, 1976) for screening λ clones. The technique is particularly applicable to screening genomic libraries since 10^6 recombinants can be screened daily, and was therefore the method of choice in this project, as described in detail in Chapter V. The method, described as plaque hybridisation, uses ^{32}P-labelled DNA or RNA as the probe and removes the tedious task of isolating a desired sequence which may be among very many inappropriate sequences, and make up no more than $2 \times 10^{-5}\%$ of the haploid genome (Grunstein and Hogness, 1975; Benton and Davis, 1977). The principal of the technique is based on the fact that the surface of a plaque produced following the lysis of host bacteria will contain some free DNA which was not packaged into viable phage prior to lysis. This DNA will be identical to that contained within the mature phage particles, resulting from one initial infection. Positive hybridisation of a complementary DNA or RNA to the free DNA on the plaque

essentially identifies the recombinant phage of interest which can then
be amplified further. The phage screening procedure has a better
signal to background ratio than the procedures used to screen plasmid
containing bacterial colonies, probably because there is less debris to
interfere and increase the non-specific binding of radioactivity.

Other, less well used screening procedures include 'hybrid
arrested' and 'hybrid selected' translation (Paterson *et al*., 1977).
These methods are applicable when a pure probe is not available but the
mRNA for the sequence of interest is relatively abundant. Hybrid
arrested translation is based on the fact that mRNA cannot be translated
in vitro when hybridised to DNA. DNA extracted from clones or groups
of clones is hybridised to the mRNA which is then incubated in a cell-
free translation system. Inhibition of translation of the protein of
interest will identify the clone (or group of clones) which contain the
DNA sequence of interest. Hybrid-selected translation is more sensitive
and is based on the presence rather than absence of a particular protein.
Again, cloned DNA is hybridised to the mRNA and the hybrids are
isolated by column chromatography, melted, and the mRNA translated
in a cell free system. Successful translation will identify the DNA
arising from the clone of interest.

More recently, radioactively labelled synthetic oligonucleotides
have been used in screening methods. These are useful when it is
possible to deduce a reasonable nucleotide sequence (Agarwal *et al*.,
1981) from a known amino acid sequence for a particular protein whose
mRNA is particularly difficult to isolate in a pure form, and therefore
cannot be used as a source of specific probe. This method was partic-
ularly successful in the isolation of a messenger RNA for the hormone
gastrin (Noyes *et al*., 1979).

The chicken gene library constructed during this project was screened by the Benton and Davis (1977) plaque hybridisation technique using specific probes for the myosin heavy chain and ribosomal DNA sequences. The latter probe was included as a positive control to test the completeness of the library and to test the screening procedure.

1.3.6 Structure of Eukaryotic Gene Sequences

Recombinant DNA technology, particularly that leading to the cloning of genomic DNA has permitted the isolation of gene sequences in pure form from the DNA of higher organisms, and has made possible detailed analysis of gene structure and gene evolution. Bacteriophage λ libraries have been a constant source of pure genomic sequences over recent years and have contributed considerably to present day knowledge about eukaryotic gene structure due to the relative ease of isolation of 'single copy' genes.

One of the most important pieces of information to emerge on gene structure is that eukaryotic, but not prokaryotic, genes are split. This phenomenon, first described in 1977 by Jeffreys and Flavell for the rabbit β-globin gene and since reported for many genes from vertebrates, insects, fungi and, recently, plants (Sun et al., 1981), means that a genomic sequence is often much larger than the corresponding messenger RNA due to the existence of one or more intervening sequences within the actual coding region, as mentioned briefly in Section 1.3.1 of this chapter. The intervening sequences are called introns and those making up the coding regions, exons. The entire gene is initially transcribed (Tilghman et al., 1978b) and subsequently introns are removed by a sequential cleavage-splicing mechanism which is very precise, removing introns in a very specific order to finally release the mature mRNA. This was demonstrated by Chambon et al.

(1979) in studies on the chicken ovalbumin gene which has 7 introns
(Gannon *et al.*, 1979; Dugaiczyk *et al.*, 1978) and spans 8 kb. A similar
procedure was shown to occur by Tsai *et al.* (1978) and Nordstrom *et al.*
(1979) for the ovomucoid gene which is 5.6 kb long compared with 821
nucleotides in the mature messenger RNA (Lai *et al.*, 1979b). Again, there
are 7 introns varying from 0.22 - 1 kb in length which are removed in a
definite order. This form of gene processing is now believed to be
universal for eukaryotic gene sequences. The only genes which
definitely do not contain introns are the histones (Kedes, 1979), and
the human leucocyte interferon gene (Nagata *et al.*, 1980). However,
there are also other extreme examples such as the *Xenopus laevis*
vitellogenin sequence which contains 33 introns (Wahli *et al.*, 1980).

Therefore, between different genes there is no set pattern
emerging relating to number or length or position of intervening
sequences though genes which tend to fall into families, such as the
globins, have similar numbers of introns at similar positions in the
genomic sequence and many genes show similar patterns between species.
In general, it has been noticed that the intron sequences are more
divergent between species than the exons (Van Den Berg *et al.*, 1978),
though an exception to this is the intron-exon splice junction where
there is considerable sequence conservation which probably act as
recognition sites for splicing enzymes. It has been suggested (Lerner
et al., 1980) that a small species of nuclear RNA is involved in the
splicing reaction. Yeast tRNA precursors contain a 15-base sequence
which is not present in the mature tRNA. Extracts of yeast cells
have been shown to be capable of removing these extra nucleotides
and rejoining the remaining segments (Knapp *et al.*, 1978). In general,
the sequences on either side of the introns are not complementary

(Breathnach *et al.*, 1978) which implies that a theory suggesting that splicing takes place by homologous pairing of small pieces of DNA causing looping out of the intron which is then removed, is not valid.

Evidence about the actual rôle of intervening sequences is very confused. Initial speculation was concerned with a possible rôle in the control mechanisms of gene expression but due to the inconsistency in number, position and sequence, this now seems unlikely. The pre-proinsulin gene, for example, exists as two non-allelic copies in many species. Preproinsulin I has two introns at identical positions in all species studied, but preproinsulin II has two introns in most species, but only one in the rat which is analogous to the first intron in gene I. Similarly, the actin gene contains one intron in yeast, sea urchin and drosophila, but has three introns in the chicken (Ordahl *et al.*, 1980). The presence of two introns in the β-globin sequence appear necessary for the correct processing of transcripts (Hamer and Leder, 1979a,b) and some level of gene expression (Fritsch *et al.*, 1979), but the actual sequences are very diverse between species when compared with the actual coding sequence, for example, between the rabbit and the mouse (Van Den Berg *et al.*, 1978). The introns seem to have diverged by all possible means, that is, base substitution, deletion and insertion, with the coding sequences remaining highly conserved. Gilbert (1978) suggested that the exons code for stable, discrete, domains of proteins which fold into a single functional unit such as the region containing the active site of an enzyme. The boundaries of protein domains are not known with enough precision for these theories to be rigorously tested, but there is evidence that some proteins consist of such regions. The structure of the immunoglobulin gene, for example, reflects the arrangement of the variable and constant regions in the heavy and light chains

(Brack *et al.*, 1978). These domains are separated by intervening sequences in the genomic sequence. Similarly, the central coding block of the globins corresponds to the haem binding region of the protein while the other areas are involved in polypeptide chain interactions. If this theory is true, then during evolution, exons are probably "shuffled" about in the genome without giving rise to haphazard peptide sequences, but possibly producing a protein with a selective advantage to the organism. Therefore, introns may not play a direct rôle in gene expression but may be advantageous as raw material for the evolution of new genetic functions.

Many genes exist as families in the genome, including myosin heavy chain, and, in fact, the true single copy gene sequence may not actually exist. Ribosomal and histone sequences are present as long tandem repeats in all eukaryotes so far studied, and others are repeated to a lesser extent, such as 8 non-allelic copies of interferon have been isolated from a human genomic library (Nagata *et al.*, 1980), some of which are clustered. However, the clustering of gene families is not universal. Fyrberg *et al.* (1980) used chicken and Dictostellium actin gene probes to screen a Drosophila genomic library and identified six sequences widely dispersed sites on the chromosome. All the genes contained intervening sequences which are transcribed but do not appear in the mature mRNA as determined by the presence of three size classes in embryos and larvae. Dictostellium has 15 actin genes (Kindle and Firtle, 1978) and sea urchin has between 5 and 20 (Durica *et al.*, 1980). The chorion genes are even more complex with over 100 different genes being expressed sequentially during oogenesis in the chicken (Kafatos *et al.*, 1978).

The most intensively studied gene family, made possible by recombinant DNA technology is the globin gene family (Efstratiadis *et al.*, 1980). Human globins are coded by two unlinked clusters, the α globins which include the foetal ζ globins, and the β globins which include the foetal γ globins, embryonic ε globin and the minor adult δ globin. The individual genes are separated by substantial tracts of intergenic DNA which reflects the apparent DNA excess in higher organisms in general. They have clearly evolved by a series of duplications and have at some stage become unlinked to form the two separate clusters which have then further diverged from each other. The genes have retained an ancestral pattern of intervening sequences though there is little sequence homology between the introns. All have two introns, one major (approximately 570 base pairs) and one minor (approximately 120 base pairs) in the protein coding region at homologous positions, including the chicken (Dodgson *et al.*, 1979), mouse (Konkel *et al.*, 1978; Tiemeier *et al.*, 1978), human and rabbit (Hardison *et al.*, 1979a,b). The primary transcript of the β-globin gene is approximately 1400 nucleotides, though the mature mRNA is less than half this size at 675 nucleotides. A 900 nucleotide species is believed to be a processing intermediate following the removal of the major intervening sequence.

Recombinant DNA technology has recently led to a better understanding of the arrangement of multiple gene sequences. It has become clear over the past few years that some small peptides (7-40 amino acids in length) are synthesised as large precursor polypeptides which appear immunologically to incorporate several functional peptides. The first of such complex "multi-genes" to be identified was proopiomelanocortin (POMC) which appeared initially to be made up of the two pituitary

hormone sequences, adrenocorticotrophic hormone and β-lipotrophin, among much redundant protein (Roberts and Herbert, 1977; Nakanishi *et al.*, 1979). However, when the gene for POMC was isolated, it was shown to contain, in addition, the sequences of several melanotrophins, endorphins and met-enkephalin (Roberts *et al.*, 1979; Cochet *et al.*, 1982), and in fact gives rise to a single polypeptide which is the precursor for all these small neuropeptides. Figure 1.10 shows the arrangement of these peptides in a single precursor molecule. Similarly, oxytocin and vasopressin are synthesised with their binding proteins, neurophysin I and II respectively (Schmale and Richter, 1980; Schmale and Richter, 1981) and met- and leu-enkephalins are synthesised as a long tandem repeat which contains the peptides in various proportions depending on the species (Noda *et al.*, 1982).

Following the sequencing of some of these multigene complexes, they appear to contain non-coding sequences which do not appear in the mature mRNA and also coding sequences which do not appear in the active peptide, as shown diagrammatically in Figure 1.10. They, therefore, have redundant material at two levels of gene expression, transcription and translation.

Following the isolation of many structural genes, questions are now being asked about the sequences which flank the genes, both on the 3' and 5' sides. A requisite for answering such questions is the isolation of overlapping clones, upstream and downstream from a given gene or cluster. This is carried out by the generation of a terminal fragment by restriction enzyme digestion of a given clone, which is then labelled by nick translation and used to re-screen the library from which the original clone was obtained. This will isolate overlapping sequences which extend upstream or downstream from the original clone. When
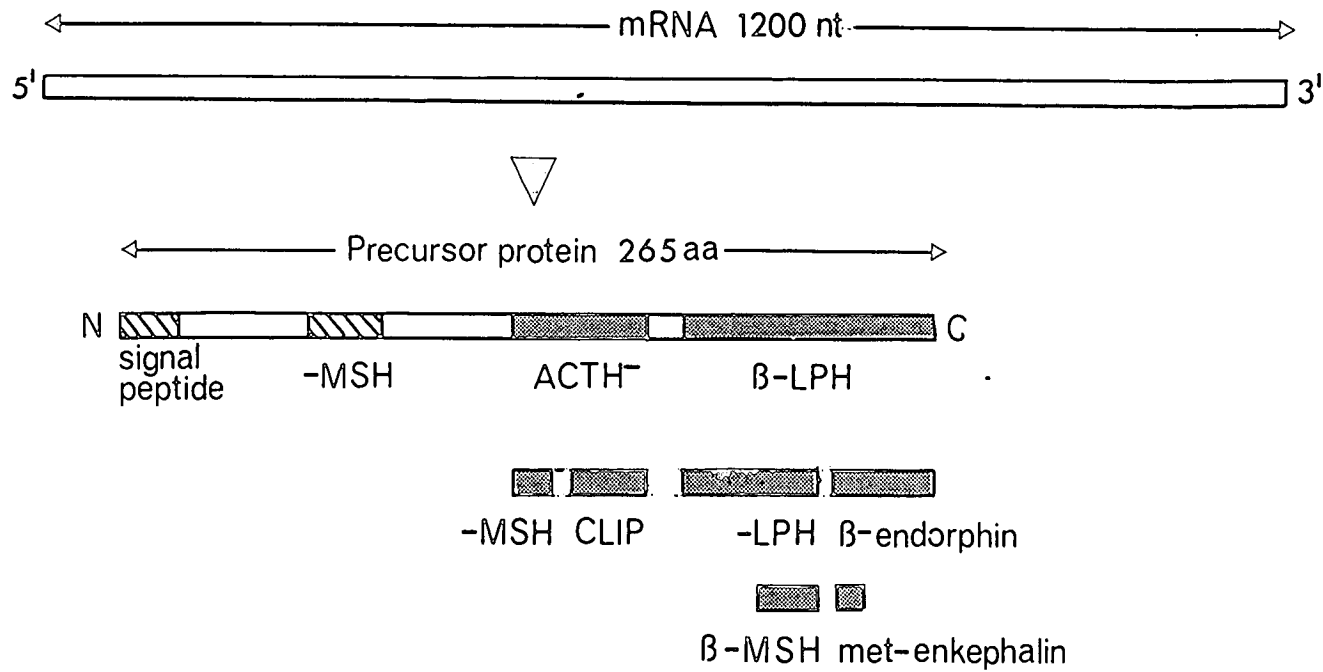
FIGURE 1.10: The Structure of the Proopiomelanocortin Molecule.

This figure shows the arrangement of the nine peptides incorporated in the proopiomelanocortin molecule. They are transcribed into a single messenger RNA molecule and then translated into a single, large precursor protein of 30,000 molecular weight. The precursor is later processed by a series of enzymatic cleavages within the anterior pituitary gland to yield the various active peptides.

flanking sequences are compared between genes and between species, in general there is little sequence homology. However, just as exons can be recognised as isolated regions of conserved sequences, there are some conserved elements in the non-coding regions around the genes. The first of these, an A-T rich region, was noticed by Goldberg in 1978 and is commonly called the TATA box. Its actual sequence is

$$T \ A \ T \ A \ {}^A_T \ A \ {}^A_T$$

and it is present approximately 30 bases before the coding sequence of many eukaryotic genes which are transcribed by RNA polymerase II which generally synthesises messenger RNA (Corden *et al.*, 1980), but is absent from genes transcribed by polymerases I and III. Its structure resembles that of the Pribnow box which is the prokaryotic promotor site (Pribnow, 1975a,b) and because of this homology, its sequence conservation and position, there is much speculation that the TATA box is the eukaryotic RNA polymerase II promotor.

About 80 nucleotides upstream from most eukaryotic genes is another conserved DNA sequence, the "CCAAT box" (Efstratiadis *et al.*, 1980) which is also postulated to play a rôle in the initiation of transcription, by analogy with the prokaryotic promotor which has two functional sites at 7 (Pribnow box) and 35 nucleotides upstream from the gene coding sequence, but evidence for this is not conclusive.

At the 3' end of many genes is the sequence AATAAA which also appears in the mRNA and may be involved with polyadenylation (Proudfoot and Brownlee, 1976). Also at the 3' end are T clusters (TTGT or TTTT) in generally AT rich regions which are preceded by GC rich regions. These have been identified in the 3' flanking sequences of the genes for ovomucoid (Lai *et al.*, 1979b), Xenopus 5S rRNA (Korn and Brown, 1978),

mouse β-globin (Konkel *et al.*, 1978) and rat insulin (Efstratiadis 1979). Brown (1979) showed that a single base change in these four base pairs will cause a read-through of the Xenopus 5S rRNA gene and hence it is thought that such a sequence plays a major role in termination of transcription.

1.3.7 *Structure of the Myosin Heavy Chain Gene*

It is unknown at present whether the structural features of the genes, briefly described in the previous section, can be applied to the myosin heavy chain gene. However, recombinant DNA technology is enabling questions concerning intron-exon structure, flanking sequences, regulatory sequences, the number of genomic sequences and the organisation of these sequences, to be asked directly about the myosin heavy chain gene. Several laboratories have recently isolated myosin heavy chain cDNA clones constructed from mRNA isolated from various sources such as rat heart (Mahdavi *et al.*, 1982), rat leg muscle (Hornig and Nadal-Ginard, 1981), chick embryo leg muscle (Umeda *et al.*, 1981; Umeda *et al.*, 1982) and mouse leg muscle (Caravatti *et al.*, 1982). These studies have provided limited insights into the number of myosin heavy chain genes and their expression in different tissues, though the most detailed investigations so far are into their coding sequences and degrees of sequence homology.

Mahdavi *et al.* (1982) selected four cDNA clones which represented two original mRNA species isolated from rat cardiac muscle. The largest of these inserts, however, only represented 23% of the mRNA molecule, comprising a portion of the polyA track, the complete untranslated 3' end and representing 430 carboxyl terminal amino acids. A most detailed sequence comparison over more than 1000 nucleotides was carried out between the two different DNA clones demonstrating a 97% homology

in amino acid sequence and a 95% homology in nucleotide sequence in the coding region. The untranslated 3' end of the mRNA's showed very divergent sequences. When these clones were compared with an embryonic skeletal muscle myosin heavy chain cDNA clone, regions of homology interspersed with regions of less homology were revealed. This pattern was also clear when two adult rat fast muscle myosin heavy chain sequences were compared (Hornig and Nadal-Ginard, 1982). Thus, it is becoming clear that there are highly conserved regions and more divergent regions between the various myosin heavy chain types from different tissues and species which may indicate that different evolutionary pressures are acting on different parts of the molecule which may reflect areas of functional importance which are able to resist amino acid changes.

When the sequences of cDNA clones from rat heart (Mahdavi *et al.*, 1982), rabbit heart (Umeda *et al.*, 1982a,b) and mouse leg muscle (Caravatti *et al.*, 1982) are compared, an 80% sequence homology is apparent which confirms the theory that "hot spots" for amino acid substitution exist between the various MHC sequences. It is also clear that the last 10 amino acids at the carboxyl terminal are highly diverged in sequence and that the 3' untranslated part of the mRNA shows no sequence homology and varies considerably in length (50-120 nucleotides).

The isolation of these clones and the sequencing studies are beginning to provide information on the expression of the various myosin heavy chain genes during development, the controversies of which were discussed in Section 1.2.4 and 1.2.5. Mahdavi *et al.* (1982) showed that their cDNA clones from adult rat cardiac muscle will hybrid-ise preferentially with mRNA from adult heart muscle but not with foetal muscle mRNA, indicating that foetal MHC is coded for by a mRNA which does not prevail into adult life. However, the two MHC cDNA

clones isolated from rabbit heart muscle (Sinha *et al*., 1982; Umeda *et al*., 1982) probably correspond to an adult and an embryonic form since one is a major form in normal adult ventricles and is replaced by the other in thyrotoxic hearts.

Evidence suggesting differences in fast myosin heavy chain in adult leg and breast muscles of the chicken were demonstrated by Umeda *et al*. (1982) who isolated two MHC cDNA clones with 83% nucleotide homology in the carboxyl terminal region. One of those sequences is represented in adult and newborn leg muscle but not in the corresponding breast. It is also present to a greater extent in the adult tissue, though the extent to which any of the clones discussed above will cross hybridise with sequences from smooth muscle and non-muscle sources is unknown.

These recent studies using MHC cDNA clones have provided some important information on gene structure, particularly the fact that all clones of the 3' and of the mRNA so far studied have a highly variable, both in size and sequence, untranslated 3' end which suggests that a specific control function in this region is unlikely. However, these clones will be most useful in determining amino acid sequences of the various myosin heavy chain isotypes and as hybridisation probes for developmental gene expression studies. It is a study of the genomic sequences which will reveal the most about gene structure and control.

Nudel *et al*. (1980) isolated several EcoR1 fragments from total rat genomic DNA which were tentatively identified as parts of five different myosin heavy chain genes through their ability to hybridise with a cDNA clone synthesised originally from rat skeletal muscle total polyA$^+$ RNA. The polyA$^+$ RNA was not shown to synthesise MHC only, though the cDNA was shown to hybridise with a particular polyA$^+$

species which in a cell free system gave a product which co-migrated with purified MHC. However, many other translation products were also produced, though these were probably breakdown products or partial translation proudcts of MHC. The cDNA clone was used to screen a rat genomic library and three different patterns of hybridisation resulted corresponding to the EcoR1 fragments previously isolated of 8.9, 8.0 and 4.6 kilobases. Electromicroscopy of hybridisation between two of the fragments showed homologous regions interrupted by regions of non-homology, thus confirming the results obtained from sequencing cDNA clones that the MHC genes possess areas where the nucleotide sequence has been allowed to diverge during evolution of the gene family. It is not possible from the study of Nudel *et al*. (1980) to speculate on the existence of introns in the MHC gene though they do imply that the areas of non-homology represent intervening sequences on the basis of the very close similarity in amino acid sequence between various MHC proteins. However, more recently Wydro *et al*. (1983) have shown, also in the rat, that very large parts of the MHC gene may actually represent introns, covering up to 3 kilo bases of DNA.

This project is concerned with the construction of a chicken genomic library with the aim of isolating a sequence from this library which contains all or part of a myosin heavy chain gene. The approach has been different to that of Nudel *et al*. (1980), who worked with rat, in that a great effort was made to purify the myosin heavy chain mRNA from embryonic chick skeletal muscle so that a pure cDNA probe could be constructed for probing the library. This made the initial cloning of a total cDNA population unnecessary in this instance.

A description of the methods employed in the project can be divided into three broad categories: firstly, the preparation of a com-

plementary DNA probe for the myosin heavy chain mRNA; secondly, the construction of a chicken recombinant DNA library; and thirdly, the screening of the library using the probe. Details of the methodology are provided in the following Chapters II - V with further particulars in the Appendix.

CHAPTER II

Preparation of Probes

CHAPTER II

2.1 Introduction

There are several widely used methods for radioactively labelling DNA which thus allows the construction of useful hybridisation probes.

End-labelling involves the transfer of a ^{32}P-phosphate group from ATP to either the 5' end (Chaconas and Van de Sande, 1980) or 3' end (Chang *et al.*, 1977) using the enzymes T4 polynucleotide kinase and terminal transferase, respectively. A more uniform distribution of label throughout the length of a DNA molecule is achieved by the two methods described in this chapter: reverse transcription (Kacian *et al.*, 1972; Ross *et al.*, 1972; Verma *et al.*, 1972), and nick-translation (Kelly *et al.*, 1970; Rigby *et al.*, 1977). It was necessary to employ both methods due to the availability of two different starting materials: a purified messenger RNA which is the template required by reverse transcriptase and a purified DNA sequence which is most efficiently labelled by nick-translation.

The specific probe for the myosin heavy chain sequence was prepared by the reverse transcription (Kacian *et al.*, 1972; Ross *et al.*, 1972; Verma *et al.*, 1972) of a purified myosin heavy chain mRNA yielding a radioactively labelled complementary DNA. The mRNA was prepared from a preparation of heavy polysomes (the myosin heavy chain message is large at 2 x 10^6 daltons) isolated from embryonic chick leg muscle by a modification of the methods of Heywood *et al.* (1967) and Sarkar *et al.* (1973). The mRNA was shown to predominantly direct the synthesis of the myosin heavy chain in an *in vitro* translation system, the products of which were analysed by SDS polyacrylamide gel electrophoresis as described in Chapter III.

The second DNA probe was prepared by the nick-translation of a cloned ribosomal gene sequence. This is a process where unlabelled,

double-stranded DNA is labelled by the replacement of existing deoxy-nucleotides by radioactive ones, there being no actual net synthesis of DNA. This second probe was used to test the screening methods described in Chapter V prior to the use of the more valuable myosin heavy chain cDNA probe.

All experiments involving RNA, either directly or indirectly, were performed under strict ribonuclease free conditions. The precaution of wearing surgical gloves was always taken to ensure no contamination from RNase on the skin. All glasswear was heat sterilised at 150°C for 12 hours, solutions were autoclaved for 30 min at 15 pounds/sq. in. and where heating was impossible, solutions and equipment were treated with fresh 0.1% diethylpyrocarbonate (DEP) to destroy any contaminating RNase. Equipment, such as plastic tubing and centrifuge tubes, was washed for 10 min in the DEP solution followed by exhaustive washing with sterile distilled water. Solutions were treated with 3 drops of 0.1% DEP followed by incubation at 60°C for 30-60 min when DEP is evolved as carbon dioxide and ethanol.

2.2 Construction of a Complementary DNA Probe to the Myosin Heavy Chain mRNA Sequence

2.2.1 Isolation of Heavy Polysomes

It has been demonstrated (Patrinou-Georgoulas and John, 1977) that the second fraction from the bottom of a polysome gradient prepared as described below contains predominantly polysomes which direct the synthesis of myosin heavy chain and therefore this method was employed as the initial step in the purification of myosin heavy chain mRNA.

The leg muscles of 14-day embryonic chicks were dissected free of skin and bone and were homogenised in pairs in an equal volume of

M-buffer, containing 50 µg/ml dextran sulphate as an RNase inhibitor, in a loose fitting Dounce homogeniser. The homogenate was centrifuged at 10,000 g for 10 min at 4°C and the supernatant was layered on to a 35 ml linear 15-40% sucrose gradient on a 1 ml 60% sucrose cushion. The supernatant from 5 embryos was layered on to 6 gradients. The gradients were centrifuged at 25,000 rpm for 2 hours in a Beckman L5 centrifuge, SW27 rotor at 4°C. 1 ml fractions were collected from the bottom of the tube and the optical density at 254 nm of each was monitored. The first six fractions (those containing the heaviest material) were pooled with the identical fraction collected from each of 30 gradients and the polysomal material was pelleted by centrifugation at 35,000 rpm for 3 hours in the L5 centrifuge, SW40 rotor at 4°C.

The pellets were rinsed with ice cold sterile water. All operations were carried out at 0°C unless otherwise stated.

The typical profile of a polysome gradient prepared as described and monitored at 254 nm using an isco UV monitor is shown in Figure 2.1.

2.2.2 *Preparation of Total Polysomes* (Lee and Braweriman, 1971)

Unfractionated polysomes were prepared at described for heavy polysomes in Section 2.2.1 but 4-5 ml of the post-mitochondrial super-natant was layered onto 4.5 ml of 20% sucrose and centrifuged at 100,000 g in an MSE 10 x 10 fixed angle rotor at 4°C for 2 hours during which time the polysomes formed a gelatinous pellet. The pellet was rinsed with ice cold sterile water.

2.2.3 *In Vitro Translation of Polysomes*

The polysomes isolated from chick skeletal muscle were translated in a homologous *in vitro* system since the factors necessary for translation were obtained from a fairly crude chick skeletal muscle homogenate
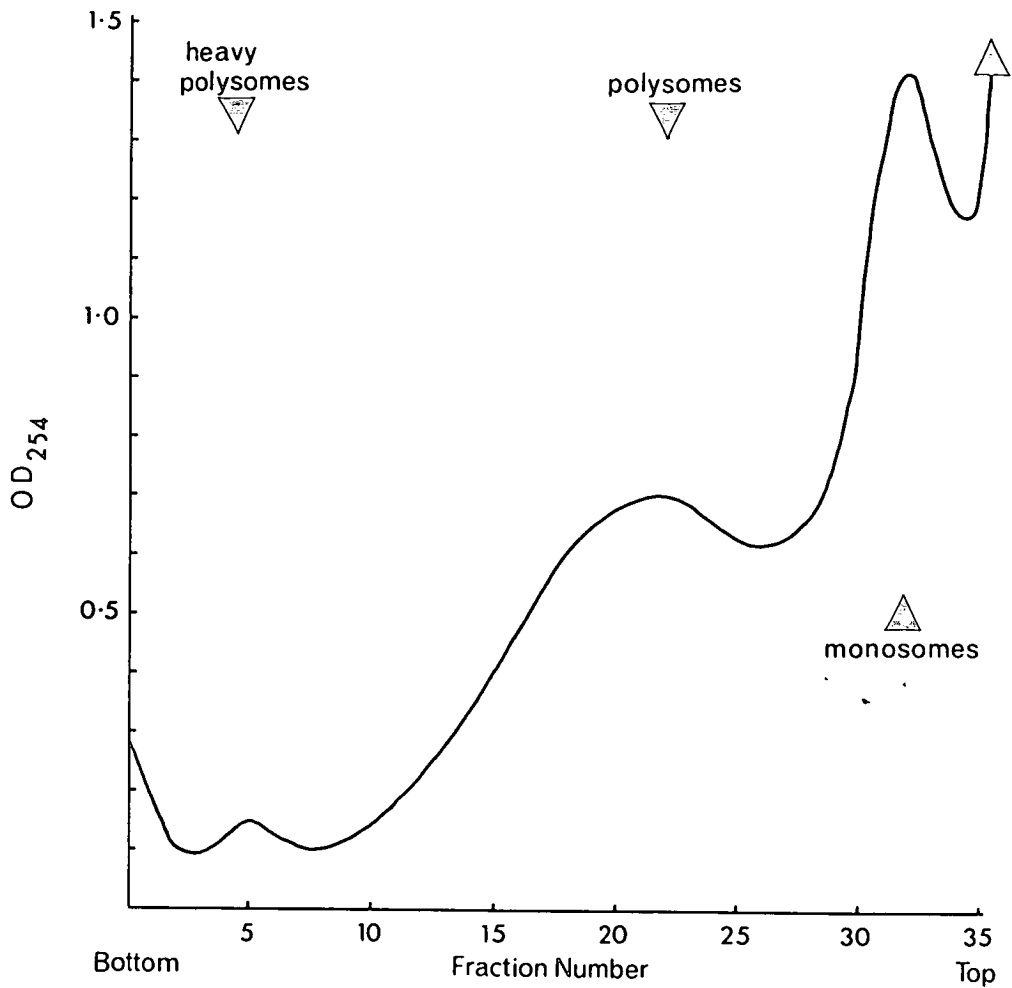
FIGURE 2.1:   Polysome Profile After Centrifugation Through a
              Sucrose Gradient.

The post-mitochondrial supernatant of the homogenate from 5 pairs
of leg muscles was layered onto six 35 ml sucrose gradients
(15 - 40%) and centrifuged as detailed in Section 2.2.1.   30 gradients
were prepared per experiment.   1 ml fractions were collected from
the bottom of the tube and the OD_{254} was monitored using an Isco
UV monitor.   Fractions 1 - 6, containing the heavy polysomes, were
retained.

(S-150 fraction). Although the homologous system is not ideal and is subject to valid criticism, it was used here as a rapid method to assess the complexity of the individual polysome fractions.

Preparation of the S-150 fraction (Heywood and Nwagwu, 1969):

14-day chick embryonic skeletal muscles were homogenised in 2 volumes of ice cold M1B buffer and centrifuged for 10 min at 10,000 g. The supernatant was filtered through fine gauze and centrifuged for 2 hours at 150,000 g at 4°C. The lower two-thirds of the supernatant were then dialysed against M1B buffer at 0°C for 12 hours and then against M1B buffer containing 50% glycerol for 4 hours. The dialysis buffers were changed regularly. The dialysate was stored at -20°C in small aliquots.

The protein concentration of the dialysate was determined by scanning the solution at 260 nm and 280 nm and using the equation:

$$\text{Protein concentration (mg/ml)} = E_{280} \times f \times \frac{1}{d}$$

<div align="right">(Warburg and Christian,
1941)</div>

where d is the light path in cm and f is a factor dependent on the ratio of E_{280}/E_{260} which also gives an indication of the percentage contamination by nucleic acids.

The protein concentration of the S-150 was normally between 6 and 12 mg/ml and a typical result is shown in Table 2.1.

TABLE 2.1: Protein Concentration of S150 Fraction.

Absorbance at 280 nm (E_{280})	11.2
Absorbance at 260 nm (E_{260})	8.6
f (Warburg and Christian, 1941)	0.969
Light path in cm (d)	1.0
Protein concentration in mg/ml	10.85
Nucleic acid contamination in %	1.26

The S150 fraction was extracted from a homogenate of embryonic chick leg muscle, as described in Section 2.2.3. This was used as the source of enzymes and co-factors necessary for the *in vitro* translation of polysomes in a homologous system. 0.6 mg of protein from the S150 fraction were used per 1 ml translation mix, which corresponds to 55 µl of this preparation.

Translation:

0.5 OD_{260} units of polyribosomes were resuspended in 1 ml of a solution containing:

0.15 M	potassium chloride	
0.01 M	magnesium chloride	
0.01 M	tris pH 7.4	
2 mM	ATP	
0.5 mM	GTP	
0.01 M	creatine phosphate	
5 µM	each of 19 amino acids (excluding leucine)	
6 mM	2-mercaptoethanol	
50 µg	creatine kinase	
0.6 mg	S-150 (55 µl, see Table 2.1)	
50 µCi	of ^{3}H-leucine was added to start the reaction	

The reaction mixes were incubated at 37°C for 50 min. 50 µl of each mix were spotted on to 3MM Whatman filter discs, dried and placed in 10% trichloroacetic acid containing 5 mM unlabelled leucine for 10 min. The discs were transferred to 5% trichloroacetic acid at 90°C for 15 min and were then washed in 3 : 1 ethanol : ether. The incorporation of ^{3}H-leucine into protein was determined by counting in toluene PPO POPOP scintillant. The radioactivity was concentrated by centrifuging the polypeptides at 150,000 g for 2 hours and then dissolving the pellet in an appropriate volume of 8 M urea – SDS solution for electrophoretic analysis as described in Chapter III, Section 3.2.1.

Figure 2.2(a) shows that translation of polysomes using the S-150 fraction is dependent on the concentration of polysomes added. Three different concentrations of total polysomes were tested and 0.5 OD_{260}

units/ml were shown to have incorporated the maximum amount of radioactivity into protein under these conditions. Therefore, this same concentration was used in the translation of fractionated polysomes as shown in Figure 2.2(b). The actual acid precipitable counts per minute incorporated following the translation of the 6 fractions of heavy polysomes are shown in Table 2.2. Fraction 1 had very low activity and therefore was not included in the calculation of the mean incorporation which is plotted in Figure 2.2(b).

2.2.4 *Isolation of Polysomal RNA*

Polysome pellets from a total of 30 gradients were resuspended in P buffer to give a final concentration of 3 OD_{260} units. An equal volume of phenol : chloroform : isoamyl alcohol (50 : 50 : 1) was added and the mixture was gently shaken for 15 min at room temperature. Following centrifugation at 10,000 g for 10 min at 4°C, the upper aqueous phase was stored on ice while the lower phenolic phase was re-extracted with an equal volume of P buffer. The first and second aqueous phases were pooled and re-extracted twice more with the phenol mixture, or until the protein interphase had been completely removed. Finally, the aqueous phase was extracted with chloroform and made 0.2 M sodium acetate pH 5.0. RNA was precipitated by adding 2 volumes of ice cold ethanol and leaving to stand at −20°C for 16 hours and then pelleted by centrifugation at 15,000 g for 30 min at −10°C in a siliconised glass tube. The precipitation step was repeated twice before the RNA was used further.

TABLE 2.2: Incorporation of ^{3}H-leucine into Protein During
in vitro Translation of Heavy Polysome Fractions.

| Incubation time (min) | Acid precipitable counts (cpm) | | | | | | Mean of fractions 2 - 6 |
| | *Fraction number* | | | | | | |
	1	2	3	4	5	6	
10	225	562	429	543	584	548	544
20	394	1154	1026	1162	1203	1163	1159
30	512	1733	1694	1785	1849	1752	1779
40	549	2201	2192	2226	2285	2256	2240
50	537	2536	2405	2552	2421	2471	2488

The incorporation of ^{3}H-leucine (expressed as counts per minute)
incorporated into acid insoluble protein during the *in vitro*
translation of heavy polysome fractions 1 - 6. The mean values of
f2 to f6 are plotted in Figure.2.2b. 50 µl aliquots from total
translation mixes of 1 ml were assayed at each time point.

Figure 2.2 →

FIGURE 2.2:

(a) Incorporation of ^{3}H-leucine into protein during *in vitro* transla-
tion of total polysomes.

This figure shows that the incoporation of ^{3}H-leucine into protein,
as directed by total polysomes in a homologous cell free system is
dependent on polysome concentration. Incubation mixes consisted
of a total volume of 1 ml, as described in Section 2.2.3, to which
either 0.25 OD_{260} units, 0.5 OD_{260} units, or 1.0 OD_{260} units of
total polysomes were added. Of these three concentrations, the
0.5 OD_{260} units of polysomes showed the maximum incoporation of
cpm into acid insoluble protein. 50 μl samples were removed from
the translation mixes at each time point, precipitated in trichloro-
acetic acid, and collected on glass fibre filters prior to scintillation
counting.

(b) Incorporation of ^{3}H-leucine into protein during *in vitro* transla-
of fractionated heavy polysomes.

This figure shows the incorporation of ^{3}H-leucine into acid
insoluble protein as directed by the heavy polysome fractions
2-6 in a homologous cell free system. The cpm values shown are
the means of those presented in Table 2.2. On the basis of the
evidence presented in (a), 0.5 OD_{260} units of polysomes were
translated per 1 ml total mix throughout.

a. Translation of Total Polysomes
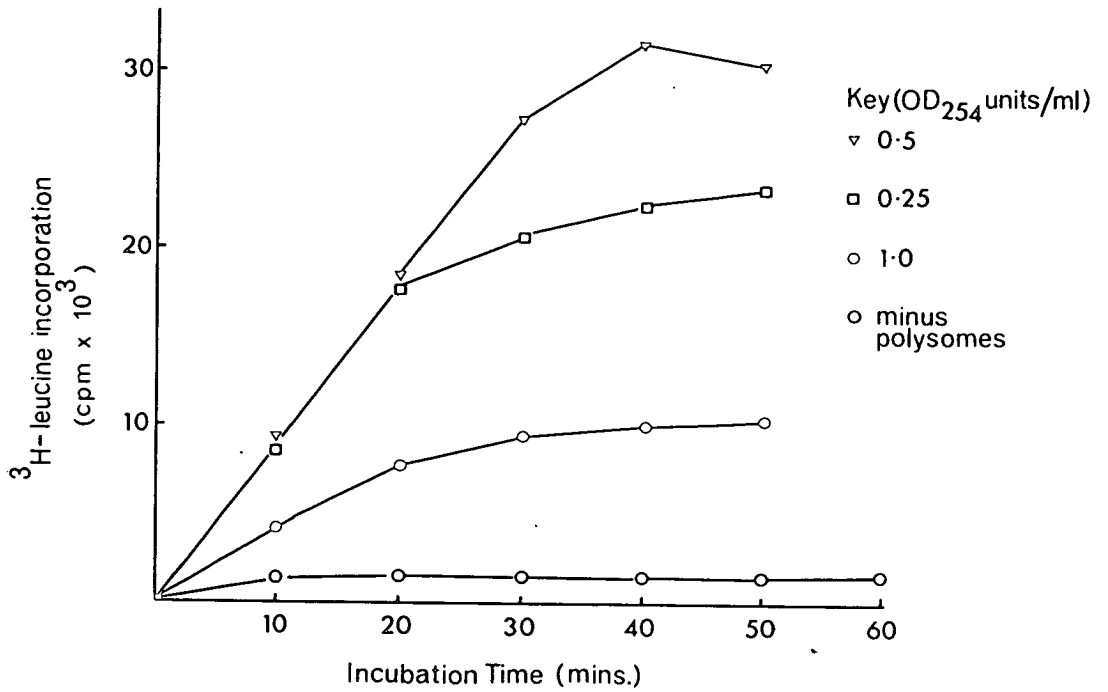
b. Translation of Heavy Polysomes
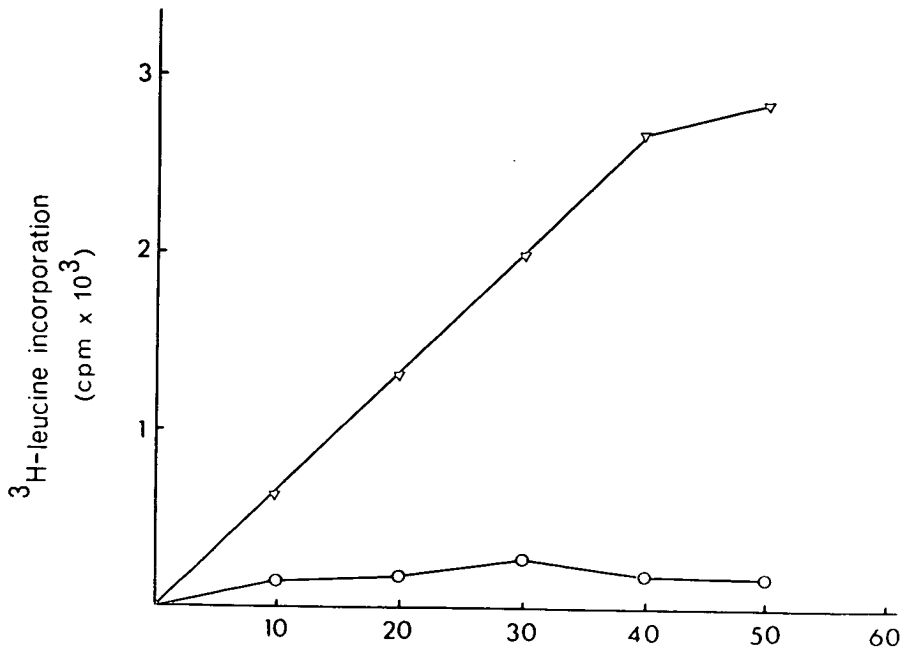
2.2.5 *Oligo dT-cellulose Chromatography of RNA Samples*
(Aviv and Leder, 1972)

There is considerable evidence that most mRNA species carry a polyadenylated tail of about 150 nucleotides at their 3' ends (Adesnik *et al.*, 1972; Greenberg and Perry, 1972; Molloy *et al.*, 1972) which is added post-transcriptionally. This feature forms the basis of the chromatography technique developed by Aviv and Leder (1972) for the separation of mRNA from all other cellular RNA types. Oligo thymidylic acid residues are linked to an insoluble support medium such as cellulose so that when an RNA solution with a high salt concentration is passed over the cellulose any RNA molecules with long stretches of polyA are retained. The A-T base pairing can then be disrupted by lowering the salt concentration when an enriched mRNA sample is eluted.

Polysomal RNA samples were dissolved at a concentration of 1 mg/ml in a small volume (100-200 μl) of 1 mM EDTA pH 7.0 which was then heated to 70°C for 30 sec to ensure disruption of any internal base pairing. The samples were cooled immediately on ice. An equal volume of 2 x concentrated binding buffer was added and the RNA concentration was adjusted to 100 μg/ml with 1 x binding buffer. This was layered on to a 500 mg oligo dT-cellulose bed equilibrated with the same buffer, mixed gently with the cellulose and left for 15 min to allow annealing to take place. The sample was eluted slowly and re-passed over the column twice more. The column was then washed to a background OD_{260} with binding buffer (2-2½ hours) and the RNA retained by the column at this stage was regarded as $polyA^+$ RNA. This was eluted from the column using 10 ml of eluting buffer. 0.5 ml fractions were collected and those with significant OD_{260} values were pooled as shown in Figure 2.3. The RNA was precipitated as described previously in Section 2.2.4.
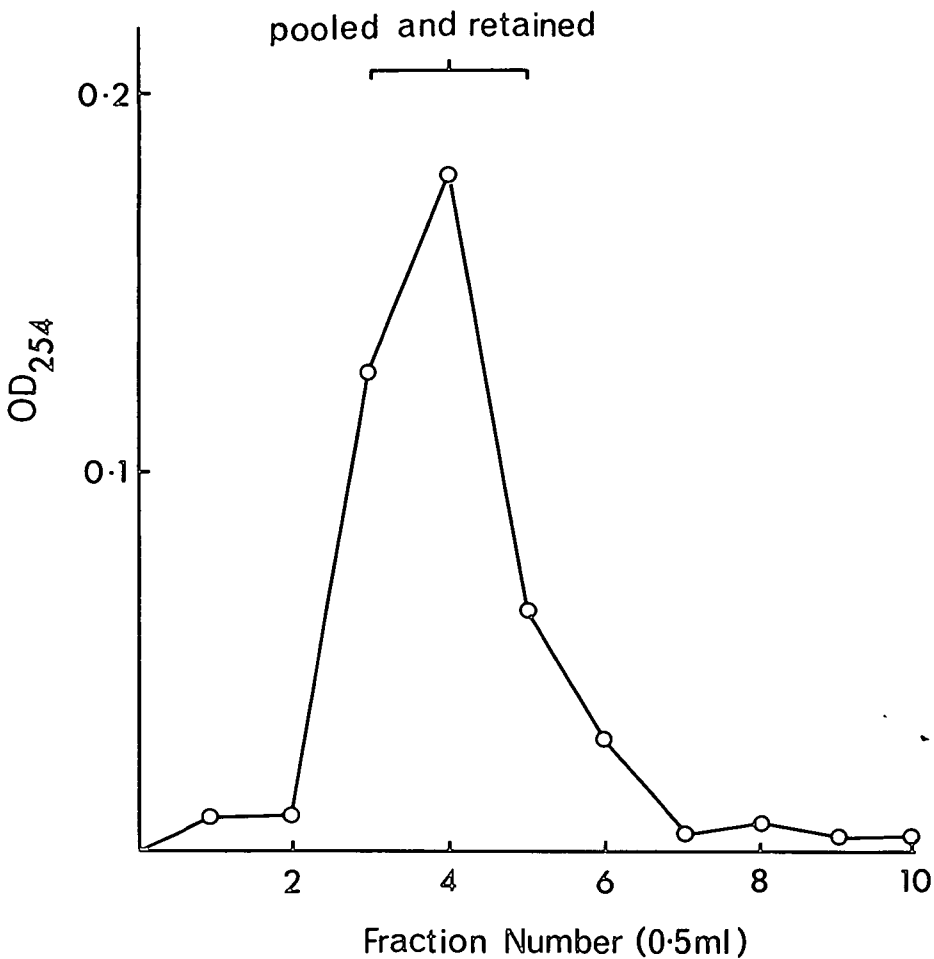
FIGURE 2.3: Elution Profile of f2 mRNA from an Oligo dT-Cellulose
Chromatography Column.

Oligo dT-cellulose chromatography of total and polysomal RNA
samples was carried out as described in Section 2.2.5. The polyA$^+$
fraction of the RNA sample is bound to the cellulose in high salt
buffer, and then eluted by lowering the salt concentration. This
elution profile was obtained when the f2 polysomal RNA was
fractionated as described. The polyA$^+$ RNA eluted in a single,
sharp peak from the oligo dT-cellulose column. Fractions 3, 4 and
5 were pooled and mRNA obtained by ethanol precipitation at -20°C.

Occasionally, the polyA$^+$ containing fractions were repassed over the column but this made little difference to the translational capacity of the RNA but significantly reduced the final yield so the procedure was not carried out routinely.

2.2.6 Poly-U Hybridisation to Determine mRNA Concentration

This method (Gillespie et al., 1972) is based on the ability of the polyA tail of the mRNA molecule to form a 1:1 hybrid with ^{3}H-polyU. Determination of the radioactivity (in counts per minute) which is ribonuclease resistant after a suitable hybridisation time gives an accurate estimate of the mRNA concentration in any particular sample when compared with the standard curve (Figure 2.4).

Incubation mixes contained:

2 µl	^{3}H-poly rU (17 Ci/mmol)
20 µl	polyA$^+$ RNA (0.5 µg) or poly dA at concentrations ranging from 75 ng/ml to 1.5 µg/ml
10 µl	10 x SSC
20 µl	water

Mixes were incubated at 37°C for 20 min and chilled for 10 min. 1 ml 2 x SSC was added with 2 µl RNase. Samples were gently mixed and further incubated at 0°C for 20 min. 10 µl of 10 mg/ml BSA was added as carrier with 1 ml ice cold 20% trichloroacetic acid and after chilling for a further 10 min precipitated hybrids were collected on GF/C filter discs. The radioactivity was determined by scintillation counting in toluene PPO POPOP scintillant.

The amounts of mRNA obtained from each polysome fraction following oligo dT-cellulose chromatography were acurately determined using this poly-U hybridisation method. The standard curve shown in Figure 2.4
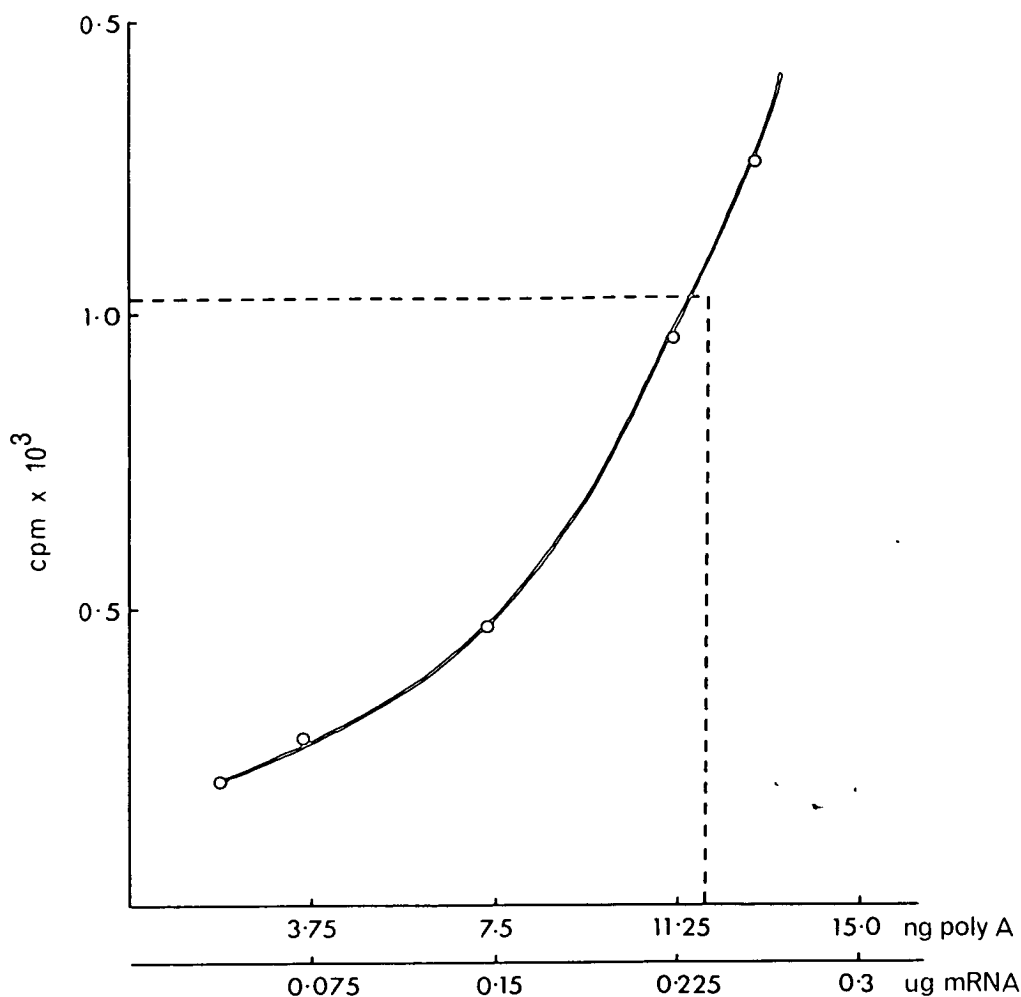
FIGURE 2.4: Calibration Curve of ^{3}H-polyU Hybridisation with polydA Used for Estimations of mRNA Concentrations.

The concentration of mRNA was determined accurately by hybridising the polyA tail of the messenger to ^{3}H-polyU as described in Section 2.2.6. The calibration curve, shown here, was firstly prepared by hybridising known amounts of polydA (1.875 ng to 13.125 ng) with ^{3}H-polyU and then determining the amount of radioactivity (measured as cpm) remaining in double stranded form following ribonuclease treatment. The dashed line (----) indicates the result obtained when the f2 mRNA concentration was determined in this way: 1060 cpm were recovered in double stranded form following hybridisation to ^{3}H-polyU. This corresponds to 11.875 ng polyA$^+$, which, assuming a 5% polyA content in messenger RNA, gives an mRNA value of 0.2⊂⊃ μg. Allowing for a 1 : 20 dilution of the original mRNA preparation, the total recovery of polyA$^+$ RNA from the f2 polysome fraction (30 gradients) was 4.75 μg.

TABLE 2.3:   Messenger RNA Yields from Polysome Fraction f2.

Initial amount of tissue (14-day chick embryo skeletal leg muscle).	25 g	(layered onto 30 gradients)
Total RNA extracted from polysome fraction f2.	150 µg	
Poly(A$^+$) RNA recovered by oligo-dT chromatography.	7.3 µg*	
mRNA recovered as determined by polyU/polydA hybridisation.	4.75 µg*	

The yields of RNA at each stage in the extraction process are shown for the f2 heavy polysome fraction.

* The significant difference between these two values indicates that the RNA recovered from the oligo-dT cellulose column chromatography step is not 100% poly(A$^+$) RNA and is, therefore, probably highly contaminated with ribosomal RNA species.

was constructed from a series of known concentrations of poly-dA

included in the hybridisation reactions, plotted against the cpm retained

by the filter following RNase treatment to remove unhybridised polyU.

It was then possible, assuming a 5% polyA content in mRNA, to estimate

the concentration of mRNA in a given sample with reasonable accuracy.

The values obtained for the polysome fraction f2 is indicated in Figure

2.4, and is compared with the value obtained from optical density

measurements in Table 2.3.

2.2.7 *Cell Free Translation of mRNA*

mRNA translation was performed in two heterologous cell free

systems. Early in the project, the mRNA samples were translated

routinely in a wheat germ system using ^{3}H-labelled amino acids. However,

in later experiments, the rabbit reticulocyte lysate system became the

method of choice as reliable preparations became commercially available

and higher radioactive incorporation into protein was achieved, as shown

in Figure 2.7, particularly if ^{35}S-methionine was included as the labelled

amino acid, thus reducing the time required for autoradiography.

Wheat germ translation system (Roberts and Paterson, 1973)
Preparation of S-30 fraction (Marcu and Ducock, 1974)

5 g of wheat germ (General Mills) was ground with 5 g ground glass

in a cold mortar for a few minutes. 13 ml of grinding buffer were added

and the paste was centrifuged at 30,000 g for 15 min. The supernatant was

passed over a sephadex G25 coarse column and eluted at 3 ml/min. 1 ml

fractions were collected and those with an OD_{650} greater than 0.5/5 μl were

pooled and centrifuged for 20 min at 30,000 g. The supernatant was dialysed

against grinding buffer for 16 hours and then centrifuged again at 30,000 g

for 20 min. The final supernatant was frozen instantly by gently pipetting

drops into liquid nitrogen. The frozen spheres were stored at -70°C.

Translation:

Each translation mix contained in a total volume of 50 µl:

20	µl	S-30
0.02 M		Hepes pH 7.6
0.16 M		potassium chloride (see Figure 2.5)
2.8	mM	magnesium acetate
60	µM	spermine
1	mM	ATP
200	µM	GTP
8	mM	creatine phosphate
40	µg/ml	creatine phosphokinase
2	mM	dithiothreitol
25	µM	each of 19 amino acids (excluding leucine)
50	µCi	^{3}H-leucine at 1 µCi/µl (specific activity 187 Ci/mmol)

1 µg of the mRNA sample was added in 1 µl sterile water to initiate the reaction. Reaction mixes were incubated at 25°C for 90 min. Aliquots of 5 µl were spotted onto Whatman GF/C filter discs, dried and placed in 10% trichloroacetic acid containing 5 mM unlabelled leucine for 30 min. The filters were then heated at 90°C for 15 min in 5% trichloroacetic acid, rinsed in ice cold 5% trichloroacetic acid, ethanol, ethanol : ether (3 : 1) and, finally, ether. Incorporation of radioactivity into acid insoluble protein was determined by scintillation counting.

It has been reported (Roberts and Paterson, 1973) that the efficiency of the wheat germ system varies for different mRNA types, depending on the potassium ion concentration. 1 µg of polyA$^+$ RNA from the heavy polysome fractions was translated with potassium chloride
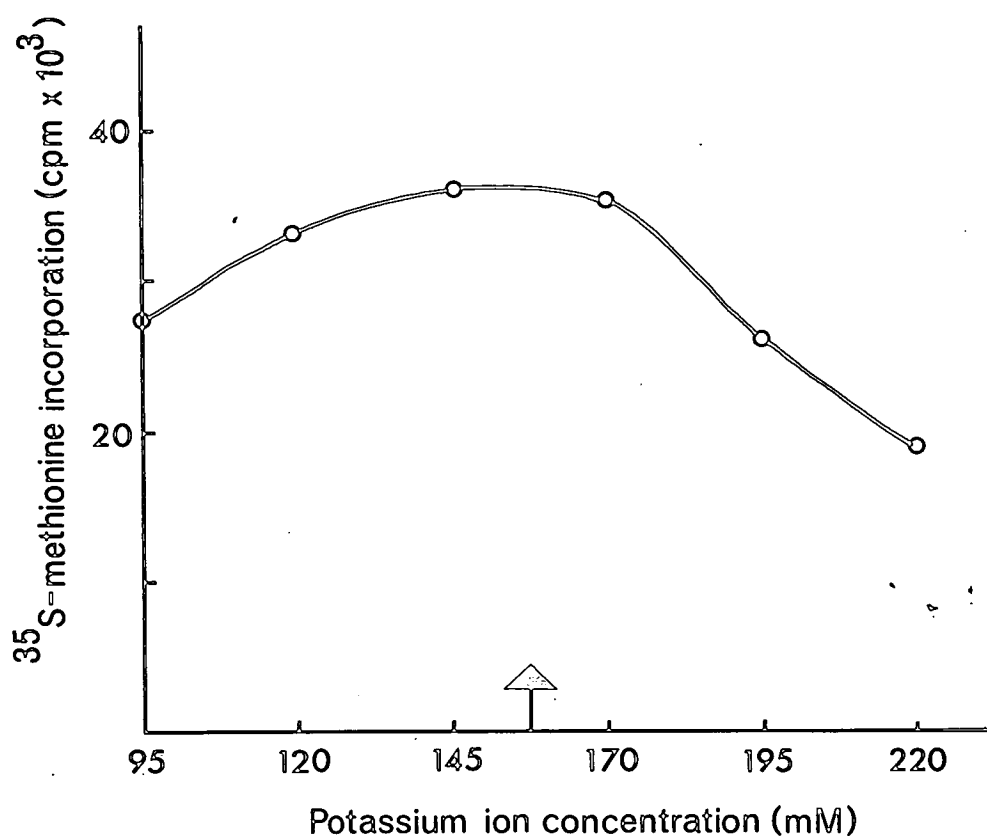
FIGURE 2.5: Effect of Changes in Potassium Ion Concentration on
the Translation of Myosin Heavy Chain mRNA (f2)
in Wheat Germ Cell Free System.

The *in vitro* translation of mRNA was carried out using the wheat germ
cell free system as described in Section 2.2.7. The incorporation of
radioactive amino acids into acid insoluble protein by this system
varies with the potassium ion concentration, as shown in this figure.
Concentrations of potassium chloride, varying from 95 mM to 220 mM
were tested and it was found that f2 mRNA translates most efficiently
with a potassium ion concentration of 160 mM, and therefore this
concentration was used routinely.

concentrations varying from 95 - 220 mM. A concentration of 160 mM
was shown to stimulate the maximum incorporation of radioactivity into
protein when all other conditions were kept constant as shown in
Figure 2.5.

The translation products were analysed by SDS-polyacrylamide
gel electrophoresis as described in Chapter III.

The rabbit reticulocyte lysate translation system:

This was purchased from the Radiochemical Centre. It is prepared
by a modification of the method described by Pelham and Jackson (1976)
in which endogenous mRNA is destroyed by treatment with a calcium
activated nuclease under carefully controlled conditions, and when
incubated with an optimal amount of exogenous mRNA is capable of
maintaining a high level of protein synthesis for over an hour. The
lysate contains:

1	mM	calcium ions
2	mM	EGTA
0.02	mM	hemin
2	mM	dithiothreitol
10	mM	creatine phosphate
50	μg/ml	creatine kinase
100	mM	potassium acetate
1.6	mM	magnesium acetate

and all other factors necessary for the mRNA directed
incorporation of labelled amino acids into protein.

The rabbit reticulocyte lysate also contains an endogenous pool
of 20 essential amino acids. When a labelled amino acid is added its
specific activity will be reduced by the cold amino acid present in this

pool. This effect is negligible for an amino acid with a low endogenous pool size, such as L-methionine, and small for L-leucine but can be considerable for an amino acid such as glycine or alanine for which the endogenous pool is rich.

Routine translation assays were carried out at 30°C for 1 hour using a mix containing 80% lysate and 20% labelled amino acid by volume. The final radioactive concentration of either ^{3}H-leucine (specific activity 187 Ci/mmol) or ^{35}S-methionine (specific activity 1050 Ci/mmol) was 1 μCi/μl. The reactions were initiated by adding mRNA in a volume not exceeding 10% of the assay mix used. This was normally 0.75 μg in 1.5 μl for a 20 μl assay mix. The optimum amount of mRNA to use in each translation was determined using mRNA extracted from total poly-somes as shown in Figure 2.6(c). Similarly, the potassium and magnesium ion concentrations were shown to considerably affect the protein synthes-ising ability of mRNA extracted from skeletal muscle polysomes as shown in Figures 2.6(a) and 2.6(b), respectively, where various concentrations were tested in an attempt to determine optimum conditions. This was necessary as it has been reported (Sonenshein and Braweriman, 1977; Tse and Taylor, 1977) that the number of incomplete polypeptide chains may increase if the concentrations of these ions are sub-optimal.

During the incubation period, 2 μl samples were removed from the reaction mixes at intervals into 0.5 ml of 1N NaOH, 5% hydrogen peroxide and kept at room temperature for 10 min. This completely decolourised the sample. Proteins were precipitated by adding 3 ml of ice cold 25% trichloroacetic acid containing 2% casein hydrolysate as a carrier. The incorporated radioactivity was collected onto pre-wetted GF/C filter discs, washed with 8% trichloroacetic acid, dried and counted by liquid scintillation.

FIGURE 2.6: Effect of Changes in Potassium Ion, Magnesium Ion, and Messenger RNA Concentration on the Translation of Myosin Heavy Chain (f2) mRNA in the Rabbit Reticulocyte Lysate System.

The *in vitro* translation of mRNA was carried out using the rabbit reticulocyte lysate cell free system, as described in Section 2.2.7. The optimum conditions for the efficient translation of f2 mRNA were determined, as recommended by the manufacturers, with respect to potassium ion, magnesium ion and mRNA concentration.

(a) shows that the maximum incorporation of radioactive methionine into protein occurs at a potassium acetate concentration of 150 mM. This required the addition of potassium acetate to 50 mM to raise the endogenous concentration of 100 mM to the necessary level.

(b) shows that the addition of magnesium acetate reduces the translational efficiency of the mRNA. The highest level of incorporation of ^{35}S-methionine into protein was achieved with no addition of extra magnesium, the endogenous concentration in the lysate being 1.6 mM.

(c) shows the effect of increasing amounts of mRNA on *in vitro* translation in the rabbit reticulocyte lysate system. In the case of f2 mRNA, translational efficiency increased sharply up to 1 µg of RNA per 22 µl translation mix, but then tended to be inhibited slightly if the RNA concentration exceeded 1.5 µg per 22 µl.

a. Potassium ion concentration

added K^+ (mM)

final K^+ (mM)

b Magnesium ion concentration

Mg^{2+} conc. (mM)
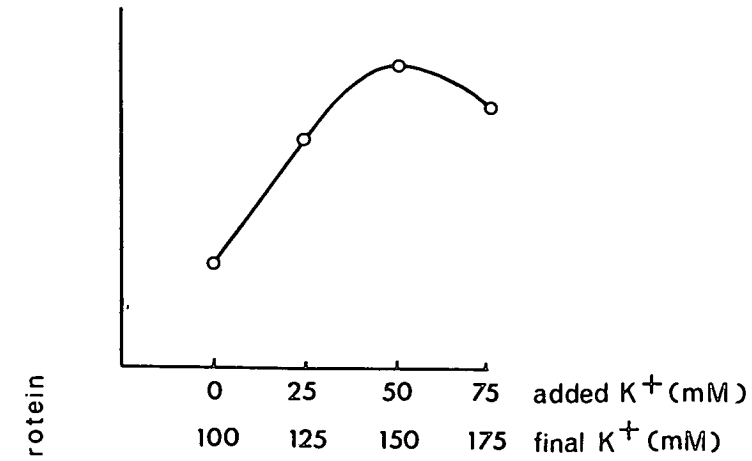
^{35}S – methionine incorporation into protein
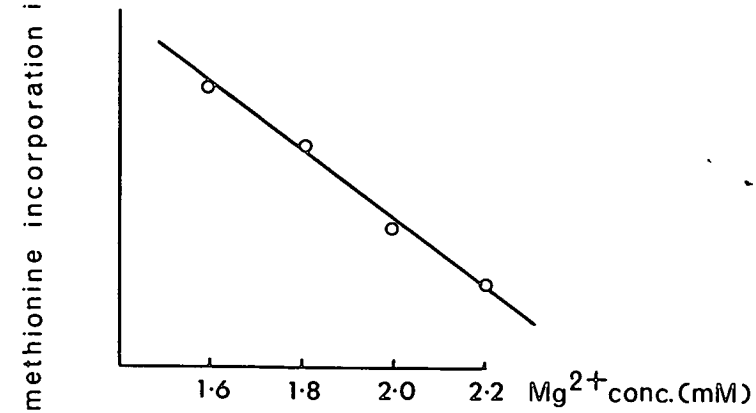
c. mRNA concentration
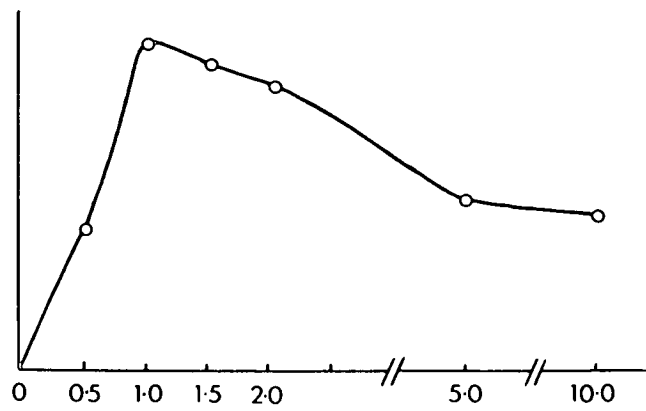
FIGURE 2.7:   Incorporation of Radioactively Labelled Amino Acids
into Protein, Directed by f2 mRNA in the Rabbit
Reticulocyte Lysate and Wheat Germ Cell Free Systems.

The incorporation of radioactively labelled amino acids into TCA
precipitable protein, directed by f2 mRNA in two cell free
systems. ^{35}S-methionine was included as the labelled amino
acid when f2 mRNA was translated in the rabbit reticulocyte
lysate system as shown by the solid line (———). In a total
volume of 22 μl, the translation mix included 150 mM potassium
acetate, 1.6 mM magnesium acetate and 1 μg of mRNA. Generally,
incubations were terminated at 60 min due to the apparent
protease activity which arises during longer incubation periods.

^{3}H-leucine was the labelled amino acid used during mRNA
translation in the wheat germ system, as shown by the dashed
line (----). In a total volume of 50 μl, the potassium chloride
concentration was 160 mM, and included 1 μg of f2 mRNA.
Incubations were continued for 90 min as there was no detectable
breakdown of newly synthesised proteins prior to this.
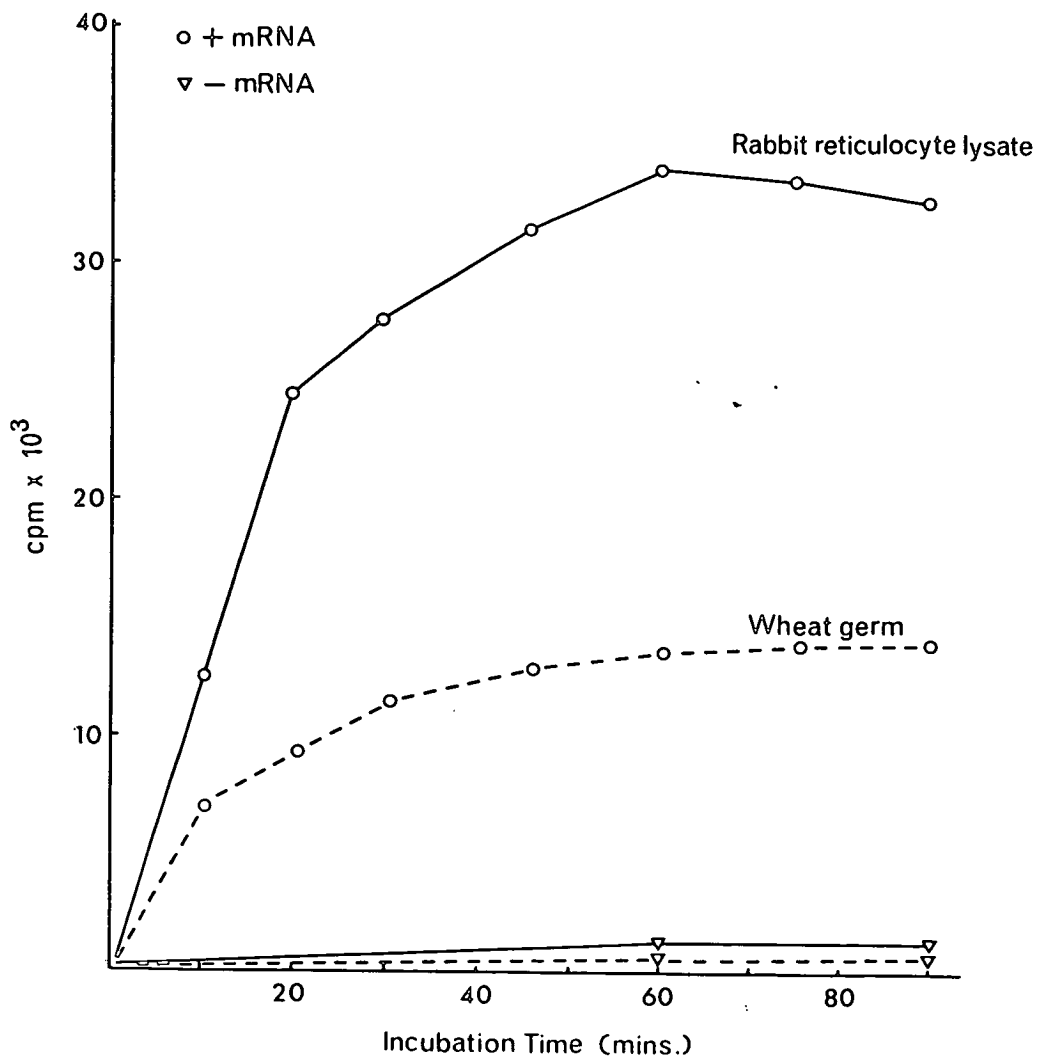
Following *in vitro* translation of mRNA samples in both the rabbit reticulocyte lysate and wheat germ systems, samples were retained for SDS-polyacrylamide gel electrophoretic analysis of the newly synthesised products. The ability of the mRNA preparation to direct the synthesis, in these *in vitro* systems, of only one protein, which was shown to co-electrophorese with a myosin heavy chain marker in SDS-polyacrylamide gels, was the main criterion used to assess the purity of the messenger preparation.

As is described in detail in Chapter III, the mRNA obtained from the second fraction (f2) of the polysome gradients fulfilled this criterion.

2.2.8 *Preparation of Complementary DNA*

A specific hybridisation probe, to be used in the screening of the recombinant library was prepared by the reverse transcription of the purified myosin heavy chain f2 mRNA preparation.

Reverse transcription is a natural process, common to RNA viruses which initiate their replication within the host cell by producing a DNA copy of the RNA genome which is subsequently used as a template for the synthesis of many RNA copies. Purified avian myeloblastosis virus reverse transcriptase is able to direct the synthesis of a DNA copy of eukaryotic mRNA molecules under appropriate conditions, and by the simple inclusion of one or more labelled deoxynucleotide triphosphates, the cDNA copy will be radioactive.

Reaction mixes contained in 100 µl:

50	mM	Tris-HCl pH 8.3
60	mM	sodium chloride
6	mM	magnesium chloride
4	mM	dithiothreitol
0.5	mM	each of dGTP, dATP, dCTP

100	μCi	^{3}H-dCTP (specific activity 21 Ci/mmol)
or 50	μCi	α-^{32}P-dCTP (specific activity 800 Ci/mmol)
100	μg/ml	oligo dT
5	μg/ml	mRNA
600	units/ml	AMV reverse transcriptase

The radioactive and non-radioactive deoxynucleotide triphosphates were freeze dried completely and the remaining ingredients of the reaction were added. The mRNA was pre-incubated with the oligo dT primer for 2 min at 37°C before adding to the cocktail. The oligo dT serves to initiate the reaction by hybridising to the polyA tail of the mRNA. The reaction will not proceed without this step, and therefore because other cellular RNA's lack a polyA tail, any minor contaminants of the mRNA (particularly ribosomal RNA) will not participate in the reverse transcription.

Incubation was at 40°C for 1 hour after which the reaction was terminated with the addition of 400 μl 10 mM EDTA and 600 μl 0.5M potassium hydroxide to hydrolyse the mRNA. Hydrolysis was completed by heating at 90°C for 10 min, and the mix was then neutralised with 600 μl 1N sodium dihydrogen orthophosphate. 5 μg of bacteriophage λ DNA was added as carrier and the mix was applied to a sephadex G100 column (10 ml bed volume). 0.75 ml fractions were collected in 3 mM sodium chloride, 0.01 M Tris-HCl pH 8.0, 0.01 M EDTA and monitored by counting 10 μl samples. The cDNA was contained in the first peak of radioactivity to elute from the column as shown in Figure 2.8. The second peak corresponded to free deoxycytidine triphosphate since this was not precipitable with TCA. The size of the cDNA was estimated following agarose gel electrophoresis as described in Chapter III.
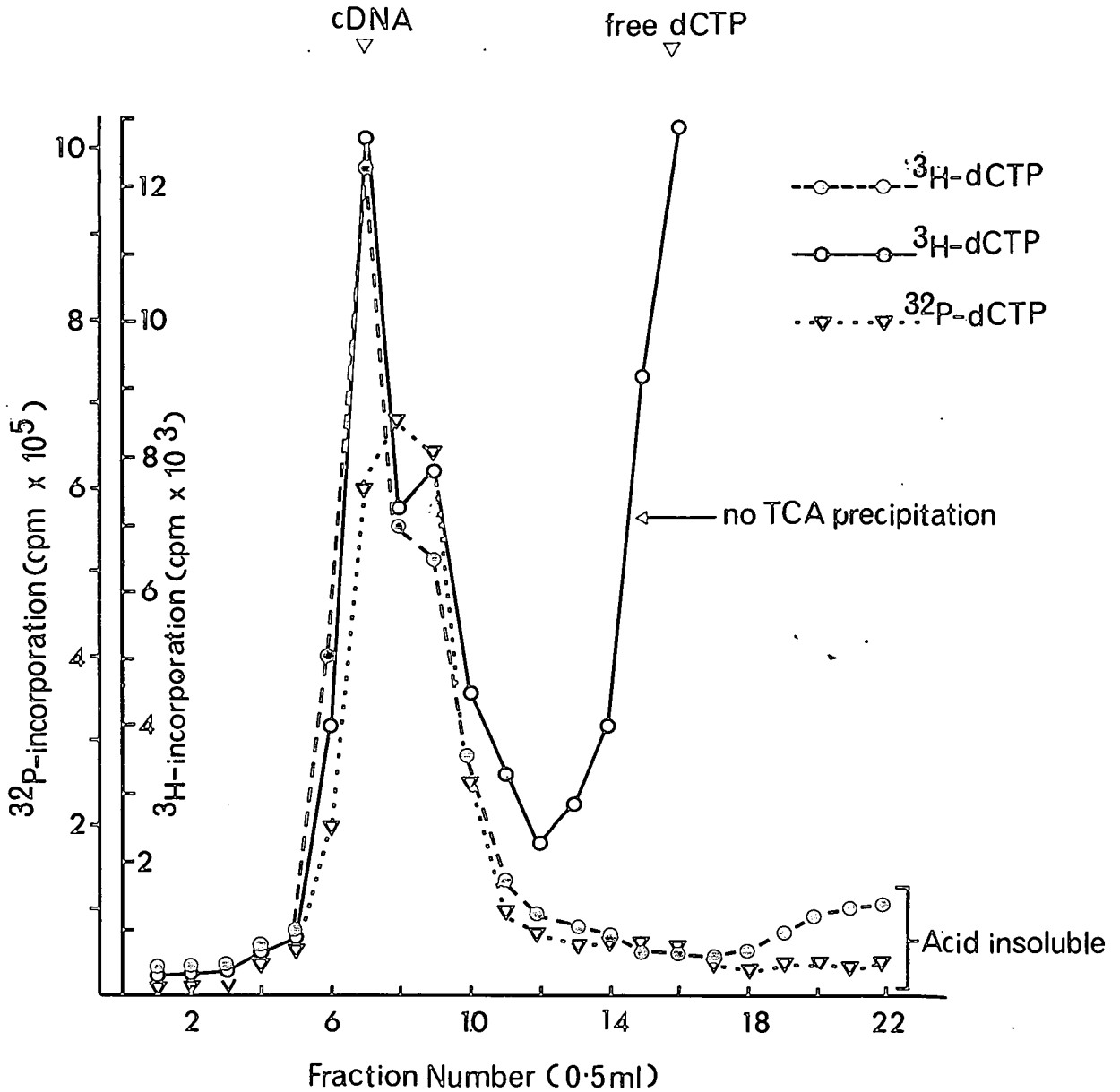
FIGURE 2.8: Isolation of cDNA by Sephadex G100 Chromatography.

Reverse transcription of f2 mRNA was carried out, as described in
Section 2.2.8, using 0.5 µg of the mRNA as template and either
100 µCi ^{3}H-dCTP or 50 µCi ^{32}P-dCTP as the radioactive deoxy-
nucleotide. The cDNA was separated from free deoxynucleotide
by Sephadex G100 chromatography. 0.75 ml fractions were collected
and 10 µl samples were allowed to precipitate in 2 ml ice cold 20%
TCA. The samples were collected on glass fibre filters and counted.
The peak fractions which contain TCA precipitable material were
pooled, freeze dried and taken up in water to give 2000 cpm/µl.

2.2.9 *Purity of the Complementary DNA Probe*

Single-stranded DNA is able to hybridise and form duplexes
very precisely with either complementary RNA or complementary DNA
molecules in solution. Such hybridisation depends not only on the
extent of the complementary sequences within the two populations, but
also on the collision reactions between them, and therefore follows simple
kinetics based on nucleic acid concentration and the time allowed for
annealing. Hybridisation between RNA and DNA, where the RNA is
in excess and therefore drives the reaction, is measured in terms of
$R_0 t$ values where R_0 is the initial RNA concentration expressed in moles/l
and t is the annealing time expressed in seconds. When mRNA drives
the hybridisation reaction to complementary DNA from the same source,
transitions of hybridisation occur corresponding to the abundance classes
of the mRNA molecules within the population. Such reactions have thus
been used for measuring the complexity of RNA populations (Bishop
et al., 1974). These homologous reactions will theoretically proceed to
100% hybridisation of the cDNA, but the $R_0 t$ required for this will
depend on the frequency of the lowest abundance class which may require
very extended time periods or high mRNA concentrations to be able to
anneal. The $R_0 t_{\frac{1}{2}}$ is the $R_0 t$ value at which 50% of the molecules within
a particular abundance class have annealed and is a measure of the
sequence complexity of that particular abundance group.

Therefore, hybridisation of a cDNA back to its purified mRNA is
a measure of the faithfulness of the copy. The number of minor transi-
tions as hybridisation approaches 100% will indicate the extent of
contamination of the mRNA by other species and the $R_0 t_{\frac{1}{2}}$ of both major
and minor transitions will provide an indication of the number of dif-
ferent species within these abundance classes. Therefore, the MHC

cDNA was hybridised back to its mRNA template as a method of obtaining an indication of the purity of the probe.

Hybridisation:

MHC cDNA was hybridised to an excess of MHC mRNA using mRNA at two different concentrations: 1 µg/ml and 20 µg/ml.

The mRNA and cDNA (approximately 30,000 cpm) were mixed in a total volume of 200 µl of 0.012M sodium phosphate pH 7.0 in a small siliconised tube. The mixture was heated for 3 min in a boiling water bath and a 20 µl sample was taken immediately and diluted into ice cold stop buffer. The hybridisation mixture was transfered to a 70°C bath and the salt concentration was increased to 0.24M PEB by the addition of 20 µl of 2.3M PEB. Following removal of the 2 min sample, liquid paraffin was layered over the reaction mixes to prevent evaporation. 20 µl aliquots were withdrawn at various times from 1 min to 5 hours following incubation at 70°C (see Table 2.4) and diluted into 380 µl of ice cold stop buffer.

The exent of cDNA annealing was measured by resistance to S1 nuclease which digests single-stranded nucleic acids but leaves duplexes intact.

S1 nuclease digestion:

All hybridisation samples were rapidly frozen at -20°C so that all those for a given experiment could be treated at the same time with S1 nuclease. One-tenth volume of concentrated S1 buffer was added to each sample and each was divided into 2 x 200 µl aliquots, one of which was treated with 10 µl (100 units) of S1 nuclease. The final reaction mix consisted of 48 mM PEB, 0.2 mM EDTA, 20 mM sodium chloride, 30 mM sodium acetate pH 4.5, 0.6 mM zinc sulphate, 0.052 N acetic acid.

Incubation was at 45°C for 40 min following which the samples were placed on ice and 200 μg of BSA and 2.5 ml of 10% trichloroacetic acid were added. They were allowed to stand for 30 min at 0°C and then collected on Millipore GF/C filters, washed with 5% TCA, dried and counted in toluene PPO-POPOP scintillant.

The ratio of the cpm value of the nuclease treated sample to the untreated sample, after subtraction of the S1 resistant background determined immediately after the denaturation step is expressed as a percentage in Table 2.4 and plotted against R_0t in Figure 2.9. In the calculation of the R_0t values, the nucleotide molecular weight was taken as 320 daltons.

A single transition of hybridisation between the cDNA probe and its template mRNA is shown in Figure 2.9, indicating the presence of a single abundance group of mRNA molecules. The transition is sharp and the S1 nuclease resistance is greater than 85%, indicating that the hybridisation is almost complete. Taken together, these observations indicate that the MHC mRNA is not contaminated to any great extent with other RNA species which are able to interfere in the reverse transcription, and thus further indicate that the cDNA is a faithful copy.

The $R_0t_{\frac{1}{2}}$ value of 7.1×10^{-3} (log R_0t - 2.15) is in close agreement, though slightly higher, than two previously determined values of 5.3×10^{-3} (Robbins and Heywood, 1976) and 6.3×10^{-3} (Patrinou-Georgoulas and John, 1977).

Patrinou-Georgoulas and John (1977) showed that highly purified globin mRNA hybridised to its cDNA with a similar sharp transition approaching 100% hybridisation with an $R_0t_{\frac{1}{2}}$ of 5.3×10^{-4}. Using the molecular weight value of globin mRNA as determined by Bishop *et al.* (1974) of 4×10^5, and of myosin heavy chain mRNA of 2×10^6 as

TABLE 2.4: Hybridisation of Muosin Heavy Chain mRNA to
Its Homologous cDNA.

Expt	RNA concentration (R_o) mol/l	Time (t) sec	$R_o t$ mol sec/l	Log $R_o t$	% S1 nuclease resistance
1	3.2×10^{-6}	60	1.92×10^{-4}	-3.72	11.2
	$(=1\mu g/ml)$	120	3.84×10^{-4}	-3.42	12.7
		600	1.92×10^{-4}	-2.72	27.9
		1.2×10^3	3.84×10^{-4}	-2.42	41.5
		1.8×10^3	5.76×10^{-3}	-2.24	44.8
		2.7×10^3	8.64×10^{-3}	-2.06	54.1
		3.6×10^3	1.15×10^{-2}	-1.94	59.6
		7.2×10^3	2.30×10^{-2}	-1.64	65.8
2	6.25×10^{-5}	60	3.75×10^{-3}	-2.43	31.5
	$(=20\mu g/ml)$	120	7.50×10^{-3}	-2.12	46.2
		300	1.88×10^{-2}	-1.73	63.2
		600	3.75×10^{-2}	-1.43	73.7
		1.5×10^3	9.38×10^{-2}	-1.03	77.5
		3.0×10^3	1.88×10^{-1}	-0.73	84.2
		1.08×10^4	6.75×10^{-1}	-0.17	85.0
		1.80×10^4	1.13	-0.05	86.6

This table shows the detailed data which is expressed in Figure
2.9 as a $R_o t$ curve. Two hybridisation experiments were carried
out, as described in Section 2.2.9, between f2 mRNA and its cDNA
(in RNA excess) to allow a wide $R_o t$ range to be covered. It was
necessary to use two different initial concentrations of mRNA (R_o)
to achieve this of 1 $\mu g/ml$ and 20 $\mu g/ml$.
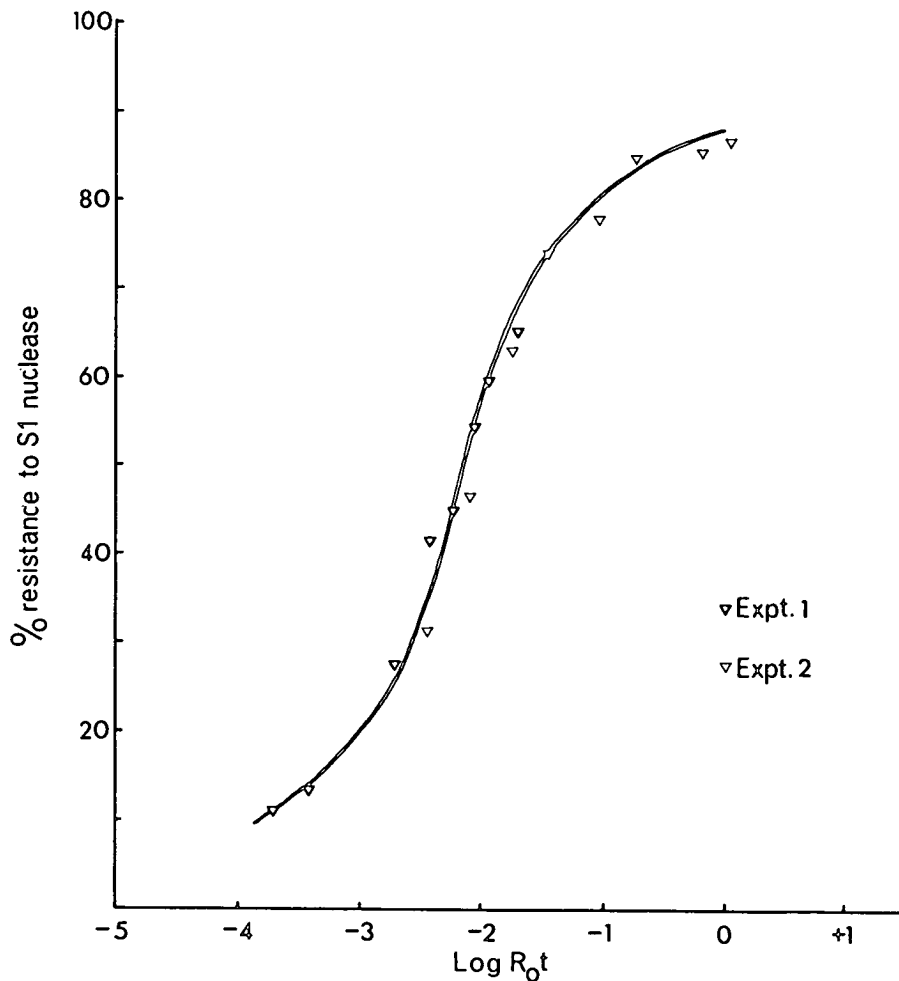
FIGURE 2.9: $R_o t$ Curve Showing the Hybridisation of Myosin Heavy
Chain mRNA to its cDNA.

Myosin heavy chain mRNA, isolated from heavy polysome fraction f2,
was hybridised in RNA excess to its complementary DNA as
described in Section 2.2.9. A single transition of hybridisation
is evident, indicating the presence of a single abundance group of
mRNA molecules. The transition is sharp, and hybridisation of
the available cDNA is almost complete since the S1 resistance is
greater than 85%. The $R_o t_{\frac{1}{2}}$ of 7.1 x 10^{-3} of the transition (cor-
responding to a log $R_o t$ of $^{\frac{1}{2}}$-2.15) is high when compared with
standard estimations for purified globin mRNA (Bishop *et al.*, 1974),
and therefore suggests the presence of two or three different myosin
heavy chain sequences, present in similar numbers.

determined by Sarkar *et al.* (1973), they calculated the $R_0 t_{\frac{1}{2}}$ for myosin heavy chain mRNA/cDNA hybridisation to be 2.7×10^{-3} which is approximately half their experimental value and therefore concluded that there were two myosin heavy chain sequences within their preparation. From the observed $R_0 t_{\frac{1}{2}}$ value, shown in Figure 2.9, a similar situation would appear to exist in the probe preparation described here as the $R_0 t_{\frac{1}{2}}$ value suggests the sequence complexity is greater than would be expected for a single species.

2.3 Nick Translation as a Method of Synthesising Hybridisation Probes

The nick translation reaction (Kelly *et al.*, 1970; Rigby *et al.*, 1977), catalysed by *Escherichia coli* DNA polymerase I (Rigby *et al.*, 1977; Maniatis *et al.*, 1975a; Mackey *et al.*, 1977), is used to replace existing unlabelled nucleotides in DNA with radioactive ones, and in this way DNA which has been purified and stored in an unlabelled form can be labelled to a high specific activity *in vitro*, as shown diagrammatically in Figure 2.10. The quality and specific activity of labelled DNA prepared in this way depends on several factors including the quality of the enzymes used, the extent to which the existing nucleotides are replaced and the specific activity of the newly introduced nucleotides. Very high specific activities can be achieved which is particularly important since the sensitivity of nucleic acid hybridisation techniques depends largely on the specific activity of the labelled probe. For instance, a 200-fold increase in the specific activity of a probe can permit a 200-fold reduction in the amount of an unlabelled DNA which is being 'probed' (Rigby *et al.*, 1977). This is an important consideration when probing a genomic library for "single copy" gene
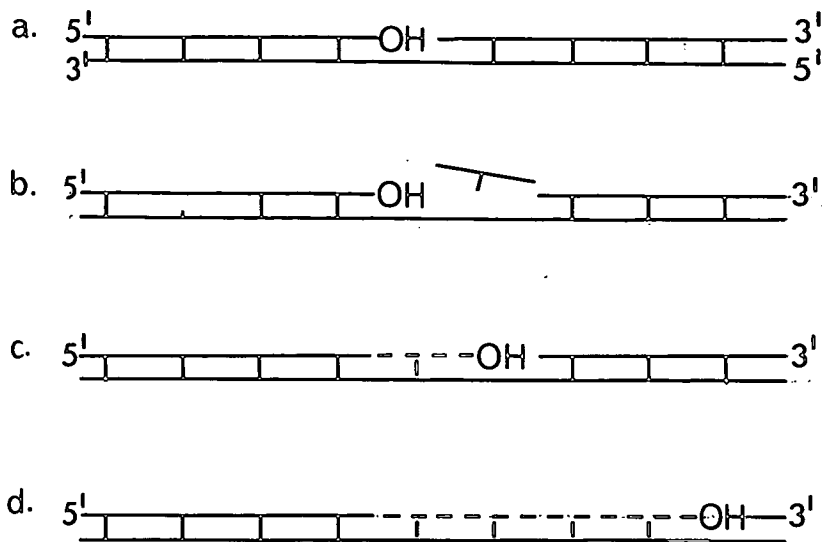
FIGURE 2.10: Diagrammatic Representation of the Labelling of
Double Stranded DNA by Nick Translation.

Double stranded DNA can be radioactively labelled *in vitro* by nick
translation, as described in detail in Section 2.3, and shown
diagrammatically in this figure.

(a)   Single stranded nicks are introduced randomly in the DNA
      by the action of DNaseI.

(b)   Nucleotides on the 5'-side of the nick are removed due to the
      exonuclease activity of DNA polymerase I.

(c)   Labelled nucleotides are added to the 3'-OH side of the nick,
      also by the action of DNA polymerase I, to replace those
      which have been removed.

(d)   The nick is 'translated' in a 5' to 3' direction along the DNA
      chain.

sequences but is not so important here where the nick translated probe
will be used to search for the repetitive sequences of the ribosomal
RNA genes.

The mechanisms of nick translation have been the subject of much
investigation, particularly by Kornberg, the discoverer of *E. coli*
DNA polymerase I. In all published procedures (Maniatis *et al.*, 1975;
Mackey *et al.*, 1977; Rigby *et al.*, 1977; Tsai *et al.*, 1978; Balmain
and Birnie, 1979) the principals are the same: single-strand breaks or
'nicks' are introduced randomly into the unlabelled DNA by an endo-
nuclease (usually pancreatic deoxyribonuclease I) which generates
3'-hydroxyl termini. DNA molecules containing such termini serve as
template-primers for DNA polymerase I. Nucleotides are added to the
3'-hydroxyl side of the nick at the same time as existing nulceotides
are removed from the other side by the 5' to 3' exonuclease activity
which is an integral part of the enzyme (Klett *et al.*, 1968; Deutscher
and Kornberg, 1969). Consequently, the nick is 'translated' along the
DNA molecule in the 5' to 3' direction (see Figure 2.10), the sequence of
nucleotides added to the 3' hydroxyl side of the nick being determined by
the opposite strand of the DNA duplex and is therefore a copy of the
sequence of nucleotides being removed from the 5' side of the nick.
Thus, there is no net synthesis of DNA, and if the nucleotides introduced
during this process are radioactive, the DNA will become radioactively
labelled. At the relatively low temperature used (15°C), the *in vitro*
DNA synthesis is restricted to one round of replication, and the reaction
proceeds no further than one complete renewal of the existing nucleotide
sequence. At higher temperatures (22°C) the synthesis continues
beyond the first round of replication (Dumlas *et al.*, 1971), and there
is overall synthesis, probably by a mechanism involving strand displace-

ment. However, such a process is difficult to control such that some sections of the DNA become represented more frequently than others and thus is not suitable for the synthesis of probes to be employed in kinetic hybridisation reactions. DNA nick translated at the lower temperature, however, is uniformly labelled (Mackey *et al.*, 1977; Rigby *et al.*, 1977) and has the expected renaturation kinetics (Mackey *et al.*, 1977; Balmain and Birnie, 1979).

Nick translation was used in this project as a method of synthesising a probe which would be ideal as a positive control during the screening of the library. Due to the repetitive nature of the ribosomal DNA sequences within the genomes of higher organisms (Spear, 1974; Vlad, 1977), including the chicken (Sinclair and Brown, 1971), the number of recombinant phage carrying such sequences, even in an · incomplete library, should be high (as discussed in Chapter V), and therefore should be readily detectable using a specific probe.

A ^{32}P-ribosomal DNA probe was prepared by the nick translation of a purified ribosomal DNA sequence provided by Ian Purdom in this laboratory.

Incubation mixes of 125 µl contained:

1	µg	DNA
50	µCi	α-^{32}PdCTP (specific activity 800 Ci/mmol)
0.008	mM	each of dATP, dGTP, dTTP
0.05	M	potassium phosphate pH 7.4
4.8	mM	magnesium chloride
10	µg/ml	deoxyribonuclease I
15	units	*E. coli* DNA polymerase I

The labelled and unlabelled deoxynucleotide triphosphates were lyophylised to dryness and the DNA and salt buffers were added in 100 µl. The DNase I was added, mixed and left at room temperature

for 1 min. Finally, the DNA polymerase I was added and the mix was incubated at 14°C for 3 hours. Samples of 2 µl were removed at 0, 1, 2 and 3 hours onto GF/C filter discs which were washed in ice cold 10% TCA for 10 min, washed in ethanol, dried and counted by scintillation counting. The reaction was terminated by the addition of 60 µl of 0.25M EDTA and 10 µl of 5% sodium dodecyl sulphate. The DNA was extracted and enzymes removed by gentle mixing with 200 µl of water saturated phenol, centrifuged for 2 min and the aqueous phase was applied to a sephadex G-50 column. 0.7 ml fractions were collected in 0.1 x SSC and the position of the radioactivity was determined by taking 2 µl samples and counting directly in aquasol. The nick translated probe was found in the first peak of radioactivity eluting from the column. The peak tubes were stored at -70°C.

Figure 2.11 shows a time course of the incorporation of ^{32}P-dCTP into ribosomal DNA by nick translation and Figure 2.12 shows the separation of the nick translated probe from the free deoxynucleotides by sephadex chromatography.

2.4 Conclusions

Radioactively labelled DNA probes were prepared using two different methods as described in this chapter.

The nick translation of purified, unlabelled ribosomal DNA produced a uniformly, highly labelled ^{32}P-probe with an efficiency of 3.5 x 10^7 cpm/µg. This probe is representative of a highly repetitive sequence within the chicken genome, and therefore, even in an incomplete genomic library has a high possibility of being represented.

The reverse transcription of f2 (myosin heavy chain) mRNA, conversely, yielded a specific probe for a sequence which is present
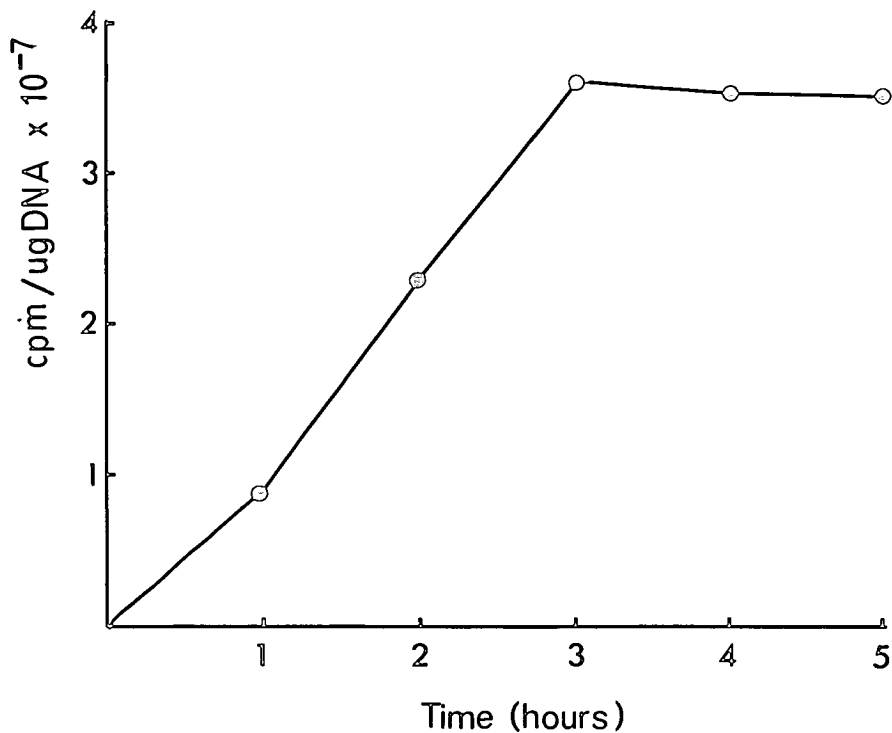
FIGURE 2.11: Nick Translation of Ribosomal DNA - Incorporation of ^{32}P-dCTP into DNA with Time.

Nick translation of a cloned ribosomal DNA was carried out as described in Section 2.3. The incorporation of ^{32}P-dCTP into the DNA was monitored over a 5 hour incubation period by removing 2 µl samples at intervals of 0, 1, 2, 3, 4 and 5 hours, spotting onto glass fibre filters, washing in TCA and counting. The maximum incorporation of radioactivity into acid insoluble material had occurred after 3 hours of incubation, but there was no significant degradation following this.
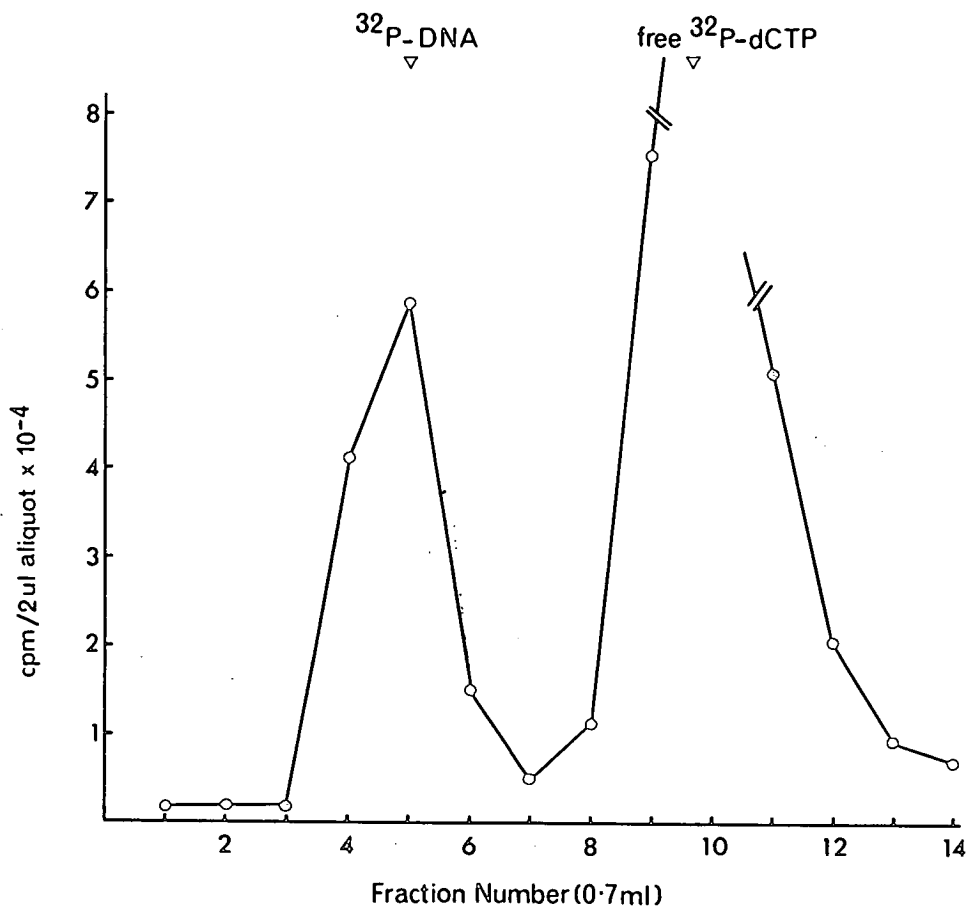
FIGURE 2.12:   Separation of Nick-translated Ribosomal DNA by
              Sephadex G 100 Chromatography.

The nick-translated DNA was separated from free ^{32}P-dCTP
by sephadex G 150 chromatography in 0.1 x SSC. 0.7 ml fractions
were collected and the peak of radioactive DNA was identified
by directly counting 2 μl samples in aquasol. The first peak of
radioactivity was pooled and retained at -70°C. The second,
large peak consists of free deoxynucleotides.

in only a few copies per genome. The validity of this statement depends on a demonstration of the purity of the mRNA preparation and that it codes exclusively for the myosin heavy chain protein. It has been shown previously (Patrinou-Georgoulas and John, 1977) that the mRNA isolated from an f2 polysome fraction is highly enriched in a mRNA which codes for myosin heavy chain and this is confirmed in Chapter III where the *in vitro* translation products of both mRNA and polysomes from f2 are analysed by polyacrylamide gel electrophoresis.

Using nucleic acid hybridisation techniques, Patrinou-Georgoulas and John (1977) concluded that there are two myosin heavy chain sequences represented in the f2 polysome fraction and that each of these is represented twice in the chicken genome. The R_0t value of 7.1×10^{-3} presented in Section 2.2.9, following the hybridisation of f2 cDNA to its template mRNA, is slightly higher than their value of 6.3×10^{-3} but it also suggests the presence of more than one sequence though it is not possible to propose a definite number without further experiments. However, the purpose of the hybridisation experiment presented in Section 2.2.9 was not to further speculate on the number of myosin heavy chain genes which has been well studied, but to assess the purity of the mRNA preparation.

The suitability of the cDNA probe for the eventual screening of the library depends entirely on the purity of the mRNA preparation. The possibility was considered that slight contamination of the mRNA preparation by a few other templates which could not be detected by *in vitro* translation of the mRNA, may give rise to a mixed cDNA preparation if these templates were particularly good substrates for reverse transcriptase. Hybridisation of the cDNA to its template

should reveal the presence of any minor contaminating species as minor transitions of hybridisation which would arise as resistance to S1 nuclease approached 100%. Such minor transitions were not observed and since the nature of the single transition was sharp and approached complete hybridisation it is strongly indicated that the mRNA (and therefore the cDNA) is not contaminated to any major extent with minor abundance groups and that the cDNA is indeed a faithful copy of the mRNA.

Therefore, the methods described in this chapter, taken together with the evidence provided by polyacrylamide gel electrophoresis presented in Chapter III, present a reasonable scheme for the isolation of purified myosin heavy chain mRNA. By virtue of its large size and abundance in developing chick skeletal muscle, it was possible to isolate the mRNA in a pure form from the heavy fraction of polysomes extracted from embryonic chick skeletal muscle and fractionated by sucrose gradient centrifugation. From this, a cDNA probe was synthesised by reverse transcription though it probably represents more than one particular heavy chain sequence. Previous evidence provided by Patrinou-Georgoulas and John (1977) suggests that a particular heavy chain sequence, of which there are probably two, is present in very few copies per haploid genome. Therefore, the number of positive signals arising following the screening of a genomic library with this cDNA probe will depend not only on the completeness of the sequence representation of the library but also on the cross hybridisation (that is, extent of sequence similarity) between the different heavy chains.

CHAPTER III

Gel Electrophoresis

CHAPTER III

3.1 Introduction

Gel electrophoresis is a widely used technique which allows the separation of proteins and nucleic acids according to their molecular size (Davis, 1964; Laemmli, 1970; McDonnell et al., 1977). The technique, in its many forms, was very important throughout this work and therefore will be discussed in this separate chapter thought it is referred to often throughout the text.

Generally, proteins are separated in polyacrylamide gels at a pH around neutral and nucleic acids are separated in agarose since it forms a gel with a larger pore size, though small fragments of nucleic acids are ideally electrophoresed in polyacrylamide. Single-stranded forms should be separated under denaturing conditions, such as alkaline pH, and double-stranded forms under non-denaturing conditions at neutral pH. By passing a current through the gel, the molecules will move from^to the anode (negative electrode) to^from the cathode (positive electrode). The speed at which they move at a given current depends on individual molecular size.

Proteins may also be separated by isoelectric focusing according to their charge (O'Farrell, 1974; O'Farrell et al., 1977). In this procedure the polyacrylamide gel has a pH gradient extending throughout its length and under the influence of an electrical current, the proteins migrate to their isoelectric point at which they have no net charge. This latter technique is more applicable to the actual characterisation of proteins where two-dimensional electrophoresis may be desirable – involving isoelectric focusing in one dimension followed by separation by molecular weight in a second. This level of separation was not required here so electrophoresis based on molecular weight only was used.

The choice of polyacrylamide or agarose gel concentration is very important as it is this which influences the molecular weight range over which a good separation can be achieved. Plots of relative mobility versus log molecular weight usually show a linear relationship over a restricted (approximately five-fold) range of molecular weight and become sigmoidal at high and low molecular weight values. A standard curve of this type is shown in Figure 3.1. Low polyacrylamide gel concentrations (5-10%) are useful for the separation of large proteins and high gel concentrations (15-20%) are useful for the separation of small proteins and peptides. A complex mixture is therefore best separated on a gel consisting of a concentration gradient. This principal similarly applies to the separation of nucleic acids in polyacrylamide or agarose gels. A low concentration of agarose is 0.3 - 0.7% and a high concentration is 1.5 - 2.0%.

Gel electrophoresis played a major part in all sections of this work. Firstly, following the *in vitro* translation of polysomal mRNA fractions by either the wheat germ or rabbit reticulocyte lysate systems, the newly synthesised products were analysed by electrophoresis on sodium dodecyl sulphate (SDS) polyacrylamide gels. The homogeneous nature of the mRNA extracted from each individual polysome fraction was assessed according to the number of polypeptides appearing as bands on the gel following *in vitro* translation. The aim was to obtain a mRNA preparation which directed the synthesis, in these cell-free systems, of a single protein which co-electrophoresed with the myosin heavy chain marker.

A dissociating gel system was used throughout. Native proteins under such conditions are dissociated into their constituent sub-units and post-translational modifications such as methyl, acetyl or glycosyl
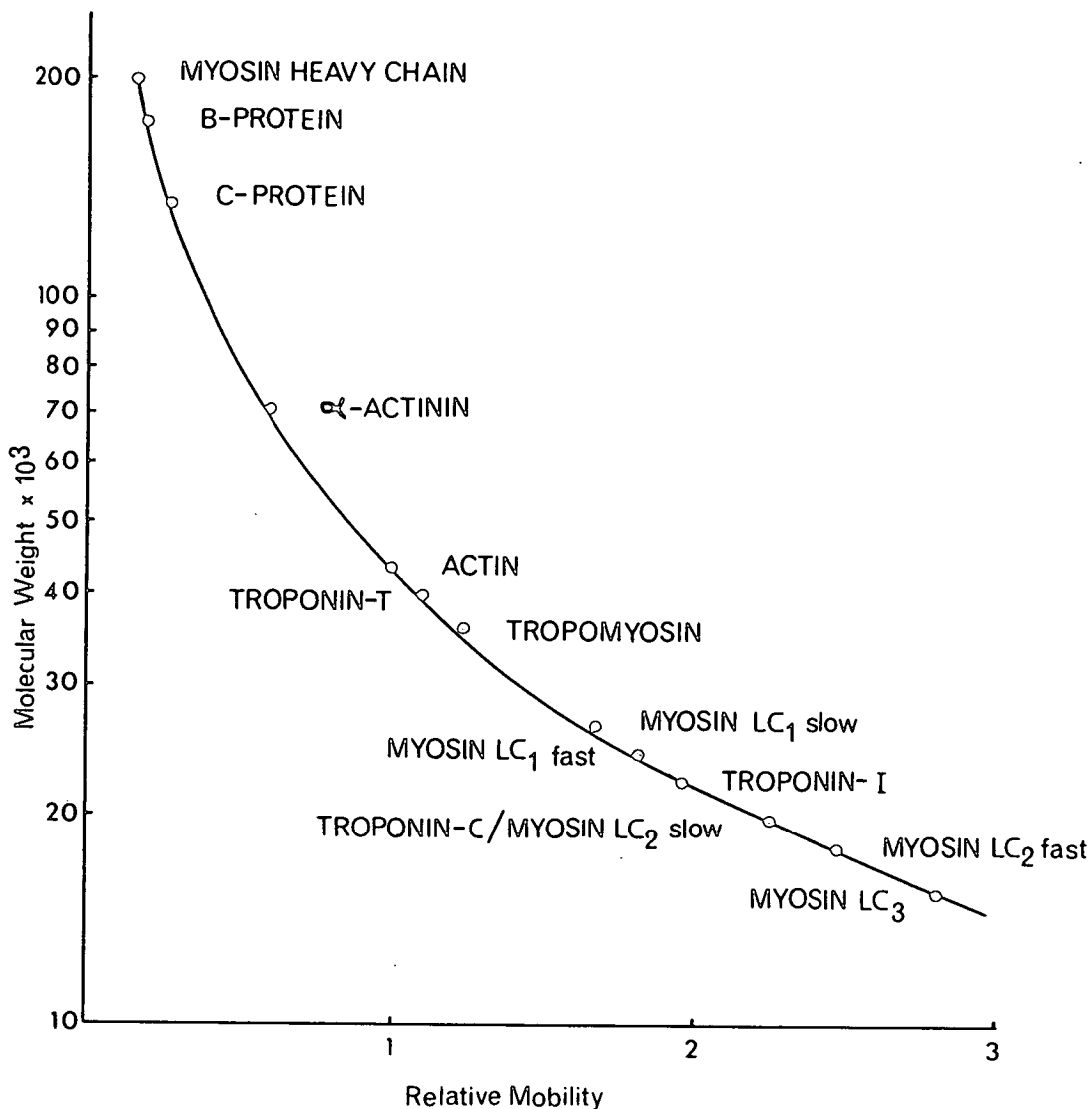
FIGURE 3.1: Standard Protein Mobility Curve.

The mobility of a series of proteins following SDS polyacrylamide gel electrophoresis, relative to that of actin, were plotted against their known molecular weights on a log scale as shown. This provides a curve which is extremely useful for the determination of unknown protein molecular weights, using actin as a marker.

side groups are removed. Since, following *in vitro* translation, it is unlikely that post translational modifications or assembly of polypeptide sub-units into active proteins take place due to the absence of necessary enzyme systems and co-factors, many of which are membrane associated, the dissociating conditions included in this gel system are to prevent non-specific aggregation of newly synthesised polypeptides.

Secondly, agarose gel electrophoresis was widely used during the construction of the DNA library and during the construction of the probes. mRNA was electrophoresed to determine the level of contamination of any particular myosin heavy chain preparation with other RNAs, particularly 28S ribosomal RNA. The size of the cDNA was determined, and all the stages involved in the construction of the library, where both bacteriophage and chicken DNAs were manipulated, were carefully monitored by agarose gel electrophoresis.

3.2 Polyacrylamide Gel Electrophoresis

Translated mRNA and polysome fractions were analysed by electrophoresis on two types of polyacrylamide gels, each having different advantages.

3.2.1 *Electrophoresis of Polysome Translation Products in 8M Urea*
    (Davis, 1964)

Initially, translation products were analysed in polyacrylamide gels by a method based on that developed by Weber and Osbourn (1969) where 8M urea is included in the sample buffer as a denaturant.

A 10% acrylamide solution was prepared containing 0.13% bis and 0.2% SDS in 0.01M sodium phosphate pH 7.0. The polymerization agents, ammonium persulphate and TEMED were added to a final concentration of 0.3% and 0.075%, respectively, and the gels were cast in 0.3 cm x

8 cm glass tubes to within 0.5 cm of the top. Water was layered over the surface prior to setting to form a perfectly level interface.

Protein samples usually consisted of 50-75 µl of the incubation mixes following the *in vitro* translation of polysomes. Often it was necessary to partially lyophylise these to concentrate the radioactive protein into a suitable volume. Two volumes of warmed dissociation buffer were added to give a final concentration of 8M urea, 1% SDS, 1% β-mercaptoethanol, 0.01M sodium phosphate pH 7.0, and samples were incubated at 37°C for 5 min to completely dissociate proteins into their sub-unit components and to disrupt the weak covalent bonds which often form cross links within polypeptide chains. In this way, the secondary structure of a protein does not affect its electrophoretic properties and thus separation in SDS-polyacrylamide gels is a function of the molecular weight of individual chains. The sample was then made 10%, with respect to glycerol, and 0.025%, with respect to bromophenol blue, 5 µg of myosin was added as marker and the whole sample was applied to the top of the gel beneath the tank buffer. Prior to application of the samples, the gel surface was rinsed with dissociation buffer containing 8M urea as this treatment was found to enhance the ability of the large myosin heavy chain to initially enter the gel. The tank buffer was 0.005M sodium phosphate pH 7.0. Electrophoresis was carried out at 8 mA constant current per tube (up to 8 gels could be electrophoresed together) for 4½ hours.

Following electrophoresis, the gels were removed from the tubes and stained in 0.2% coomassie blue stain prepared in 45% ethanol, 10% acetic acid at 60°C for 2 hours. Background stain was removed by destaining in 25% ethanol, 10% acetic acid followed by 10% acetic acid alone at room temperature.

The position of the myosin heavy chain marker was noted and the gels were then sliced neatly into 1 mm sections. Each section was individually solubilised in 0.5 ml of 100 volumes hydrogen peroxide at 50°C for 6 hours and counted directly in aquasol.

The advantage of this gel system is that marker proteins and samples can be electrophoresed together. This allows the location of protein bands by both staining and slicing (followed by scintillation counting) to be performed on the same gel. Since co-electrophoresis is the main criterion for the identification of newly synthesised myosin heavy chain, it was felt that this degree of accuracy was necessary.

Figures 3.2a-f are plots of counts per minute against gel slice number. Figure 3.2a shows the complex profile obtained following the translation of total polysomes and electrophoresis as described above. However, as is shown in Figures 3.2b-f, when the polysome fractions collected from the bottom of the gradients were translated individually the resulting protein profiles were much simpler. There was no translational activity in the first fraction (Figure 3.2b) but there was one major peak of radioactive incorporation in fraction 2 (Figure 3.2c) and this co-electrophoresed with the myosin heavy chain. This peak was also present in fractions 3, 4 and 5 (as shown in Figures 3.2d, 3.2e and 3.2f), but to a lesser extent and it did not form an exclusive peak in these fractions.

When a polysome translation mix to which no exogenous polysomes had been added was treated in an identical manner to that described in Chapter II, Section 2.2.3 and as above, the sliced gel profile was similar to that shown for polysome fraction 1 in Figure 3.2b. This indicates that the endogenous activity of the S-150 fraction is very low, such that the peaks of radioactivity arising from fractions 2-5 are due to the *in vitro* translation of the exogenous polysomes.

FIGURE 3.2:

Total polysomes, extracted from 14-day embryonic chick leg muscle
were translated and the ^{3}H-labelled translation products were
electrophoresed on disc gels, as described in Section 3.2.1,
according to the method of Weber and Osborn (1969). The gels
were sliced into 2 mm slices, solubilised and counted. As shown
in (a), a complex mixture of ^{3}H-proteins resulted from the *in vitro*
translation of total polysomes.

Individual heavy polysome fractions 1 to 5, collected from the
bottom of a 15-40% sucrose gradient were translated and electro-
phoresed by identical procedures. The profiles of the sliced disc
gels are shown in (b) to (f). Purified marker myosin heavy
chain was co-electrophoresed with each sample and its position
on the gel, determined by coomassie blue staining, is shown (▽).
This position corresponded to a peak of radioactivity following
translation of both total polysomes (a) and fractions 2 to 5 [(c) -
(f)]. However, only fraction 2 showed a single major peak of
incorporation which co-electrophoresed with the myosin heavy chain
marker.

a. Total Polysomes

b. Fraction 1.

c. Fraction 2.

Myosin Heavy Chain
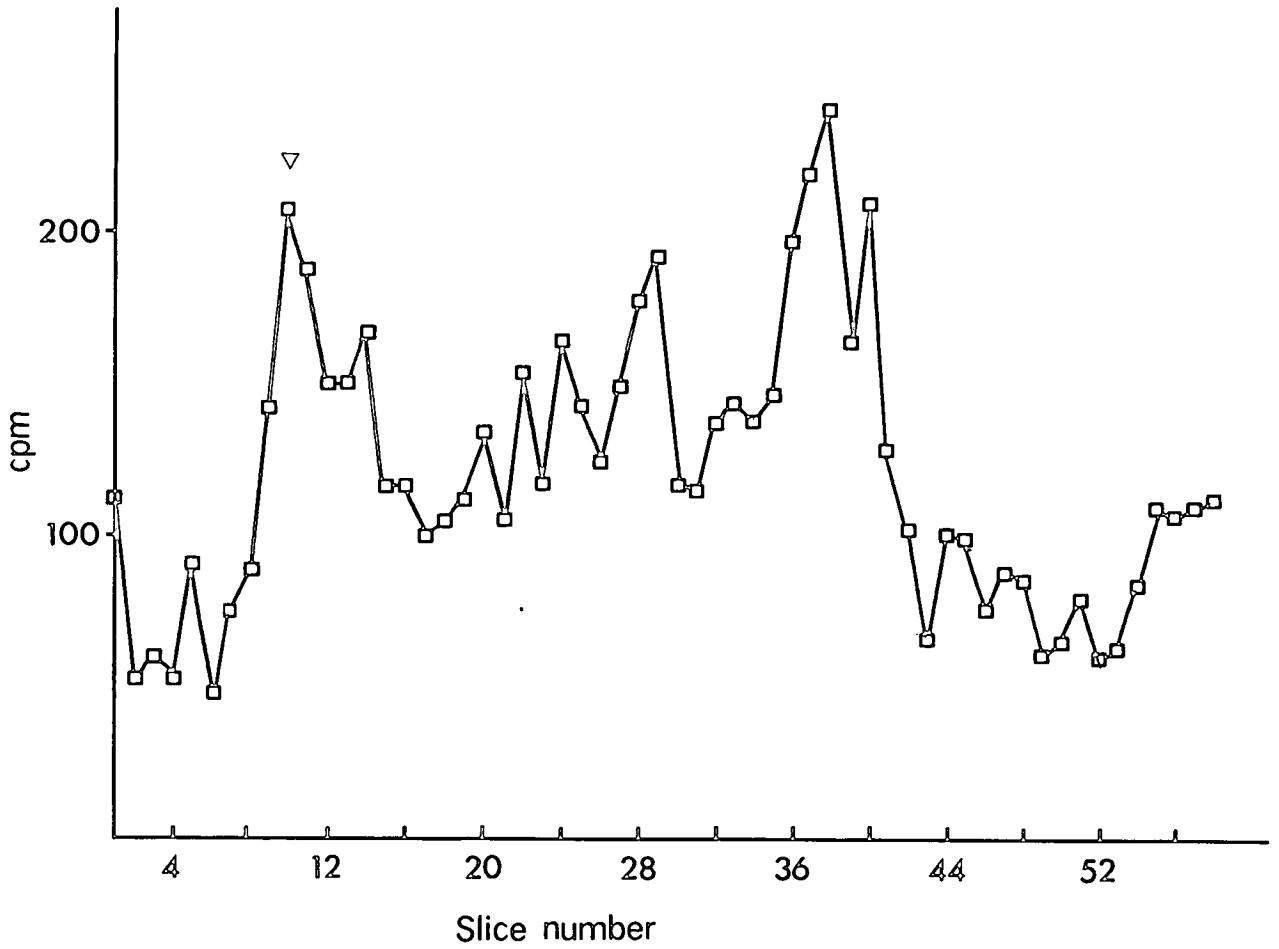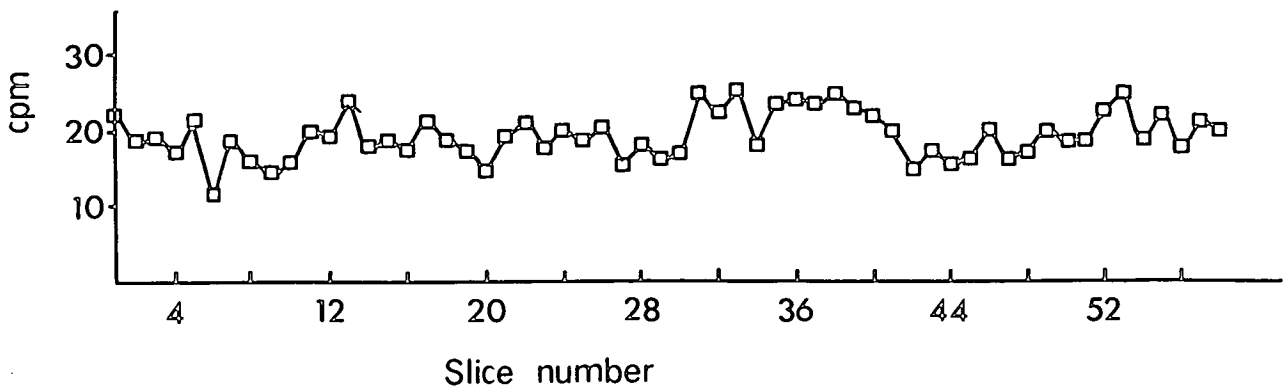
d. Fraction 3.
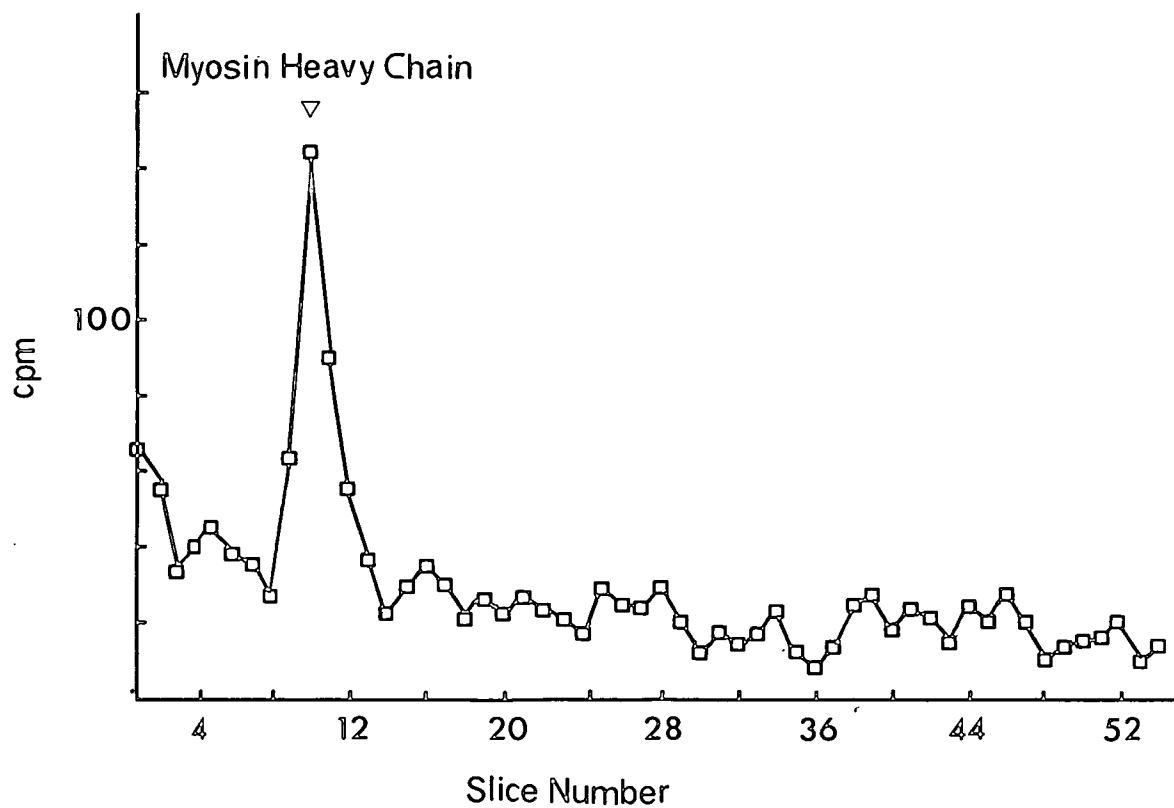
e. Fraction 4.

f. Fraction 5.
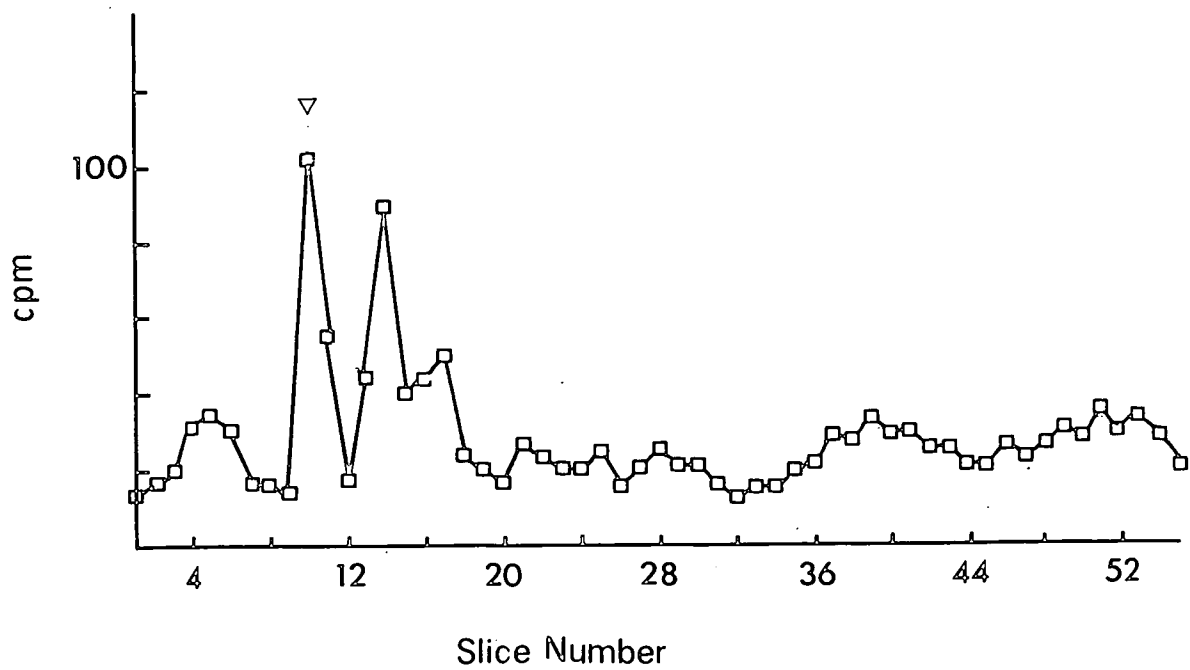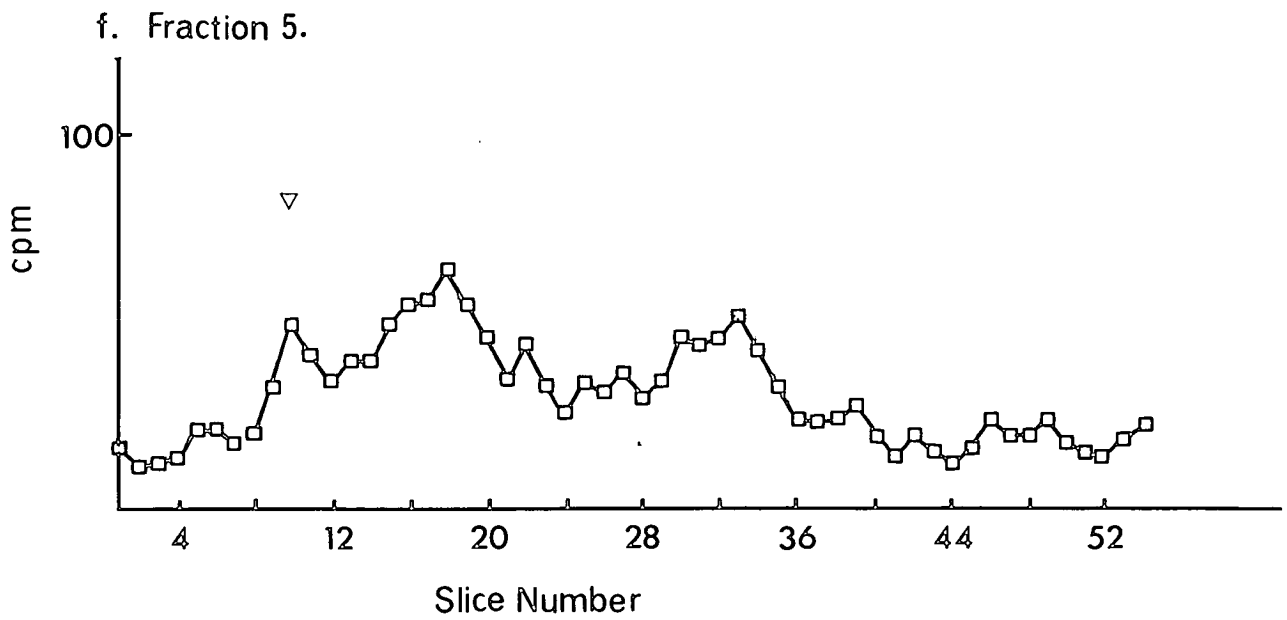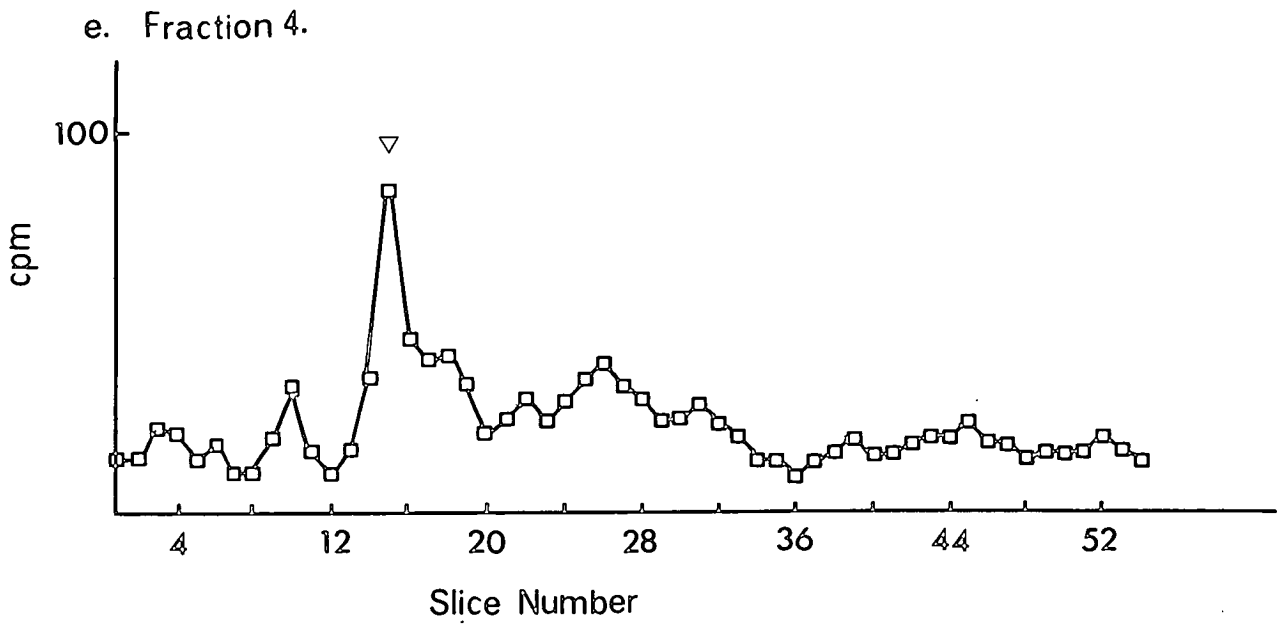
3.2.2 *Electrophoresis of Proteins by the Laemmli Discontinuous Buffer System* (Laemmli, 1970)

This system involves electrophoresis through two types of poly-acrylamide gel: an upper stacking gel at pH 6.8 and a lower separating gel at pH 8.8. The short stacking gel serves to concentrate the sample so that the effects of diffusion which inevitably takes place in the separating gel are reduced to a minimum. For this concentration effect to be successful, it is necessary for the pH difference between the two phases to be two units.

An advantage of this system is that the gel is easily dried with no distortion since it is only 1 mm thick and is therefore an excellent system for the autoradiographic detection of newly synthesised proteins. Several samples can be electrophoresed in parallel and the final profiles can be compared easily.

10% polyacrylamide slab gels (18 cm x 16 cm x 1 mm) supported between two glass plates, consisted of a 2 cm plug of 22.5% polyacryl-amide over which was layered the separating gel. This consisted of 9.7% acrylamide, 0.3% bis as the cross linking agent, 0.2% SDS, 1M tris pH 8.8 with 0.3% ammonium persulphate and 0.05% TEMED as the poly-merisation catalysts. Following polymerisation of the separating gel, a 4.75% stacking gel, approximately 3 cm deep, was layered on top, consisting of 4.6% acrylamide, 1.75% bis, 1% SDS, 0.3% ammonium per-sulphate, 0.05% TEMED in 0.125M Tris pH 6.8. A row of sample wells were formed in the stacking gel capable of holding a volume of 50 µl.

Protein samples usually consisted of suitable volumes of the incuba-tion mixes directly following the *in vitro* translation of mRNA (10 µl following translation in the rabbit reticulocyte lysate system and 25 µl following translation in the wheat germ system). These were incubated

for 5 min at 100°C in an equal volume of 2 x sample buffer which contained SDS and 2-mercaptoethanol as protein denaturants for those reasons described in Section 3.2.1. Samples were made 10% with respect to glycerol and 0.025% with respect to bromophenol blue, and applied to the gel. Molecular weight marker proteins, both labelled and unlabelled, were treated in an identical manner. The tank buffer was 0.025M Tris, 0.192M glycine, 0.1% SDS. Gels were electrophoresed ·· at a constant current of 8 mA for 16 hours or 25 mA for 5 hours.

Staining of the gels was carried out as described in Section 3.2.1 above, using coomassie blue. However, staining of these gels was not always necessary as the newly synthesised proteins were not present in sufficient quantity to be directly visualised and autoradiographic detection was always necessary. Figure 3.3 shows a stained and dried gel upon which were electrophoresed translation samples from mRNA fractions 1-5. Under coomassie blue staining, all the tracts appear identical but the stained bands simply represent the endogenous rabbit reticulocyte lysate proteins. These bands are not represented on the autoradiograph of a similar gel which is shown in Figure 3.4a.

3.2.3 Fluorography

Labelled proteins in the slab gels were detected by a scintillation autoradiography method called fluorography (Bonner and Laskey, 1974). The gel is impregnated with the scintillant, 2,5-diphenyloxazole (PPO), and film placed adjacent to the gel is exposed by light generated by interaction of β particles with the PPO. 3000 dpm of tritium can be detected in 24 hours by this procedure and when applied to ^{35}sulphur or ^{14}carbon the method is 10 times more sensitive than conventional autoradiography. 130 dpm of these isotopes can be detected in 24 hours.
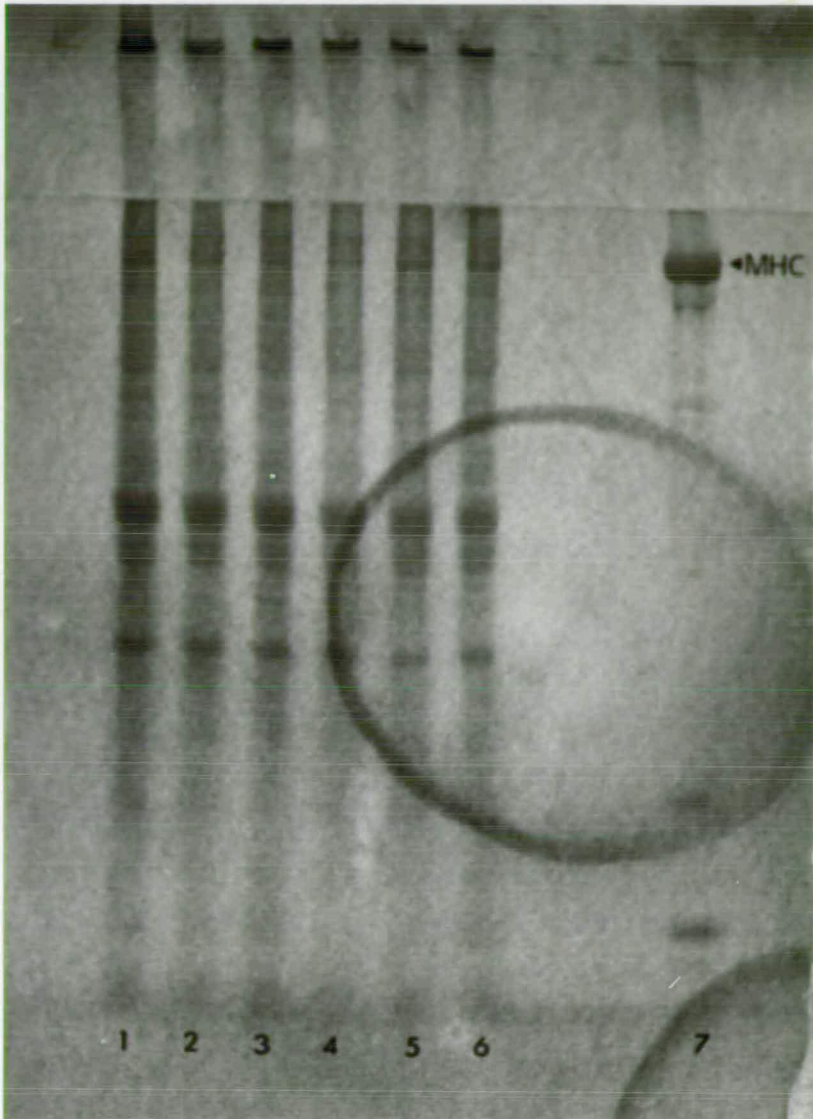
FIGURE 3.3: Stained Polyacrylamide Slab Gel of mRNA Translation Products.

mRNA extracted from polysome fractions 1-5 were translated in the rabbit reticulocyte lysate cell free system (Chapter 2, Section 2.2.7). The translation products were electrophoresed on 10% polyacrylamide slab gels and stained using coomassie blue. All tracts appear identical, including that to which no exogenous mRNA was added (lane 6) due to the predominant staining of the reticulocyte lysate proteins. The *in vitro* products are not synthesised in sufficient quantity to be stained by coomassie blue and therefore must be detected by autoradiography as shown in Figure 3.4. Lanes 1 to 5 represent translation products from fractions f1 to f5, respectively. Lane 7 is purified myosin marker protein, and the position of the heavy chain is indicated.

PPO is highly insoluble in water, and therefore to allow its penetration into the gel, it is necessary to dehydrate the gel without causing distortion. Dimethylsulphoxide is a suitable solvent which will both dehydrate polyacrylamide and dissolve PPO.

Directly after electrophoresis or after staining, the gels were soaked in 20 volumes of dimethylsuphoxide (DMSO) for 30 min followed by a second 30 min soak in fresh DMSO. The gels were then immersed in 4 volumes of 22.2% (w/v) PPO and DMSO for 3 hours followed by 1-2 hours in a large volume of water which was regularly changed. This final re-hydration of the gel greatly aids the drying process since DMSO is difficult to remove under vacuum. However, this does lead to the *in situ* precipitation of PPO which turns the gel opaque, but the fluorographic efficiency is not affected by this. The gels were dried in approximately 3 hours under vacuum at 90°C. The dried gel was then placed in contact with pre-flashed RP Royal X-omat medical x-ray film (Laskey and Mills, 1975) and exposed at -70°C. The exposure times varied from 5 days to 3 weeks depending on the specific activity of the proteins and the isotope used in the *in vitro* translation.

It was shown (Laskey and Mills, 1975) that the pre-exposure of the film to a brief flash of light greatly increases the sensitivity of fluorography and allows quantitative interpretation of the film image. Without the pre-flashing treatment, the absorbance of the fluorographic image is not proportional to the concentration of radioactivity in a particular band or proportional to the time of exposure. Pre-flashing corrects this non-linear relationship.

Figure 3.4 shows an autoradiograph of a fluorographed poly-acrylamide slab gel following electrophoresis of polysomal mRNA translation products. This figure indicates that the newly synthesised
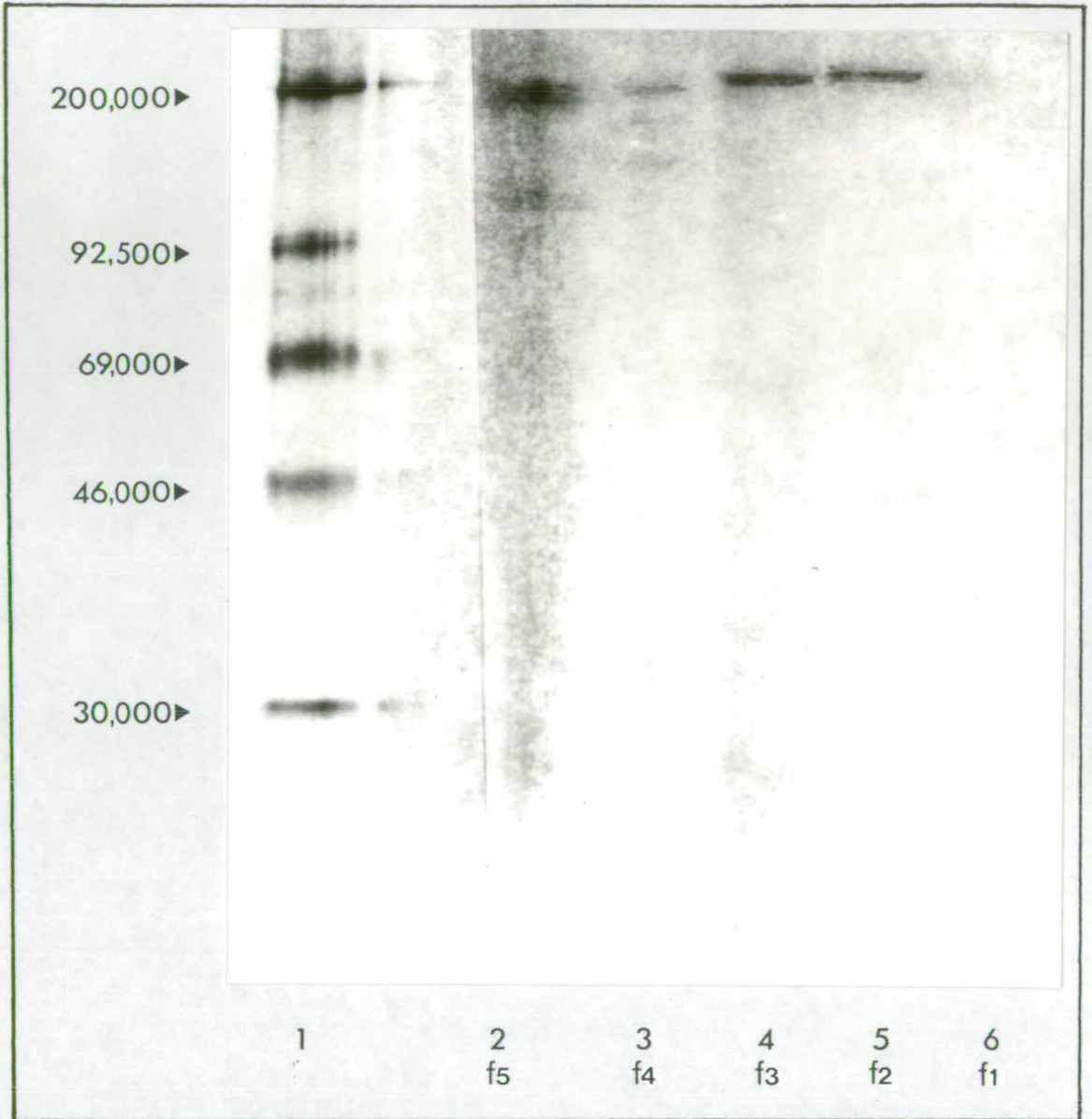
FIGURE 3.4: Autoradiograms of mRNA Translation Products.

(a) Autoradiogram of a similar gel to that presented in Figure
     3.3. Lanes 2–6 show the ^{35}S-labelled *in vitro* translation
products of mRNA extracted from polysome fractions f1 to f5.
All fractions except f1 synthesise a 200,000 molecular weight
protein which co-electrophoreses with MHC, but only f2 mRNA
is able to direct the synthesis of this polypeptide alone. Lane 1
shows a series of ^{14}C-molecular weight markers: myosin heavy
chain (200,000), phosphorylase b (92,500), bovine serum albumin
(69,000), ovalbumin (46,000) and carbonic anhydrase (30,000).
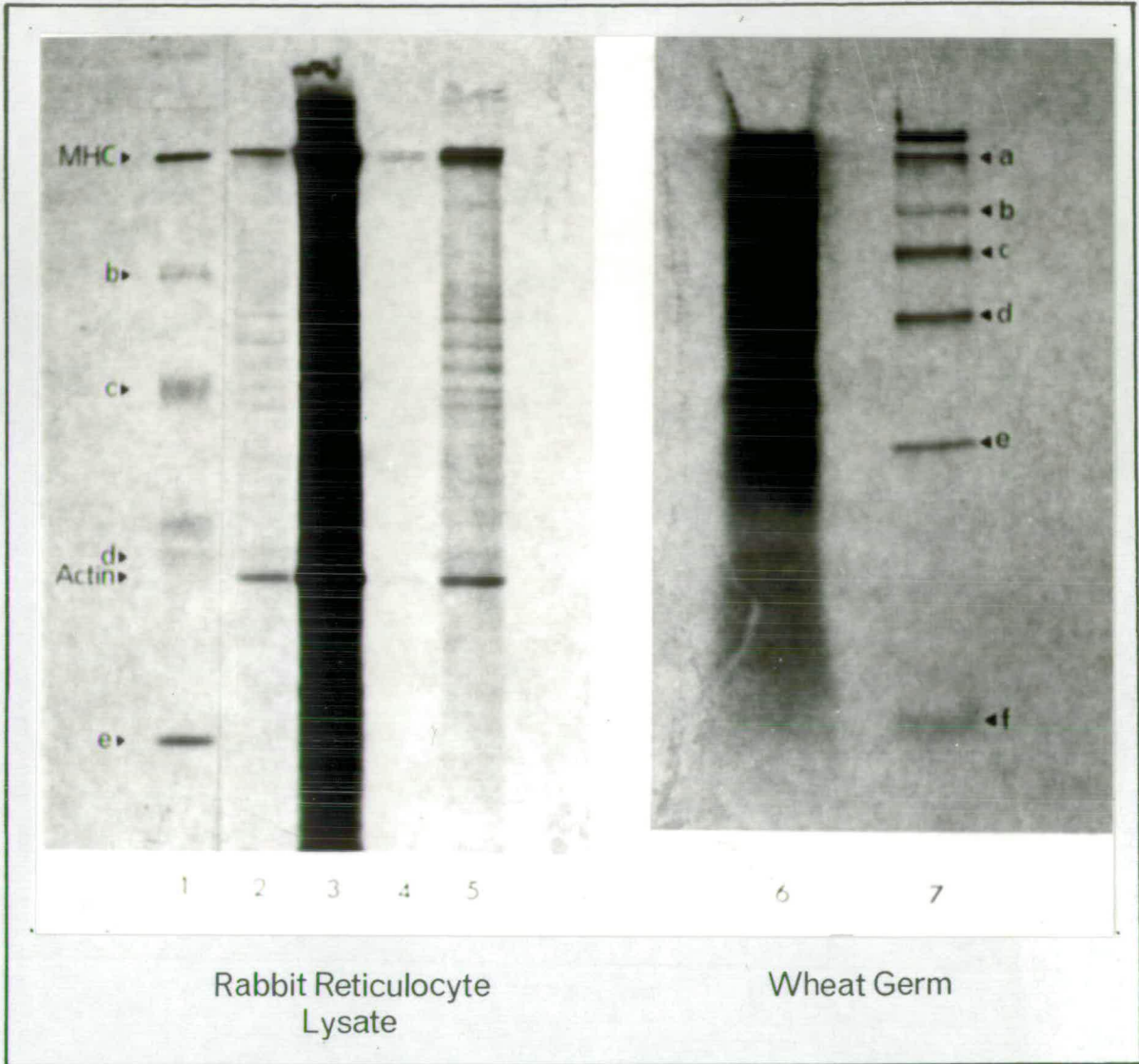
FIGURE 3.4: Autoradiograms of mRNA Translation Products.

(b) Autoradiograms of the *in vitro* translation products of mRNA
    extracted from total chick embryo skeletal leg muscle poly-
somes directed by the rabbit reticulocyte (lanes 2 - 5 with various
loadings) and the wheat germ systems (lane 6). Lanes 1 and 7 show
a series of molecular weight markers: a) myosin heavy chain
(200,000), b) phosphorylase b (92,500), c) bovine serum albumin
(69,000), d) ovalbumin (46,000), e) carbonic anhydrase (30,000),
and f) lysozyme (14,300). The positions of actin, and myosin heavy
chain which appear very predominantly in the translation products
of total polysomes are shown.

proteins arising from the translation of f2 mRNA consist predominantly of myosin heavy chain.

3.2.4 *Polyacrylamide Gel Electrophoresis of Small Double-Stranded DNA Fragments*

Small fragments of double-stranded DNA (10 to 100 base pairs) can be separated electrophoretically on polyacrylamide gels (Maniatis *et al.*, 1975b) with better resolution than can be achieved with agarose gels. The concentration of acrylamide depends on the size of the DNA to be separated. As discussed in Chapter IV, Section 4.7.1, this electrophoretic method was the method of choice for the separation of small double-stranded DNA fragments ranging from 16 to 104 base pairs resulting from the ligation of a hexamer sequence. A very low acrylamide concentration was therefore required.

5% polyacrylamide slab gels (18 cm x 16 cm x 1 mm) were prepared, consisting of 4.75% acrylamide, 0.25% bis as the cross linking agent in tris-borate buffer (0.09M Tris pH 8.3, 0.025M EDTA, 6.5 mM boric acid). Ammonium persulphate was added to 0.3% and TEMED to 0.075% as polymerisation catalysts.

The DNA samples were prepared in the same tris-borate buffer with the addition of 5% glycerol and 0.025% bromophenol blue. Electrophoresis was continued at a constant voltage of 200 volts for 5 hours.

The DNA bands were visualised by staining with ethidium bromide solution. Following electrophoresis the gel was soaked for 30 min in 10 mM tris-HCl pH 7.4 containing 0.5 μg/ml ethidium bromide after which the DNA bands could be visualised under ultra violet light.

3.3 Agarose Gel Electrophoresis of Nucleic Acids

The manipulation of DNA which was carried out during the con-
struction of the chicken gene library as described in Chapter IV, was
continually monitored by agarose gel electrophoresis at all stages, and
is therefore referred to throughout that chapter. However, the
experimental details are presented here since the technique was also
important during the construction of the hybridisation probes.

3.3.1 Electrophoresis of RNA Samples

1.5% vertical agarose gels were prepared by gently refluxing
agarose in water followed by addition of an equal volume of warmed
2 x concentrated RNA gel buffer. The gel plates were heat sterilised
prior to the casting of the gel which was carried out before the plates
had completely cooled to prevent irregular polymerisation of the agarose.
The gel was then chilled at 4°C until the agarose turned opaque.

1 µg of f2 mRNA was added to 15 µl of RNA gel buffer (50 mM
boric acid, 5 mM sodium borate, 10 mM sodium sulphate, 1 mM EDTA),
containing 10% glycerol and 0.025% bromophenol blue. 1 µg of ribosomal
RNA (collected in the unbound fraction during oligo dT cellulose
chromatography) was treated in the same way and electrophoresed in
parallel. Electrophoresis was continued at a constant voltage of 80 V
for 3 hours. The tank buffer was also RNA gel buffer.

The gel was stained by soaking for 30 min in 10 mM tris-HCl pH
7.4 which neutralised the denaturing conditions set up in the gel by
the boric acid followed by a further 30 min in the same buffer containing
0.5 µg/ml ethidium bromide. This causes nucleic acids to fluoresce
under ultra violet (UV) light which thus allows visualisation of the
pattern of bands on the gel when safety spectacles are worn to protect

the eyes. Gels were photographed under UV light using polaroid film.

Figure 3.5 shows the photograph of such a gel. The f2 mRNA appears to electrophorese as one major band which is slightly larger than the 28S rRNA species.

3.3.2 Sizing of cDNA by Agarose Gel Electrophoresis

Single-stranded cDNA prepared by the reverse transcription of f2 mRNA as described in Chapter II, Section 2.2.8, was sized on agarose gels under denaturing alkaline conditions according to the method of McDonnell et al. (1977).

A solution of 2% agarose was prepared by gently refluxing in water and was then cooled to 50°C. This was made 30 mM sodium hydroxide (NaOH) and 2 mM EDTA and diluted to an agarose concentration of 1%. The gel was cast in a horizontal electrophoresis apparatus due to the low concentration of agarose which would not have sufficient rigidity to form a vertical gel. It was then chilled at 4°C until the agarose turned opaque.

Approximately 2000 cpm of the f2 cDNA was prepared in 15 μl of 30 mM NaOH, 2 mM EDTA, left to stand at room temperature for 15 min to allow disruption of any base pairing, and then made 10% glycerol, 0.025% bromophenol blue. The sample was applied to the gel and electrophoresis was continued at 80 V for 3 hours, or until the stain had reached the end of the gel. The tank buffer was 30 mM NaOH. 2 μg samples of standard molecular weight markers were treated in an identical way and electrophoresed parallel to the cDNA sample. The markers used were denatured Hin dIII restriction fragments of bacteriophage lambda which have sizes ranging from 100 to 23,000 nucleotides, a very broad range ideal for sizing unknown DNA fragments.
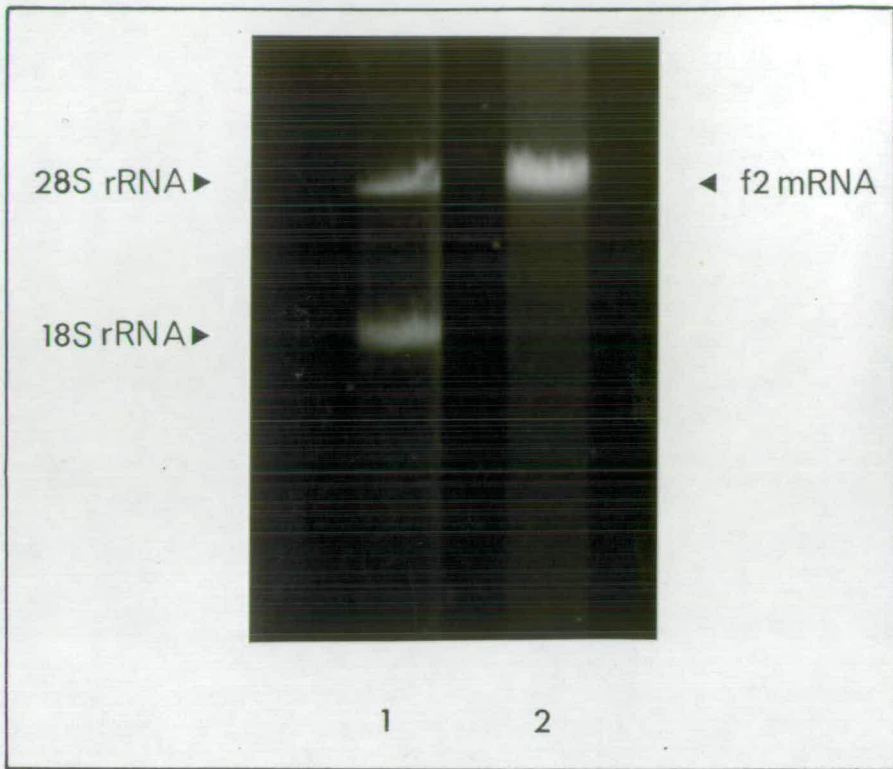
FIGURE 3.5: Electrophoresis of RNA.

Ribosomal RNA (lane 1) and f2 mRNA (lane 2) were electrophoresed
in parallel on 1% denaturing agarose gels as described in Section
3.3.1. The f2 mRNA (MHC mRNA) electrophoresed as a single
species with a molecular weight similar to that of 28S ribosomal RNA,
in agreement with several previous studies which have identified
MHC mRNA as a 26S molecule.

Following electrophoresis, the gel was neutralised by soaking for 30 min in 1M Tris-HCl pH 7.4 and was then stained using ethidium bromide, as described in Section 3.3.1, for the RNA gels. The positions of the marker bands were carefully noted and the tract containing the cDNA sample was sliced into 2 mm slices. These were solubilised in 0.5 ml hydrogen peroxide at 80°C for 2 hours, and counted directly in 10 ml of aquasol.

Figure 3.6a shows the position of the peak of radioactivity corresponding to the f2 (myosin heavy chain) cDNA following alkaline agarose electrophoresis. There was one major peak of ^{3}H-activity which corresponds to a single-stranded cDNA of approximately 550 nucleotides when compared with the electrophoretic mobility of the molecular weight markers represented diagrammatically in Figure 3.6b and plotted as mobility against log molecular weight in Figure 3.6c. When ^{32}P-dCTP was used as the radioactive label during the reverse transcription of f2 mRNA, again a single peak appeared on dectro-phoresis of the resulting ^{32}P-cDNA, but the size was calculated to be 10% smaller than the tritiated form at 500 nucleotides.

3.3.3 *Sizing of Double-Stranded DNA by Agarose Gel Electrophoresis*

Gels were usually prepared as described previously for cDNA gels in Section 3.3.2, except that the gel buffer was a neutral one (36 mM tris, 30 mM sodium dihydrogen orthophosphate, 1 mM EDTA pH 7.2), since denaturing conditions were not necessary. However, occasionally, netural agarose gels were cast in small tubes, 0.3 cm diameter x 8 cm, when only one or two samples were to be analysed. It was also possible under neutral conditions to include ethidium bromide in the gel at 0.5 µg/ml which enabled the DNA bands to be visualised immediately following electrophoresis.

Figure 3.6 →

.

FIGURE 3.6:   Sizing of Myosin Heavy Chain (f2) ^{3}H-cDNA and
^{32}P-cDNA by Alkaline Agarose Gel Electrophoresis.

(a)   2000 cpm of ^{3}H-cDNA and 10,000 cpm of ^{32}P-cDNA were applied
       to 2% alkaline agarose gels as described in Section 3.3.2.   2 mm
       gel slices were solubilised and counted.   The positions of the
       peaks of radioactivity (slices 43 and 44) are shown.   The size
       of the ^{3}H-cDNA was calculated to be 550 bases, approximately
       10% of the myosin heavy chain mRNA, and the ^{32}P-cDNA was
       slightly smaller at 520 bases.

(b)   The HindIII restriction fragments of bacteriophage λ which
       range from 22 kb to 0.1 kb were electrophoresised in parallel
       with the cDNA samples, as molecular weight standards.   A
       diagram of the position of the bands, viewed by ethidium bromide
       staining under UV illumination, is shown.

(c)   The mobility, measured from the top of the gel, of the
       bacteriophage λ HindIII restriction fragments were plotted
       against molecular weight as shown.   The sizes of the cDNA
       preparations were determined accurately from this plot by
       directly comparing the positions of the ^{3}H and ^{32}P radioactive
       peaks.

a.

b.

Molecular size in kilo bases
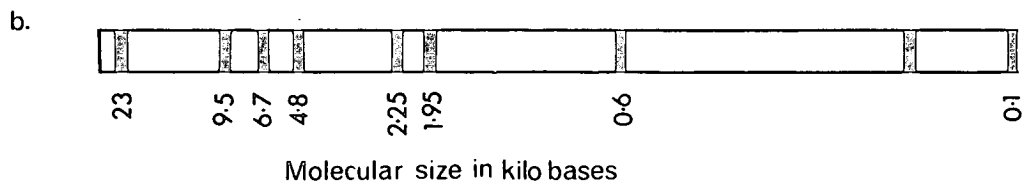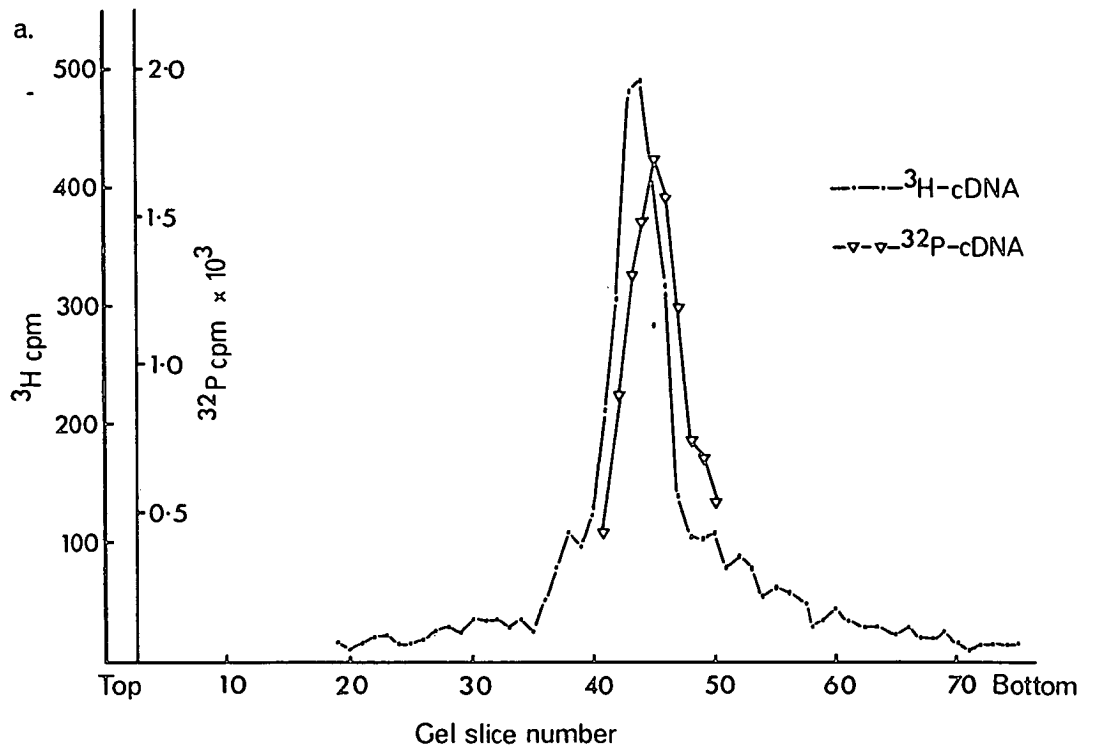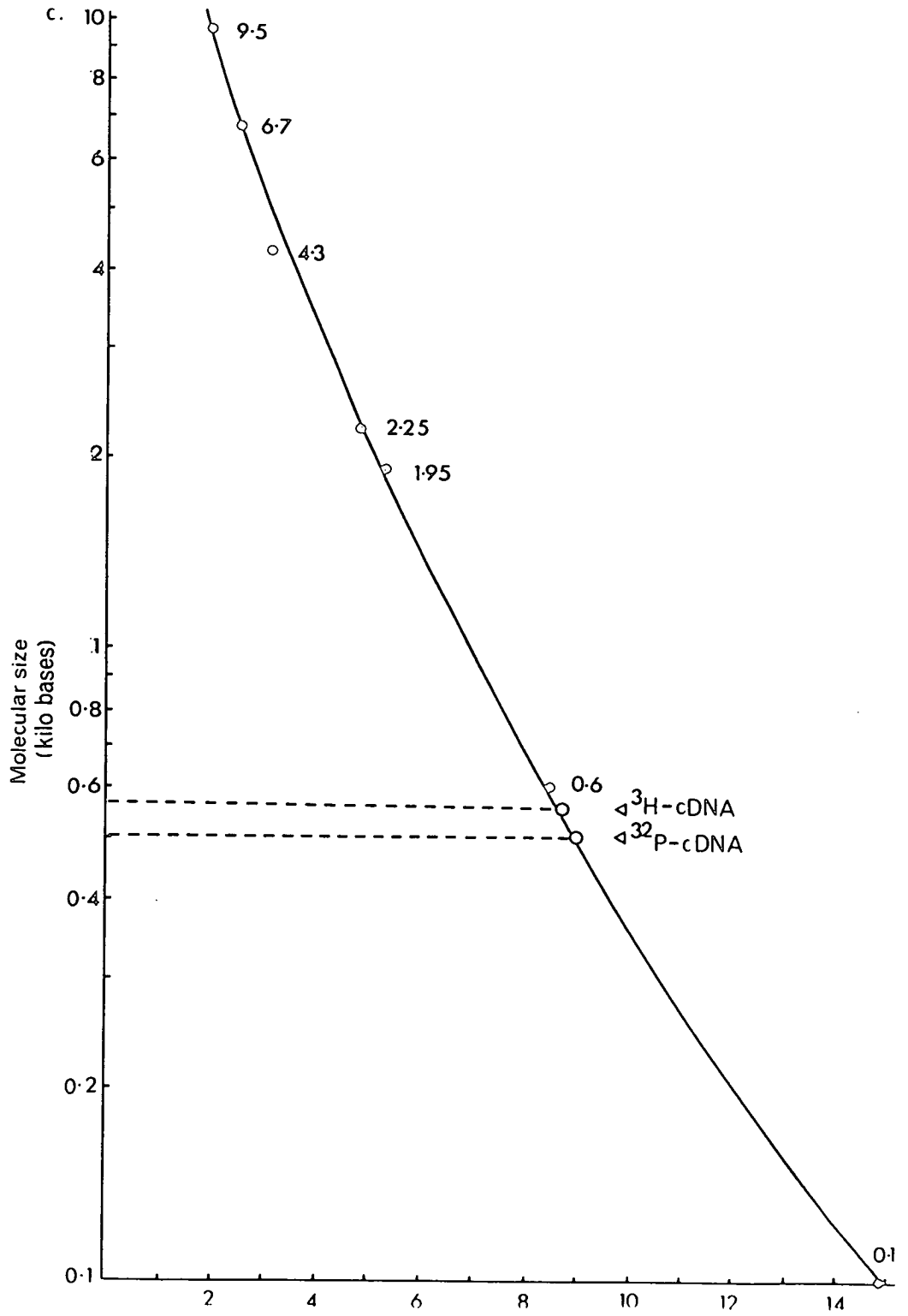
1 µg DNA samples were prepared in 10-15 µl of the above gel buffer containing 10% glycerol and 0.025% bromophenol blue. Electrophoresis was carried out using the same buffer as tank buffer for 6 hours at 60 V or for 16 hours at 20 V. The gels were viewed and photographed under an ultra violet lamp with no further staining required.

Figure 3.7 shows the electrophoretic pattern of the EcoR1 restriction enzyme fragments of wild type bacteriophage lambda (λ) compared with those of charon 4A, a derivative of λ to be used as the cloning vector for the chicken genomic library, described in Chapter IV. The molecular weights of the fragments are indicated.

3.4 Conclusions

Gel electrophoresis was a widely used technique throughout this work and therefore the various methods which are referred to only briefly elsewhere are presented in detail in this chapter. However, the evidence presented here must be considered in parallel with that presented in Chapter II.

The purity of the f2 mRNA preparation was assessed using two main criteria: the synthesis of a single protein which co-migrated with purified myosin heavy chain marker protein during SDS-polyacrylamide gel electrophoresis, following the *in vitro* translation of a small aliquot of the mRNA or polysomes; and the back hybridisation of the cDNA to the original mRNA which occurred with a single, complete transition. This latter criterion was fulfilled and discussed in Chapter II, though to be meaningful it must be complemented with the polyacrylamide gel electrophoresis evidence described in this chapter which fulfils the former criterion.
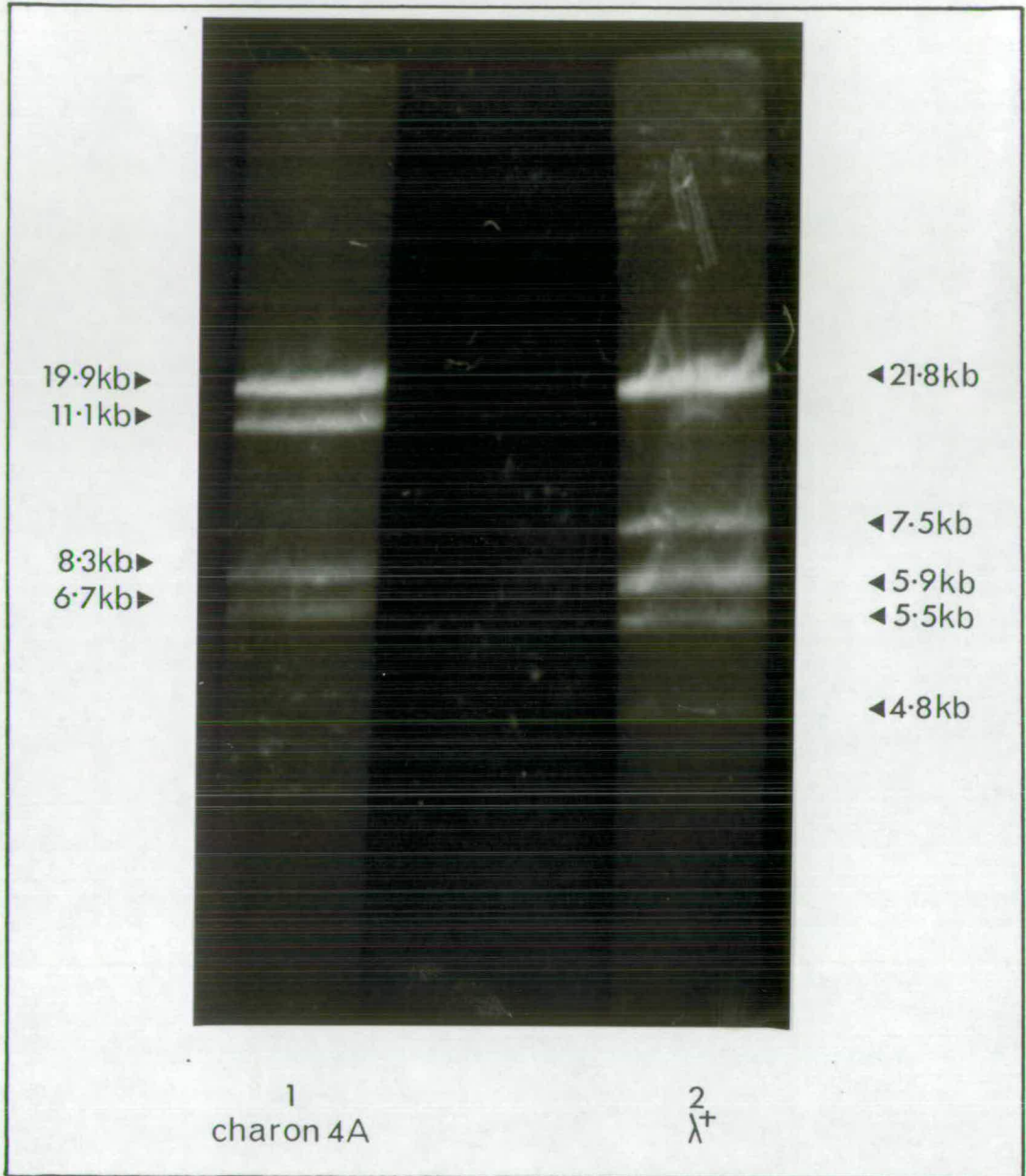
FIGURE 3.7: Separation of DNA Fragments by Agarose Gel Electrophoresis.

Double stranded DNA fragments of a wide molecular weight range
can be separated by electrophoresis on 1% neutral agarose gels and
visualised by staining with ethidium bromide. This figure shows
separation of the fragments obtained following EcoR1 digestion of
charon 4A DNA (lane 1) and λ⁺ DNA (lane 2).

It is shown in Figure 3.2c that the polysomes isolated from a single fraction (f2) taken from a polysome gradient, as described in Section 2.2.1, predominantly directed the synthesis of a single radioactively labelled polypeptide of 200,000 molecular weight, which co-electrophoresed with a myosin heavy chain marker, in a cell-free translation system. The translation system was a homologous system. That is, it was developed from chick skeletal muscle, the same source as the polysomes. However, it was shown that no endogenous polysomes were contributing to the translation assay as there was no incorporation of radioactivity into protein if water replaced the exogenous polysomes. Also, the patterns arising from the translation of several different heavy polysome fractions were very different, as shown in Figures 3.2a-f.

The purity of the polysome fraction f2 was further confirmed when mRNA was purified and translated in two heterologous systems developed from wheat germ and rabbit reticulocytes. Figure 3.4 demonstrates that several of the heavy polysome fractions are able to direct the synthesis of a protein which co-electrophoreses with myosin heavy chain, but only in the f2 fraction is this the major protein. Faint minor bands were often identified on the autoradiographs after electrophoresis of the f2 mRNA translation products, especially if the translation had been performed in the rabbit reticulocyte lysate system. The low molecular weight bands represent endogenous mRNA activity of the lysate since they were always present, even when water replaced the mRNA. The presence of the higher molecular weight contaminants was never consistent between different batches of f2 mRNA and, therefore, these are most likely to represent various degradation products of both the large mRNA molecule and the newly synthesised myosin heavy chain polypeptide.

Patrinou-Georgoulas and John (1977), in this laboratory, have
previously assessed the purity of the f2 polysome fraction, isolated
in an identical manner to that described here, more rigorously than was
carried out in this project: firstly, the mRNA isolated from the f2
fraction was visualised as a single band in formamide-polyacrylamide
gels where it runs at 32S, and was shown to migrate as a single peak
at 26S in sucrose gradients as detected by hybridisation to ^{3}H-polyU.
This enabled polyA containing material to be completely distinguished
from any contaminating ribosomal RNA which has a component which
migrates at 28S in such gradients. It is shown in Figure 3.5 that
f2 mRNA also migrates as a single species on agarose gels under
denaturing conditions with a molecular weight similar to that of 28S
rRNA. Secondly, following the *in vitro* translation of the f2 mRNA,
the translation products were not only shown to co-electrophorese with
myosin heavy chain marker protein, but 90% of the incorporated radio-
activity was precipitated with carrier at low ionic strength, and co-
chromatographed with myosin heavy chain on DEAE sephadex 50 after
precipitation at 38-50% saturation ammonium sulphate.

There are other high molecular weight proteins, present in
considerable amount in skeletal muscle, such as M-line protein (180,000
daltons) and C-protein (150,000 daltons), and, as expected, on trans-
lating a larger portion of the heavy polysomes (Figures 3.2a-f).
several high molecular translation products were observed. Myosin
heavy chain appears to make up a large proportion of the translation
products of total chick embryonic skeletal muscle polysomes (Figure
3.2a) and this may be as great as 36% (Patrinou-Georgoulas and John,
1977). However, since the f2 polysomes only gave rise to a single
protein possessing the characteristics of myosin heavy chain, as

determined here and previously, this was used as the source of myosin heavy chain mRNA.

The size of the cDNA probe was calculated following denaturing agarose gel electrophoresis in parallel with markers of known size. The final probe was estimated, as shown in Figure 3.6, to be 550 nucleotides in length when labelled with tritium, but 10% shorter at 500 nucleotides when labelled with ^{32}P. This difference may be explained if the ^{32}P had a mild destructive effect on the mRNA causing earlier chain termination during reverse transcription and thus a shorter average lengthened cDNA. However, this difference in size is negligible when compared with the length of the myosin heavy chain mRNA which has been estimated at 5830 nucleotides (Sarkar *et al.*, 1973). However, even the 10% copy of 530 nucleotides obtained by Patrinou-Georgoulas and John was shown to be quite specific since it cross reacted to an extent of 85% with myosin heavy chain mRNA, but to only 6% with purified globin mRNA.

Attempts were made to increase the length of the cDNA by increasing the concentration of the deoxynucleotide triphosphates to 600 μM, as described by Efstatiadis *et al.* (1975) and Monahan *et al.* (1976), and by employing milder hydrolysis procedures than those described in Section 2.2.8 which, it was suspected, may lead to some cDNA degradation. None of the modifications tried, such as lower temperature (Robbins and Heywood, 1976), use of RNase instead of alkali to remove the template (Sullivan *et al.*, 1973) and a combination of both, was found to increase the cDNA length significantly. Therefore, the 10% copy of 500 nucleotides was accepted as the probe to be used in following experiments.

CHAPTER IV

Construction of a Chicken Genomic Library

CHAPTER IV

CHAPTER IV (cont)

4.1 Introduction

This chapter describes in detail the methods employed in the construction of the chicken genomic library. The theory behind the manipulation and development of the cloning vector and the general methodology which has made the production of genomic libraries possible (such as the use of restriction enzymes and rapid screening techniques) have been discussed in general terms in Chapter I, Section 1.3.

The library was constructed according to the method developed by Maniatis and his co-workers (1978) for the construction of libraries of the drosophila, silkmoth and rabbit genomes. Where modifications were introduced they are discussed in the appropriate section. Briefly, the general strategy required three developments: firstly, the isolation of clonable DNA fragments from chicken which were representative of the total genome. Secondly, the covalent joining of these fragments to a suitable cloning vector, and, thirdly, introduction into a host cell to obtain a large number of independent recombinants.

The detailed basic scheme is illustrated diagrammatically in Figure 4.1. Genomic chicken DNA was extracted from the liver and fragmented by performing a mixed restriction enzyme digestion using two enzymes, Hae III and Alu I, both of which generate blunt-ended molecules. The reaction was not allowed to go to completion, thus generating a series of random fragments. Maniatis *et al*. (1978) attempted to fragment DNA by mechanical shearing followed by trimming single-stranded ends using S1 nuclease. The enzymatic method, however, was shown to be more efficient and hence was employed here. DNA fragments of approximately 20 kb in length are ideal for cloning in the vector, Charon 4A. These were, therefore, selected by sucrose gradient centrifugation and gel

High molecular
weight chicken DNA   Charon 4A DNA

digest with restriction enzymes   anneal cohesive ends

△ R1

size fractionate

20 kb DNA fragments

R1   R1

methylate EcoR1 sites   treat with EcoR1

CH₃   CH₃

CH₃ CH₃   +

attach synthetic linkers   size fractionate
retain large fragments

CH₃   CH₃

CH₃ CH₃

treat with EcoR1

CH₃   CH₃

CH₃ CH₃   Anneal and ligate cohesive ends

CH₃ CH₃   CH₃   CH₃

CH₃   CH₃ CH₃

In vitro packaging into
viable phage particles

FIGURE 4.1:   Diagrammatic scheme to show the strategy which was
followed in the construction of a random genomic
library of chicken DNA fragments, cloned in the
phage λ vector, charon 4A.

electrophoresis, and were then rendered resistant to the restriction endonuclease, EcoR1, by treatment with EcoR1 methylase. It was now possible to attach synthetic DNA linkers, which carry the EcoR1 recognition site, to these fragments and to generate cohensive ends by treatment with EcoR1 without destruction of the chicken DNA fragment.

Vector DNA was acquired in large quantities by propagation of charon 4A phage through its lytic growth cycle in an appropriate host bacterium. Under appropriate conditions the cohesive ends (cos site) of the charon 4A DNA molecules were annealed and the circles subjected to EcoR1 restriction endonuclease digestion. This results in three fragments (31 kb, 8 kb and 7 kb) which were fractionated on a sucrose gradient. The two smaller, internal fragments contain genes which are not essential for phage viability and, therefore, were discarded to provide 'physical space' for a eukaryotic DNA insert. Since all DNA fragments had been manipulated to contain EcoR1 staggered ends, it was possible to covalently join the chicken DNA fragments, via the attached linkers, to the 31 kb charon 4A fragment. The chicken DNA had replaced the two smaller charon 4A fragments but the total length of the recombinant molecule was approximately equal to the native phage genome. It is possible to replace these internal EcoR1 fragments of charon 4A with DNA fragments in the size range 8.2 to 22.2 kb without hindering the packaging efficiency and viability, though packaging will not take place if the molecule is above or below these size limits due to certain physical restrictions. Thus, a positive selection method for recombinant phage is endogenous to the packaging procedure. Packaging of recombinant molecules into viable phage particles was carried out *in vitro* using the Hohn and Murray (1977) procedure.

The essential feature of this scheme developed by Maniatis *et al*. (1978) for gene isolation is that a permanent library is established which can be repeatedly screened. Although the aim of a cloning experiment may be to isolate a specific gene and prior enrichment of this sequence, if possible, will make the cloning procedure simpler and faster, when constructing a gene library, in general it is not desirable to enrich for a specific gene, since this will eliminate other DNA sequences which may be required in the future.

The incorporated chicken DNA fragments were, therefore, amplified by growing the recombinants in a suitable bacterial host and storing as a plate lysate. In this way, a permanent library of the chicken genome was established from which, in principle, any gene sequence could be isolated with a specific hybridisation probe. In Chapter·V, the screening procedure is described with specific reference to the myosin heavy chain gene using the specific ^{32}P-labelled cDNA probe.

All techniques involving the growth of bacteria and phage must be carried out under sterile conditions to prevent bacterial and phage cross contamination from the immediate environment. The methods described in this chapter, therefore, were all carried out using auto-claved, or filter sterilised (through 0.22 μm filters), solutions and glasswear pre-heated to 150°C for several hours, usually overnight. The tops of flasks, bottles and tubes and all pipettes were rigorously flamed using a strict microbiological technique when in use, especially if bacterial growth was involved. Bacteriological grade plastic petri dishes (Sterilin) were purchased pre-sterilised and used throughout.

4.2 Growth of Bacterial Strains

The host bacterium for all strains of bacteriophage λ is
Escherichia coli, and in almost all recombinant DNA research, mutants
of *E. coli* K12 have been used. These hosts lack a restriction system
and therefore do not degrade foreign, unmodified DNA when it is
introduced into the cell. Two strains of *E. coli* were used here: C600
for the propagation of the wild type phage $\lambda(\lambda^{+})$ and DP50supF for the
propagation of the vector phage, charon 4A.

Both strains were stored at a concentration of 10^9 cell/ml at $-20°C$
in L-broth containing 12% glycerol.

To grow the C600 strain, 50 ml of L-broth, supplemented with
1 mM magnesium sulphate were inoculated with a loopful of the bacterial
stock in a 500 ml conical flask. This was incubated in an orbital shaker
with moderate shaking for approximately 14 hours, or overnight. A very
cloudy bacterial suspension was evident after this incubation period
which contained in the region of 2×10^9 cells/ml.

The DP50supF strain was grown in an identical manner, except
that it was necessary to supplement the L-broth with 40 µg/ml thymidine
and 10 µg/ml diaminopimelic acid (DAP) as well as 1 mM magnesium sulphate.

4.3 Growth of Bacteriophage λ

Large quantities of bacteriophage, both λ^+ and charon 4A, were
obtained, as a source of DNA, by propagation through the normal lytic
cycle in their respective hosts.

4.3.1 *Preparation of Phage Liquid Lysates*

An overnight bacterial culture was diluted by 1 : 50 into L-broth supplemented with 1 mM magnesium sulphate (and 40 μg/ml thymidine, 10 μg/ml DAP if charon 4A or recombinant phage were to be propagated) to give an optical density reading at 650 nm (OD_{650}) of 0.1, corresponding to 4×10^7 cells/ml. It was necessary to use a large conical flask with a cotton wool stopper to allow for vigorous aeration of the culture, which is essential for phage growth. Typically, 100 ml of L-broth were contained in a 1 litre flask. The culture was incubated at 37°C in an orbital shaker while the OD_{650} was followed until it reached 0.5 (or 2×10^8 cells/ml).

Bacteriophage were added to the bacterial culture, either wild type or charon 4A to their respective hosts to give a multiplicity of infection of one (m.o.i.) (i.e. one viable phage/bacterial cell). The incubation was continued with vigorous shaking to ensure rapid, continuous aeration of the culture for approximately $2\frac{1}{2}$ hours. The OD_{650} was monitored throughout this period as it rose (often as high as 1.8) and then, as the lytic cycle of the phage neared its end, the OD_{650} would fall as the bacterial cells were lysed. It was important to follow the OD_{650} during this period as the lysis occasionally occurred after $1\frac{1}{2}$ hours of phage infection, but could sometimes take up to 4 hours if the aeration of the culture was not optimum. When the OD_{650} was at a minimum chloroform was added to 0.2% to complete the cell lysis and the culture was shaken rapidly for 10 min.

Figure 4.2 is a graphic representation of the lytic cycle of both λ^+ and charon 4A when grown in their respective *E. coli* hosts, C600 and DP50supF, as described here.
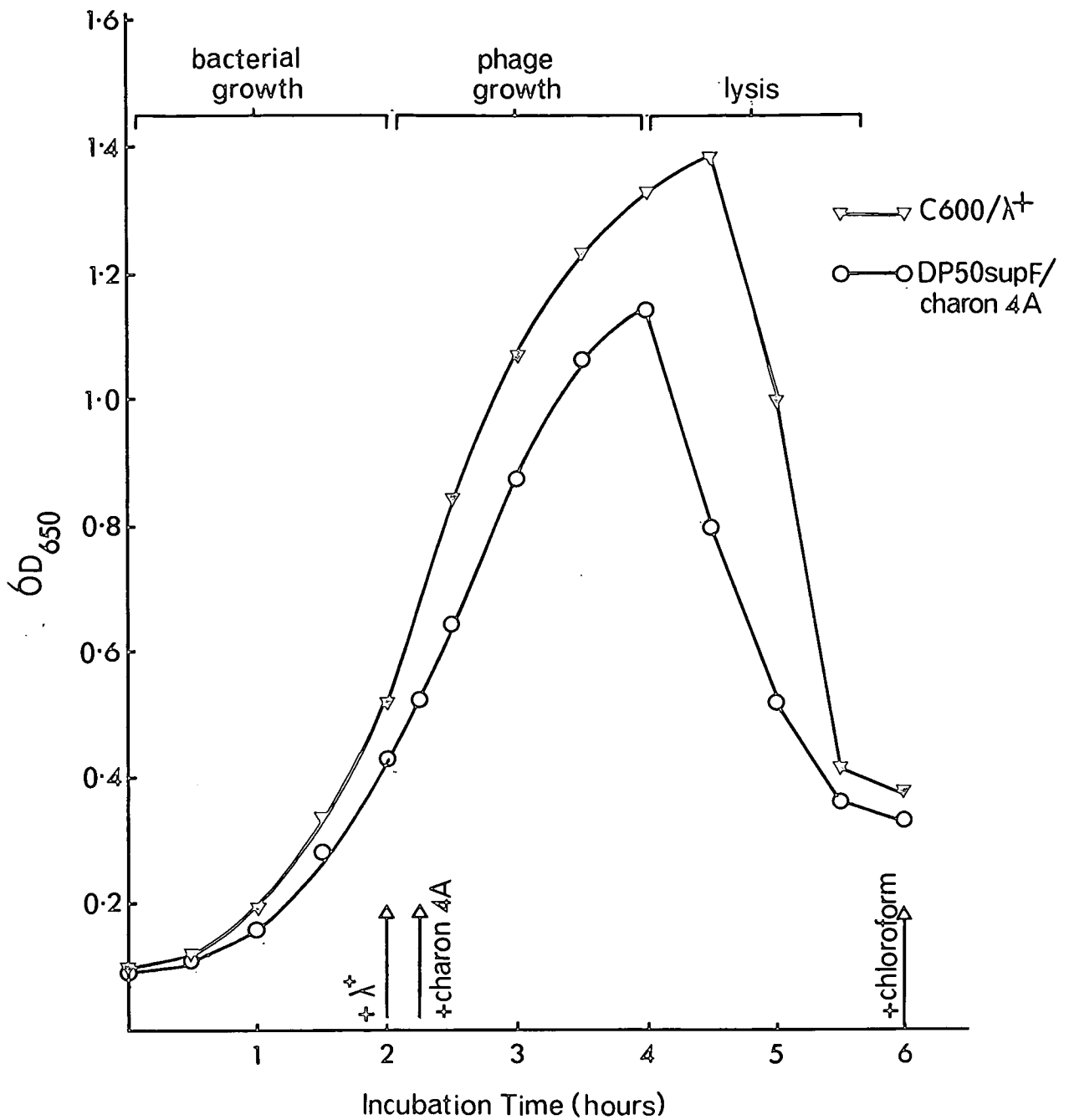
FIGURE 4.2: Growth of Wild Type Bacteriophage λ and the Vector Charon 4A.

λ^+ and charon 4A were propagated through their respective hosts, *E. coli* C600 and DP50SupF, as described in Section 4.3.1. The growth, and subsequent lysis of the host cells was monitored at 650 nm both before and during 'phage infection. The lytic cycle is represented graphically in this figure.

4.3.2 *Precipitation of Phage with Polyethylene Glycol 6000*

The chloroform treated lysate was made 4% with sodium chloride. 10 µg/ml DNase and RNase were added to remove bacterial nucleic acids and the lysate was left to stand at room temperature for 1 hour. The liquid was clarified by centrifugation at 10,000 rpm for 10 min in the Sorval SS30 rotor when bacterial debris pelleted leaving the bacterio- phage in suspension.

Solid polyethylene glycol 6000 was added to the supernatant to give a concentration of 10% w/v, and was dissolved with gentle swirling. The solution was then left to stand at 4°C for at least 2 hours, but often overnight, as it turned more and more opaque as the phage particles precipitated. The bacteriophage were recovered by centrifugation at 10,000 rpm for 20 min. The supernatant was now clear and the pellet was resuspended in phage buffer by gentle rotary shaking at 4°C. Typically, the phage pellet obtained from 100 ml of culture was re- suspended in 4 ml phage buffer.

Concentration of bacteriophage through caesium chloride step gradients:

After polyethylene glycol 6000 precipitation, the phage suspension was still contaminated with bacterial and phage debris. This was removed by centrifugation of the suspension through caesium chloride step gradients when the various components band at different points in the gradient according to their densities.

A three-step gradient was prepared in a 12 ml nitrocellulose tube consisting of 1.5 ml steps of caesium chloride in phage buffer of 1.3 g/ ml, 1.45 g/ml and 1.65 g/ml, as shown in Figure 4.3. The lowest density solution was placed in the tube initially and the denser steps were gently layered underneath using a syringe and fine needle. 7 ml of the phage sample was layered onto one gradient and centrifugation was at 35,000 rpm for 1½ - 2 hours at 20°C in a Beckman SW4 rotor.
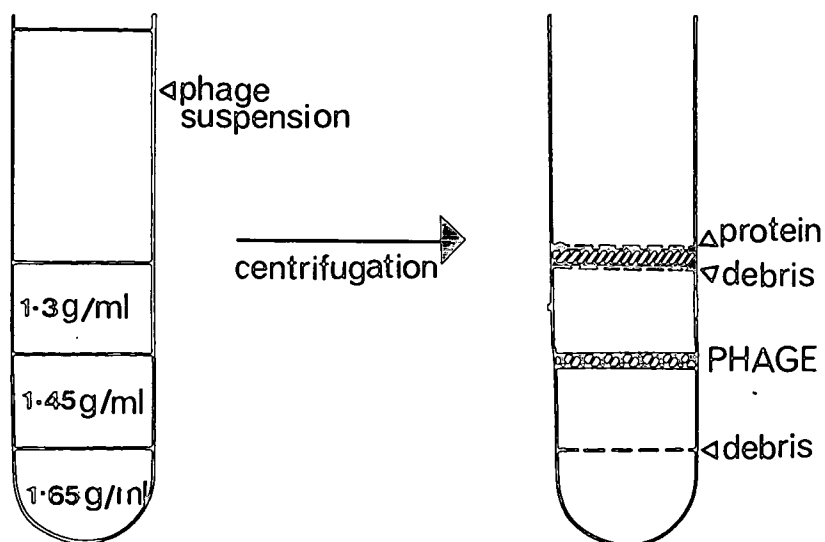
FIGURE 4.3: Partial Purification of Bacteriophage by Caesium
Chloride Centrifugation.

Bacteriophage were routinely purified from contaminating bacterial
debris by centrifugation through a three-step caesium chloride
gradient, as described in Section 4.3.2. The bacteriophage form a
white, sharp band at the interface between the 1.3 g/ml and the
1.45 g/ml caesium chloride steps. Protein debris and DNA remain
above the caesium chloride, and occasionally some further debris was
observed near the bottom of the tube, but this did not contain viable
phage particles.

The phage band was readily identified since it was always sharp and white in contrast to the two bands of debris above and below it, which tended to be more difuse and discoloured.

The white phage band was carefully removed with a hooked pasteur pipette and could be stored indefinitely in the caesium chloride solution at 4°C.

4.3.3 Titration of Phage

It was necessary throughout the whole cloning procedure, from the initial preparation of λ^+ and charon A4 liquid lysates to the amplification and screening of the library, to be familiar with the numbers of phage produced at each stage.

The titration of phage numbers is a standard procedure carried out using the *plaque assay* technique which is based on the principal that plaques corresponding to areas of phage growth will arise in lawns of dense bacterial growth on nutrient agar surfaces. The plaques are regions where bacteria have lysed due to the growth of phage colonies descended from parent phages present in the initial bacterial inoculum. Since only one phage may infect one bacterium, a plaque is generated from one single infective unit so that the number of plaques on a bacterial lawn is directly proportional to the number of infective units contained in the original phage suspension, and therefore is useful in a quantitative assay.

Fresh overnight bacterial cultures of the strain necessary for the propagation of the phage to be titred were used. 10^8 cells (i.e. about 50 µl of the overnight suspension) were shown to form a thick bacterial lawn on a 10 cm petri plate containing L-agar supplemented with 1 mM magnesium sulphate to encourage phage growth, so this number of bacteria were considered adequate for routine titration assays.

Various dilutions of the phage stocks were made in 1 ml aliquots of phage buffer. A typical range was from 10^{-2} (i.e. 10 µl of the stock) to 10^{-7}. 10^8 bacterial host cells were added and the infection was allowed to take place by incubating at 37°C for 15 min. 3 ml of moulton soft top agar at 45°C was added and the contents gently mixed by inversion of the tubes. Immediately, the mixture was poured over a prepared petri-plate containing L-agar and allowed to solidify. The L-agar plates were always prepared at least 1 day in advance to allow moisture which always accumulated on the surface to dry. Plates were incubated at 37°C, overnight in an inverted position to prevent spreading of the phage by the accumulation of moisture. A control plate minus added phage was always included to test for contamination of the bacterial stock.

The infective titre of the phage suspension per millilitre was calculated by choosing a plate with a reasonable number of plaques, usually around 100 and multiplying that number by the dilution made before plating.

The recovery of viable phage from a single caesium chloride step gradient was usually in the region of 3-2 x 10^{12} λ^+ infective units and an order of magnitude less at about 3.5 x 10^{11} charon 4A infective units.

4.4 Preparation of DNA

4.4.1 Extraction of Phage DNA from Liquid Lysates

The phage suspension which had been purified by step gradient centrifugation was dialysed against 10 mM Tris pH 8, 1 mM EDTA for at least 1 hour at 4°C to remove caesium chloride. The phage were then diluted to a concentration of 2 x 10^{10} plaque forming units per ml, with the same Tris/EDTA buffer. If the phage concentration is too high, it is difficult to achieve efficient phenol extraction.

Freshly distilled phenol was pre-equilibrated with 0.5M Tris-HCl pH 8, and the phage suspension was extracted with an equal volume of the phenol by gentle inversion and rolling of the tube (but not shaking) for 15 min at room temperature. The aqueous and phenol phases were separated by centrifugation at 10,000 g for 10 min at 4°C and the aqueous phase was extracted with fresh phenol three more times. Finally, the aqueous phase was dialysed against 10 mM Tris-HCl pH 8, 1 mM EDTA to remove traces of organic solvents, with several changes of buffer over a 24-hour period. The optical density was measured at 260 nm, 280 nm, 235 nm and 320 nm. For a clean DNA preparation, the ratio $OD_{260}: OD_{280}$ should be 2 : 1, and the $OD_{260}: OD_{235}$ ratio should be higher. The absorbance at 320 nm is ideally zero since the reading at this wavelength is an indication of particulate protein matter in the solution. Under these conditions, the absorbancy at 260 nm gives the concentration of the DNA according to the relationship

$$1 \ OD_{260} = 50 \ \mu g/ml \ DNA$$

The DNA solution was made 0.2M sodium chloride and was concentrated by precipitation at -20°C in 66% ethanol for 24 hours, followed by centrifugation at 10,000 g for 30 min at -10°C. The DNA pellet was taken up in 10 mM Tris pH 7.5, 0.1 mM EDTA to a concentration of 1 mg/ml.

4.4.2 *Preparation of Phage DNA from Plate Lysates*

10 cm plates of L-agarose supplemented with 1 mM $MgSO_4$ were prepared. Agarose was used in preference to agar as the growth medium here because contaminants of agar are known to inhibit restriction enzyme activities, and such contaminants may be carried through the DNA extraction process to inhibit such reactions later. Phage were plated at a concentration of 5×10^5 pfu/plate (this number can be obtained from a single plaque of wild type λ) in soft top agarose in a similar manner to that used for titration of phage, except that the plates could be used fresh as there was no need to wait for surface moisture to evaporate. Plates were incubated, inverted at 37°C for approximately 8 hours or until confluent lysis was achieved. Well cleared L plates are necessary for DNA preparations. A plate without phage was always included to check for contaminants and as an indication of the amount of bacterial growth in the absence of phage.

Cleared plates were overlayed with 5 ml 10 mM Tris, 10 mM EDTA pH 7.5 and left overnight at 4°C. The solution was pipetted off and DNase and RNase were added to give a final concentration of 10 μg/ml to remove the *E. coli* DNA and RNA which was released following cell lysis. The solution was allowed to stand for 30 min at room temperature. To each 4 ml of supernatant were added 0.4 ml of 0.5M EDTA pH 8.5, 0.2 ml 2M tris base and 0.2 ml 10% SDS, mixed and placed on ice. A white precipitate formed. 10 μl of diethylpyrocarbonate was added and the mixture was heated at 65°C for 30 min in open tubes in a fume cupboard. The precipitate dissolved. The mixture was chilled and 1 ml of 5M potassium acetate (unbuffered) was added and chilled for 1 hour when the white precipitate formed again. Centrifugation was at 25,000 g in an 8 x 50 rotor for 10 min when the nucleic acids remained in the

supernatant. This was decanted and precipitated at -20°C overnight in 70% ethanol. The DNA was pelleted by centrifugation at 10,000 g for 30 min and dissolved in 10 mM Tris, 1 mM EDTA pH 7.5 at a concentration of 1 mg/ml.

The DNA prepared in this way tended to be slightly contaminated with *E. coli* DNA (and probably RNA) as determined by EcoR1 restriction endonuclease analysis, in spite of the enzyme treatment. DNA prepared from liquid lysates was much purer and therefore was the method of choice for the preparation of vector DNA to be used in the construction of the library.

4.4.3 Preparation of Chicken Liver DNA

The chicken DNA was prepared by a modification of a procedure, described originally by Marmur (1961), which isolates DNA in a high molecular weight form. The method basically involves the disruption and lysis of cells, the removal of protein and debris by denaturation and centrifugation, removal of RNA using ribonuclease and the selective precipitation of DNA with isopropanol. Chelating agents and sodium dodecyl sulphate (SDS) help to reduce deoxyribonuclease activity.

10 g of fresh chicken liver were washed in 50 ml saline-EDTA at room temperature, and then homogenised in a glass/teflon motor driven homogeniser in small pieces (2-3 g) in a total volume of 100 ml of fresh saline-EDTA. The homogenate was made 2% with respect to SDS to ensure complete lysis of cells and nuclei and the mixture was placed in a 60°C waterbath for 10 min and then cooled to room temperature. The mixture was now extremely viscous. Sodium perchlorate (5M) was added to a final concentration of 1M to help dissociate proteins from the nucleic acids. The mixture was shaken with an equal volume of phenol/chloroform/isoamyl alcohol (50 : 50 : 1) for 30 min at room temperature

and then centrifuged at 5,000 rpm in a Sorvall SS-34 rotor. This separated the mixture into three layers: an upper aqueous layer containing the nucleic acids, a middle protein layer and a lower phenol layer. The upper aqueous layer was gently pipetted off and re-extracted with the phenol mix several times until the middle protein layer was not present following centrifugation. Nucleic acids were precipitated by gently layering two volumes of ethanol over the aqueous phase and stirring gently with a glass rod. The nucleic acids precipitated at the interphase of the aqueous and ethanol phases and spooled onto the rod as a thread-like precipitate. This was drained free of excess ethanol by pressing against the side of the beaker. The spool was gently removed from the rod and allowed to dissolve in 50 ml of dilute saline citrate (DSC) at 4 °C. This took several days.

The solution was adjusted to standard saline citrate (SSC) and RNase was added to final concentration of 50 µg/ml. Incubation was at 37°C for 30 min. The mixture was again treated with phenol/chloroform/ isoamyl alcohol as before and the supernatant was treated with protein-ase K at 20 µg/ml for 1 hour at 37°C. Deproteinisation was completed by several more treatments with the phenol mixture until no protein was visible at the interphase following centrifugation. A final extraction was then carried out using chloroform/isoamyl alcohol (24 : 1). The aqueous phase was precipitated with ethanol by spooling as described and was dissolved in 25 ml DSC. 1 ml of acetate-EDTA was added and while stirring with a glass rod 13.5 ml (0.54 vols) of isopropyl alcohol were added. RNA, oligoribonucleotides and polysaccharides remained in solution during this treatment while DNA precipitated through a gelatinous to a thread-like state and wound onto the glass rod.

The final precipitate was washed free of acetate, salt and organic solvents by stirring the adhered precipitate in increasing (70-95%) concentrations of ethanol. The DNA was stored in DSC at -20°C at a concentration of 1 mg/ml. The concentration was determined according to the absorbance at 260 nm and using the relationship 1 OD_{260} = 50 μg/ml of DNA. The yield was 10-15 mg of DNA from 10 g of tissue.

4.5 Restriction of DNA

Restriction endonucleases are a group of enzymes, extremely useful in nucleic acid research which have an endonuclease activity in that they are able to cut double-stranded DNA molecules internally to produce many fragments. The phenomenon was first observed by Meselson and Yuan in 1968, though the first specific enzyme was discovered by Smith and Wilcox in 1970. There are now 250 reported examples, though it is not certain that these are all in fact different enzymes.

Restriction enzymes occur naturally in bacteria which are able to modify their own DNA by methylation as protection against such endo-nucleolytic attack. Smith and Wilcox (1970) observed that a cell extract from *Heamophilus influenzae* degraded DNA from the bacteriophage P22 but had no effect on the DNA from *Heamophilus* itself. Therefore, infection of a bacterium by a piece of foreign DNA, such as that of a phage, generally results in destruction of that DNA unless it also carries the modified sequences.

Restriction enzymes are normally named according to their bacterial source, a convention developed by Smith and Nathans (1973). For example, Eco from *Escherichia coli*, Hae from *Heamophilus aegyptius*, and Alu from *Arthrobacter luteus*. They are classified as type I, II

or III. Types I and III do not cleave DNA at specific sequences and therefore are not particularly useful in genetic engineering. Type II enzymes, however, cleave at, or very close to a defined recognition sequence and have been exhaustively used in recombinant DNA work. It is three particular type II enzymes which were very important in the construction of the chicken genomic library described here. Their specific recognition sequences and sites of cleavage are shown on Figure 4.4.

Most restriction enzymes, though not all, recognise a specific palindromic sequence of 4 to 8 base pairs and cut the DNA of both strands within this sequence. Thus, in a random DNA molecule which will contain 50% A-T and 50% G-C base pairs, a particular tetranucleotide sequence will occur every 256 base pairs and a hexanucleotide sequence will occur every 4096 base pairs. However, most DNA is not 'random' and therefore the number of fragments produced from a restriction enzyme digestion which is allowed to proceed to completion will not be random and will depend exclusively on the number of sites which the enzyme recognises within that molecule.

Some restriction enzymes cut both strands of the double-stranded DNA molecule at the same position and thus leave fragments with blunt ends, while others cut the two strands at positions which are not opposite and thus form fragments with staggered but complementary ends (with either 5 or 3' ends protruding). These two groups of enzymes have special uses in the field of genetic engineering. Staggered ends allow the simple joining of molecules from different sources but which have been produced by digestion with the same enzyme, by virtue of the complementary nature of their protruding single-stranded ends. However, as will be discussed in Section 4.5.3, this may not

Eco R1

3'A/T-C-T-T-A-A-G-T/A 5'
5'T/A-G-A-A-T-T-C-A/T 3'

Alu 1

3'T-C-G-A 5'
5'A-G-C-T 3'

Hae III

3'C-C-G-G 5'
5'G-G-C-C 3'
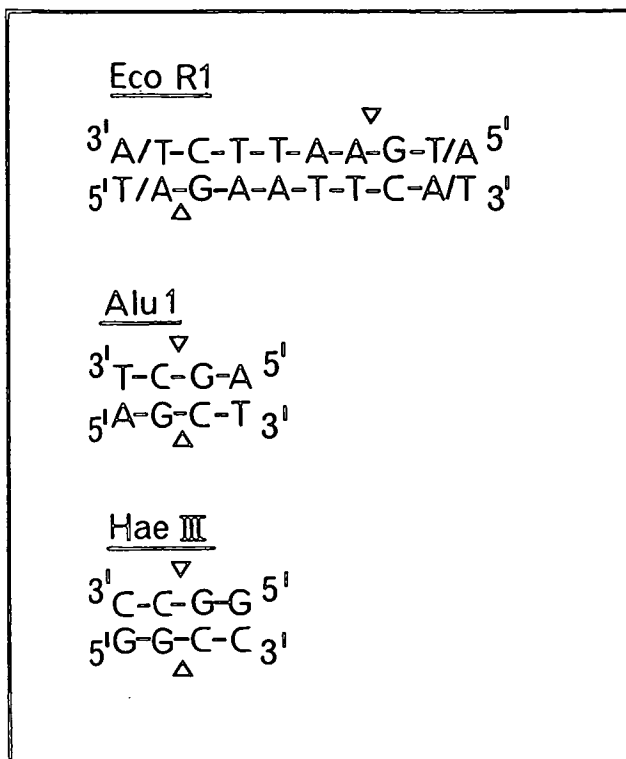
FIGURE 4.4: Recognition Sequences of the Restriction Endonucleases EcoR1, AluI and HaeIII.

This figure shows the recognition sequences of the three restriction endonucleases used extensively throughout this project. EcoR1 (from *Escherichia coli*) cuts to produce staggered 5' protruding ends (Hedgpath *et al.*, 1972; Green *et al.*, 1974) at the points indicated (▽). AluI (from *Arthrobacter luteus*) and HaeIII (from *Haemophilus aegyptius*) give rise to blunt ended molecules (Roberts *et al.*, 1976; Middleton *et al.*, 1972). The significance of each type of enzyme is discussed in Section 4.5.

always be the ideal way of recombining DNA molecules, and often blunt ends are preferable so that DNA fragments produced by the action of different enzymes can be linked together and ligated using T4 ligase.

The enzymes usually have broad pH optima (for example, from 6.5 to 8.5) and require only magnesium. However, salt concentration is often very critical as high ionic strengths tend to be inhibitory, and therefore it is important to follow the suppliers' directions for optimum activity.

4.5.1 Restriction of Bacteriophage λ DNA by EcoR1

DNA was prepared from phage liquid lysates of λ⁺ phage and restricted using the enzyme EcoR1 to provide a ready supply of molecular weight markers for routine use in agarose gel electrophoresis of DNA fragments during the cloning experiments.

Usually 1 μg of λ⁺ DNA was digested to completion in 10 μl of EcoR1 restriction buffer with 2 units of EcoR1 for 1 hour at 37°C. Sodium chloride is present in the buffer at high concentration to inhibit 'R1 star' activity, which is a secondary, much less active enzymatic site of EcoR1.

Reactions were terminated by heating the digests at 70°C for 10 min followed by rapid cooling on ice. The addition of excess EDTA will also stop the EcoR1 digestion but the heat step was considered more suitable to ensure dissociation of the λ DNA cohesive ends which may have annealed during the incubation. Samples were electrophoresed on neutral agarose gels as described in Section 3.3.3.

Bacteriophage λ⁺ genome has a molecular weight of 3.1×10^7 daltons which corresponds to a size of 49,000 base pairs (Schroeder and Blattner, 1978). The molecule contains 5 EcoR1 recognition sites so that a complete digestion produces 6 fragments as shown in Figure 4.5. A more

◀B 19·9 kb
◀C 11·1 kb

◀D 8·3 kb
◀E 6·7 kb

◀A 21·8 kb

◀B 7·5 kb
◀C 5·9 kb
◀D 5·5 kb

◀E 4·8 kb

1

charon 4 A

2

λ⁺
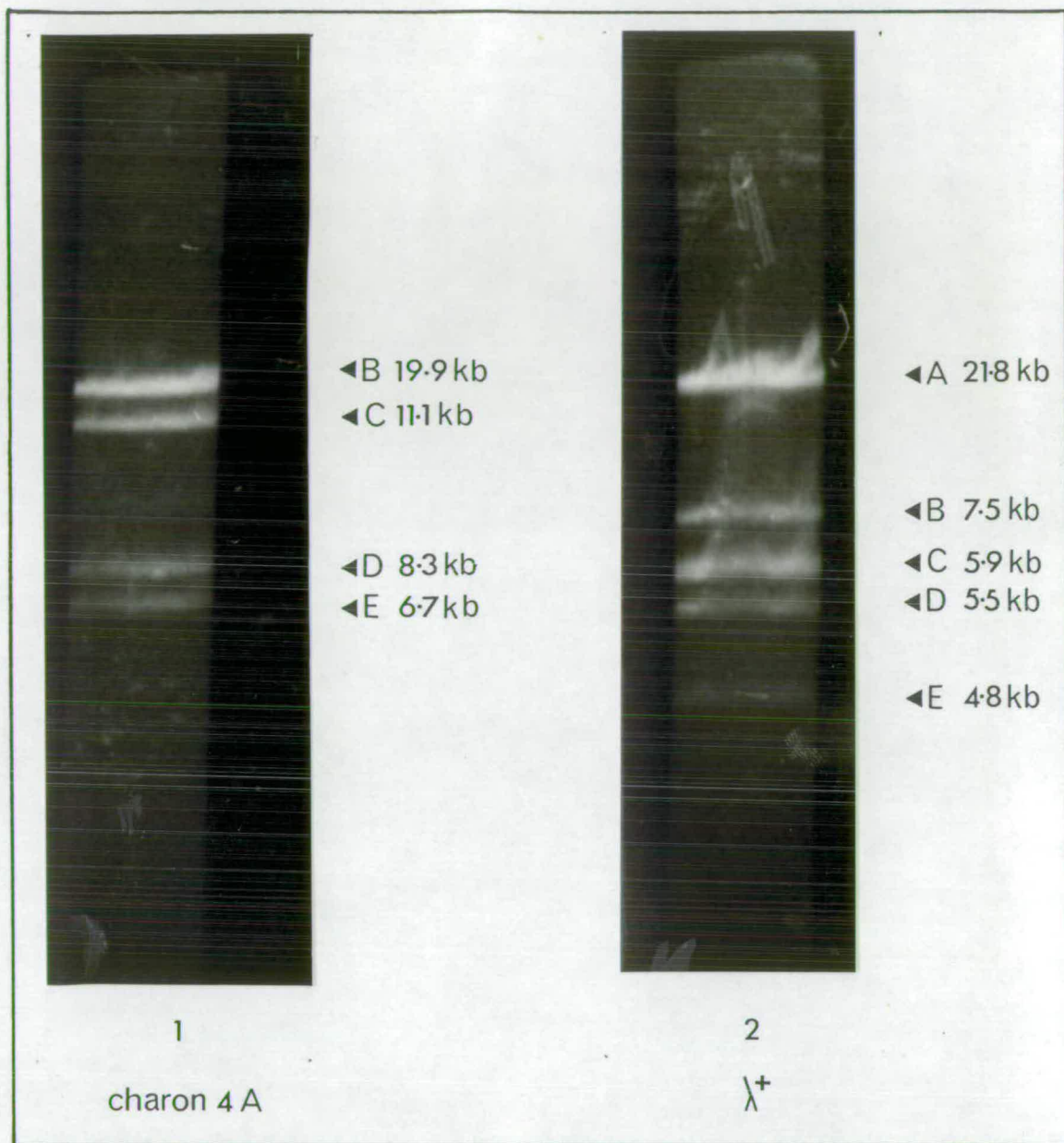
FIGURE 4.5: EcoR1 Digestion of λ^+ and Charon 4A DNA.

1 µg of charon 4A DNA (lane 1) and λ^+ DNA (lane 2) were digested
to completion by the restriction enzyme EcoR1. The fragments
produced were analysed by neutral agarose gel electrophoresis and
their sizes are indicated in this figure with further details presented
in Figure 4.6 and Table 4.1. The letters in the figure refer directly
to those in Figure 4.6 and Table 4.1.

TABLE 4.1: EcoR1 Digestion of λ^+ and Charon 4A Genomes.

Fragment (see Fig 4.6)	Molecular weight (daltons)	Molecular weight (base pairs)	% of λ^+ genome
(a) $\underline{\lambda^+}$			
λ^+	3.10×10^7	49,000	100
A	1.39×10^7	21,800	44.5
B	4.83×10^6	7,500	15.4
C	3.78×10^6	5,900	12.1
D	3.52×10^6	5,500	11.3
E	3.07×10^6	4,800	9.8
F	2.18×10^6	3,400	6.9
(b) Charon 4A			
Charon 4A	2.94×10^7	46,000	100
A	1.98×10^7	31,000	67.4
B	1.02×10^7	19,900	34.7
C	9.60×10^6	11,100	32.6
D	5.10×10^6	8,300	17.4
E	4.50×10^6	6,700	15.2

The information obtained following the EcoR1 restriction endonuclease digestion and neutral agarose gel electrophoresis of the λ^+ genome is presented in (a) and of the charon 4A genome in (b).
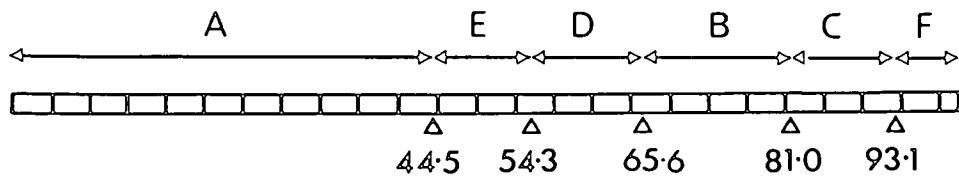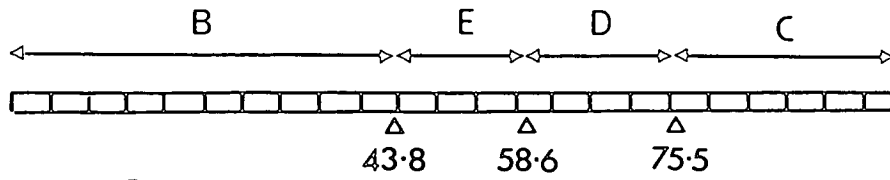
FIGURE 4.6:   EcoR1 Restriction Maps of λ^+ and Charon 4A Genomes.

The positions of the EcoR1 restriction sites (Δ) within the λ^+ (a) and the charon 4A (b) genomes are shown diagrammatically. The restriction sites are in map units where the total DNA molecule is 100 map units. The vertical lines in the figure are equivalent to 2 kb. More details are presented in Table 4.1.

detailed diagrammatic restriction map is presented in Figure 4.6 and Table 4.1a.

4.5.2 *Restriction of Charon 4A DNA*

(a) *Annealing of cohesive ends:*

Charon 4A DNA was prepared as described in Section 4.4.1. The cohesive ends of the linear molecules were first annealed by incubation of 100 µg of the DNA at a concentration of 1 µg/µl in 0.1M Tris-HCl pH 8.0, 10 mM magnesium chloride for 1 hour at 42°C. This gave rise to a series of circles and concatamers, as shown diagrammatically in Figure 4.7c, and following agarose gel electrophoresis, as shown in Figure 4.7a.

(b) *EcoR1 digestion:*

2-mercaptoethanol and sodium chloride were added to the concatamer and circle mixture to concentrations of 10 mM and 50 mM, respectively, with 200 units (an excess of EcoR1). The reaction mix was incubated at 37°C for 3 hours to ensure complete digestion. An aliquot was reserved for neutral agarose gel electrophoretic analysis. The DNA was extracted with water saturated distilled phenol by gentle inversion of the tube, and finally extracted with ether and precipitated at -20°C with two volumes of ethanol.

EcoR1 was the enzyme of choice for manipulation of the vector DNA because it has few target sites (only three) within the charon 4A genome and none of these lie within the gene sequences essential to the viability of the phage, and as discussed previously, the two internal fragments can be removed, also without loss of viability. In addition, the EcoR1 single-stranded protruding termini of the large, 31 kb, fragment which results from the enzyme digestion are utilised in the

final ligation to chicken DNA fragments to be discussed in Section
4.7.2.

Figure 4.7b shows an agarose gel of the EcoR1 fragments of
charon 4A in which the cos sites have not annealed (lane 1) and where
they have annealed (lane 2). Lane 1 contains four bands corresponding
to fragments of 16, 15, 8 and 7 kb and lane 2 contains only three
bands corresponding to 31 kb (due to the annealing of the cos site),
8 and 7 kb. Figure 4.6b is a diagrammatic representation of the
EcoR1 restriction map of charon 4A, the details of which are presented
in Table 4.1b.

The internal fragments of 7 and 8 kb were separated from the
31 kb fragment by sucrose gradient centrifugation. 70 µg of the
EcoR1 cleaved DNA in 10 mM Tris-HCl pH 8, 1 mM EDTA were layered
onto a 17 ml 10-40% linear sucrose gradient in 1M sodium chloride,
20 mM Tris-HCl pH 8, 10 mM EDTA and centrifuged in a Beckman L6
ultracentrifuge, SW27 rotor, at 27,000 rpm for 24 hours at 20°C.
0.5 ml fractions were collected from the bottom of the tube, and the
OD_{260} of each fraction was monitored as shown in Figure 4.8a. 25 µl
samples from those fractions corresponding the peak OD_{260} units were
electrophoresed on neutral 1% agarose gels to check the molecular weight
of the DNA fragments and then fractions 5, 6 and 7 were pooled as these
were shown to contain the 31 kb annealed arms of charon 4A (Figure
4.8b).

4.5.3 *Restriction of Chicken DNA*

It was necessary to fragment the chicken DNA, prepared as
described in Section 4.4.3, into lengths of approximately 20 kb.
charon 4A internal fragments can be replaced by sequences of between
8.2 and 22.2 kb, but it was decided to use inserts towards the upper

FIGURE 4.7: Annealing of Charon 4A DNA at the Cos Site.
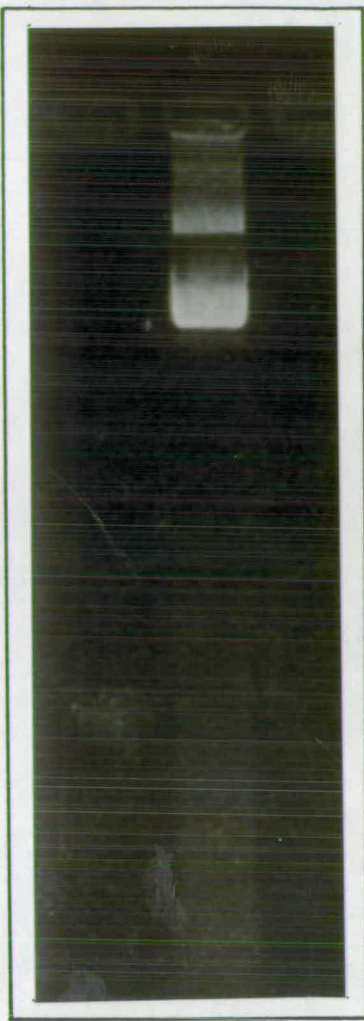
(a) The cos sites of the charon 4A DNA molecules were annealed as described in Section 4.5.2. This figure shows that molecules larger than intact charon 4A were present following annealing and therefore confirms that concatamers, as shown diagrammatically in c) had formed.

◄ 31 kb
◄ 19.9 kb
◄ 11.1 kb
◄ 8.3 kb
◄ 6.7 kb

1   2
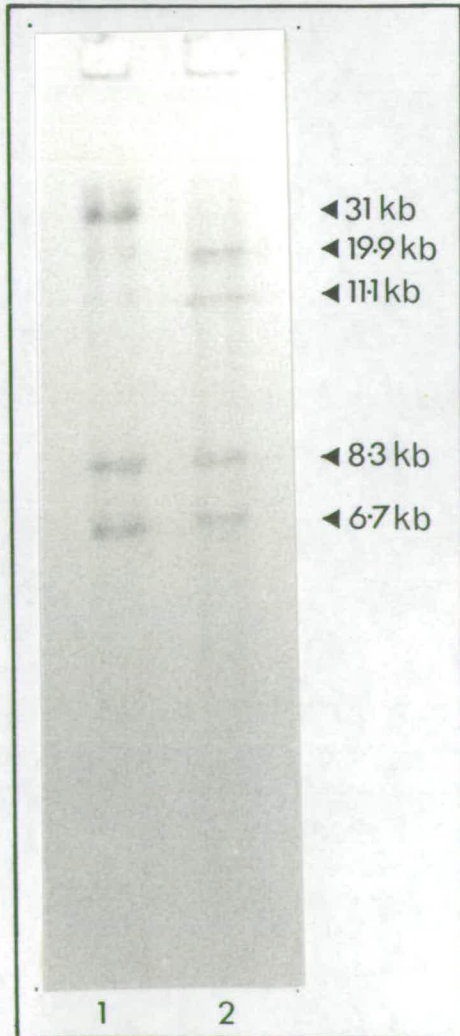
(b) Following EcoR1 digestion of the annealed charon 4A cos sites, a 31 kb fragment is evident as shown in lane 1. Lane 2 shows the digestion, under the same conditions, of charon 4A DNA which had not been annealed. There are traces of the 19.9 and 11.1 kb fragments in the annealed DNA sample indicating that not all molecules had formed concatamers or circles.

c.
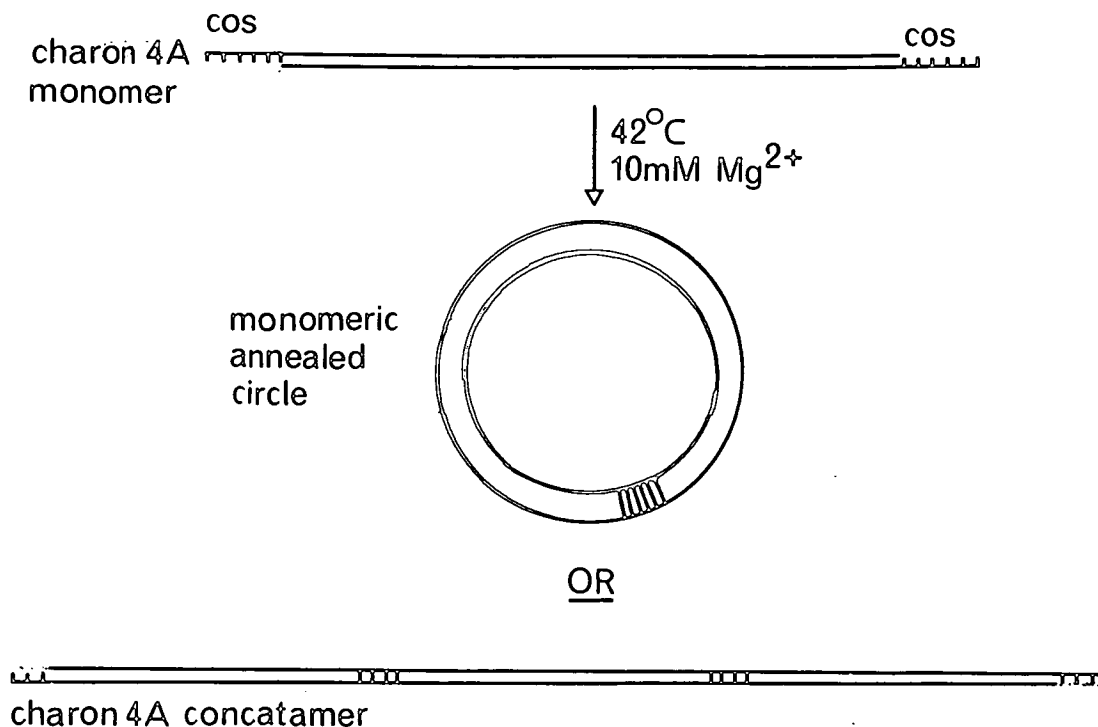
FIGURE 4.7: Annealing of Charon 4A DNA Molecules at the Cos Sites.

Diagrammatic representation of the charon 4A genome following
annealing of the cos sites. The two possible conformations are shown.
Circles will tend to form when the DNA is very dilute, but the
formation of concatamers will be encouraged in more concentrated
DNA solutions when molecules will be more likely to make contact with
each other.

a.
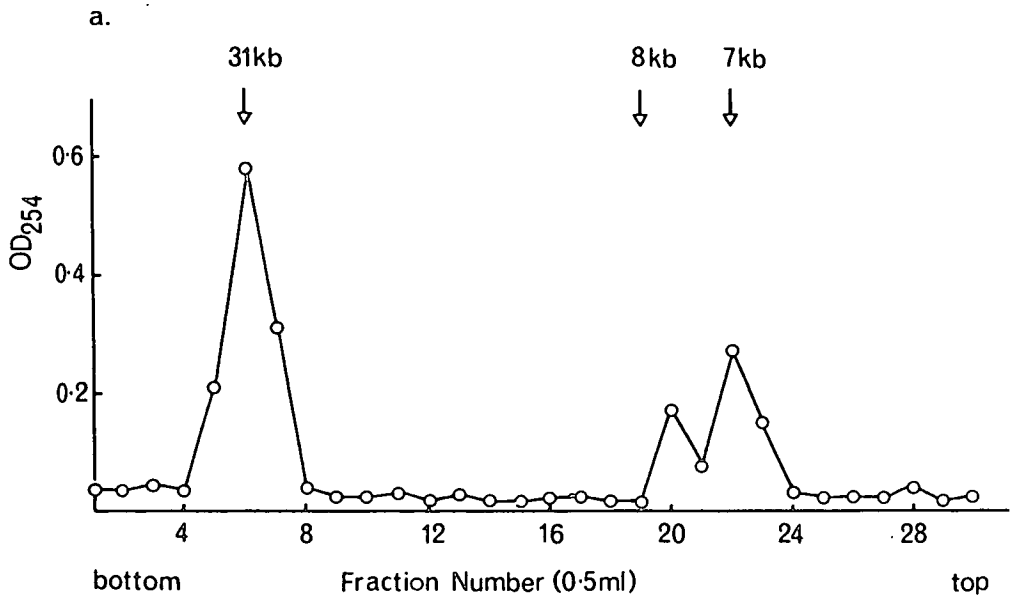
FIGURE 4.8: Sucrose Gradient Fractionation of Annealed and
EcoR1 Restricted Charon 4A DNA.

(a) Following complete EcoR1 endonuclease digestion of annealed
charon 4A DNA, 70 µg of the DNA digest was fractionated by
sucrose gradient centrifugation, as described in Section 4.5.2.
0.5 ml fractions were collected from the bottom of the tube and the
position of the charon 4A EcoR1 fragments located by monitoring
the absorbance of each fraction at 260 nm. These positions are
shown in the figure. The 31 kb fragment was isolated from fractions
5, 6 and 7 by ethanol precipitation at -20°C.
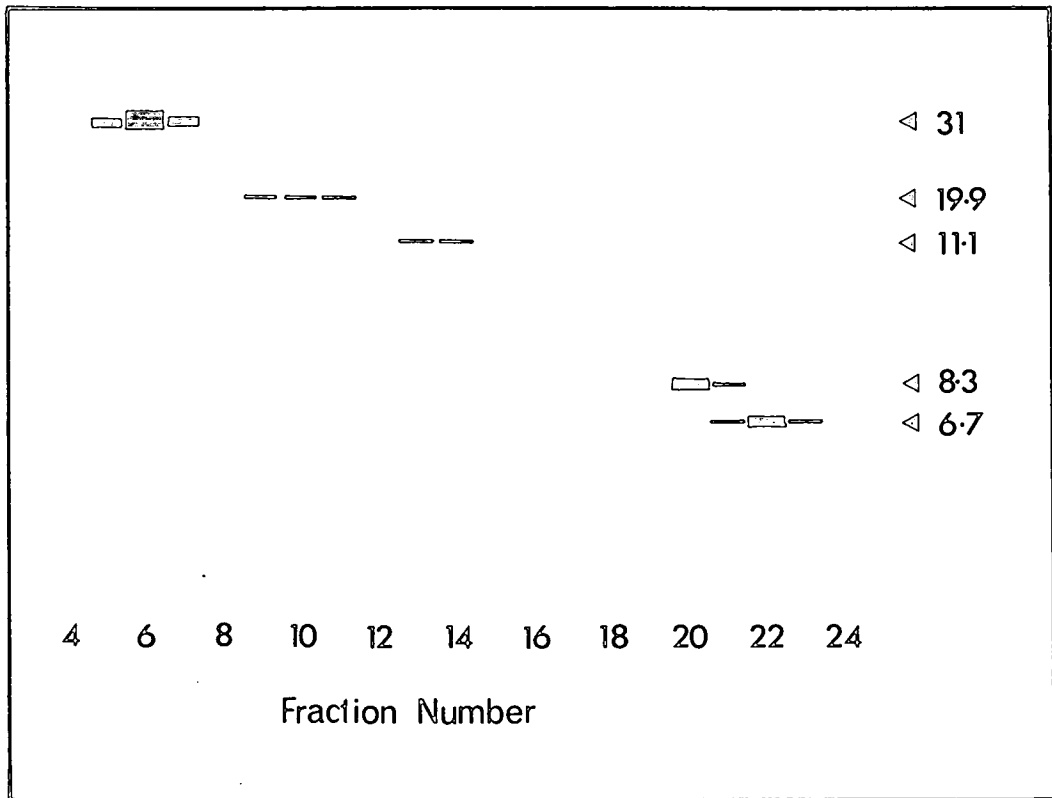
FIGURE 4.8:

(b) Analysis of Sucrose Gradient Fractionated Charon 4A DNA
Fragments by Agarose Gel Electrophoresis.

A small aliquot (25 μl) from selected fractions of the sucrose
gradient was electrophoresed on a neutral agarose gel. It is clear
from this figure that in this experiment not all the charon 4A DNA
molecules were annealed at the cos site due to the presence of 19.9
and 11.1 kb fragments in some fractions. The efficiency of annealing
tended to vary between DNA batches.

limit of this range to minimise the number of plaques required to screen a complete genome and to reduce the chance of eukaryotic fragments joining to each other prior to joining to the vector. In order that a complete genomic library could be established, it was important that the DNA fragments should be produced as randomly as possible to eliminate any bias for or against a particular sequence.

For this reason, EcoR1 was not used as the restriction enzyme (even though it would have been the most simple as then both chicken and vector DNA would have possessed cohesive complementary ends) due to the possibility that EcoR1 sites may be non-randomly distributed and specific sites within and adjacent to structural gene sequences may be selectively lost during the size fractionation steps. Regions where EcoR1 sites were greater than 20 kb apart would not be represented in the library.

Maniatis *et al.* (1978) found that a non-limit restriction endo-nuclease digestion with two enzymes was a very efficient method for producing fragments in a random fashion and therefore this method was followed. The restriction endonucleases HaeIII and AluI were chosen because they cleave frequently (each approximately every 256 or 4^4 nucleotides) and generate blunt ended fragments (Figure 4.4). The greater the number of possible cleavage sites, the larger the number of possible ways of producing 20 kb fragments by non-limit digestion of a high molecular weight molecule, and thus the more random the collection of fragments will be, and the risk of losing a sequence of interest will be very small.

The best conditions for partial endonuclease digestion with HaeIII and AluI were established by diluting the enzymes from a concentration of 1 unit/μl to concentrations of 0.5, 0.1, 0.05, 0.01 and 0.001 units/μl

in 6 mM Tris-HCl ph 7.5, 6 mM MgCl$_2$, 6 mM β-mercaptoethanol. 1 µl of each enzyme at each dilution was incubated with 1 µg chicken DNA in 10 µl of the same buffer at 37°C for 1 hour. The digestion mixes were electrophoresed on 1% neutral agarose gels and as shown in Figure 4.9 the dilution of the enzymes to 0.05 units/µg DNA yielded a high number of fragments in the region of 20 kb (i.e. some digestion had occurred but the fragment size was not too small).

On the basis of this information, six large scale digests were set up containing 330 µg of DNA per reaction in 6 mM Tris-HCl pH 7.5, 6 mM MgCl$_2$, 6 mM β-mercaptoethanol. Two mixes contained HaeIII and AluI at 0.1 unit/µg DNA (i.e. 33 units of each enzyme), two contained the enzymes at 0.05 units/µg (i.e. 16.5 units of each enzyme) and the final two contained the enzymes at 0.025 units/µg (i.e. 8.25 units of each enzyme). Incubation was for 1 hour at 37°C and then the DNA fragments were extracted by gentle mixing with fresh water saturated phenol followed by precipitation with ethanol.

The fragments were then dissolved in 10 mM Tris-HCl pH 8, 10 mM EDTA, heated to 68°C for 20 min to dissociate aggregates and 250 µg were then layered onto and sedimented through a 17 ml 10-40% linear sucrose gradient as described previously in Section 4.5.2(b). 20 µl aliquots of each 0.5 ml fraction from the gradients were electrophoresed on a 0.5% agarose gel using λ$^+$ EcoR1 fragments as molecular weight markers, as shown in Figure 4.10. Those fractions containing 19-21 kb DNA were pooled and dialysed against 10 mM Tris pH 8, 1M NaCl, 1 mM EDTA for 24 hours at 4°C. The DNA fragments were then concentrated by ethanol precipitation.
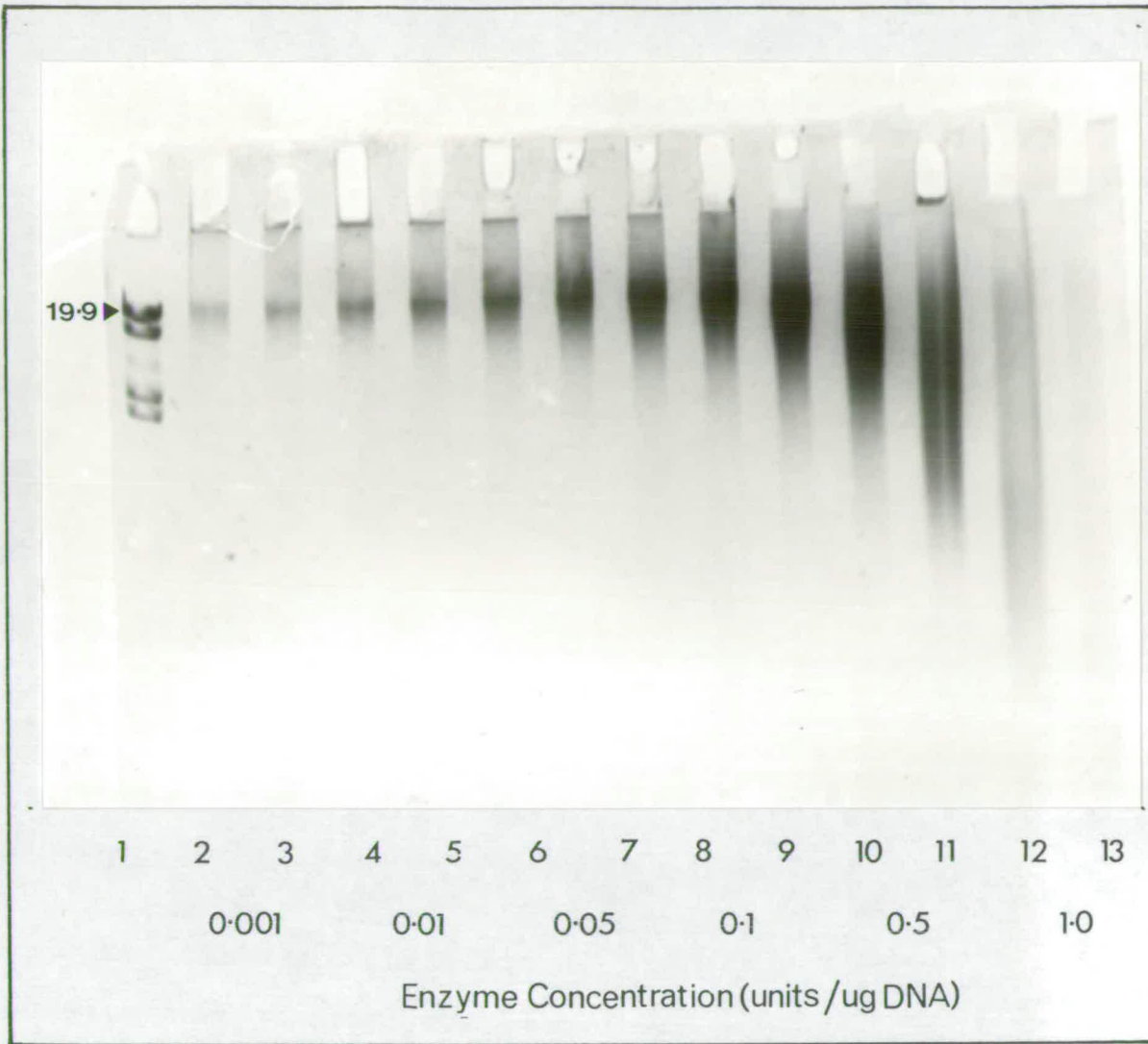
FIGURE 4.9: Partial Restriction Digests Using HaeIII and AluI of Chicken
Genomic DNA.

1 µg of chicken DNA was digested with HaeIII and AluI restriction
enzymes at 6 different dilutions in an attempt to determine the
optimum concentration of enzyme required to produce a maximum of
20 kb fragments as described in Section 4.5.3. This figure shows
that the two enzymes at a concentration of 0.05 units/µg of DNA is
sufficient to yield fragments in the required size range. Lane 1
shows EcoR1 digested charon 4A DNA, electrophoresed as molecular
weight standards.

◄ 21·8 kb

4   5   6   7   8   9   10   11

Fraction Number

FIGURE 4.10:   Sucrose Gradient Fraction of 20 kb Chicken DNA
Fragments.

Following HaeIII and AluI partial restriction enzyme digestion of
chicken DNA, the 20 kb fragments were isolated by centrifugation
through a sucrose gradient as described in Section 4.5.3.   20 μl
aliquots of each 0.5 ml fraction from the gradient were electro-
phoresed on neutral agarose gels to determine the position of the
20 kb fragments using EcoR1 digested λ⁺ DNA as molecular weight
markers (right hand lane).   This figure shows fractions 4 to 11 of
such a gradient.   Fractions 4 and 5 were pooled and retained.

4.6 Methylation of 20 kb Chicken DNA Fragments

Any EcoR1 cleavage sites within the 20 kb chicken DNA fragments were methylated by treatment with an EcoR1 modification methylase to render the sites resistant to EcoR1 attack. This procedure was necessary because the synthetic linkers (which were to be attached to the chicken DNA to form the link between the chicken and charon 4A fragments), contained an EcoR1 site which must be cleaved to generate the necessary cohesive ends.

Charon 4A DNA was mixed with aliquots of the methylation mix to monitor the extent of methylation. If discrete bands were visible on a gel following EcoR1 restriction and agarose gel electrophoresis, then the methylation was incomplete.

115 µg of 20 kb chicken DNA fragments were made up to 1 ml in 0.1M Tris-HCl pH 8.9, 10 mM EDTA, and 6 µmoles of S-adenosyl-L-methionine were added. Two 10 µl aliquots were taken. 20 units of EcoR1 methylase were added and a second two 10 µl aliquots were taken. All aliquots were mixed with 0.5 µg of charon 4A DNA and were incubated with the 1 ml mix at 37°C for 1 hour. Each 10 µl aliquot was mixed with an equal volume of 2 x concentrated EcoR1 buffer and two of them (one withdrawn before, and one after, the addition of methylase) were mixed with 2 units EcoR1. All four aliquots were incubated for a further 1 hour at 37°C and then electrophoresed on a horizontal 0.5% neutral agarose gel as shown in Figure 4.11. This figure demonstrates that the charon 4A DNA which was incubated with EcoR1 methylase was subsequently resistant to EcoR1 endonuclease. From this it was assumed that the chicken DNA would also have been successfully methylated as is indicated by the apparent lack of degradation of the 20 kb fragments following incubation with the restriction enzyme.
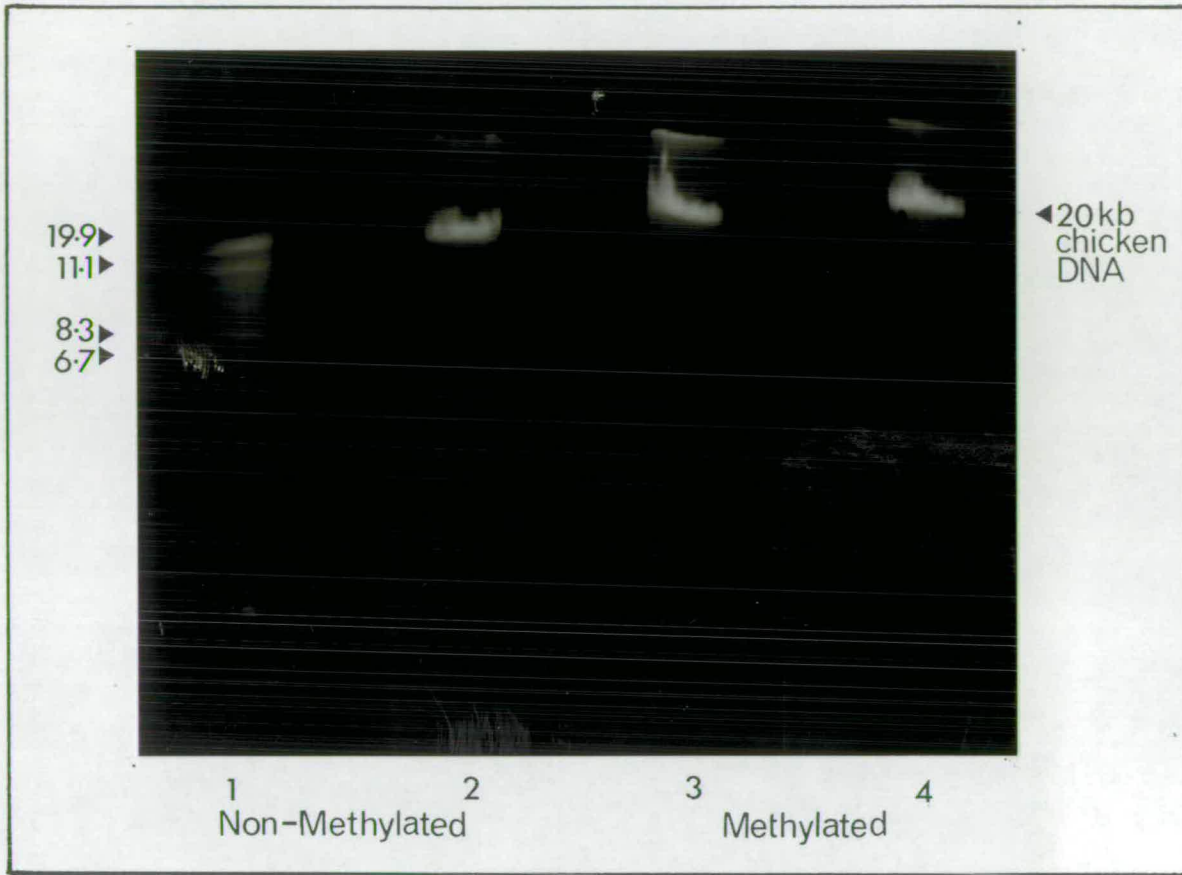
FIGURE 4.11: Methylation of Chicken DNA.

The 20 kb chicken DNA fragments were rendered resistant to EcoR1 digestion by methylation as described in Section 4.6. 10 µl aliquots were mixed with charon 4A DNA and incubated with EcoR1 both before and after methylation.

> Lane 1 - before methylation + EcoR1
>
> Lane 2 - before methylation - EcoR1
>
> Lane 3 - following methylation + EcoR1
>
> Lane 4 - following methylation - EcoR1

The presence of charon 4A DNA EcoR1 restriction fragments in Lane 1 and their absence in Lane 2 demonstrates that methylation was successful. Electrophoresis was in a 1% neutral agarose gel.

The methylated DNA in the 1 ml mix was extracted with an equal volume of water saturated distilled phenol in an eppendorf tube by gentle inversion and rolling. The DNA in the aqueous phase was precipitated by the addition of two volumes of ethanol and incubation at -20°C for 20 hours. The DNA was pelleted by centrifugation and dissolved on 40 μl 5 mM Tris-HCl pH 5.

4.7 Production of Recombinant DNA Molecules

Both donor and vector DNA molecules have been suitably manip- ulated into a form where they can now be joined together to produce a recombinant before re-introduction into viable phage particles.

When vector and donor DNA are both cleaved by an enzyme which generates sticky ends, the DNA termini can be annealed by base pairing and use of DNA ligase. However, as was discussed previously in Section 4.5.3, it was not possible to create fragments of chicken DNA which possessed the EcoR1 sticky ends complementary to those of the charon 4A 31 kb fragment. As shown in Figure 4.3, the enzymes HaeIII and AluI, used to generate the 20 kb chicken DNA fragments leave blunt ends, and though T4 DNA ligase can be used to join blunt ends under appropriate conditions, the efficiency is much lower than that of cohesive end ligation (Sugino *et al.*, 1977). Therefore, it was necessary to generate EcoR1 cohesive ends on the chicken DNA fragments rather than creating blunt ends on the charon 4A fragments, by use of S1 nuclease, for example.

This was achieved by the blunt and ligation of synthetic DNA linkers to the chicken DNA which contain the recognition sequence for EcoR1 endonuclease. After ligation, the linker was treated with EcoR1 and the fragments were then used in cohesive end ligation.

The synthetic linkers are duble-stranded oligonucleotides containing specific enzyme recognition sequences which have been chemically synthesised by several groups (Bahl *et al.*, 1977; Scheller *et al.*, 1977) following the development of the triester method of polynucleotide synthesis (Itakura *et al.*, 1975) and used to insert various DNA sequences into plasmids (Heynecker *et al.*, 1976; Ullrich *et al.*, 1977; Shine *et al.*, 1977).

A similar approach has been adopted for the insertion of eukaryotic DNA fragments into bacteriophage λ (Maniatis *et al.*, 1978). For this project, the EcoR1 site containing linkers, GGAATTCC, were bought commercially as they are now readily available from several companies.

4.7.1 *Joining of Synthetic Linkers to Eukaryotic DNA*

The first step in the procedure is the blunt end ligation of the linkers to each end of the chicken DNA fragments. Since the method of synthesis produces oligonucleotides with 5' hydroxyl ends (Scheller *et al.*, 1977) and the DNA ligase requires deoxyribonucleotide 5' phosphate ends to covalently link to a 3' hydroxyl group, the linkers were first phosphorylated using T4 kinase.

5 μg of EcoR1 linker DNA was taken up in 10 μl of 66 mM Tris-HCl pH 7.6, 10 mM magnesium chloride, 1 mM ATP, 1 mM spermidine, 15 mM dithiothreitol, 200 μg/ml gelatin, and added to 10 units T4 kinase, and incubated at 37°C for 1 hour. This mixture was then added directly to 100 μg of methylated chicken DNA fragments in 100 μl of the same buffer. 5 units of T4 ligase was added and the mixture was left in at a cool room temperature (approximately 16°C) for 6 hours.

5 μl of the reaction mix was electrophoresed on a 12% Tris-borate-EDTA gel (Maniatis *et al.*, 1975b), as described in Chapter III, Section 3.2.4. This method of nucleic acid gel electrophoresis gives successful

resolution of very small DNA fragments. The DNA was visualised by ethidium bromide staining as previously described, and the ligation was considered successful if the linkers had formed oligomers with each other. These could be seen as a series of linker oligomers from dimers up to 13-mers, as shown in Figure 4.12.

Following ligation, the mixture was diluted with 400 μl of 10 mM EDTA to stop the enzymatic reaction, incubated at 68°C for 15 min and centrifuged through a 10-40% linear sucrose gradient as described in Section 4.5.2(b). This step was important to remove unincorporated linkers and linker oligomers which would otherwise compete for the EcoR1 restriction endonuclease during digestion of the terminal linkers to produce sticky ends. The OD_{260} of the gradient fractions were monitored and a broad peak of absorbance was found, as shown in Figure 4.13. Fractions 10, 11, 12, 13 and 14 were pooled and dialysed against the 0.1M Tris-0.2M sodium chloride-1mM EDTA buffer to remove sucrose. The DNA was recovered by precipitation with 70% ethanol at -20°C.

The DNA pellet was dissolved in 100 μl of 5 mM sodium chloride and an equal volume of 2 x concentrated EcoR1 restriction buffer was added with 100 units of EcoR1. 5 μl of the reaction mix was mixed with 0.5 μg of charon 4A DNA and both were incubated at 37°C for 1 hour. The small digestion mix was electrophoresed on a 0.5% neutral agarose gel. The complete digestion of the charon 4A DNA, as shown in Figure 4.14, was taken as evidence of a similar complete digestion of the linkers attached to the chicken DNA fragments. By this procedure, the randomly produced, 20 kb chicken DNA fragments were provided with EcoR1 cohesive ends which were complementary to those of the charon 4A 31 kb fragment.
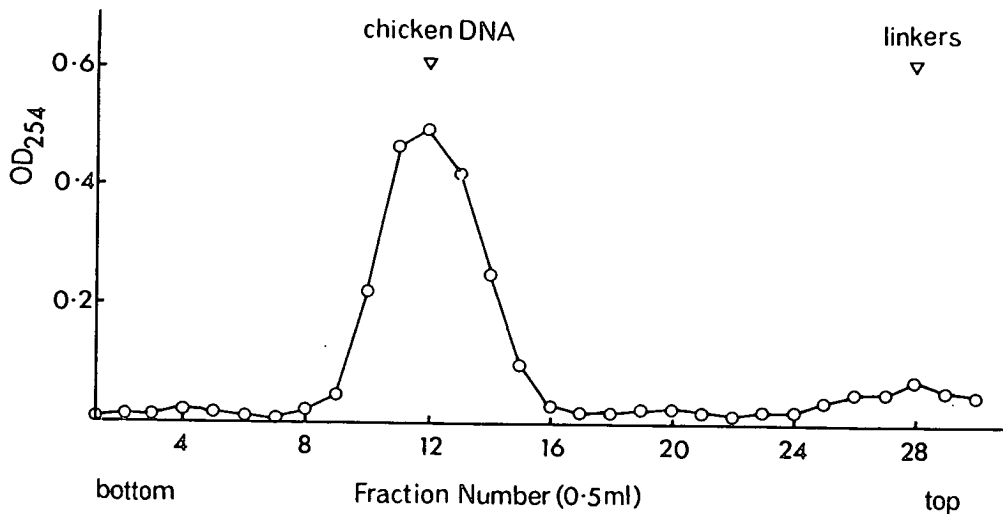
FIGURE 4.13: Sucrose Gradient Fractionation to Separate Chicken
DNA Fragments from Free Linker Molecules.

Linker oligomers incorporating the EcoR1 recognition site were
ligated to the 20 kb chicken DNA fragments as described in
Section 4.7.1. The chicken DNA fragments were then separated
from free manomeric and oligomeric linkers by centrifugation
through a 10-40% sucrose gradient. 0.5 ml fractions were collected
from the bottom of the tube and the OD_{260} was monitored. These
fractions containing the 20 kb material were retained and the
DNA recovered as described. The linkers remained near the top
of the gradient, and this figure demonstrates that the separation
was effective.

20 kb DNA fragments at origin

Range of linker oligomers

14-nt linkers at dye front
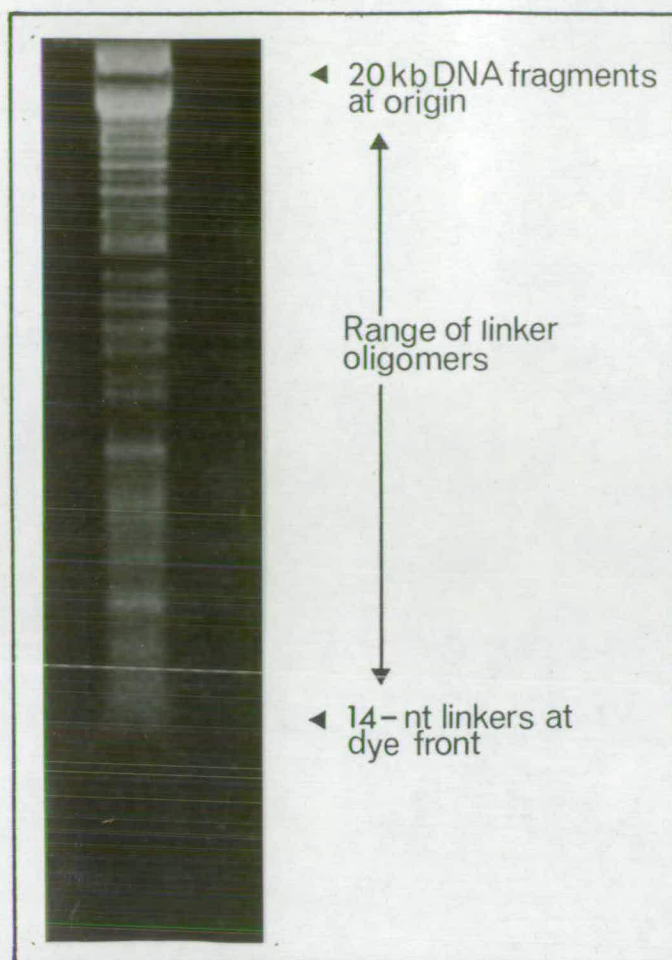
FIGURE 4.12: Ligation of Synthetic EcoR1 Linkers to Chicken DNA Fragments.

Synthetic linkers containing the EcoR1 restriction site were ligated to chicken DNA fragments as described in Section 4.7.1. An aliquot (5 µl) of the reaction mix was electrophoresed on a Tris-Borate-EDTA polyacrylamide slab gel and stained using ethidium bromide. This figure shows that linker oligomers were present in the mix and therefore present evidence of successful ligation. The large 20 kb chicken DNA fragments cannot enter the high concentration poly-acrylamide gel, but the small 14-nucleotide linker electrophoreses with the bromophenol blue dye front.
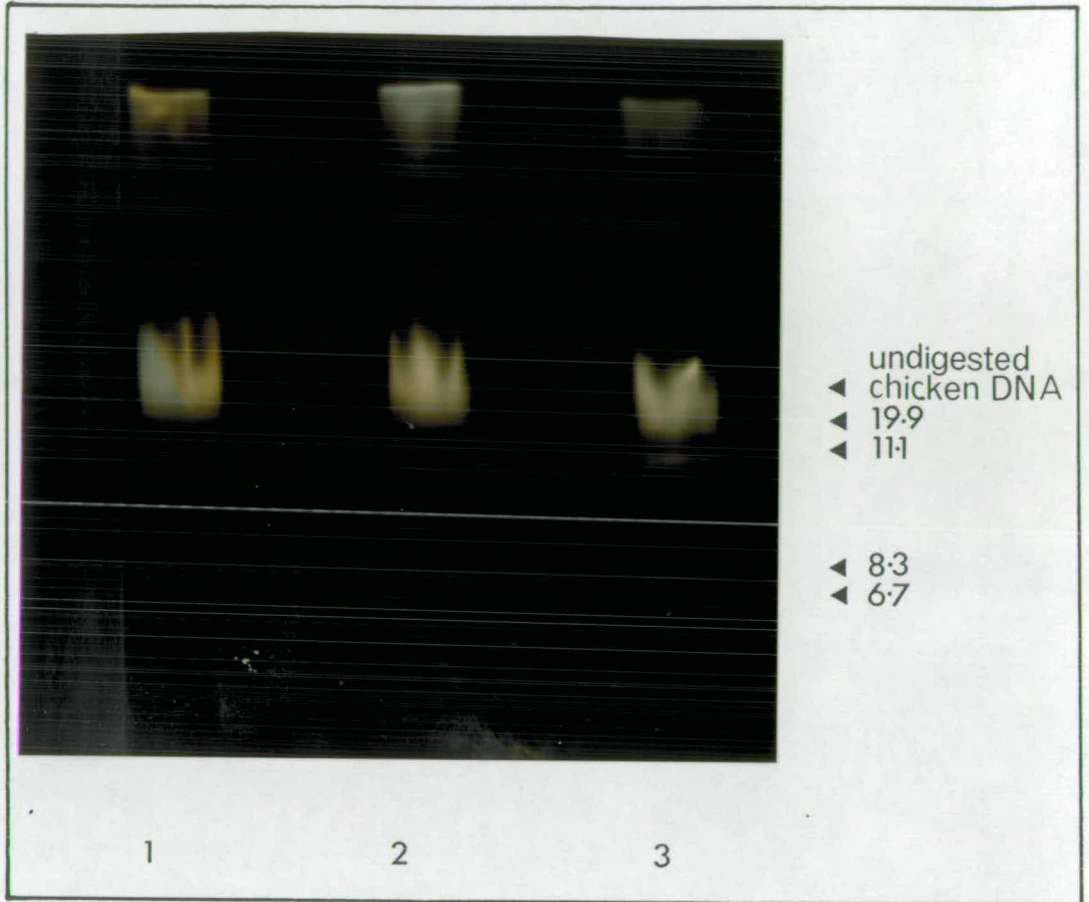
FIGURE 4.14: EcoR1 Digestion of Linkers Attached to Chicken DNA
Fragments.

The linkers carrying the EcoR1 site were digested with EcoR1 to
produce staggered ends on the chicken DNA fragments as described
in Section 4.7.1.

Lane 1 – 20 kb chicken DNA fragments – EcoR1

Lane 2 – 20 kb chicken DNA fragments + EcoR1

Lane 3 – 20 kb chicken DNA fragments + EcoR1
+ 0.5 µg charon 4A DNA

The presence of charon 4A DNA EcoR1 restriction fragments in
lane 3 was taken as evidence that the concentration of EcoR1 used in
the large digestion mix was sufficient to digest the linkers at the
EcoR1 site. The single band formed by the chicken DNA, even in
the presence of EcoR1 is evidence that the 20 kb fragments are still
resistant to endonculease attack. Electrophoresis was in a 1% neutral
agarose gel.

4.7.2 Ligation of Chicken and Charon 4A DNA Fragments

The vector : donor DNA ratio during ligation is very important. The *in vitro* λ packaging system is to be used for transduction of the recombinant molecules, and therefore the aim of this ligation is to produce concatamers where both donor DNA ends are joined to vector molecules. This is favoured by a high molar vector : donor DNA ratio. Also, the total DNA concentration in the *in vitro* ligation reaction influences the type of molecules which are formed. Low concentrations (less than 20 µg/ml) favour circularisation due to reduced intermolecular interaction, but high concentrations (300-400 µg/ml) encourage formation of concatamers.

The conditions employed here were as described by Maniatis *et al.* (1978) who showed that, using charon 4A, a vector : donor ratio of 2.5 : 1 gave high cloning efficiencies. Also, to encourage the formation of concatamers, the substrate for *in vitro* packaging, relatively high concentrations of DNA were used.

The charon 4A arms were incubated at 42°C for 1 hour in 66 mM Tris-HCl pH 7.6, 10 mM magnesium chloride, 200 µg/ml gelatin to ensure the cos sites were annealed and that the vector DNA did exist as a 31 kb fragment. For every 50 µg of charon 4A arms, 20 µg of chicken DNA fragments were added with 19 units of T4 ligase in a total volume of 300 µl (giving a DNA concentration of 230 µg/ml). The mixture was incubated at 12°C for 12 hours and a 5 µl aliquot was electrophoresed on a 0.3% neutral agarose gel. As shown in Figure 4.15, molecules larger than intact charon 4A DNA were present in the mixture which provides evidence that the ligation had been successful, and a suitable recombinant series of concatamers were available for *in vitro* packaging. Several ligation reactions were performed, and the DNA

◄ Intact Charon 4A
DNA
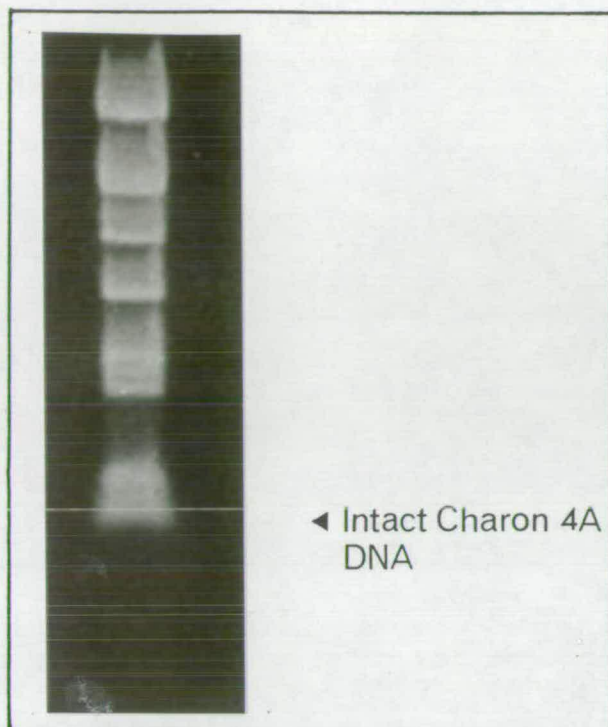
FIGURE 4.15: Production of Recombinant DNA Molecules.

The 31 kb charon 4A DNA fragment was ligated to the 20 kb
chicken DNA fragments by base pairing of the complementary
EcoR1 staggered ends at each end of both sets of molecules. This
figure demonstrates that ligation has been successful in producing
a series of different sized molecules larger than intact charon 4A
DNA which represent concatamers suitable for *in vitro* packaging.
Electrophoresis was in a 0.2% neutral agarose gel.

from each was recovered by ethanol precipitation at -20°C followed by pelleting at 10,000 g for 30 mins.

4.8 *In vitro* Packaging of Recombinant DNA Molecules

The actual cloning step depends on the transfer of the recombinant DNA into a host cell which must be able to replicate that DNA to enable the eventual isolation of one species of recombinant.

Until recently, the most commonly used method of cloning recombinant phage DNA was by transfection (Mandel and Higa, 1970). This involves treating *E. coli* with calcium chloride to make them susceptable to DNA uptake, mixing with the recombinant DNA at 0°C followed by a short incubation at 42°C (heat shock), and then immediately plating on agar. Eventually, plaques will appear in a bacterial lawn due to phage DNA molecules having been taken up by the bacteria in which they have replicated and given rise to infectious phage which cause cell lysis. The number of plaques provides a measure of the number of transfected phage DNA molecules.

However, Maniatis *et al*. (1978) showed that cloning efficiencies using *in vitro* packaging procedures employed in this project were increased by a factor of 0.15×10^3 when compared to those achieved *in vivo* by the calcium chloride transfection procedure of Mandel and Higa (1970). Cloning efficiencies of 2 and 0.15×10^7 plaque forming units/μg were obtained for intact and religated λ DNAs, respectively, using both the Hohn and Murray (1977) and Sterberg *et al*. (1977) packaging techniques.

Theoretical considerations of the *in vitro* technique have been discussed in Chapter I, Section 1.3.4, so they will only be mentioned briefly here.

In vitro packaging of λ DNA, either wild type or vector, depends on the temperature induction of prophage which carry amber mutations in genes necessary for the packaging of DNA into viable phage, using pre-formed head and tail components. Due to the genetic block, therefore, viable phage cannot be produced by a single lysogenic strain, but if two strains, carrying different amber mutations, are mixed following induction, the induced products will complement and packaging can take place.

The possibility exists that hybrid DNA, when added to the mixed lysogen extract, will recombine with the endogenous prophage DNA to produce a molecule bearing wild type markers of the prophage. In order that the precise biological containment features of the vector DNA are not altered by the occurrence of such a recombination event, the lysogenic strains have special features also.

Sternberg *et al.* (1977) developed packaging strains carrying the λb2 mutation which removes part of the attachment site (att) of the prophage and thus prevents excision from the host chromosome. In addition, both bacteria and prophage in these strains are recombination deficient (that is, are rec A$^-$ and red A$^-$, respectively) as a precaution against recombination.

The Hohn and Murray (1977) packaging strains, which were used in this project, must be irradiated with ultra violet light to ensure damage of all endogenous DNA prior to the addition of the hybrid vector DNA.

4.8.1 *Preparation of the Packaging Extracts*

The lysogenic strains used were BHB2671 and BHB2673 which carry prophages bearing amber mutations in genes E and D, respectively.

A single colony of each lysogen was picked from an L-agar plate and streaked over two fresh L-agar plates. The plates were incubated, inverted at 32°C overnight. The bacteria were then scraped off the plates taking care not to remove any agar. Keeping the two strains entirely separate, the bacteria were resuspended in 2 ml L-broth and added to 100 ml of pre-warmed L-broth in a 1 l flask. This generally gave an OD_{650} value of $0.1 - 0.15$. The cultures were grown to an OD_{650} of 0.5 at 32°C with rapid shaking to ensure adequate aeration and then were induced by incubating in a water bath at 45°C for 15 min. During the induction period, the flasks were shaken by hand every 2-3 min. The cultures were then replaced in the orbital shaker and incubated at 37°C for 3 hours with very fast shaking. During this time, phage coded proteins were accumulating in the bacterial cell, but no release of viable phage was possible due to the lack of one vital protein. Following the 3-hour incubation period, the two cultures were mixed and centrifuged at 9,000 g for 10 min. The cell pellet was resuspended in M9 buffer to give a two-fold dilution of the original volume (that is, 400 ml). The suspension was divided equally between four sterile glass staining troughs and was irradiated with ultra violet light for 25 min at a distance of 20 cm (see Section 4.8.2, below). The cell suspensions were gently stirred during this time. It was shown that bacteriophage could not survive this treatment and therefore it was assumed that all DNA had been damaged and so the risks of recombination between the prophage and charon 4A were eliminated.

The cells were pelleted (9,000 g, 10 min) and the phage proteins were concentrated by a series of steps as follows: the cells were washed in 50 ml concentration buffer and pelleted again; they were then resuspended in 8 ml of concentration buffer, centrifuged again

and finally resuspended in 1 ml of concentration buffer. The final packaging extracts were prepared by centrifuging for 2 min in an eppendorf centrifuge and resuspending in 0.8 ml of concentration buffer. At this stage, the suspension was very viscous following lysis of the bacterial cells during the concentration process. This viscous suspension was stored in 20 μl aliquots, snap frozen in liquid nitrogen and maintained below -60°C. The extracts remained active for 4-5 months when stored in this manner.

4.8.2 _Safety Precautions_

(a) _Calibration of the ultra violet lamp:_

The killing efficiency of the ultra violet light source was calibrated using the vector phage, charon 4A. The phage were transfered in phage buffer, to four glass staining dishes, at a concentration of 10^6 pfu/ml with 100 ml per dish. They were irradiated at a distance of 20 cm for 60 min. 1 ml aliquots were removed after 5, 10, 20, 30, 40, 60 mins, and the numbers of surviving phage were determined by titrating on DP50supF bacteria, as described in Section 4.3.3. It was shown that no phage survived a 20 min exposure to UV light as can be seen from the data presented in Figure 4.16 and Table 4.2, and all phage were probably destroyed after 15 min irradiation. This dose was considered adequate for use in the preparation of the packaging extracts, but to make absolutely certain that all endogenous DNA would be destroyed, an irradiation period of 25 min was routinely used.

(b) _Mock packaging experiments:_

Every preparation of packaging extracts was tested for prophage excision and packaging by performing 'mock' packaging experiments, as described in Section 4.8.3 below, without the addition of exogenous

180.

TABLE 4.2: Calibration of the Ultra-violet Light Source.

| Irradiation time (min) | Dilution from 1 ml sample | | | | | | Total number of surviving phage/ml |
	10^0	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	
0	–	–	–	approx 10^3	92	12	10^6
5	–	–	82	9	2	0	8.5×10^3
10	89	11	0	0	0	–	10
20	0	0	0	0	0	–	0
30	0	0	0	0	0	–	0
40	0	0	0	0	–	–	0
60	0	0	0	0	–	–	0

During the preparation of *in vitro* packaging extracts, as described in Section 4.8, endogenous phage were destroyed by ultra-violet irradiation. The data presented here shows the numbers of surviving viable phage after various periods of irradiation. The initial concentra-was 10^6 pfu/ml, and titration on DP50supF at various dilutions after periods of irradiation showed that all viable particles were destroyed after 20 min. This data is plotted in Figure 4.16.
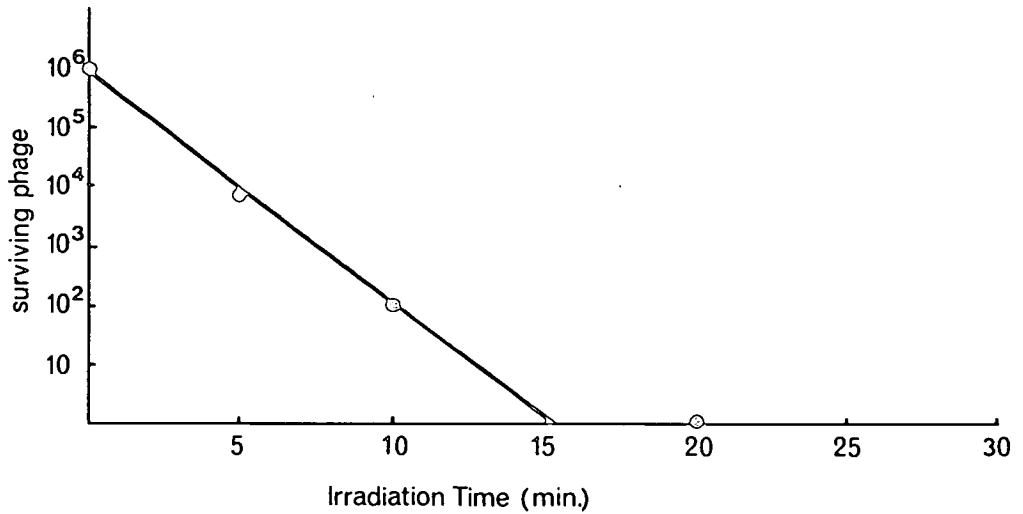
FIGURE 4.16: Calibration of the Ultra-violet Light Source.

Charon 4A at 10^6 pfu/ml were irradiated at a distance of 20 cm.
1 ml aliquots were removed at suitable time intervals and the
survival rate was determined by titrating on DP50SupF. The
survival rate of the phage plotted against irradiation time, as
shown above, can be reduced to zero in 15 min. Certainly
no phage survived a 20 min exposure to ultra violet, and 25 min
was chosen as the routine exposure time in the preparation of
the packaging extracts, as described in Section 4.8.2a.

DNA, and plating on DP50supF as this is the strain which would be used to propagate the recombinants. Five 'mock' experiments were normally carried out for each batch of extract, using 20 μl of the extract and one 10 cm L-agar plate supplemented with 1 mM magnesium sulphate, 40 μg/ml thymidine and 10 μg/ml DAP, per experiment. Plaques were observed very rarely, and the numbers observed were acceptable to the Genetic Manipulation Advisory Group which states that the ratio of plaques formed in a mock experiment, without the addition of exogenous DNA, to plaques formed when recombinant DNA is added must be less than $1 : 10^6$.

4.8.3 *Use of the Packaging Extracts*

The 20 μl aliquots of extract were transfered from −60°C to liquid nitrogen and then thawed very slowly on ice. The extracts performed more efficiently if this procedure was followed. ATP was added to 5 mM. A $\frac{1}{20}$th volume of a 100 mM ATP solution in water, pH 7.0, was a sufficient volume to add without causing dilution of the extract. 1 μl of DNA, either wild type or recombinant, was added to give a concentration of between 0.1 and 1 μg per 20 μl extract. The contents were mixed carefully and collected with a short spin in an eppendorf centrifuge and then incubated for 60 min at 37°C. The packaging reaction was stopped by the addition of a further 20 μl of extract which had been incubated on ice for 30 min with DNase at 10 μg/ml. Incubation was continued at 37°C for 45 min followed by the addition of 0.5 ml of phage buffer and a drop of chloroform to each incubation mix. In this way, any free bacteria having escaped previous treatments were destroyed and non-packaged DNA was removed.

4.8.4 Construction of the Chicken Gene Library

Following the ligation of the chicken DNA fragments to the charon 4A arms as described in Section 4.7.2, approximately 80 µg of DNA were recovered. Although the evidence presented in Figure 4.15 demonstrates that ligation between the two sets of molecules had successfully taken place, there is also evidence that the ligation was not 100% complete due to the presence of bands smaller than intact charon 4A which probably correspond to free 31 kb charon 4A and 20 kb chicken fragments. Therefore, the 80 µg does not represent a total of recombinant DNA. However, since the 31 kb fragment, and anything smaller cannot be packaged, it was decided not to attempt to fractionate the DNA sample further, but to use it at its highest possible concentration in the *in vitro* packaging procedure on the assumption that only recombinant DNA molecules would emerge as viable phage particles.

Therefore, to construct the chicken gene library, a total of 65 packaging reactions were performed as described in Section 4.8.3 above. Three different batches of packaging extracts were used with approximately 20 reactions per batch. The cloning efficiencies of each batch were similar, as shown in Table 4.3. 1 µg of recombinant DNA was used per reaction.

Following dilution with phage buffer and chloroform treatment, each batch of packaging experiments (approximately 20) were pooled and mixed with solid caesium chloride to give a concentration of 0.5 g/ml. This was layered on to caesium chloride step gradients composed of 1.5 ml steps of 1.45, 1.5 and 1.7 g/ml caesium chloride in phage buffer. The gradients were centrifuged at 32,000 rpm in an SW40 rotor for 1.5 hours at 4°C. 0.4 ml fractions were collected from the bottom of the

TABLE 4.3:   Efficiency of Pakcaging Recombinant Charon 4A DNA Molecules

Packaging extract number	Efficiency of packaging intact charon 4A DNA (pfu/μg)	Total number of recombinant phage recovered	Efficiency of packaging recombinant charon 4A DNA (pfu/μg)
1. (20 reactions)	5.6×10^6	1.4×10^5	0.7×10^4
2. (23 reactions)	8.3×10^6	3.3×10^5	1.4×10^4
3. (22 reactions)	4.1×10^6	2.1×10^5	0.9×10^4
		6.8×10^5	

Three batches of packaging extracts were prepared, as described in Section 4.8.1.   The packaging efficiencies were determined using intact charon 4A DNA (five reactions of 1 μg for each extract) and were shown to be similar, as presented in column 2 above.   Recombinant DNA was packaged by an identical procedure.   65 reactions, each containing 1 μg of DNA were performed and a total of 6.8×10^5 recombinants were recovered.   The efficiency of packaging recombinant DNA is shown in column 4 and is consistantly between 10^2 and 10^3-fold lower than the efficiency of packaging intact DNA.

tube and the phage were located by spotting 50 μl of 10^{-3} dilutions on to a lawn of DP50supF bacteria. The phage banded in fractions 4, 5 and 6. These fractions were pooled and dialysed against phage buffer for 24 hours at 4°C with three changes of buffer. The gelatin concentration in the phage suspension was increased to 0.02% to stabilise the phage.

After removal of the caesium chloride by dialysis, the recombinant phage suspension was accurately titred as described previously in Section 4.3.3.

Dilutions of 10^{-1}, 10^{-2} and 10^{-3} were titrated for each batch of packaging reactions in order to obtain an accurate estimate of the number of recombinants as shown in Table 4.3, the total number of recombinant phage recovered from all three packaging batches was 6.8×10^5 and the mean efficiency was 1×10^4 plaque forming units per μg of recombinant DNA.

4.9 Amplification of the Gene Library

On constructing a gene library by *in vitro* packaging of phage DNA molecules, any particular sequence will not be represented in more than a few copies. In order that the library can be screened for a particular sequence, therefore, it must be amplified many times to:

(a) increase the chance of locating a particular sequence;

(b) to establish a permanent library which can be screened repeatedly for the same or other sequences;

(c) to provide many recombinant phage for restriction enzyme analysis, for example, when confirming the existence and size of the inserts.

Assuming that a complete sequence representation of the chicken genome had been achieved by the *in vitro* packaging of recombinant charon 4A DNA molecules, it was necessary to amplify every phage particle in the library to ensure that this situation remained. It was, therefore, also important to reduce the chances of competitive growth between the phage which may have led to the elimination from the population of recombinant phage which may have a growth disadvantage. For this reason, amplification was achieved using low density plate lysates.

Fresh 15 cm L-agar plates supplemented with 1 mM magnesium sulphate, 40 µg/ml thymidine and 10 µg/ml DAP were used. 10,000 recombinant phage particles were diluted to 2.5 ml in phage buffer and 10^6 DP50supF cells were added. Pre-adsorption was allowed to continue for 20 min at 37°C and then the phage were plated with 7.5 ml of soft top agar. To amplify the whole library it was necessary to set up seventy 15 cm plates. The plates were incubated overnight (usually 14-16 hours) at 37°C wrapped in tin foil to prevent photoreactivation. Almost complete lysis had occurred during this time.

The top agar was scraped into 500 ml sterile beakers and the plates were rinsed with 4 ml phage buffer. Each 300 ml of lysate was stirred with 10 ml chloroform for 20 min and then centrifuged in 200 ml bottles at 5,000 g for 30 min at 4°C to pellet the top agar. The supernatant was withdrawn and the phage were concentrated by polyethylene glycol precipitation as described in Section 4.3.2. This treatment removed bacterial debris and allowed the original 600 ml of lysate to be concentrated to a phage pellet. The pellet was resuspended in 21 ml of phage buffer and centrifuged through a caesium chloride step gradient as described in Section 4.3.2. The phage band was recovered and

titrated at 10^{-7}, 10^{-8}, 10^{-9} and 10^{-10} dilutions. The total number of phage following amplification of the library was 2.3×10^{10} representing an amplification of 3×10^{4}-fold. The library was stored in caesium chloride, in aliquots at 4°C.

4.10 Conclusions

Three important factors were stated in the introduction to this chapter as being necessary for the production of a genomic DNA library. These have been further discussed and evidence to support them has been presented.

The isolation of clonable DNA fragments was achieved by the use of non-limit restriction enzyme digests which aimed to produce even sized fragments of approximately 20 kb as randomly as possible. The size of the fragments was chosen towards the upper limit of the maximum possible insert which could be made to reduce the number of clones required to represent the whole genome. The random nature of the fragments means that cloned gene sequences would carry extensions of various lengths away from the gene in both directions, and independently isolated clones of a given gene are unlikely to be identical.

The chicken DNA fragments were joined to the vector via EcoR1 linker oligomers, and finally the library was completed by the *in vitro* packaging of the recombinant molecules into viable phage which were propagated on plate lysates to give large numbers of recombinants. The assumption was made here that all plaque forming units were, in fact, recombinants due to the positive selection pressure exerted by the lambda system where only molecules of a certain size class are packaged into viable phage. Any concatamers arising from the ligation of charon 4A 31 kb arms with each other would not possess cos sites at the correct distance apart to be recognised by the packaging enzyme, terminase.

The covalent linking of vector and chicken DNA was not 100% efficient and this may account for relatively low cloning efficiencies, though the efficiency also depends on the packaging extract itself. The efficiency of the various batches of extract were tested using intact Charon 4A DNA and were found to vary between $4.1 - 8.3 \times 10^6$ plaque forming units per µg of input DNA. Maniatis *et al*. (1978) obtained ten-fold higher efficiencies using the Sternberg *et al*. (1977) packaging strains of $2 - 20 \times 10^7$ pfu/µg, but does not indicate the source of lambda DNA used for this test. However, these efficiencies are both higher than have been reported for the calcium chloride transfection procedure, which are close to 1×10^6 pfu/µg.

The efficiency of packaging recombinant DNA was found to be consistently lower than for intact DNA. Maniatis *et al*. (1978) quote values of 3.8×10^4 to 5.6×10^5 pfu/µg, depending on the source of eukaryotic DNA. This project gave efficiencies of $0.7 - 1.4 \times 10^4$ pfu/µg for chicken DNA, a value which is probably a reflection of both the quality of the packaging extracts and the efficiency of ligation of the eukaryotic and vector DNAs. However, even these efficiencies were at the upper end and above the scale recorded for efficiencies of transfection which range from $2 - 10 \times 10^3$ pfu/µg.

The total number of recombinants obtained using 65 µg of recombinant DNA was 6.8×10^5. Assuming the chicken genome is 1×10^9 base pairs, an excess of 5×10^4 recombinant phage are required to be certain of complete genomic sequence representation. A ten-fold increase on this number was obtained here. McClements and Skalka (1977) estimated that the equivalent of 20 genomes is required to be certain of complete sequence representation in a library containing 10,000 kb inserts of chicken DNA. The inserts in this library were 20,000 kb on

average, so it was assumed that all sequences have a good chance of being represented and the possibility of detecting a myosin heavy chain recombinant is a reasonable expectation.

CHAPTER V

Screening of the Chicken Gene Library

CHAPTER V

5.1 Introduction

The original aim of the project was to attempt to isolate genomic DNA sequences which correspond to the skeletal muscle myosin heavy chain gene, using a specific cDNA hybridisation probe constructed by the reverse transcription of a purified messenger RNA for this gene. For this reason, the emphasis of the project has been divided into two distinct parts:

(a) the construction of the probe;

(b) the manipulation of genomic DNA into the form of a
 library which could be maintained in this state.

It is here, however, that the two parts complement each other in an attempt to use the probe to identify specific sequences within the recombinant DNA library.

The screening procedure was developed by Benton and Davis (1977). It allows a total of 10^6 recombinants to be screened daily. As discussed in Chapter I, Section 1.3.5, the technique depends on the fact that not all the phage DNA produced during a lytic infection is packaged prior to lysis of the bacteria. Therefore, when plaques are produced in a bacterial lawn, there will be free recombinant DNA, as well as viable recombinant phage on the surface of the plaque. This free DNA is immobilised on a nitrocellulose filter and the cDNA probe is allowed to hybridise to any complementary sequences on that filter. Positive hybridisation, following autoradiography, identifies a recombinant phage of interest.

This 'plaque hybridisation' method of screening is particularly appropriate to a situation where one may wish to isolate a single copy gene from a whole genome, since the desired sequence will be present

among very many inappropriate ones, yet many recombinants can be screened rapidly - indeed, a whole genome can be screened at one time.

Maniatis *et al.* (1978) examined the possible variables in the Benton and Davis screening procedures in an attempt to optimise the screening efficiency. The type of growth medium, the host bacterial strain, the plating concentration of the phage and the method of preparing the filters were tested with little effect on the final number of positive hybridisation signals observed for any particular probe. The most obvious differences were observed following variations in the concentration of plating bacteria. It was shown that 3×10^8 bacteria per 10^4 phage particles per 15 cm dish were optimum plating conditions since confluent lysis was never achieved during an overnight incubation. It was also shown that duplicate filters should be applied sequentially rather than by stacking together.

An attempt was made in this project to screen the library on large 30 cm x 30 cm plates, but it was difficult to achieve a smooth surface and the handling of the filters was extremely difficult. There-fore, the more conventional 15 cm plates were used routinely, even though many more were required.

To screen the complete chicken genomic library, it was necessary to screen in excess of 5×10^4 plaques. To increase the chance of detecting a sequence which is present at very low concentration, as was expected for the myosin heavy chain gene, a total of 2.5×10^5 phage particles were screened, a total of 25 plates.

5.2 The Screening Method

5.2.1 Preparation of Plates

15 cm L-agar plates, supplemented with 1 mM magnesium sulphate, 40 µg/ml thymidine and 10 µg/ml DAP, were dried in a 37°C incubator for four hours to minimise the possibility of agar adhering to the nitrocellulose filters later in the procedure. Once on the filter, agar is very difficult to remove and it causes an increase in the background hybridisation signal.

10^4 recombinant phage were mixed with 3.1×10^8 DP50supF bacterial cells in 2.5 ml phage buffer, pre-adsorbed for 20 min at 37°C and plated on to the 15 cm plates with 7.5 ml soft top L-agarose. Agarose was used as the top layer in preference to agar as it has a lesser tendency to adhere to nitrocellulose. The plates were incubated at 37°C overnight (14-16 hours) by which time the plaques were fairly large, and many were making contact, but lysis was not confluent. The plates were incubated at 4°C for 1 hour before proceeding, to harden the top agarose.

5.2.2 Preparation of Filters

Millipore HA nitrocellulose filters, pore size 0.45 µm, were cut to fit the 15 cm agarose surface exactly. The filters were marked for orientation by placing three small 'V' shaped cuts asymmetrically in the edges which corresponded to marks placed on the edges of the plates.

The filters were placed onto the chilled plates very carefully. It was important that air bubbles were not allowed to form between the filter and the agarose, that the filter fitted onto the plate perfectly, and that it was not moved at all to prevent 'smudging' the plaques over the plate. Adsorption of DNA to the filter was allowed to continue for

3 min and then it was removed very carefully to avoid lifting the agarose layer. A second filter was applied in the same way and allowed to adsorb DNA for 5 min.

Immediately following adsorption the filters were placed individually into a solution of 0.1M sodium hydroxide, 1.5M sodium chloride for 30 sec to denature the DNA and were then neutralised in 0.2M Tris-HCl pH 7.5 for 30 sec. They were then washed in 2 x SSC for 30 sec, blotted on to Whatman 3MM paper, and placed on fresh 3MM sheets. between glass plates to keep the filters flat as they dried. Drying took place in a vacuum oven at 80°C for 4 hours. If hybridisation was not to be carried out immediately, the filters were stored in sealed polythene bags between 3 MM paper at -20°C.

The used plates were stored at 4°C for a few days until the results of the autoradiography were known.

5.2.3 Hybridisation

Filters were washed in 3 x SSC and then incubated in 10 x concentrated Denhardt solution (Denhardt, 1966) (0.2% polyvinylpyrolidone, 0.2% ficol, 0.2% bovine serum albumin in 3 x SSC), containing 50 µg/ml denatured E. coli DNA. Incubation was carried out in sealed polythene bags, with two filters per bag from which all bubbles had been removed, at 68°C, in a shaking water bath overnight.

The polythene bags were prepared from cut sheets of heavy gauge polythene which were thoroughly cleaned before use by washing in warm water, two changes of cold distilled water and ethanol before drying.

The probe of interest was denatured by heating at 100°C for 5 min followed by immediate chilling on ice to prevent renaturation. A hybridisation solution of 3 x SSC, 1 x concentrated Denhardt solution

containing 50 µg/ml denatured *E. coli* DNA was prepared. Filters were incubated in sealed polythene bags with 20 ml of hybridisation solution and $10^5 - 10^6$ cpm of ^{32}P-labelled probe for 18 hours at 68°C in a shaking water bath. Again, it was important to remove all air bubbles, and for this reason the polythene bags were cut long to ease the teasing out of air bubbles when ^{32}phosphorus was present.

The filters were removed from the bags and rinsed in 2 x SSC. They were then resealed in bags containing a large volume (approximately 75-100 ml) of 2 x SSC and washed by shaking at 60°C for 1 hour. This washing procedure was repeated with fresh 2 x SSC. It is important that the washing is thorough to ensure removal of all non-hybridised probe which otherwise would given rise to high background hybridisation. The filters were then placed between Whatman 3MM paper and two glass plates and dried in a vacuum oven at 80°C for 4 hours.

5.2.4 *Autoradiography*

The dried, hybridisation filters were fixed to filter paper sheets and covered with 'cling film'. They were exposed to Kodak medical X-ray film in Kodak cassettes, with intensifying screens, for 3-4 days. The films were developed by immersing in D19B developer (Kodak) for 5-10 min, rinsed in water and then fixed in FX80 fixer (Kodak) for 5 min, in a dark room. The films were exhaustively rinsed in tap water before air drying.

5.3 Screening the Library with Specific Hybridisation Probes

5.3.1 *Nick-translated Ribosomal DNA*

A specific hybridisation probe for ribosomal DNA sequences was prepared by nick translation of cloned ribosomal genes as detailed in Chapter II, Section 2.3.

McClements and Skalka (1977) screened a series of chicken genomic DNA fragments cloned in the vector λgtWES for ribosomal sequences. They estimated that the fragments screened represented 5-10% of the total chicken genome and from this concluded that there are approximately 200 copies of the ribosomal gene per chicken genome, representing 0.12% of the total DNA. The sequences are clustered (Sinclair and Brown, 1971), as has been shown for many other eukaryotic species, and are cleaved into fragments of two sizes of 5×10^6 and $12\text{-}14 \times 10^6$ daltons which suggests, as is also the case for the other eukaryotic species so far studied, that the 18S, 28S and 5.8S ribosomal genes are present as tandem repeats.

(a) *Sequence representation of the library:*

Due to the highly repetitive nature of the ribosomal sequences in the genome it was considered reasonable to expect, even in an incomplete library, to be able to detect a number of clones carrying rDNA. The absolute number which were detected was expected to provide an indication of the completeness of the sequence representation of the library and thus give an estimate of the number of clones which need to be screened in order to be sure of detecting a sequence belonging to the 'unique abundance class', such as myosin heavy chain. McClements and Skalka (1977) estimated that the cloning of a single copy gene would require the screening of 20 genomes of DNA. However,

their eukaryotic DNA inserts averaged 10,000 base pairs, that is, half the size of those cloned here, so using this estimate, the equivalent of ten genomes should be screened to increase the probability of detecting a single copy sequence.

A total of 10 recombinant phage plates were screened by the procedure described in Section 5.2 using the ^{32}P-labelled ribosomal DNA. A total of 205 spots were identified on the autoradiographs, of which 188 probably represented true recombinant phage. 17 spots were discounted as representing non-specific adsorption of the ^{32}P-probe to the nitrocellulose filters. An example of such a non-specific spot is shown in the autoradiograph in Figure 5.1. These were identified by their very intense appearance and very sharp edges and arose due to adsorption of the nucleic acid probe by the nitrocellulose at sites·in the membrane which had not been completely blocked by the Denhardt treatment. Such spots are never present on the duplicate filter. Spots representing true hybridisation of probe to complementary recombinant DNA have less well defined edges, often are not completely circular, and of course are present on the duplicate filter.

The identification of 188 positive plaques from a total of 10 re-combinants carrying 20 kb chicken DNA inserts indicates that the ribosomal gene sequences represent 0.19% of the chicken genome. This is in close agreement with the previous estimate of McClements and Skalka (1977) of 0.12% and therefore suggests that the genomic library produced in this project did not contain any major deficiencies, and therefore the sequence representation of the library must be approach-ing completion.
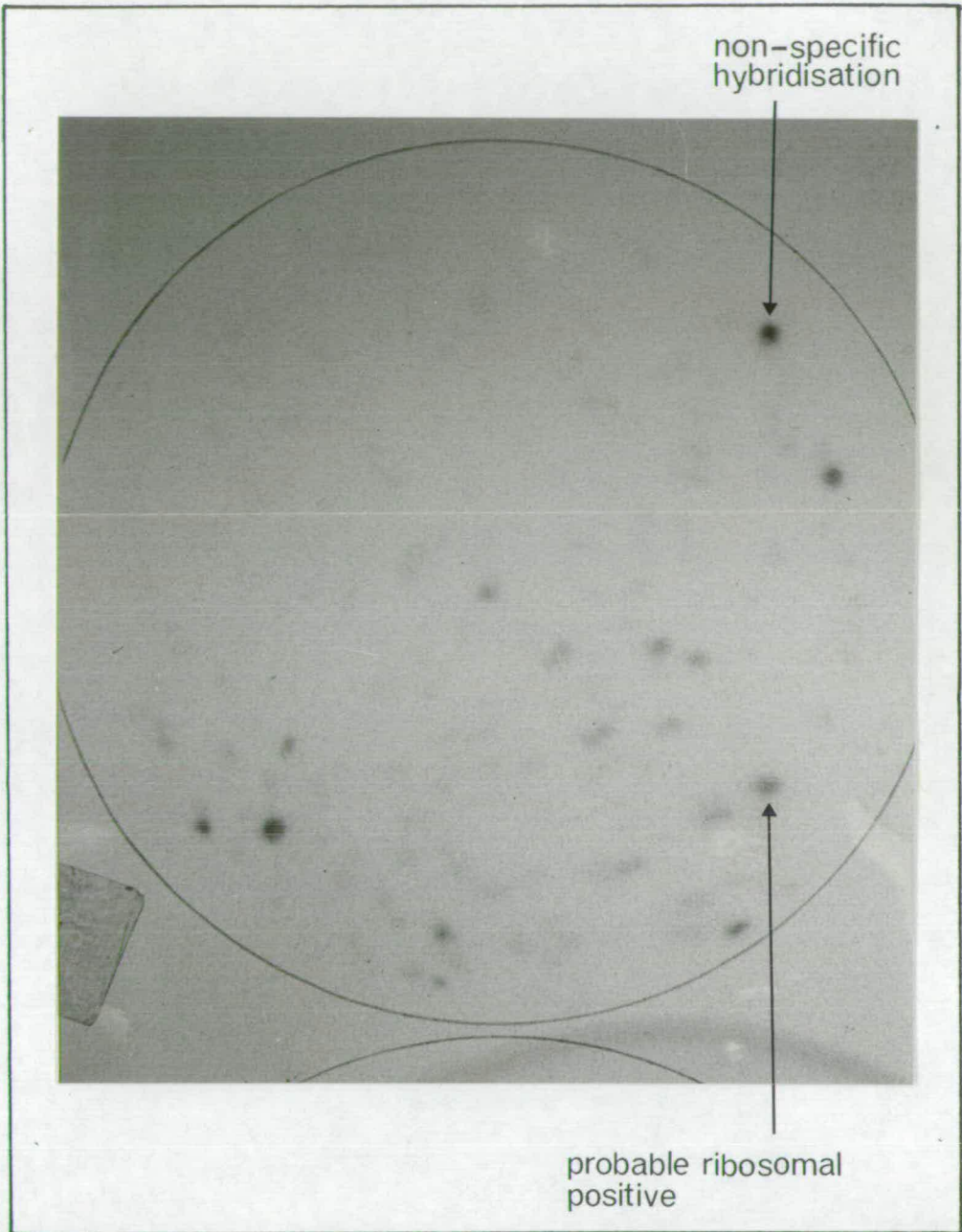
FIGURE 5.1: Autoradiogram of Recombinant Charon 4A Plaques
Screened with the Nick Translated ^{32}P-ribosomal DNA
Probe.

(b) *Contamination of the myosin heavy chain cDNA probe*

The ribosomal DNA probe was also used to test the possibility of contamination of the myosin heavy chain cDNA with ribosomal sequences. As shown in Chapter II, Table 2.2, the myosin heavy chain mRNA preparation was contaminated, to an extent of 35%, with rRNA. In theory, rRNA should not have been a substrate for reverse transcriptase due to the absence of a primer sequence, but it has been shown (King *et al.*, 1979; Williams and Lloyd, 1979) that low levels of reverse transcription of rRNA can take place. This did not occur to a very high degree, as shown by the hybridisation studies presented in Chapter II, Section 2.2.9, between the cDNA and its template. However, even a very low level of reverse transcription can lead to an unrepresentatively high number of false positive plaques during screening due to the repetitive nature of the ribosomal genes. This was indeed shown to be the case.

To determine the extent of such possible contamination, DNA from the 10 plates previously screened with the rDNA probe was adsorbed onto a third filter. This was carried out in the same way as described for the duplicate set of filters, except that adsorption was allowed to continue for 15 min to ensure that sufficient DNA was transferred from the plaques to the nitrocellulose. This third set of filters was screened with the ^{32}P-cDNA probe. Eleven positive plaques were detected and their positions were compared with the plaques picked up by the ribosomal sequence probe, and of these two were possibly detected by both probes. It was impossible to be absolutely certain without re-screening due to the fact that the boundaries of some plaques were touching and there was no duplicate filter as a double check. However, this suggests that approximately 20% of the recombinants recognised by

the myosin heavy chain cDNA probe will be due to probe contamination which arises due to a very low level of reverse transcription of contaminating ribosomal RNA in the original mRNA preparation. However, due to the repetitive nature of the ribosomal sequences and the fact that the cDNA only cross hybridised with a possible 2 out of 188 recombinants, this level of contamination was not considered to be excessive.

5.3.2 ^{32}P-labelled Myosin Heavy Chain cDNA

A total of 2.5×10^5, the equivalent of five chicken genomes of DNA, were screened by the Benton and Davis method, as described in Section 5.2, with a specific myosin heavy chain ^{32}P-cDNA probe prepared, as described in detail in Chapter II. This was expected to provide a 96.9% chance of detecting a single copy gene sequence.

An autoradiogram of 15 cm filters which had been adsorbed with recombinant DNA and then treated with the ^{32}P-cDNA probe is shown in Figure 5.2. The regions of non-specific hybridisation are obvious by their high intensity and definite boundaries as was discussed for similar spots occurring in Figure 5.1. True myosin heavy chain sequence recombinants are represented by the more diffuse 'plaque-sized' areas of positive hybridisation as shown in the autoradiogram (Figure 5.2).

A total of 75 areas of hybridisation were detected, of which 22 were immedaitely discounted as areas of non-specific adsorption of the probe by the introcellulose. It can be estimated, from the data presented in Section 5.3.1 above, that 20% of the 53 positive recombinants may have arisen due to hybridisation by contaminating sequences in the probe which recognise ribosomal DNA sequences within the library. It is impossible to identify which of these positive clones represent ribosomal

probable myosin heavy
chain positive
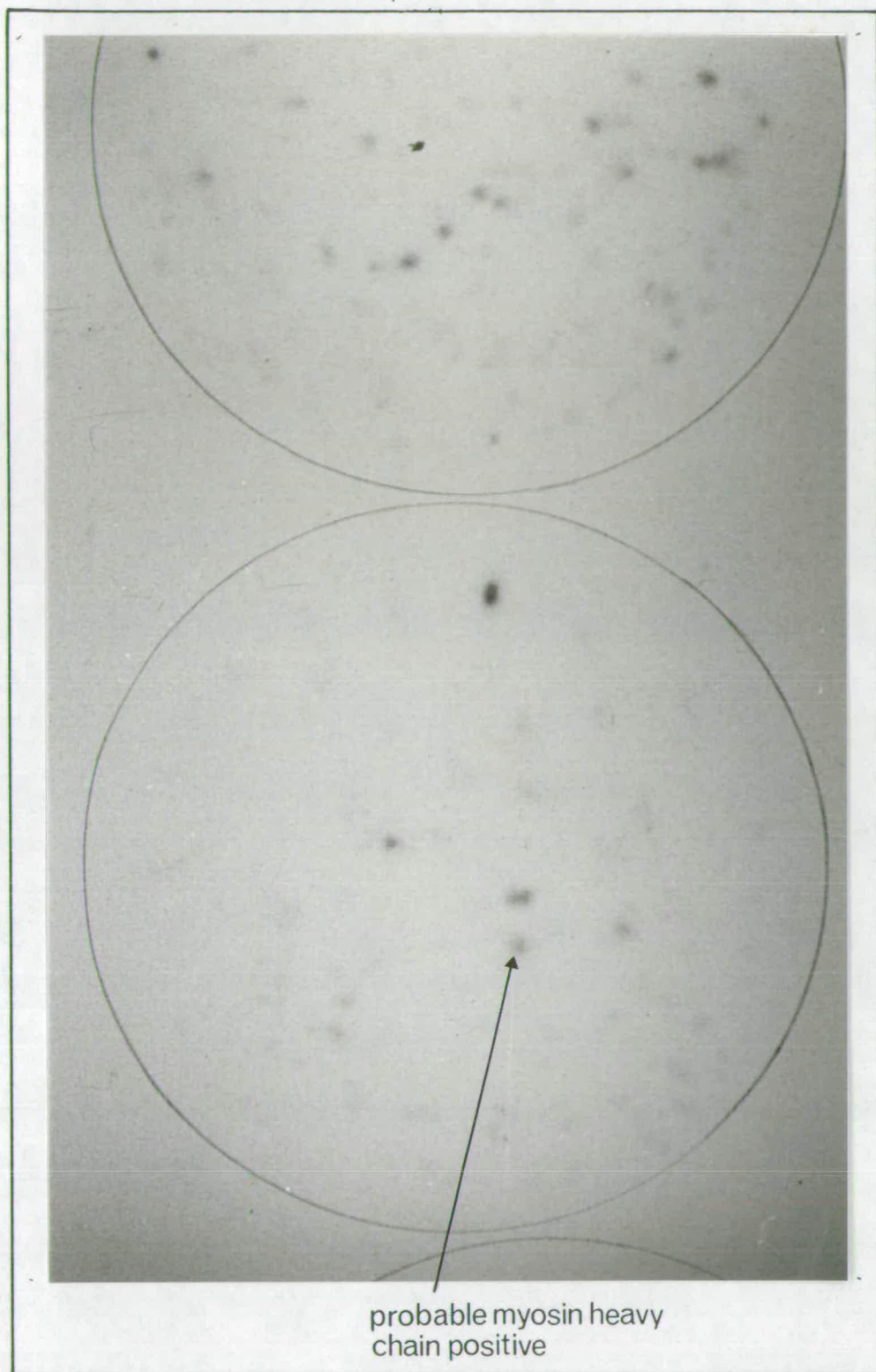
FIGURE 5.2:   Autoradiogram of Recombinant Charon 4A Plaques
              Screened with the ^{32}P-myosin Heavy Chain cDNA Probe.

The diffuse areas of hybridisation which lie directly over plaques
when the autoradiogram is compared with the original plate are those
most likely to represent phage carrying the mysoin heavy chain
gene sequence.

genes without screening identical filters with the ^{32}P-rDNA probe.
Therefore, it can be concluded that within a total of 2.5 x 10^5 recombin-
ant phage, approximately 43 were expected to contain myosin heavy
chain gene sequences.

5.4 Plaque Purification

The isolation of a recombinant phage of interest is achieved by
picking plaques from the region of a plate corresponding to an area of
positive hybridisation on the autoradiograms. This is usually performed
with a sterile pasteur pipette which removes a complete plug of agar
containing the plaque of interest. The phage are then resuspended
in 0.5 ml phage buffer and the suspension should be titred as described
in Chapter IV, Section 4.3.3. A titration plate containing about 100
plaques is then rescreened with the specific ^{32}P-labelled probe. The
plaques will now be more widely distributed on the plate, enabling a
positive to be repicked by the same procedure, and again screened until
more than 90% of the plaques on a plate give positive signals. It is
then acceptable to prepare the recombinant phage DNA for further study
from a single positive plaque by growing the phage as a plate lysate,
as described in Chapter IV, Section 4.4.2, and as was carried out to
amplify the library (Chapter IV, Section 4.9).

5.5 Conclusions

A summary of the numbers of positive recombinants identified
following screening of the library is presented in Table 5.1.

The nick-translated ribosomal DNA probe was prepared and
successfully used to screen the chicken genomic library purely as a
control. The aim was to test the completeness of the sequence representa-

TABLE 5.1:

Probe	Number of recombinants screened	Number of positive clones identified			Contamination with other sequences (%)	Estimated proportion of chicken genome (%)
		non-specific hybridisation	cross reaction with rDNA	true positives		
Nick-translated ^{32}P-ribosomal DNA	10^5	17	N/A	188	0†	0.19
Myosin heavy chain ^{32}P-cDNA	i) 10^5 ⎤ 3.5 x 10^5 ii) 2.5 x 10^5 ⎦	14 22	2 10*	9 ⎤ 52 43 ⎦	20 ⎤ 20* ⎦	0.0015

N/A   not applicable

†   the extent of contamination of rDNA probe was not determined in this project, but since the original material was obtained from cloned DNA, the sequence should be pure.

*   these figures were not directly determined in this project, but were inferred from the data presented in line i) of this table.

tion, the purity of the myosin heavy chain cDNA probe and the Benton

and Davis (1977) screening procedure. It was shown by screening the

equivalent of two genomes of chicken DNA that recombinants containing

ribosomal sequences make up 0.19% of the genome. This implies that

the library is complete – that is, all sequences in the chicken genome

are likely to be represented on a statistically random basis, assuming

that no positive selection was operating against particular sequences

at any stage during the construction of the library. The rDNA probe

also showed, by cross hybridisation with some recombinant plaques,

that the myosin heavy chain cDNA probe was contaminated with ribosomal

sequences to an extent of 20%, and therefore before embarking on plaque

purification of single recombinant phage, these ribosomal clones should

be positively identified by rescreening.

Allowing for contamination and non-specific hybridisation, 43 out

of 2.5×10^5, and 9 out of 10^5 recombinant phage were detected following

hybridisation with the myosin heavy chain probe. Several inferences can

be drawn from these results. Since the equivalent of seven genomes

were screened in total, the myosin heavy chain gene must be present

at a frequency of greater than one copy per genome, unless the genomic

sequence is so large, and incorporates sufficient AluI and HaeIII

recognition sites, as to be distributed throughout many clones. However,

even if this possibility were true, it is unlikely that the relatively

short cDNA probe of 500 nucleotides would contain sufficient sequence

homology to be able to hybridise with all the AluI and HaeIII restriction

fragments which were generated, since it is prepared against 10% of

the mRNA only and probably corresponds to the 3' end.

The cDNA was shown in Chapter II, Section 2.2.9 to consist of

more than one myosin heavy chain sequence. The $R_0t_{\frac{1}{2}}$ value obtained

204.

from the kinetics of hybridisation to the template mRNA indicated that 2-3 different myosin heavy chain sequences were represented, and therefore if these sequences are present as single copy in the genome, then 14-21 positive plaques would be expected. However, since more than twice this number (that is, 43) areas of true probe hybridisation were observed, it can be concluded that either the myosin heavy chain sequences are duplicated within the genome of the chicken, or that they exist as a family with considerable sequence homology, representing myosin heavy chain for various types of muscle, with which this short cDNA is able to cross react.

CHAPTER   VI

Discussion

CHAPTER VI

6.1 Introduction

The original aim of this project, as described in Chapter I, was to further investigations into muscle function genes at the level of the nucleic acids as part of an on-going scientific programme being conducted in this laboratory with the support of the Muscular Dystrophy Group of Great Britain. The reasons for choosing the myosin heavy chain gene in particular were also discussed in Chapter I with an outline of the strategy behind the work presented in this thesis. As previously mentioned, it was decided to construct a genomic DNA library with a view to isolating the myosin heavy chain gene sequences with their corresponding flanking and intervening sequences using a purified hybridisation probe. Such specific sequences should be invaluable tools for investigations into controversies arising from current research on muscle development.

The achievements of the project can be classified under three headings:

(a) a specific hybridisation probe for the chick embryonic skeletal muscle myosin heavy chain gene was prepared and characterised, as described in Chapters II and III.

.(b) A chicken genome DNA library was constructed, as described in Chapter IV.

(c) Using plaque hybridisation techniques, clones carrying myosin heavy chain gene inserts were identified, though the actual structure of these sequences was not investigated. However, one advantage of genomic DNA libraries is that they can be continually screened for any sequence for which a probe is available and therefore such clones can be isolated in the future for sequence determinations.

Since this project was initiated others have adopted similar schemes which will be discussed later in this chapter.

206.

6.2 Conclusions

6.2.1 Purity of the cDNA Probe and Its Cross Hybridisation with Cloned Genomic Sequences

The reliable screening of the DNA library, in an attempt to identify recombinant bacteriophage carrying myosin heavy chain gene sequences, was primarily dependant on two qualities of the cDNA probe. These were, most importantly, the purity of the probe, and, secondly, the cDNA length, which would partially dictate the final number of recombinants identified due to cross reactivity with different myosin heavy chain gene sequences, the extent of which is unknown. The more complete the copy, the more specific it will be as a hybridisation probe.

It was shown in Chapters II and III that myosin heavy chain mRNA could be isolated in a very pure form from a single 1 ml fraction, f2, collected from the bottom of a heavy polysome sucrose gradient. Other fractions collected from the heavy polysome region also contained myosin heavy chain mRNA, but never in a form uncontaminated with other messengers. The purity was determined initially by *in vitro* translation of the mRNA in three cell-free systems: one homologous, requiring addition of polysomes and, two heterologous, the wheat germ and rabbit reticulocyte lysate systems, requiring addition of mRNA. Polyacrylamide gel electrophoresis in the presence of denaturing agents, SDS and β-mercaptoethanol, demonstrated that a single protein was synthesised by the f2 polysomes which co-migrated with purified myosin heavy chain with an apparent molecular weight of 200,000 daltons.

Such *in vitro* translation studies indicated that the mRNA preparation was not contaminated to any detectable extent with other mRNA

species. However, these gave no indication as to the possible extent of contamination with ribosomal RNA sequences, which would not be identified in this way. Following reverse transcription, therefore, the sequence complexity of the probe was further examined by the hybridisation of the ^{3}H-cDNA to its template mRNA. The kinetics of hybridisation showed a single transition approaching 100% duplex formation. The absence of minor transitions further reinforced the conclusion that the mRNA, and thus the cDNA, was not contaminated to any detectable level with minor RNA species, though the $R_0 t_{\frac{1}{2}}$ value obtained did indicate the possible existence of more than one myosin heavy chain sequence in the mRNA population as was suggested previously by Patrinou-Georgoulas and John (1977). However, the likely possibility of contamination by rRNA cannot be eliminated since rRNA is not efficiently reverse transcribed due to the absence of a suitable primer sequence and therefore may not be detected by the back hybridisation procedure. This argument, however, can be similarly used in a more positive way: if rRNA is not reverse trans- cribed, then even if it is present in the mRNA, it will not be represented in the cDNA and therefore will not interfere with the screening process since it is the cDNA which is used as the probe. The copying efficiency of rRNA by reverse transcriptase has been reported to be very low (King *et al.*, 1979; Williams and Lloyd, 1979) and does not generally present a problem but it is important to be aware that mRNA prepara- tions are normally contaminated to a significant extent with rRNA, and therefore false positives can arise if screening is performed with a cDNA as in this project. Such effects can be further amplified because of the multiplicity of ribosomal genes. Ideally, cDNA cloned in a bacterial plasmid is the better probe, since, by definition, it represents

only a single sequence and this reduces the chances of detecting false positives. Thus, the genomic sequence of an abundant mRNA can be detected more convincingly if the cDNA is first cloned and then isolated from a single, well characterised recombinant plasmid and then employed as a probe for screening genomic libraries. This sort of scheme would be very feasible for myosin heavy chain due to the abundance of the mRNA, and would allow the use of a single sequence probe, rather than a mixed one, as was probably employed here. Myosin heavy chain cDNA clones have been isolated recently by other workers and will be discussed in a later section of this chapter.

The degree of contamination of the myosin heavy chain mRNA preparation was estimated to be as high as 35% as shown in Table 2.3, though the extent to which this was transfered to the cDNA probe following reverse transcription was not determined until late in the screening process. The same clones were screened with both the cDNA and a specific nick-translated ribosomal DNA probe to estimate the number of clones recognised by both probes. This was too late to make any adjustments to the composition of the probe, but did show that the cDNA was much less contaminated with ribosomal sequences than its mRNA template. However, a significant number of common recombinants were recognised by the two probes which is probably a reflection of the repetitive nature of the ribosomal genes which are represented in approximately 50 times as many phage as the myosin heavy chain sequences. However, awareness of these findings made it possible to use the cDNA probe with an acceptable degree of confidence.

The ^{32}P-labelled myosin heavy chain cDNA sequence was used to screen a total of 2.5 x 10^5 recombinant phage, which corresponded to

5 genomes of chicken DNA since the inserts were close to 20 kb in length. If the probe was able to recognise a single copy gene only, and the library was a complete sequence representation of the chicken genome, then an absolute maximum of five positive clones would be expected. However, a tenfold increase on this expected value was obtained, and 53 positive recombinants were identified by plaque hybridisation. Ten of these, that is 20%, may be attributed to contaminants in the probe which recognised recombinant phage carrying ribosomal sequences. The fact that a further 43 recombinant phage were able to hybridise with the probe suggests three possibilities:

(a) There is a relatively high degree of non-specific hybridisation to the plaques, or non-specific adsorption by the nitrocellulose filters.

(b) The two or three myosin heavy chain genes represented in the cDNA probe are present in several copies in the genome.

(c) The probe is able to cross react with the genomic sequences of different myosin heavy chain genes which, therefore, may exist as a family of sequences even though they may not be directly represented in the probe, such as cardiac and non-muscle myosin heavy chain gene sequences.

The first of these possibilities is a technical problem which was discussed in detail in Chapter V. Spots on the autoradiographs corresponding to non-specific adsorption of ^{32}P-cDNA by the nitrocellulose filter have definite characteristics which make them readily identifiable, and, therefore, such artifacts are not included in the 43 positives referred to here. The degree of non-specific hybridisation to recombinant phage can only be tested by rescreening isolated plaques.

In this study, it is impossible to distinguish between possibilities (b) and (c). Since the cDNA is only a partial copy (10%) of the mRNA, as shown in Figure 3.6, it is likely, due to the similar amino acid sequences of various myosin heavy chains that it will not be completely specific for one sequence only. The several forms of the myosin heavy chain polypeptide have some similar properties, such as electrophoretic mobility, the ability to complex with other muscle proteins, such as actin, and some have similar ATPase activities. Therefore, they may share common amino acid sequence domains which confer these properties and which will be represented in the nucleotide sequence, to a certain extent, also. It is therefore probable that a partial DNA copy of a myosin heavy chain gene will recognise several related, but not identical, genomic sequences, and the extent of any cross hybridisation will depend partially on the length of the cDNA probe. The more complete the copy, the more unstable hybrids will be under stringent hybridisation conditions, if they are not entirely complementary, and therefore the more specific the probe will be for one or very few sequences. However, these arguments may not be entirely justified as Mahdavi et al. (1982) have shown in some recent sequencing studies on cardiac muscle myosin heavy chain that the 3' end of the mRNA molecule consists of a short nucleotide sequence which is highly specific to one particular myosin heavy chain type, and Rozek and Davidson (1983) also indicate that the 3' end of the primary transcript is important in determining myosin heavy chain type in Drosophila.

It would, therefore, be desirable to obtain a complete cDNA probe rather than the 10% copy only. Much effort has been directed towards defining conditions which optimise cDNA synthesis due to its importance in molecular biological research, but unfortunately there are

many discrepancies between different laboratories. The two main criteria are the yield of cDNA per amount of input mRNA, and the length of the cDNA; and the optimal conditions for the former are not necessarily those for the latter. It is, of course, the latter criterion which is important here, and some effort was made during this work to devise a method of increasing cDNA length as mentioned in Chapter III. However, since then an attempt has been made to rationalise the various reaction conditions used in many laboratories and has shown that much of the confusion arises from the use of reverse transcriptase contaminated with ribonuclease (Retzel *et al.*, 1980). By using highly purified reverse transcriptase, the other modifications to the basic procedure have no effect, and a high proportion of full length transcripts can be obtained from the avian myceloblastosis virus (AMV) genome, an RNA which is 3-4-fold longer than the average eukaryotic mRNA at 7500 nucleotides, and is also longer than a very long eukaryotic mRNA, such as myosin heavy chain which contains 7100 nucleotides. However, it must be realised that reverse transcription is a natural process to AMV, from which the enzyme is obtained, and therefore efficient copying is to be expected. The eukaryotic mRNA used in the studies of Retzel *et al.* (1980) and which was copied completely, was conalbumin, consisting of 2750 nucleotides which is considerably shorter than the myosin heavy chain messenger.

Therefore, the number of positive plaques identified would be reduced if a full length cDNA probe was available for screening, but this would still depend very much on the degree of sequence homology between the various myosin heavy chain genes. The only satisfactory way to distinguish between the clones would be to grow them in isolation to obtain enough DNA for sequencing studies. This, of course, would

be highly desirable, since specific probes for the various myosin heavy chain genes may then become available. Knowledge of the complete genomic sequence may reveal particular regions of the molecule unique to one heavy chain type and if of sufficient length, this oligonucleotide region could be synthesised *in vitro* and employed as a specific probe in studies on muscle development using the technique of *in situ* hybridisation (Jones and Murray, 1975) with tissue slices. A synthetic oligonucleotide of just 13 nucleotides was sufficient to identify the mRNA for the hormone, gastrin (Noyes *et al.*, 1979), by Northern blotting of total mRNA isolated from the intestine onto nitrocellulose filters followed by hybridisation of this probe with the immobilised mRNA. Such a technique is becoming more widely used and could be adapted to hybridisation studies on tissue sections. A small synthetic probe could represent an intron where the sequences may be more diverse between different genes and the problems of extensive similarities in the cDNA sequences would then become irrelevant. Such small, but specific oligonucleotides for the different myosin genes would be extremely useful in research into muscle development and associated disorders.

6.2.2 *Sequence Representation within the Library*

A further very important factor determining the number of positive clones identified in this study is the completeness of the sequence representation of the library. Several factors are to be considered here:

(a) Were sufficient 20 kb fragments, randomly generated by partial restriction enzyme digests, available for cloning?

(b) Were sufficient clones generated following *in vitro* packaging and what are the selection pressures acting on this process?

(c) Do selection pressures act on the individual re-
combinants during amplification of the library?

(d) Were the final numbers of clones obtained and
screened sufficient to statistically have a very high
chance of representing all single copy gene sequences?

When the library was screened with a ^{32}P-nick translated ribo-
somal DNA probe, 188 positive plaques were identified from the 10^5
which were screened, as described in Chapter V, Section 5.3.1.
This indicated that the ribosomal gene sequences make up 0.19% of
the total genome which is in close agreement with previous estimates
(McClements and Skalka, 1977) and therefore suggests a complete
chicken genome sequence representation in this library. However,
due to the much less abundant nature of the myosin heavy chain
sequences, the factors listed above must also be given careful con-
sideration when attempting to determine the expectation of detecting
such a sequence amongst so many inappropriate ones.

The chicken DNA was fragmented using a double, partial digest
with the restriction endonucleases HaeIII and AluI, as described in
Chapter IV, Section 4.5.3. The reaction conditions were suitably
manipulated so that a large proportion of the resulting fragments were
in the size range 19 to 20 kb in an attempt to prevent elimination of
particular sequences at this initial stage. Both enzymes have the
minimum recognition sequence of four bases which results in a more
random collection of fragments following partial digestion. A total of
65 μg of recombinant DNA was employed in the packaging reaction
which corresponds to 26 μg of chicken DNA and represents 2.08 x 10^7
chicken genomes (the haploid DNA content of the chicken is 1.25 pg;
Mirsky and Ris, 1949; Vendrely and Vendrely, 1949). Therefore,

if the restriction enzymes operated in a truely random fashion, all single copy sequences should be represented in this amount of recombinant DNA (Clarke and Carbon, 1976). However, the experience of some workers has shown that libraries constructed by the methods described in this thesis may lack certain sequences. For example, Maniatis and his co-workers (Lawn *et al*., 1978) constructed a human genomic library using the restriction endonucleases HaeIII and AluI, but were unable to identify clones carrying the γ globin gene or the region between the γ- and δ-globin genes, though several β-globin gene recombinants were isolated. The γ-globin sequence was eventually isolated from a library constructed with the restriction endonuclease, EcoR1, and the region separating the γ- and δ-globin genes from a library constructed with MboI digested DNA (Van Der Ploeg *et al*., 1980). It is not understood why some sequences are absent or under-represented following some enzyme digests. Some sites may be preferentially cut, even in partial digests giving small fragments which will never be packaged due to the lower limit size restriction, or there may be isolated regions in the genome of several thousand base pairs (greater than 20 kb in these cases) which simply do not contain the necessary recognition sequences, even though this is highly unlikely on a purely statistical basis, giving large fragments which are also eliminated prior to packaging. Of course, some recombinants may be lost during the cloning or amplification processes for other reasons to be discussed later.

A total of 6.8×10^5 recombinant phage particles were recovered from the packaging reaction, which represents greater than 10 chicken genomes. All sequences should be represented in this number (Clarke and Carbon, 1976), though it is towards the lower limit for certainty

of single copy sequence representation. The efficiency of packaging
(10^4 recombinant phage were recovered per μg of input recombinant

DNA) was also low when compared with the values of 3.1 x 10^4- 5.6 x

10^5 obtained by Maniatis *et al.* (1978) though 10^4 is equivalent to the

maximum recorded for transfection efficiencies. The low packaging

efficiencies may be due to the quality of the packaging extracts since

the efficiency of packaging intact charon 4A DNA (4.1 - 8.3 x 10^6 pfu

per μg) was found to be, on average, ten-fold lower than that recorded

by Maniatis *et al.* (1978) of 2 - 20 x 10^7 pfu per μg. The packaging

extracts used in this study were prepared from different bacterial

strains to those used by Maniatis and co-workers, but without comparing

the two directly it is impossible to determine whether one set of strains

has some advantage, or whether an unknown factor contributed to

the lower packaging efficiencies obtained here. However, it is necessary

to remain aware of the fact that these low packaging efficiencies may

contribute to the under-representation of sequences in the final library.

Though, if this arises, it will be an entirely random process since

packaging does not rely on phage growth within a host cell, as does

transfection, and therefore within certain size limits, the characteristics

of the DNA being packaged are of little consequence.

Several factors may contribute to the further loss of sequences

during library amplification. Slow growth of some recombinant phage

due to a disadvantage such as slow replication, the production of toxins

by certain clones, or by further recombination of the recombinant DNA

molecule, is one possible reason for selective loss, though efforts were

made to minimise these competitive growth effects and any other selection

pressures of a genetic type, by amplifying the library on low density

plate lysates, as described in Chapter IV, Section 4.9. The eukaryotic

DNA insert size may also exert selection pressures on the growth of a particular recombinant phage. Fiess *et al*. (1977) showed that low yields of recombinants were obtained following amplification if the recombinant DNA molecules which are of a size towards the upper or lower limits of the packaging range. Therefore, since recombinants are continually repackaged every generation during amplification, those with a size approaching these limits will be selected against which may result in selective elimination of certain sequences from the complete library. For this reason, 20 kb were chosen as the insert size in this project as a compromise between the maximum permissible size of 22.5 kb, which would subject many clones to the possibilities of elimination, and a smaller size which would unnecessarily increase the number of clones required to detect a single copy sequence when screening. The loss of single copy sequences during library amplification was tested in a simple experiment. Tritiated single copy tracer DNA from *Drosophila* was hybridised to both sheared *Drosophila* genomic library DNA and embryo DNA. The re-association rates were found to be identical and therefore the single copy complexity in the library was the same as in the intact organism. In this case, there was no appreciable loss of single copy DNA during library amplification, though the loss of a few selective sequences would not be detected by this method.

The stability of clones has also been shown to present a problem in certain cases and is worth consideration here. In a few cases, specific deletions in DNA cloned in phage λ have been detected and shown to be artifacts in the cloning procedure. For example, Lacy *et al*. (1979) detected a 4.35 kb deletion in a rabbit β-like globin sequence, Fritsch *et al*. (1980) showed a deletion in a clone spanning the region between the $^{G}\gamma$-and $^{A}\gamma$-globin genes, and Lauer *et al*. (1980)

showed that deletions occur at high frequency in phage λ clones containing the α-like genes. A common factor in these examples seems to be the occurrence of deletions in genomic areas showing a high degree of sequence homology, and therefore perhaps arise by unequal crossing over during growth of the phage. This is reinforced by the finding that cosmids grown in a recA$^-$ host have a very low incidence of such deletions. Since it is probable that different myosin heavy chain genes share some common sequences, the possibility of crossing over occurring between different genes and thus introducing deletions is very real and should be considered when sequencing.

The haploid DNA content of evolutionary diverse organisms varies considerably, as shown in Table 6.1. For example, *Escherichia coli* has the potential to code for 3000 – 4000 different gene products, whereas the corresponding number in humans is 10^6, though only a small proportion of this DNA does code for protein. Table 6.1 also shows the theoretical number of clones required to represent the total genomes of four very diverse types of organism. Statistically, this means that there is a 50% chance of detecting a single copy sequence when screening a library containing this minimum number of clones, and when screening twice this number there is a 75% chance, and so on. Therefore, 3-10 times the minimum number of clones should be screened to ensure a reasonable chance of identifying a particular gene. As stated previously, 6.8×10^5 clones were obtained following *in vitro* packaging which represents greater than 13 times the minimum number, and following amplification, 2.5×10^5 recombinants were screened representing five times the number required for a 50% probability of isolating a particular gene which is equivalent to a 96.9% chance of detecting a single copy sequence. Since 43 clones were recognised by the cDNA probe, it was

TABLE 6.1: The Theoretical Number of Recombinants Required
to Represent a Complete Genome.

Size of fragment (kilobase pairs)	Organism			
	Bacteria 2 x 10^6 bp	Fungi 2 x 10^7 bp	Chicken 1 x 10^9 bp	Mammals 3 x 10^9 bp
5	400	4,000	200,000	600,000
10	200	2,000	100,000	300,000
20	100	1,000	50,000	150,000
40	50	500	25,000	75,000

The theoretical, minimum number of clones required when cloning
genomic DNA into a bacteriophage λ library in order that the total
genome of four very evolutionary diverse species will be represented.
Figures are given for DNA fragment sizes of 5, 10, 20 and 40 kb.

considered that one or more of these initially positive recombinants would contain a myosin heavy chain sequence, and therefore that the number of clones obtained and the screening procedure were adequate. A second, and possible third, screening of these first-time positives will be necessary to determine unambiguously how many represent true myosin heavy chain genes and how many are false positives due to unavoidable non-specific hybridisation of the probe, the extent of which is entirely unknown.

6.3 Other Studies of this Type

As mentioned at the beginning of this chapter, several workers have adopted a similar strategy to that outlined in this thesis. The lines of investigation can be broadly divided into two categories, though these are very much inter-related. The first is the structure and number of the myosin heavy chain gene(s) and the second is the regulation of expression of these genes during development and in specific muscle cell types. As discussed in Chapter I, Section 1.3.7, most information so far has been gained from cDNA clones, and a large amount of interesting sequence data has emerged. However, genomic myosin heavy chain gene clones, in combination with cDNA clones, must be used to study the actual structure of the gene sequence and to identify possible control regions within that sequence. An initial study of this nature has been presented by Nudel *et al*. (1980).

6.3.1 *Myosin Heavy Chain Gene Number*

A series of cDNA clones which hybridised with myosin heavy chain mRNA were constructed by Nudel *et al*. (1980) and a rat genomic library, constructed in charon 4A, was screened using one of these

clones as a hybridisation probe. The positive genomic sequences fell into three categories based on their restriction patterns following digestion with the enzymes EcoR1, HindIII and BamHI. Two of the isolated genomic sequences were hybridised to each other and studied by heteroduplex analysis under the electron microscope. This revealed the existence of several homologous regions interspersed with regions of non-homology. Such non-homologous areas may simply represent non-hybridising sequences within two different genes; or they may represent intervening sequences. Nudel *et al.* (1980) suggest the latter due to the difference in size between the observed loops. The myosin heavy chain proteins are all of a similar size and therefore may be expected to contain similar amounts of DNA within the coding regions of their genes, though this argument may not necessarily be true, as some splicing of information may take place at the level of a precursor RNA, following transcription. The size and number of these intervening sequences is not suggested by Nudel and co-workers, but more recent, similar studies by Wydro *et al.* (1983) suggest introns which total 30 kb of DNA within a gene whose mRNA is only 7100 nucleotides in length, and Rozek and Davidson (1983) have tentatively identified 9 introns within a single myosin heavy chain gene of *Drosophila melanogaster*. The mRNA length also suggests the possible existence of further large regions of non-homology within the myosin heavy chain genes since only 5500 nucleotides are required to code for the 2×10^5 dalton protein. This implies that one or more large areas of the gene exist which are represented in the mRNA but are not translated and whose nucleotide sequence may have diverged considerably.

The suggestion that the myosin heavy chain genomic sequence of the rabbit is of the order of 37 kb (Wydro *et al.*, 1983) implies that

a single charon 4A recombinant, containing the maximum insert of 22.5 kb, will never contain the whole gene. Indeed, the DNA library prepared in this project consisted of charon 4A recombinants carrying 20 kb inserts, and therefore it is likely that a complete genomic sequence will be detected in a minimum of two or three individual recombinants depending on the exact positions of AluI and HaeIII restriction sites. This uncertainty makes the estimation of gene number from the number of positive clones identified within a library very difficult. A *Drosophila* genomic library was recently screened by Rozek and Davidson (1983) and two clones representing a single MHC gene were identified. The genomic length of the gene is estimated to be 20-22 kb, with a transcriptional unit of 19 kb, carrying at least 9 introns. Chromosome walking techniques can be used to isolate recombinants carrying pieces of adjacent, overlapping DNA. Such techniques, employed here to piece together a complete copy of the myosin heavy chain gene, may answer such questions as to what lies upstream or downstream from a given gene or cluster on a given chromosome. The principal of this very useful molecular mapping technique involves fragmentation of a particular positive clone with a restriction endonuclease to generate a terminal fragment which contains single copy DNA. This is then labelled *in vitro* by nick translation and hybridised back to the original library, making it possible to isolate more clones containing this terminal fragment, some of which will extend further upstream if a 5' fragment was used as the probe, or downstream if the probe was a 3' fragment. This process can then be repeated, and theoretically a whole myosin heavy chain gene and its flanking sequences could be isolated. However, since many investigations indicate that the myosin heavy chain gene family

has several members in birds and mammals, then similarities in sequence
may make the isolation of sufficient lengths of single copy DNA, suitable
as probes in the chromosome walking technique, very difficult. There-
fore, if the myosin heavy chain gene sequences are shown to be very
large, it may be wiser to use another cloning system which accepts
much larger DNA fragments as inserts, such as the cosmid system.
The chances of isolating a gene sequence in a single recombinant are
then very much increased. This would simplify structural comparisons
of individual members of a gene family due to increased certainty that
a particular sequence is not a hybrid of various members of the family.

Cosmids are cloning vectors which accept foreign DNA inserts of
up to 45 kb. They are derived from plasmids which also contain the
bacteriophage λ cohesive ends (Collins and Brüning, 1978; Collins and
Hohn, 1979) and are therefore recognised by the λ packaging system as
described in Chapter IV. They are transduced at high efficiency
into *E. coli* and once within the bacterial cell they circularise and
resemble plasmid vectors, rather than λ vectors, by confering some
antibiotic resistance. They are frequently less than 5 kb in size, and
therefore, because of the size restrictions imposed by the packaging
system, the foreign DNA insert must be of a minimum size. Hence,
the cosmid system is an efficient way of cloning large fragments of DNA
and therefore may be advantageous for the isolation of complete myosin
heavy chain genes.

The number of different myosin heavy chain genes which exist
is still highly speculative, despite the application of recombinant DNA
technology. Nudel *et al*. (1980) show that at least five different genes
are represented in rat skeletal muscle but do not suggest at what stage
of development these are expressed. Similar estimates of between 5 and 8

different genomic sequences have been made by other workers in the rat (Nguyen *et al.*, 1982) and the chick (Robbins *et al.*, 1982), though these may be underestimates since the probes employed may not detect smooth muscle and non-muscle myosin sequences. Patrinou-Georgoulas and John (1977) showed that two sequences are actually expressed in embryonic chick skeletal muscle by an analysis of R_0t curves, and this is confirmed in Chapter II, Section 2.2.9, but they were unable to speculate on the number of genes represented in the genome. Similarly, two different cDNA clones were detected by Umeda *et al.* (1981) as representing the mRNAs expressed in embryonic chick skeletal muscle. Mahdavi *et al.* (1982) showed that at least two myosin heavy chain genes are expressed in the rat heart muscle, but due to sequence homology, their hybridisation data is unable to determine exactly how many genomic sequences exist.

Data concerning the number of myosin heavy chain genes is continuing to accumulate, and unfortunately this is made more speculative due to varying experimental conditions. Variations in exposure times and hybridisation stringency conditions can contribute considerably to background levels and thus to false positives or, conversely, can cause failure of detection of some positives. However, the fact which is unanimously agreed throughout the literature is that myosin heavy chain genes exist as a multigene family in mammals and birds, as has been comprehensively described for the actins (Schwatz and Rothblum, 1980; Cleveland *et al.*, 1980; Zimmer and Schwatz, 1982) and globins (Fritsch *et al.*, 1980). Recent observations on peptide mapping (Whalen and Sell, 1980), immunological cross reactivity (Whalen *et al.*, 1981) and the DNA analysis described above, suggest the family may consist of as many as ten members in the rat. However, in contrast to these findings,

Rozek and Davidson (1983) have more recently isolated two overlapping myosin heavy chain genomic clones from a *Drosophila melanogaster* DNA library which represents the only myosin heavy chain gene in the *Drosophila* genome.

An interesting and perhaps relevant point emerging from the recent cDNA sequencing data of myosin heavy chain clones is the fact that the coding sequences of the genes give rise to mRNA's which have highly unconserved 3' sequences, as shown diagrammatically in Figure 6.1. A small part of this is translated, corresponding to 5-10 amino acids, but most is untranslated. It may, therefore, be more fruitful to employ short cDNA sequences as hybridisation probes in the search for specific genomic sequences as these will consist of sequences represented at the 3' ends of the mRNA molecules which may be highly specific for particular myosin heavy chain types. It is sequence data of this type which confirmed unambiguously that at least two cardiac myosin heavy chain genes are expressed in the adult rat (Mahdavi *et al.*, 1982). It has been demonstrated with actin cDNA sequences that those clones containing 3'-non-coding sequences hybridise specific-ally to their homologous mRNA's, and cross hybridisation is very weak (Cleveland *et al.*, 1980; Shani *et al.*, 1981b).

Comparisons of gene sequences isolated by recombinant DNA techniques allow evolution to be studied at a molecular level. There is in sufficient data available on the myosin heavy chain gene sequences to be able to speculate on the nature of the genes' evolution, but such studies will be possible as more sequencing data is generated, partic-ularly from the further detailed characterisation of cloned genomic DNA as described in this thesis. The actin genes have been comprehensively studied in this way. A comparison of actin genes which are widely

Figure 6.1 →

FIGURE 6.1:

A general summary of the available structural data of the myosin
heavy chain gene is shown diagrammatically. The information
presented in (a) was accumulated from studies on cDNA clones,
so no information on the position or size of possible intervening
sequences is shown. The 3' non-translated region spans a maximum
of 120 nucleotides. (b) is a representation of the studies of
Rozek and Davidson (1983) who have tentatively identified 9 introns
in the myosin heavy chain gene of Drosophila. The exons are
shown as boxes and the introns as the lines between them.

The scales of (a) and (b) are not identical. The total length
covered by the boxes in (b) is equivalent to the coding sequence
shown in (a).

a.

50-120nt
non-translated
region
no homology

15-30 nt, no homology
at nt. or aa. level

88-97% homology
between cDNA clones

3' ←——————— coding sequence ——————→ 5'

b.

3' ←——————— coding sequence ——————→ 5'

225.
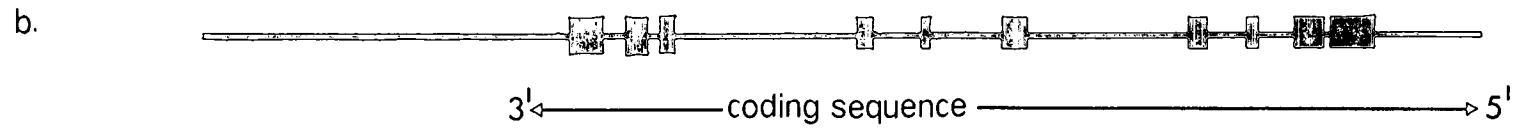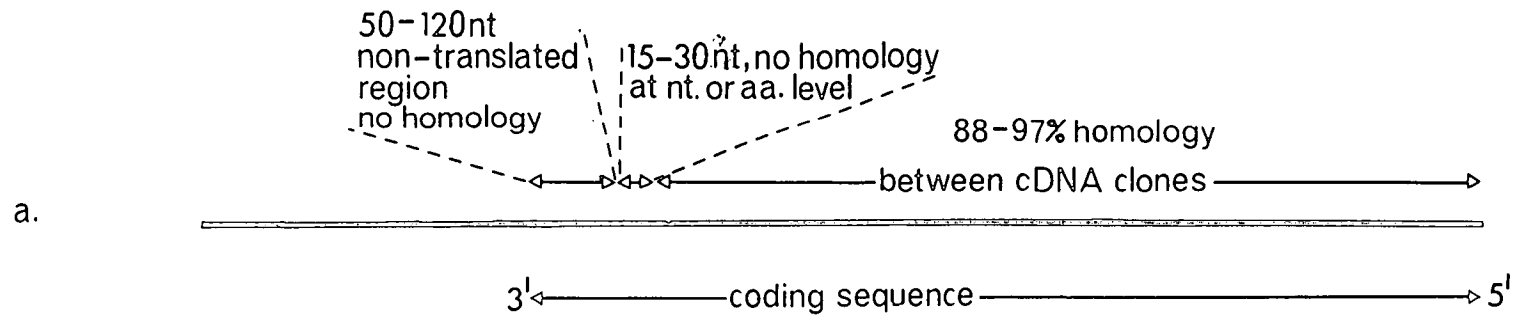
separated evolutionarily, such as sea urchin with *Drosophila* or yeast shows a moderately high percentage of nucleotide sequence divergence of the order of 14-18% (Cooper and Crain, 1982), but the actin amino acid sequence has been highly conserved throughout evolution. For example, *Physarum* actin shows only 5% amino acid divergence from mammalian actins (Vandekerckhove and Weber, 1978) and sea urchin actin only 6% divergence from that of *Drosophila* (Cooper and Crain, 1982). Therefore, in the case of this highly universal protein, evolutionary conservation appears to result from selective pressure at the protein level rather than the nucleic acid level where there is more opportunity for silent substitution in the third base of atriplet. The sequence information obtained from the myosin heavy chain cDNA clones so far isolated suggest close similarities between the genes which are expressed in a particular tissue of a single species. The two rat cardiac myosin heavy chain cDNA clones isolated by Mahdavi *et al*. (1982) show about 8% nucleotide divergence only. However, no comparisons have yet been made between evolutionary diverse species simply because the sequence data is not sufficiently complete. The isolation of genomic clones will enable the evolutionary rates of the coding and non-coding regions to be compared. In systems studied so far, such as the globins, it has been shown that selection pressures acting on the exons are far greater than those acting on the introns.

The divergence of gene sequences throughout phylogeny is just one measure of molecular evolution. Superimposed upon this is the evolution of multigene families as must have occurred with myosin heavy chain. The first of such gene families to be studied were the histones (Kedes and Birnsteil, 1971) and ribosomal RNA genes (Brown and Weber, 1968; Brown *et al*., 1971) which are the simplest cases,

consisting essentially of identical copies of tandem repetitive blocks which appear to have evolved by simple duplication due to unequal crossing-over at cell division. More complex examples have since been extensively studied, such as the globin gene family. Here, the individual sequences have diverged but remain linked in two clusters, as described in Chapter I, Section 1.3.6.

Very little is known about the linkage relationship between members of the myosin heavy chain gene family, but it seems probable that a further evolutionary procedure may have taken place - that is, dispersion of the various sequences throughout the genome such that they are now unlinked. Davidson *et al*. (1982) have suggested that various actin genes may have been dispersed throughout the genome during evolution into different regulatory modules depending on their sequence of expression in different tissues and during different developmental periods.

6.3.2 *Myosin Heavy Chain Gene Expression*

The regulation of the expression of muscle specific genes during development and in particular cell types is an area of investigation to which recombinant DNA techniques will contribute considerably in the future. The subject at present is highly controversial, as discussed in detail in Chapter I, Sections 1.2.4 and 1.2.5, where much of the *in vivo* and *in vitro* studies on gene expression were presented. These have attempted to measure the accumulation and relative levels of protein and mRNA within a particular cell type, particularly when developmental changes cause a switching from one phenotype to another. The main controversies which have arisen from these earlier studies concern:

(a) The levels of control of muscle specific genes. That
is, are they solely transcriptional, or does translation
play a major rôle too?

(b) The timing of the onset of myosin heavy chain (and
other muscle specific protein) synthesis during early
development of the muscle.

(c) The number and type of myosin heavy chain genes
expressed in the various foetal and adult muscle tissues.

(d) If an embryonic form of myosin heavy chain is
expressed, when is its synthesis repressed, and the
adult forms induced?

Simple measurement of the levels of a particular protein or a
messenger RNA, indirectly by *in vitro* translation, and analysis of
the protein products is too susceptible to inaccuracies in the methodol-
ogy employed to yield conclusive solutions to the above dilemmas.
Indeed, it is such methodology which created the controversies initially.
Hence, very precise procedures which specifically detect the expression
of a gene for a particular type of myosin are necessary in experiments
designed to determine the control mechanisms involved. Such studies,
using recombinant DNA techniques are in their very early stages at
present and the small advances which have been made recently on
myosin heavy chain gene structure, number and organisation have not
yet provided enough data to determine some fundamental points such as
the existence of 5' regulatory regions adjacent to myosin heavy chain
genes which may indicate some mechanism of activation or repression.
However, continued examination of the myosin heavy chain gene structure
and analysis of the transcription products of the different molecular
forms of the gene family may reveal some regulatory mechanisms.
Linkage relationships, as mentioned earlier, between the various muscle

specific protein genes have not been conclusively elucidated, but accumulating evidence based mainly on actin and tubulin would tend to argue against linkage of genes which are co-expressed in a given contractile muscle (Fyrberg *et al.*, 1980; Cleveland *et al.*, 1981). Therefore, regulation is probably not by some cis-acting regulatory sequence, but instead a trans-acting mechanism may be responsible for co-ordinated gene expression.

A situation which is contrary to all previous work is the finding described by Rozek and Davidson (1983) that the *Drosophila melanogaster* genome contains a single myosin heavy chain gene but from which three mRNA species of 7.2 kb, 8.0 kb and 8.6 kb are expressed. These RNAs arise through different patterns of splicing at the 3' end of the primary transcription unit which alter at the various developmental stages of the fly as it progresses through the embryo, larval and pupal stages to the adult fly. Rozek and Davidson (1983) speculate that a single gene which can be regulated in this way may provide sufficient genetic information for myofibril formation throughout the life of the fly. It is also suggested that other myosin heavy chain genes were not detected in their study due to lack of sequence homology which may have arisen by nucleotide sequence divergence throughout evolution. However, the possibility also exists that the fruit fly simply does not require multiple myosin heavy chain genes due to its muscle structure being much simpler than that of the other species which have been extensively studied.

The use of very specific cloned nucleic acid probes for the various myosin heavy chain genes in hybridisation experiments could provide an accurate measure of the levels of mRNA in the various muscle types at very specific times. Dym *et al.* (1979) adopted this

approach to investigate the important question of transcriptional versus translational control mechanisms. Previously, translational control had been suggested following several findings, including the detection of myosin heavy chain mRNA as a stored, translationally repressed free mRNP particle (Robbins and Heywood, 1978); the ability of cultured muscle cells, treated with the RNA synthesis inhibitor actinomycin D, to synthesise muscle specific proteins (Yaffe and Dym, 1972); and marked differences in the stability of mRNA molecules and the myoblast and myotube stages of development (Buckingham et al., 1974; Buckingham et al., 1976). Since actinomycin D is now believed to be much less specific in the inhibition of RNA synthesis than at the time of the early studies, some of the conclusions from this work have been questioned. Dym et al. (1979), therefore, investigated this problem with a different approach. Using an uncloned myosin heavy chain cDNA, the extent of hybridisation with mRNA extracted from sub-cellular fractions of cultured chick embryo breast muscle was measured. It was demonstrated that prior to cell fusion, before the onset of myosin synthesis, the majority of transcripts were present in a protein complex, whereas in differentiated myotube cultures, when myosin synthesis is occurring at a very high rate, the amount of stored mRNA is minimal, while a majority is localised in the polysome fraction. This apparent shift in the sub-cellular localisation of the myosin heavy chain mRNA suggests that translational control plays a major rôle, perhaps as important as transcriptional control, in the regulation of gene expression, when a definite increase in the efficiency of translation takes place during terminal differentiation. This approach, using cDNA-mRNA hybridisation is expected to yield more quantitative data since it is not limited to the detection of translatable messengers and does

not have to account for the different translational efficiencies of mRNA's as do the more indirect methods of estimation, such as *in vitro* translation.

More recently, Saidapet *et al*. (1982), recognising the quantitative importance of the cDNA-mRNA hybridisation techniques, have adopted a similar approach to the same problem. These workers studied the distribution of mRNA abundance classes for contractile muscle proteins in general, between polysomal and mRNP fractions in chick embryo skeletal muscle, developing *in ovo*. They argue that this system is more physiologically representative than a tissue culture system and conclude that translational control is of minor importance only. Total cDNA to a mRNA fraction which represents approximately 20% of the total RNA was prepared and demonstrated to be muscle specific by its inability to hybridise to liver mRNA, but highly specific for myogenic tissue culture mRNA. This abundance class was shown to be represented more in polysomes than mRNP particles in 9-day embryo muscle (rich in myoblasts). The polysomal content increased twelve-fold from 9 days to 14 days of development (as fusion proceeds) with a further 3.6-fold increase from 14 to 18 days (rich in myotubes). This 45-fold increase in the polysomal level of these muscle specific mRNAs corresponded to a 3% decrease in the mRNP population, and therefore Saidapet and co-workers conclude that mRNPs are not a reservoir of untranslated muscle specific mRNAs and hence transcriptional control must play a major rôle.

Although the approach of Saidapet *et al*. (1982) in using cDNA to quantitate the amounts of mRNA in different fractions is an excellent one and indirectly confirms the earlier work of Paterson and Bishop (1977) because a mixed cDNA probe was used, conclusions concerning

the distribution of myosin heavy chain mRNA cannot be made from this study. Therefore the results do not directly oppose those of Dym *et al*. (1979). Though translational control may not be of general importance in muscle protein synthesis, there is considerable evidence suggesting that it is very significant in the control of myosin heavy chain synthesis.

A technique which will be invaluable if very specific probes become available is that of *in situ* hybridisation. John *et al*. (1977), using a cDNA probe prepared from the myosin heavy chain mRNA of chick embryonic skeletal muscle and this technique, were able to distinguish between myogenic and non-myogenic cells in tissue culture, and demonstrate that early replicating myoblasts contained no detectable myosin heavy chain mRNA. The first cells to synthesise this message were those undergoing mitosis which indicates, firstly, that it is not withdrawal from the cell cycle which triggers differential gene expression, and, secondly, that fusion is not a prerequisite since myosin heavy chain mRNA was also detected in EGTA treated, fusion blocked cells.

The application of these recombinant DNA techniques to the study of myogenesis is in its early stages. For it to progress, probes are now required which are specific for individual members of the myosin heavy chain gene family in order that tissue specific gene expression can be unambiguously investigated. Very few of these are available in a cDNA form and it is genomic DNA clones which will prove the most useful since the non-transcribed sequences, in general, are more divergent between various members of a gene family than the coding regions. Genomic clones, therefore, will be more specific for any one particular member. The cDNA probe prepared in this

project is a mixture of two or three sequences as demonstrated by the hybridisation data presented in Chapter II, Section 2.2.9, but the genomic clones which it detected in the chicken gene library could represent several members of the gene family due to an unknown extent of cross homology.

Such highly specific probes are necessary in investigations into the critical period where embryonic muscle becomes adult type. Gene expression is influenced at this time by unknown factors which may reflect the physiological and hormonal status of the organism (Rubinstein *et al.*, 1978). The rôle, and even the existence, of embryonic type myosin, has been questioned, as have the relative quantities of fast and slow myosin in muscle during this development period (Whalen *et al.*, 1979; Whalen *et al.*, 1981; Gauthier *et al.*, 1982). Nucleic acid probes have been used to investigate these problems. Umeda *et al.* (1981) used chick embryo muscle myosin heavy chain cDNA clones to study the expression of individual myosin heavy chain genes and found two types to be expressed in agreement with this project and with Patrinou-Georgoulas and John (1977). Both cDNAs hybridised to fast muscle myosin heavy chain mRNA which prompted Umeda and co-workers to speculate that one represents a true fast form and the other an embryonic form which has some cross reactivity. This uncertainty again reinforces the problems encountered when genes have extensive sequence homology and stresses that nucleic acid hybridisation studies cannot be conclusive unless the probe is very specific.

It has been shown that there are also two myosin heavy chain genes expressed in cardiac muscle, an α and β form (Sinha *et al.*, 1982; Mahdavi *et al.*, 1982). The expression of these genes follows a distinct developmental pattern (Chizzonite *et al.*, 1982). During the second

half of gestation in the rabbit, the $\alpha:\beta$ protein ratio is $1:3$,
whereas after birth, α synthesis increases so that $\alpha:\beta$ is $1:1$ two
weeks postnatally. Following this, α synthesis decreases again and
the β form remains exclusive in the adult. The mechanisms controlling
the expression of these cardiac genes are not understood, but they
are influenced by certain external factors, such as the application
of thyroid hormone which causes continued expression of the α-gene into
adulthood and repression of the β-gene. Conversely, propylthiouracil
allows expression of the β form only. The isolation of cDNA clones by
Sinha *et al*. (1982) and Mahdavi *et al*. (1982) has demonstrated that
these two myosin heavy chains do indeed represent separate genes and
appear to be specific to cardiac muscle, since the amount of hybridisa-
tion with skeletal and myogenic cell line mRNA can be accounted for by
cross homology. The cDNAs show absolutely no affinity for smooth
muscle or non-muscle tissue mRNA.

6.4 Concluding Remarks

It is clear from this and other studies, that a genomic myosin
heavy chain clone is essential to the further progress of research into
muscle gene expression. The achievements of this study are that a
chicken genomic library has been constructed and several recombinant
phage identified by screening with a partial cDNA copy of the myosin
heavy chain mRNA from embryonic skeletal muscle. It can be concluded
that more than one myosin heavy chain gene is expressed in chick
embryonic muscle due to the fact that more than one sequence was
shown to be present in the cDNA preparation. The number of positive
recombinants identified by the cDNA in the genomic library suggests
that the myosin heavy chain genes are present as a family of sequences

with some homology between them, and that each individual sequence may be present in more than one copy per genome.

Many other workers have since cloned in bacterial plasmids, and characterised to various degrees, the myosin heavy chain cDNA, but thus far, a genomic clone has not been positively identified and sequenced from a mammal or bird. The example described by Rozek and Davidson (1983) is unusual, since it appears that a multiple family of myosin heavy chain genes do not exist in the fruit fly.

Genomic clones will be extremely useful tools for differentiating between the various muscle myosins, as discussed at length throughout this thesis. Therefore, future investigations must concentrate on characterising the positive recombinant phage which were identified by the cDNA probe employed here. The true positives should firstly be identified by re-screening, and then sequencing studies used to identify the differences and similarities between them.

REFERENCES

ADELSTEIN, R.S., POLLARD, T.D. and KNEHL, W.M. (1971). Isolation and characterisation of myosin and two myosin fragments from human blood platelets. Proc. Natl Acad. Sci. USA $\underline{68}$: 2703-2707.

ADELSTEIN, R.S., CONTI, M.A., JOHNSTON, G.S., PASTAN, N.I. and POLLARD, T.D. (1972). Isolation and characterisation of myosin from cloned mouse fibroblasts. Proc. Natl Acad. Sci. USA $\underline{69}$: 3693-3697.

ADESNIK, M., SALDITT, M., THOMAS, W. and DARNELL, J.E. (1972). Evidence that all messenger RNA molecules (except histone messenger RNA) contain poly(A) sequences and that the poly(A) has a nuclear function. J. molec. Biol. $\underline{71}$: 21-30.

AGARWAL, K.L., BRUNSTEDT, J. and NOYES, B.E. (1981). A general method for detection and characterisation of an mRNA using an oligonucleotide probe. J. Biol. Chem. $\underline{256}$: 1023-1028.

ALLET, B., JEPPESON, P., KATAGIRI, K.J. and DELIUS, H. (1973). Mapping the DNA restriction fragments produced by cleavage of λ DNA with endonuclease R1. Nature New Biol. $\underline{241}$: 120-123.

ANDERSON, J.M. and SCHIMKE, R.T. (1976). Partial purification of the ovalbumin gene. Cell $\underline{7}$: 331-338.

AVIV, H. and LEDER, P. (1972). Purification of biologically active globin messenger RNA by chromatography on oligothymidylic acid-cellulose. Proc. Natl Acad. Sci. USA $\underline{69}$: 1408-1412.

AXEL, R., FIEGELSON, P. and SCHULTZ, G. (1976). Analysis of the complexity and diversity of mRNA from chicken liver and oviduct. Cell $\underline{7}$: 247-254.

BAG, J. and BARKER, S. (1976). Studies on non-polysomal ribonucleoprotein coding for myosin heavy chains from chick embryonic muscles. J. Biol. Chem. $\underline{251}$: 7600-7609.

BAHL, C.P., MARIANS, K.J., WU, R., STAWINSKY, J. and NARANG, S. (1977). A general method for inserting specific DNA sequences into cloning vehicles. Gene $\underline{1}$: 81-92.

BALMAIN, A. and BIRNIE, G.D. (1979). Nick-translation of mammalian DNA. Biochim. Biophys. Acta $\underline{561}$: 155-166.

BENDALL, J.R. (1969). In, "Muscles, Molecules and Movement". Heinemann Educational Books Ltd.

BENOFF, S. and NADAL-GINARD, B. (1979). Most myosin heavy chain mRNA in L_6Eq rat myotubes has a short polyA tail. Proc. Natl Acad. Sci. USA $\underline{76}$: 1853-1857.

BENOFF, S. and NIDAL-GINARD, B. (1980). Transient induction of polyA short myosin heavy chain mRNA during terminal differentiation of L_6Eq myoblasts. J. Molec. Biol. $\underline{140}$: 283-298.

BENTON, W.D. and DAVIS, R.W. (1977). Screening λgt recombinant clones by hybridisation to single plaques *in situ*. Science $\underline{196}$: 180-182.

BERL, S., PUSZKIN, S. and NICKLAS, W.J. (1973). Actomyosin-like protein in brain. Science 179: 441-446.

BESTER, A.J., KENNEDY, D.S. and HEYWOOD, S.M. (1975) Two classes of translational control RNA: Their role in the regulation of protein synthesis. Proc. Natl Acad. Sci. USA 72: 1523-1527.

BETTEX-GALLAND, M. and LEUSCHER, E.F. (1965). Isolation and characterisation of myosin and two myosin fragments from human blood platelets. Advan. Protein Chem. 20: 1-35.

BILLETER, R., HEIZMANN, C.W., HOWALD, H. and JENNY, E. (1981). Analysis of myosin light and heavy chain types in single human skeletal muscle fibres. Eur. J. Biochem. 116: 389-395.

BISCHOFF, R. and HOLTZER, H. (1969). Mitosis and the process of differentiation of myogenic cells *in vitro*. J. Cell Biol. 41: 188-199.

BISHOP, J.O., MORTON, J.G., ROSBASH, M. and RICHARDSON, M. (1974). Three abundance classes in HeLa cell messenger RNA. Nature 250: 199-204.

BLATTNER, F.R., WILLIAMS, B.G., BLECHE, A.E., THOMPSON, K.D., 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 phage λ for DNA cloning. Science 196: 161-169.

BLATTNER, F.R., BLECHE, A.E., DENISTON-THOMPSON, K., FABER, H.E., RICHARDS, J.E., SLIGHTOM, J.L., TUCKER, R.W. and SMITHIES, O. (1978). Cloning human foetal γ-globin and mouse α-type globin DNA: preparation and screening of shotgun collections. Science 202: 1279-1284.

BONNER, W.M. and LASKEY, R.A. (1974). A film detection method for ^{3}H-labelled proteins and nucleic acids in polyacrylamide gels. Eur. J. Biochem. 46: 83-88.

BORCK, K., BEGGS, J.D., BRAMMER, W.J., HOPKINS, A.S. and MURRAY, N.E. (1976). The construction *in vitro* of transducing derivatives of phage lambda. Molec. Gen. Genet. 146: 199-207.

BRACK, C., HIRAMA, M., LENHARD-SCHULLER, R. and TONEGAWA, S. (1978). A complete immunoglobulin gene is created by somatic recombination. Cell 15: 1-14.

BREATHNACH, R., BENOIST, C., O'HARE, K., GANNON, F. and CHAMBON, P. (1978). Ovalbumin gene: Evidence for a leader sequence in mRNA and DNA sequences at the exon-intron boundaries. Proc. Natl Acad. Sci. USA 75: 4863-4867.

BRENNER, D.J., FOURNIER, M.J. and DOCTOR, B.P. (1970). Isolation and partial characterisation of the tRNA cistrons from *E. coli*. Nature 227: 448-451.

BREVIT, A., PINTO, E., PEACOCK, J. and STOCKDALE, F.E. (1976). Myosin synthesis is increased by electrical stimulation of skeletal muscle cells in cultures. Science 193: 1152-1154.

BROWN, D.D. (1979). In, "8th ICN-UCLA Symposium". Academic Press, New York.

BROWN, D.D. and WEBER, C.S. (1968). Gene linkage by RNA-DNA hybridisation. I. DNA sequences homologous to 4S RNA, 5S RNA and rRNA. J. Molec. Biol. 34: 661-680.

BROWN, D.D., WENSINK, P.C. and JORDAN, E. (1971). Position and characteristics of 5S DNA from X. laevis. Proc. Natl Acad. Sci. USA 68: 3175-3179.

BUCKINGHAM, M.E. (1977). Muscle protein synthesis and its control during the differentiation of skeletal muscle cells in vitro. In, "International Review of Biochemistry, Volume II: Biochemistry of Cell Differentiation". (Ed. J. Paul) University Park Press, Baltimore.

BUCKINGHAM, M.E., CAPUT, D., COHEN, A., WHALEN, R.G. and GROS, F. (1974). The synthesis and stability of cytoplasmic messenger RNA during myoblast differentiation in culture. Proc. Natl Acad. Sci. USA 71: 1466-1470.

BUCKINGHAM, M.E., COHEN, A. and GROS, F. (1976). Cytoplasmic distribution of pulse-labelled poly(A)-containing PNA, particularly 26S RNA, during myoblast growth and differentiation. J. Molec. Biol. 103: 611-626.

CAMERON, J.R., PANASENKO, S.M., LEHMAN, I.R. and DAVIS, R.W. (1975). In vitro construction of bacteriophage λ carrying segments of the Escherichia coli chromosome: selection of hybrids containing the gene for DNA ligase. Proc. Natl Acad. Sci. USA 72: 3416-3420.

CARAVATTI, M., MINTY, A., ROBERT, B., MONTARRA, D., WEYDERT, A., COHEN, A., DAUBAS, P. and BUCKINGHAM, M.E. (1982). Regulation of muscle gene expression. The accumulation of messenger RNAs coding for muscle-specific proteins during myogenesis in a mouse cell line. J. Molec. Biol. 160: 59-76.

CARBON, J. (1977). In, "Recombinant Molecules: Impact on Science and Society". (Eds, R.F. Beers and E.G. Bassatt). Raven Press, New York. pp 335-378.

CASJENS, S., HOHN, T. and KAISER, A.D. (1972). Head assembly steps controlled by genes F and W in bacteriophage λ. J. Molec. Biol. 64: 551-563.

CHACONAS, G. and VAN DE SANDE, J.H. (1980). In, "Methods in Enzymology" (Eds, L. Grossman and K. Moldave). Academic Press, New York. 65: 75-85.

CHAMBON, P., BENOIST, C., BREATHNACH, R., COCHET, M., GANNON, F., GERLINGER, P., KNIST, A., LEMEUR, M., LE PENNEC, J.P., MANDEL, J.L., O'HARE, K. and PERRIN, F. (1979). In, "Miami Winter Symposium" (Eds, R. Russell, K. Brew, H. Faber and J. Schultz). Academic Press, New York. <u>16</u>: 55-81.

CHANG, A., NURNBERG, J., KAUFMAN, R., EHRLICH, H., SCHIMKE, R. and COHEN, S. (1978). Phenotypic expression in *Escherichia coli* of a DNA sequence coding for mouse dihydrofolate reductase. Nature <u>275</u>: 617-624.

CHANG, J.C., TEMPLE, G.F., POON, R., NEUMANN, K.H. and KAN, Y.W. (1977). The nucleotide sequences of the untranslated 5' regions of human α- and β-globin mRNAs. Proc. Natl Acad. Sci. USA <u>74</u>: 5154-5149.

CHI, J.C.H., RUBINSTEIN, N., STRAHS, K. and HOLTZER, H. Synthesis of myosin heavy and light chains in muscle cultures. J. Cell Biol. <u>67</u>: 523-535.

CHIZZONITE, R.A., EVERETT, A.W., CLARK, W.A., JAKOVICIC, S., RABINOWITZ, M. and ZAK, R. (1982). Isolation and characterisation of two molecular variants of myosin heavy chain from rabbit ventricle. J. Biol. Chem. <u>257</u>: 2056-2065.

CLARKE, L. and CARBON, J. (1975). Biochemical construction and selection of hybrid plasmids containing specific segments of the *Escherichia coli* genome. Proc. Natl Acad. Sci. USA <u>72</u>: 4361-4365.

CLARKE, L. and CARBON, J. (1976). A colony bank containing synthetic ColE1 hybrid plasmids representative of the entire *E. coli* genome. Cell <u>9</u>: 91-99.

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 β- and γ-actin genes using specific cloned cDNA probes. Cell <u>20</u>: 95-105.

CLEVELAND, D.W., HUGUES, S.H., STUBBLEFIELD, E., KIRSCHNER, M.W. and VARMUS, H.E. (1981). Multiple α- and β-tubulin genes represent unlinked and dispersed gene families. J. Biol. Chem. <u>256</u>: 3130-3134.

COCHET, M., CHANG, A.C.Y. and COHEN, S.N. (1982). Characterisation of the structural gene and putative 5'-regulatory sequences for human pro-opiomelanocortiu. Nature <u>297</u>: 335-339.

COHEN, S.N., CHANG, A.C.Y., BOYER, H.W. and HELLING, R.B. (1973). Construction of biologically functional bacterial plasmids *in vitro*. Proc. Natl Acad. Sci. USA <u>70</u>: 3240-3244.

COHEN, S.N. and FISCHBACH, G.D. (1973). Regulation of muscle acetylcholine sensitivity by muscle activity in cell culture. Science <u>181</u>: 76-78.

COLEMAN, J.R. and COLEMAN, A.W. (1968). Muscle differentiation and macromolecular synthesis. J. Cell Physiol. 72: 19-34.

COLLINS, J. and BRUNING, H.J. (1978). Plasmids usable as gene-cloning vectors in an *in vitro* packaging system by coliphage λ 'cosmids'. Gene 4: 85-107.

COLLINS, J. and HOHN, B. (1979). Cosmids: A type of plasmid gene-cloning vector that is packagable *in vitro* in bacteriophage heads. Proc. Natl Acad. Sci. USA 75: 4242-4246.

COOPER, A.D. and CRAIN, W.R. (1982). Complete nucleotide sequence of a sea urchin actin gene. Nucl. Acids Res. 10: 4081-4092.

CORDEN, J., WASYLYK, B., BUCHWALDER, A., SASSONE-CORSI, P., KEDINGER, C. and CHAMBON, P. (1980). Promotor sequences of eukaryotic protein coding genes. Science 209: 1406-1414.

DABROWSKA, R., SOSINSKI, J. and DRABIKOWSKI, W. (1977). Changes in the composition of the myofibrillar fraction during development of the rabbit. FEBS Letters 79: 295-300.

DAVIDSON, E.H., THOMAS, T.L., SCHELLER, P.H. and BRITTEN, R.J. (1982). The sea urchin actin genes and a speculation on the evolutionary significance of small gene families. In, "Genome Evolution" (Eds, G.A. Dover and R.A. Flavell). Academic Press, New York.

DAVIS, B.J. (1964). Disc electrophoresis-II method and application to human serum proteins. Ann. New York Acad. Sci. 121: 404-436.

DE LA HABA, G., KAMALI, H.M. and TIEDE, D.M. (1975). Myogenesis of avian striated muscle *in vitro*: Role of collagen in myofibre formation. Proc. Natl Acad. Sci. USA 72: 2729-2732.

DENHARDT, D.T. (1966). A membrane filter technique for the detection of complementary DNA. Biochem. Biphys. Res. Commun. 23: 641-646.

DEUTSCHER, M.P. and KORNBERG, A. (1969). Enzymatic synthesis of deoxyribonucleic acid. XXIX: Hydrolysis of DNA from the 5'-terminus by an exonuclease function of DNA polymerase. J. Biol. Chem. 244: 3029-3037.

DEVLIN, R.B. and EMERSON, C.P. (1978). Co-ordinate regulation of contractile protein synthesis during myoblast differentiation. Cell 13: 599-611.

DE WET, J.R., DANIELS, D.L., SCHROEDER, J.L., WILLIAMS, B.G., DENISTON-THOMPSON, K., MOORE, D.D. and BLATTNER, F.R. (1980). Restriction maps for twenty-one charon vector phages. J. Virology. 33: 401-410.

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.

DOW, J. and STRACHER, A. (1971). Identification of the essential light chains of myosin. Proc. Natl Acad. Sci. USA 68: 1107-1110.

DRABKINA, L.E., KONEVEGA, L.V., LEGINA, O.K. and MOSEVITSKY, M.I. (1976). Dependence of transfection efficiency of calcium treated *Escherichia coli* cells on bacterial genotype and form of lambda DNA. Molec. Gen. Genet. 144: 83-86.

DUANCE, V.C., RESTALL, D.J., BEARD, H., BOURNE, F.J. and BAILEY, A.J. (1977). The location of three collagen-types in skeletal muscle. FEBS Letters 79: 248-252.

DUGAICZYK, A., WOO, S.L.C., LAI, E.C., MACE, M.L., REYNOLDS, L. and O'MALLEY, B.W. (1978). The natural ovalbumin gene contains 7 intervening sequences. Nature 274: 328-333.

DUMLAS, L.B., DARBY, G. and SINSHEIMER, R.L. (1971). The replication of bacteriophage ØX174 DNA *in vitro*. Temperature effects on repair synthesis and displacement synthesis. Biochim. Biophys Acta 228: 407-422.

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.

DYM, H.P., KENNEDY, D.S. and HEYWOOD, S.M. (1979). Subcellular distribution of the cytoplasmic myosin heavy chain mRNA during myogenesis. Differentiation 12: 145-155.

EFSTRATIADIS, A. (1979). 8th ICN-UCLA Symposium. In, "Eukaryotic Gene Regulation". Academic Press, New York..

EFSTRATIADIS, A., MANIATIS, T., KAFATOS, F.C., JEFFREY, A. and VOURNAKIS, J.N. (1975). Full length and discrete partial reverse transcripts of globin and chorion mRNAs. Cell 4: 367-378.

EFSTRATIADIS, A., POSAKONY, J.W., MANIATIS, T., LAWN, R.M., O'CONNELL, C., SPRITZ, R.A., DE RIEL, J.K., FORGET, B., WEISSMANN, 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.

ELSDALE, T. and BARD, J. (1972). Collagen substrata for studies on cell behaviour. J. Cell Biol. 54: 626-637.

EMERSON, C.P. and BECKNER, S.K. (1975). Activation of myosin synthesis in fusing and monucleated myoblasts. J. Molec. Biol. 93: 431-447.

FIESS, M., FISHER, R.A., CLAYTON, M.A. and EGNER, C. (1977). Packaging of bacteriophage λ chromosome: Effect of chromosome length. Virology 77: 281-293.

FIESS, M. and WIDNER, W. (1982). Bacteriophage λ DNA packaging: Scanning for the terminal cohesive end site during packaging. Proc. Natl Acad. Sci. USA 79: 3498-3502.

FLAVELL, R.A., KOOTER, J.M., DE BOER, E., LITTLE, P.F.R. and WILLIAMSON, R. (1978). Analysis of the β- δ-globin gene loci in normal human DNA: Direct determination of gene linkage and intergene distance. Cell 15: 25-41.

FRITSCH, E.F., LAWN, R.M. and MANIATIS, T. (1979). Characterisation of deletions which affect the expression of foetal globin genes in man. Nature 279: 598-603.

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.

FYRBERG, E.A., KINDLE, K.L., DAVIDSON, N. and SODJA, A. (1980). The actin genes of Drosophila: a dispersed multigene family. Cell 19: 365-378.

GANNON, F., O'HARE, K., PERRIN, F., LE PENNEC, J.P., BENOIST, C., COCHET, M., BREATHNACH, R., ROYAL, A., GARAPIN, A., CAMI, B. and CHAMBON, P. (1979). Organisation and sequence at the 5' end of a cloned complete ovalbumin gene. Nature 278: 428-434.

GAUTHIER, G.F., LOWEY, S. and HOBBS, A.W. (1978). Fast and slow myosin in developing muscle fibres. Nature 274: 25-29.

GAUTHIER, G.F., LOWEY, S., BENFIELD, P.A. and HOBBS, A.W. (1982). Distribution and properties of myosin isozymes in developing avian and mammalian skeletal muscle fibres. J. Cell Biol. 92: 471-484.

GELINAS, R.E. and KAFATOS, F.C. (1977). The control of chorion protein synthesis in silkmoths: mRNA production parallels protein synthesis. Dev. Biol. 55: 179-190.

GILBERT, W. (1978). Why genes in pieces? Nature 271: 501.

GILLESPIE, D., MARSHALL, J. and GALLO, R.C. (1972). RNA of RNA tumour viruses contains polyA. Nature 236: 227-231.

GLOVER, D.M., WHITE, R.L., FINNEGAN, D.J. and HOGNESS, D.S. (1975). Characterisation of six cloned DNAs from Drosophila melanogaster, including one that contains the genes for rRNA. Cell 5: 149-157.

GORDON, T., PERRY, R., SRIHARI, T. and VRBOVA, G. (1977). Differentiation of slow and fast muscles in chickens. Cell and Tiss. Res. 180: 211-222.

GREEN, P.J., BETLACH, M.C., GOODMAN, H.M. and BOYER, H.W. (1974). In, "Methods in Molecular Biology, Vol. 6". (Ed. R. Wickner). Marcel Dekker, New York.

GREENBERG, J.R. and PERRY, R.P. (1972). Relative occurrence
of polyadenylic acid sequences in messenger RNA and heterogeneous
nuclear RNA of L cells as determined by polyU hydroxylapatite
chromatography. J. Molec. Biol. 72: 91-98.

GRUNSTEIN, M. and HOGNESS, D.S. (1975). Colony hybridisation:
A method for the isolation of clonal cDNAs that contain a specific
gene. Proc. Natl Acad. Sci. USA 72: 3961-3965.

HAMER, D.H. and LEDER, P. (1979a). SV40 recombinants carrying a
functional RNA splice junction and polyadenylation site from the
chromosomal mouse β^{maj} globin gene. Cell 17: 737-747.

HAMER, D.H. and LEDER, P. (1979b). Splicing and the formation of
stable RNA. Cell 18: 1299-1302.

HANAHAN, D. and MESELSON, M. (1980). Plasmid screening at high
density. Gene 10: 63-67.

HARDISON, R.C., QUON, D. and MANIATIS, T. (1979a). The linkage
arrangement of four rabbit β-like globin genes. Cell 18: 1273-1283.

HARDISON, R.C., BUTLER, E.T., LACY, E., MANIATIS, T., ROSENTHAL,
N and EFSTRATIADIS, A. (1979b). The structure and transcription
of four linked β-like globin genes. Cell 18: 1285-1297.

HARRIS, A.J., HEINEMANN, S., SCHUBERT, D. and TARAKIS, M.
(1971). Trophic interaction between cloned tissue culture lines of
nerve and muscle. Nature 231: 296-301.

HAUSCHKA, S. and KONIGSBERG, I.R. (1966). The influence of
collagen on the development of muscle clones. Proc. Natl Acad. Sci.
USA 55: 119-126.

HEDGPATH, J., GOODMAN, H.M. and BOYER, H.W. (1972). DNA
nucleotide sequence restricted by the R1 endonuclease. Proc. Natl
Acad. Sci. USA 69: 3448-3452.

HENDRIX, R.W., ROBERTS, J.W., STAHL, F.W. and WEISBERG, R.A.
(1982) In, "Lambda II". Cold Spring Harbour Laboratory, New
York.

HERSHEY, A.D., BURGI, E. and INGRAM, L. (1963). Cohesion of
DNA molecules isolated from phage lambda. Proc. Natl Acad. Sci.
USA 49: 748-755.

HERSHEY, A.D. and BURGI, E. (1965). Complementary structure of
interacting sites at the ends of lambda DNA molecules. Proc. Natl
Acad. Sci. USA 53: 325-328.

HEYNEKER, H.L., SHINE, J., GOODMAN, H.J., BOYER, H., ROSENBERG,
J., DICKSON, R.E., NARANG, D.A. and RIGGS, A.D. (1976).
Synthetic Lac operator DNA is functional in vitro. Nature 263: 748-752.

HEYWOOD, S.M., DOWBEN, R.M. and RICHE, A. (1967). The identification of polyribosomes synthesising myosin. Proc. Natl Acad. Sci. USA 57: 1002-1009.

HEYWOOD, S.M. and NWAGWU, M. (1969). Partial characterisation of presumptive myosin messenger ribonucleic acid. Biochemistry 8: 3839-3845.

HEYWOOD, S.M., KENNEDY, D.S. and BESTER, A.J. (1975). Stored myosin messenger in embryonic chick muscle. FEBS Letters 53: 69-72.

HEYWOOD, S.M. and KENNEDY, D.S. (1976). Purification of myosin translational control RNA and its interaction with myosin messenger RNA. Biochemistry 15: 3314-3319.

HOH, J.F.Y., McGRATH, P.A. and HALE, P.T. (1978). Electrophoretic analysis of multiple forms of rat cardiac myosin. Effects of hypophysectomy and thyroxine replacement. J. Molec. Cell. Cardiol. 10: 1053-1076.

HOHN, B. (1975). DNA as a substrate for packaging into bacteriophage Lambda in vitro. J. Molec. Biol. 98: 93-106.

HOHN, B. and HOHN, T. (1974). Activity of empty head-like particles for packaging of DNA of bacteriophage λ in vitro. Proc. Natl Acad. Sci. USA 71: 2372-2376.

HOHN, B. and MURRAY, K. (1977). Packaging recombinant DNA molecules into bacteriophage particles in vitro. Proc. Natl Acad. Sci. USA 74: 3259-3263.

HOHN, T., WURTZ, M. and HOHN, B. (1976). Packaging of genomes in bacteriophages - a comparison of single stranded-RNA bacteriophages and double stranded DNA bacteriophages. Phil. Trans. Roy. Soc. Lond. B 276: 143-150.

HOLLAND, D.I. and PERRY, S.V. (1969). The ATPase and Ca^{2+} transporting activity of the sarcoplasmic reticulum in developing muscle. Biochem. J. 114: 116-170.

HOPKINS, A.S., MURRAY, N.E. and BRAMMAR, W.J. (1976). Characterisation of λ_{trp}-transducing bacteriophages made in vitro. J. Molec. Biol. 107: 549-569.

HORNIG, D. and NADAL-GINARD, B. (1981). Characterisation of adult skeletal muscle myosin heavy chain genes of rat. In, "Abstracts of CSH Meeting on 'Molecular and Cellular Control of Muscle Development'".

INGWALL, J.S., MORALES, M.R. and STOCKDALE, F.E. (1972). Creatine and control of myosin synthesis in differentiating skeletal muscle. Proc. Natl Acad. Sci. USA 69: 2250-2253.

INGWALL, J.S. and WILDENTHAL, K. (1974). Creatine stimulation of myofibrillar protein synthesis. J. Cell. Biol. 63: 151a.

ITAKURA, K., KATAGIRI, N., BAHL, C.P., WIGHTMAN, R. and NARANG, S.A. (1975). Improved friester approach for the synthesis of peutadecathymidylic acid. J. Am. Chem. Soc. 97: 7327-7331.

JACOB, F. and WOLLMAN, E.L. (1954). Eclude genetique d'un bacterio-phage tempere d'E. coli. Le systeme genetique du bacteriophage λ. Ann. Inst. Pasteur 87: 653-673.

JAKOWLEW, S.B., KHANDEKAR, P., ZARRAGA, A.M., ARNOLD, H.H. and SIDDIQUI, M.A.Q. (1982). Molecular cloning and expression of cardiac myosin heavy chain gene in chick embryo. J. Molec. Biol. 156: 673-683.

JEFFREYS, A.J. and FLAVELL, R.A. (1977). The rabbit β-globin gene contains a large insert in the coding sequence. Cell 12: 1097-1108.

JOHN, H.A. (1974). The myosin of developing and dystrophic skeletal muscle. FEBS Letters 39: 278-282.

JOHN, H.A. (1980). Comparison of the heavy chains of physiologically different myosins by isoelectric focusing. Biochem. Biophys. Res. Commun. 92: 1223-1230.

JOHN, H.A., BIRNSTIEL, M.L. and JONES, K.W. (1969). RNA-DNA hybrids at the cytological level. Nature 223: 582-587.

JOHN, H.A., PATRINOU-GEORGOULAS, M. and JONES, K.W. (1977). Detection of myosin heavy chain mRNA during myogenesis in tissue culture in vitro and in situ hybridisation. Cell 12: 501-508.

JOHN, H.A. and LAWSON, H. (1980). The effect of different collagen-types used as substrata on myogenesis in tissue culture. Cell Biol. International Reports 4: 841-850.

JONES, K.W. and MURRAY, K. (1975). A procedure for detection of heterologous DNA sequences in lambdoid phage by in situ hybridisa-tion. J. Molec. Biol. 96: 455-460.

KACIAN, D.L., SPIEGELMAN, S., BANK, A., TERADA, M., METAFORA, S., DOW, L. and MARKS, P.A. (1972). In vitro synthesis of DNA components of human genes for globins. Nature New Biol. 235: 167-169.

KAFATOS, F.C. (1978). In, "Structure and organisation of develop-mentally regulated genes" (Eds, J. Schultzand and F. Ahmad). Academic Press, New York.

KAISER, A.D. and WU, R. (1968). Structure and function of DNA cohesive ends. Cold Spring Harbor Symp. Quant. Biol. 33: 729-734.

KARN, J., BRENNER, S., BARNETT, L. and CESARINI, G. (1980). Novel bacteriophage λ cloning vectors. Proc. Natl Acad. Sci. USA 77: 5172-5776.

KEDES, L.H. (1979). Histone genes and histone messengers. Ann. Rev. Biochem. 48: 837-870.

KEDES, L.H., CHANG, A.C.Y., HOUSMAN, D. and COHEN, S.N. (1975). Isolation of histone genes from unfractionated sea urchin DNA by subculture cloning in *E. coli*. Nature 255: 523-538.

KEDES, L.H. and BIRNSTEIL, M.L. (1971). Reiteration and clustering of DNA sequences complementary to histone mRNA. Nature 230: 165-169.

KELLER, L.R. and EMERSON, C.P. (1980). Synthesis of adult myosin light chains by embryonic muscle cultures. Proc. Natl Acad. Sci. USA 77: 1020-1024.

KELLY, A.M. and RUBINSTEIN, N.A. (1980). Why are foetal muscles slow? Nature 288: 266-269.

KELLY, R.B., COZZARELLI, N.R., DEUTSCHER, M.P., LEHMAN, I.R. and KORNBERG, A. (1970). Enzymatic synthesis of deoxyribonucleic acid. XXXII: Replication of duplex DNA by polymerase at a single strand break. J. Biol. Chem. 245: 39-45.

KINDLE, K. and FIRTLE, R.A. (1978). Identification and analysis of Dictyostelium actin genes, a family of moderately repeated genes. Cell 15: 763-778.

KING, C.R., UDELL, D.S. and DEELEY, R.G. (1979). Characterisation of the estrogen-responsive domain of avian liver and cloning of double stranded cDNA derived from estrogen-inducible RNA species. J. Biol. Chem. 254: 6781-6787.

KLETT, R.P., CERAMI, A. and REICH, E. (1968). Exonuclease V1, a new nuclease activity associated with *E. coli* DNA polymerase. Proc. Natl Acad. Sci. USA 60: 943-950.

KNAPP, G., BECKMANN, J.S., JOHNSON, P.F., FUHRMAN, S.A. and ABELSON, J. (1978). Transcription and processing of intervening sequences in yeast tRNA genes. Cell 14: 221-236.

KONIGSBERG, I.R. (1971). Diffusion mediated control of myoblast fusion. Dev. Biol. 26: 133.

KONKEL, D.A., SILGHMAN, S.M. and LEDER, P. (1978). The evolution and sequence comparison of two recently diverged mouse chromosomal β-globin genes. Cell 18: 865-873.

KORN, L.J. and BROWN, D.D. (1978). Nucleotide sequence of *Xenopus borealis* oocyte 5S DNA: Comparison of sequences that flank several related eukaryotic genes. Cell 15: 1145-1156.

KRAMER, R.A., CAMERON, J.R. and DAVIS, R.W. (1976). Isolation of bacteriophage λ containing yeast ribosomal RNA genes: Screening *in situ* RNA hybridisation to plaques. Cell 8: 227-232.

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.

LAEMMLI, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227: 680-685.

LAI, E.C., WOO, S.L.C., DUGAICZYK, A. and O'MALLEY, B.W. (1979a). The ovalbumin gene: Alleles created by mutations in the intervening sequences of the natural gene. Cell 16: 201-211.

LAI, E.C., STEIN, J.P., CATTERALL, J.F., WOO, S.L.C., MACE, M.L., MEANS, A.R. and O'MALLEY, B.W. (1979b). Molecular structure and flanking nucleotide sequences of the natural chicken ovomucoid gene. Cell 18: 829-842.

LASKEY, R.A. and MILLS, A.D. (1975). Quantitative film detection of ^{3}H and ^{14}C in polyacrylamide gels by fluorography. Eur. J. Biochem. 56: 335-341.

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.

LAWN, R.M., FRITSCH, E.F., PARKER, R.C., BLAKE, G. and MANIATIS, T. (1978). The isolation and characterisation of linked δ- and β-globins from a cloned library of human DNA. Cell 15: 1157-1174.

LEDER, P., TIEMEIER, D. and ENQUIST, L. (1977). EK2 derivatives of bacteriophage lambda useful in the cloning of DNA from higher organisms: The λgt WES system. Science 196: 175-177.

LEE, S.Y. and BRAWERIMAN, G. (1971). Pulse labelled ribonucleic acid complexes released by dissociation of rat liver polysomes. Biochemistry 10: 510-516.

LERNER, M.R., BOYLE, J.A., MOUNT, S.M., WOLIN, S.L. and STEITZ, J. (1980). Are s.n. RNPs involved in splicing? Nature 283: 220-224.

LIZARDI, P.M. (1976). The size of pulse labelled fibroin mRNA. Cell 7: 239-245.

LOMPRE, A.M., MERCADIER, J.J., WISNEWSKY, C., BOUVERET, P., PANTALONI, C., D'ALBIS, A. and SCHWARTZ, K. (1981). Species and age dependent changes in the relative amounts of cardiac myosin isoenzymes in mammals. Dev. Biol. 84: 286-291.

McCLEMENTS, W. and SKALKA, A.M. (1977). Analysis of chicken ribosomal RNA genes and construction of Lambda hybrids containing gene fragments. Science 196: 195-197.

McDONNELL, W.M., SIMON, M.N. and STUDIER, F.W. (1977). Analysis of restriction fragments of T7 DNA and determination of molecular weights by electrophoresis in neutral and alkaline gels. J. Molec. Biol. 110: 119-146.

MACKEY, J.K., BRACKMANN, K.H., GREEN, M.R. and GREEN, M. (1977). Preparation and characterisation of highly radioactive *in vitro* labelled adenovirus DNA and DNA restriction fragments. Biochemistry 16: 4478-4483.

MAHDAVI, V., PERIASAMY, M. and NADAL-GINARD, B. (1982). Molecular characterisation of two myosin heavy chain genes expressed in the adult heart. Nature 297: 659-664.

MANDEL, J.L. and PEARSON, M.L. (1974). Insulin stimulates myogenesis in a rat myoblast line. Nature 251: 618-620.

MANDEL, M. and HIGA, A. (1970). Calcium dependent bacteriophage DNA interaction. J. Molec. Biol. 53: 159-162.

MANIATIS, T., JEFFREY, A. and KLEID, D.G. (1975a). Nucleotide sequence of the right ward operator of phage λ. Proc. Natl Acad. Sci. USA 72: 1184-1188.

MANIATIS, T., JEFFREY, A. and VAN DE SANDE, H. (1975b). Chain length determination of small double- and single-stranded DNA molecules by polyacrylamide gel electrophoresis. Biochemistry 14: 3787-3794.

MANIATIS, T., HARDISON, R.C., LACY, E., LAUER, J., O'CONNELL, C., QUON, D., SIM, G.K. and EFSTRATIADIS, A. (1978). The isolation of structural genes from libraries of eukaryotic DNA. Cell 15: 687-701.

MARCU, K. and DUDOCK, B. (1974). Characterisation of a highly efficient protein synthesising system derived from commercial wheat germ. Nucl. Acids Res. 1: 1385-1397.

MARMUR, J. (1961). A procedure for the isolation of deoxyribonucleic acids from micro-organisms. J. Molec. Biol. 3: 208-218.

MASAKI, T. and KINOSHITA, T. (1974). Synthesis of cardiac myosin by a polyribosome fraction from chick embryonic skeletal muscle. J. Biochem. 75: 1193-1195.

MASAKI, T. and YOSHIZAKI, C. (1974). Differentiation of myosin in chick embryos. J. Biochem. 76: 123-131.

MASAKI, T., BADER, D.M., REINACH, F.C., SHIMIZU, T., OBINATA, T., SHAFIG, S.A. and FISCHMAN, D.A. (1982). Monoclonal antibody analysis of myosin heavy chain and C-protein isoforms during myogenesis. In, "Proceedings of the CSH Meeting on Molecular and Cellular Control of Muscle Development" (Ed. M. Pearson). C.S.H., New York.

MAYNE, R., VAIL, M.S., MILLER, E.J., BLOSE, S.H. and CHACKO, S. (1977). Collagen polymorphism in cell cultures derived from guinea pig aortic smooth muscle: Comparison with three populations of fibroblasts. Arch. Biochem. Biophys 181: Res. Commun. 462-469.

MESELSON, M. and YUAN, R. (1968). DNA restriction enzyme from *E. coli*. Nature 217: 1110-1114.

MIDDLETON, J.H., EDGELL, M.H. and HUTCHINSON, C.A. (1972). Specific fragments of ØX174 deoxyribonucleic acid produced by a restriction enzyme from *Haemophilus aegyptius*, Endonuclease Z. J. Virol. 10: 42-50.

MIRSKY, A.E. and RIS, H. (1949). Variable and constant components of chromosomes. Nature 163: 666-667.

MOIR, A. and BRAMMER, W.J. (1976). The use of specialised transducing phages in the amplification of enzyme production. Molec. Gen. Genet. 149: 87-99.

MOLLOY, G.R., SPORN, M.B., KELLY, D.E. and PERRY, R.P. (1972). Localisation of poly(A) sequences in mRNA of mammlaian cells. Biochemistry 11: 3256-3260.

MONAHAN, J.J., HARRIS, S.E., WOO, S.L., ROBERTSON, D.L. and O'MALLEY, B.W. (1976). The synthesis and properties of the complete complementary DNA transcript of ovalbumin mRNA. Biochemistry 15: 223-233.

MONDAL, H., SUTTON, A., CHEN, V. and SARKAR, S. (1974). A highly purified mRNA for myosin heavy chain: Size and polyA content. Biochem. Biophys. Res. Commun. 56: 988-996.

MORRIS, G.E., BUZASH, E.A., ROURKE, A.W., TEPPERMAN, K., THOMPSON, W.C. and HEYWOOD, S.M. (1972). Myosin messenger RNA: Studies on its purification, properties and translation during myogenesis in culture. Cold Spring Harbor Symp. Quant. Biol. 35: 535-541.

MORROW, J.F. (1979). In, "Methods in Enzymology" (Ed. R. Wu). Academic Press, New York. 68: 3-26.

MOSS, P.S. and STROHMAN, R.C. (1976). Myosin synthesis by fusion-arrested chick embryo myoblasts in cell culture. Dev. Biol. 48: 431-437.

MURRAY, N.E., BRAMMER, W.J. and MURRAY, K. (1977). Lambdoid phages that simplify recovery of *in vitro* recombinants. Molec. Gen. Genet. 150: 53-61.

NAGATA, S., TAIRA, H., HALL, A., JOHNSRUD, L., STREULI, M., ESCODI, J., BÖLL, W., CANTELL, K. and WEISSMAN, C. (1980). Synthesis in *Escherchia coli* of a polypeptide with human leukocyte interferon activity. Nature 284: 316-320.

NAKANISHI, S., INONE, A., KITL, T., NAKAMURA, M., CHANG, A.C.Y., COHEN, S.N. and NUMA, S. (1979). Nucleotide sequence of cloned cDNA for bovine corticotropin-β-lipotropin precursor. Nature 278: 423-427.

NGUYEN, H.T., WYDRO, R.M., GUBITS, R. and NADAL-GINARD, B. (1982). Sarcomeric myosin heavy chain is coded by a highly conserved multigene family. Proc. Natl Acad. Sci. USA 79: 5230-5234.

NODA, M., TERANISHI, Y., TAKAHASHI, H., TOYOSATO, H., NOTAKE, M., NAKANISHI, S. and NUMA, S. (1982). Isolation and structural organisation of the human preproenkephalin gene. Nature 297: 431-434.

NORDSTRÖM, J.L., ROOP, D.R., TSAI, M.J. and O'MALLEY, B.W. (1979). Identification of potential ovomucoid mRNA precursors in chick oviduct nuclei. Nature 278: 328-331.

OBINATA, T. (1969). The myosin of developing chick embryo. Arch. Biochem. Biophys. 132: 184-197.

O'FARRELL, P.H. (1974). High resolution two-dimensional electro-phoresis of proteins. J. Biol. Chem. 250: 4007-4021.

O'FARRELL, P.Z., GOODMAN, H.M. and O'FARRELL, P.H. (1977). High resolution two dimensional electrophoresis of basic as well as acidic proteins. Cell 12: 1133-1142.

OGAWA, H. and TOMIZAWA, J. (1967). Phage lambda DNA with different structures formed in infected cells. J. Molec. Biol. 23: 265-276.

OKAZAKI, K. and HOLTZER, H. (1966). Myogenesis: Fusion, myosin synthesis and the mitotic cycle. Proc. Natl Acad. Sci. USA 56: 1484-1490.

O'NEILL, M. and STOCKDALE, F.E. (1972). A kinetic analysis of myogenesis in vitro. J. Cell Biol. 52: 52-65.

ORDAHL, C.P., TILGHMAN, S.M., OVITT, C., FORNWALD, J. and LARGEN, M.T. (1980). Structure and developmental expression of the chick α-actin gene. Nucl. Acids Res. 8: 4989-5005.

OSTLUND, R.E. and PASTAN, I. (1974). Myosin cultured fibroblasts. J. Biol. Chem. 249: 3903-3907.

PATERSON, B.M. and BISHOP. J.O. (1977). Changes in the mRNA population of chick myoblasts during myogenesis in vitro. Cell 12: 751-765.

PATERSON, B., ROBERTS, B. and KLUFF, E. (1977). Structural gene identification and mapping by DNA-mRNA hybrid-arrested cell-free translation. Proc. Natl Acad. Sci. USA 74: 4370-4374.

PATERSON, B.M. and STROHMAN, R.C. (1972). Myosin synthesis in cultures of differentiating chicken embryo skeletal muscle. Dev. Biol. 29: 113-138.

PATRINOU-GEORGOULAS, M. and JOHN, H.A. (1977). The genes
and messenger RNA coding for the heavy chains of chick embryo
skeletal myosin. Cell 12: 491-499.

PATRINOU-GEORGOULAS, M. and JOHN, H.A. (1980). Estimation
of the numbers of different sequences in myosin heavy chain messenger
RNA prepared by different methods. FEBS Letters 109: 239-243.

PELHAM, H.R.B. and JACKSON, R.J. (1976). An efficient mRNA
dependent translation system from reticulocyte lysates. Eur. J.
Biochem. 67: 247-256.

PERRY, R.P. and KELLY, D.E. (1973). Messenger RNA turnover in
mouse L cells. J. Molec. Biol. 79: 681-696.

PRIBNOW, D. (1975a). Nucleotide sequence of an RNA polymerase bind-
ing site at an early T7 promotor. Proc. Natl Acad. Sci. USA 77:
784-788.

PRIBNOW, D. (1975b). Bacteriophage T7 promotors: Nucleotide
sequences of two RNA polymerase binding sites. J. Molec. Biol.
99: 419-443.

PROUDFOOT, N.J. and BROWNLEE, G.G. (1976). 3' non-coding region
sequences in eukaryotic messenger RNA. Nature 263: 211-214.

PZYBYLA, A. and STROHMAN, R.C. (1974). Myosin heavy chain
messenger RNA from myogenic cell cultures. Proc. Natl Acad. Sci.
USA 71: 662-666.

RATZKIN, B. and CARBON, S. (1977). Functional expression of cloned
yeast DNA in *Escherichia coli*. Proc. Natl Acad. Sci. USA 74: 487-
491.

RETZEL, E.F., COLLET, M.S. and FARAS, A.J. (1980). Enzymatic
synthesis of deoxyribonucleic acid by the avian retrovirus reverse
transcriptase *in vitro* - optimum conditions required for transcription
of large ribonucleic acid templates. Biochemistry 19: 513-518.

RIGBY, P.W.J., DIECKMANN, M., RHODES, C. and BERG, P. (1977).
Labelling DNA to high specific activity *in vitro* by nick translation
with DNA polymerase I. J. Molec. Biol. 113: 237-251.

ROBBINS, J. and HEYWOOD, S.M. (1976). Preparation and characterisa-
tion of myosin copy DNA. Biochem. Biophys. Res. Commun. 68: 918-
926.

ROBBINS, J. and HEYWOOD, S.M. (1978). Quantitation of myosin
heavy chain during myogenesis. Eur. J. Biochem. 82: 601-608.

ROBBINS, J., FREGER, G.A., CHISHOLM, D. and GILLAM, T.C.
(1982). Isolation of multiple genomic sequences coding for chicken
myosin heavy chain protein. J. Biol. Chem. 257: 549-556.

ROBERTS, B.E. and PATERSON, B.M. (1973). Efficient translation of tobacco mosaic virus RNA and rabbit globin 9S RNA in a cell-free system from commercial wheat germ. Proc. Natl Acad. Sci. USA 70: 2330-2334.

ROBERTS, J.L. and HERBERT, E. (1977). Characterisation of a common precursor to corticotropin and β-lipotropin: Cell free synthesis of the precursor and identification of corticotropin peptides·in the molecule. Proc. Natl Acad. Sci. USA 74: 4826-4830.

ROBERTS, J.L., SEEBURG, P.H., SHINE, J., HERBERT, E., BAXTER, J.D. and GOODMAN, H.M. (1979). Corticotropin and β-endorphin: Construction and analysis of recombinant DNA complementary to mRNA for the common precursor. Proc. Natl Acad. Sci. USA 76: 2153-2157.

ROBERTS, R.J., MYERS, P.A., MORRISON, A. and MURRAY, K. (1976). A specific endonuclease from *Arthrobacter luteus*. J. Molec. Biol. 102: 157-165.

ROBESON, J.P., GOLDSCHMIDT, R.M. and CURTISS, R. (1980). Potential of *Escherichia coli* isolated from nature to propagate cloning vectors. Nature 283: 104-106.

ROSS, J., AVIV, H., SCOLNICK, E. and LEDER, P. (1972). *In vitro* synthesis of DNA complementary to purified rabbit globin mRNA. Proc. Natl Acad. Sci. USA 69: 264-268.

ROSS, J., GIELEU, J., PACKMAN, S., IKAWA, Y. and LEDER, P. (1974). Globin gene expression in cultured erythroleukemic cells. J. Molec. Biol. 87: 697-714.

ROWLERSON, A., POPE, B., MURRAY, J., WHALEN, R.G. and WEEDS, A.G. (1981). A novel myosin present in cat jaw-closing muscles. J. Muscle Res. and Cell Motility 2: 415-438.

ROZEK, C.E. and DAVIDSON, N. (1983). Drosophila has one myosin heavy chain gene with three developmentally regulated transcripts. Cell 32: 23-34.

RUBINSTEIN, N.A. and HOLTZER, H. (1974). Actin and myosin in a variety of myogenic and non-myogenic cells. Biochem. Biophys. Res. Commun. 57: 438-446.

RUBINSTEIN, N.A., PEPE, F.A. and HOLTZER, H. (1977). Myosin types during the development of embryonic chicken fast and slow muscles. Proc. Natl Acad. Sci. USA 74: 4524-4527.

RUBINSTEIN, N.A. and KELLY, A.M. (1978). Myogenic and neurogenic contributions to the development of fast and slow twitch muscles in rat. Dev. Biol. 62: 473-485.

RUBINSTEIN, N.A., MABUCH, K., PEPE, F.A., SALMONS, S., GERGELY, J. and SRÉTER, F. (1978). Use of type-specific anti-myosins to demonstrate the transformation of individual fibers in chronically stimulated rabbit fast muscles. J. Cell Biol. 79: 252-261.

RUBINSTEIN, N.A. and HOLTZER, H. (1979). Fast and slow muscles in tissue culture synthesise only fast myosin. Nature 280: 323-325.

SAIDAPET, C.R., MUNRO, H.N., VALGEIRSDOTTIR, K. and SARKAR, S. (1982). Quantitation of muscle specific mRNAs by using cDNA probes during chicken embryonic muscle development *in ovo*. Proc. Natl Acad. Sci. USA 79: 3087-3091.

SALMONS, S. and SRÉTER, F.A. (1976). Significance of impulse activity in the transformation of skeletal muscle type. Nature 263: 30-34.

SARGENT, T.D., WU, J.R., SALA-TREPAT, J.M., WALLACE, R.B., REYES, A.A. and BONNER, J. (1979). The rat serum albumin gene: Analysis of cloned sequences. Proc. Natl Acad. Sci. USA 76: 3256-3260.

SARKAR, S., MUKHERJEE, S.P., SUTTON, A., MONDAL, H. and CHEN, V. (1973). Isolation of messenger RNA for myosin heavy chain. Prep. Biochem. 3: 583-604.

SCHELLER, R.H., DICKESON, R.E., BOYER, H.W., RIGGS, A.D. and ITAKURA, K. (1977). Chemical synthesis of restriction enzyme recognition sites useful for cloning. Science 196: 177-180.

SCHMALE, H. and RICHTER, D. (1980). *In vitro* biosynthesis and processing of composite common precursors containing amino acid sequences identified immunologically as neurophysin I/oxytocin and as neurophysin II/vasopressin. FEBS Letters 121: 358-362.

SCHMALE, H. and RICHTER, D. (1981). Immunological identification of a common precursor to arginine vasopressin and neurophysin II synthesised by *in vitro* translation of bovine hypothalamic mRNA. Proc. Natl Acad. Sci. USA 78: 766-769.

SCHROEDER, J.L. and BLATTNER, F. (1978). Least squares method for restriction mapping. Gene 4: 167-174.

SCHWARTZ, R. and ROTHBLUM, K. (1980). Regulation of muscle differentiation: Isolation and purification of chick actin messenger RNA and quantitation with cDNA probes. Biochemistry 19: 2506-2514.

SHANI, M., ZEVLIN-SONKIN, D., SAXEL, O., CARMAN, Y., KATCOFF, D., NUDEL, U. and YAFFE, D. (1981a). The correlation between the synthesis of skeletal muscle actin, myosin heavy chain and myosin light chain and the accumulation of corresponding mRNA sequences during myogenesis. Dev. Biol. 86: 483-492.

SHANI, M., NUDEL, U., ZEVIN-SONKIN, D., ZAKUT, R., GIVOL, D. KATCOFF, D., CARMAN, Y., REITER, J., FRISCHAUF, A.M. and YAFFE, D. (1981b). Skeletal muscle actin mRNA: Characterisation of the 3' untranslated region. Nucl. Acids Res. 9: 579-589.

SHIH, T.Y. and MARTIN, M.A. (1973). A general method of gene isolation. Proc. Natl Acad. Sci. USA 70: 1697-1700.

SHINE, J., SEEBURG, P.H., MARTIAL, J.A., BAXTER, J.D. and GOODMAN, H.M. (1977). Construction and analysis of recombinant DNA for human chorionic somatomammotropin. Nature 270: 494-499.

SINCLAIR, J. and BROWN, D. (1971). Retention of common nucleotide sequences in the ribosomal deoxyribonucleic acid of eukaryotes and some of their physical characteristics. Biochemistry 10: 2761-2769.

SINHA, A.M., UMEDA, P.K., RAJAMANICKHAM, C., KAVINSKY, C., HSU, H.J., JAKOVICIC, S. and RABINOVITZ, M. (1982). Molecular cloning of mRNA sequences for cardiac α- and β-form myosin heavy chains: Expression in ventricles of normal, hypothyroid and thyrotoxic rabbits. Proc. Natl Acad. Sci. USA 79: 5847-5851.

SMITH, D.F., SEARLE, P.F. and WILLIAMS, J.G. (1979). Characterisation of bacterial clones containing DNA sequences derived from *Xenopus laevis*. Nucl. Acids Res. 6: 487-506.

SMITH, H.O. and WILCOX, K.W. (1970). A restriction enzyme from *Haemophilus influenzae*: I. Purification and general properties. J. Molec. Biol. 51: 379-391.

SMITH, H.O. and NATHANS, D. (1973). A suggested nomenclature for bacterial host modification and restriction systems and their enzymes. J. Molec. Biol. 81: 419-423.

SONENSHEIN, G.E. and BRAWERIMAN, G.E. (1977). Differential translation of rat liver albumin messenger RNA in a wheat germ cell free system. Biochemistry 16: 5445-5448.

SPEAR, B.B. (1974). The genes for ribosomal RNA in diploid and polytene chromosomes of *D. melanogaster*. Chromosome 48: 159-180.

SRÉTER, F., HOLTZER, S., GERGLEY, J. and HOLTZER, H. (1972). Some properties of embryonic myosin. J. Cell Biol. 55: 586-594.

SRÉTER, F.A. and GERGLEY, J. (1973). Synthesis by fast muscles of myosin light chains characteristic of slow muscle in response to long-term stimulation. Nature New Biol. 241: 17-18.

SRÉTER, F.A., ELZINGA, M. and MABUCHI, K. (1975). The N$^+$-methylhistidine content of myosin in stimulated and cross reinervated skeletal muscles of the rabbit. FEBS Letters 57: 107-111.

STAHL, F.W., McMILIN, K.D., STAHL, M.M., MALONE, R.E. and NOZU, Y. (1972). A role for recombination in the production of "free-loader" lambda bacteriophage particles. J. Molec. Biol. 68: 57-67.

STARR, R. and OFFER, G. (1973). Polarity of the myosin molecule. J. Molec. Biol. 81: 17-31.

STREHLER, B.L., KONIGSBERG, I.R. and KELLEY, J.E.T. (1963). Ploidy of myotube nuclei developing *in vitro* a determined with a recording double beam micro-spectrophotometer. Exptl Cell. Res. 32: 232-241.

STERNBERG, N., TIEMEIER, D. and ENQUIST, L. (1977). *In vitro* packaging of a λD_{am} vector containing EcoR1 DNA fragments of *Escherichia coli* and phage P1. Gene 1: 255-280.

STOCKDALE, F.E., SADEN, H. and RAMAN, N. (1981). Slow muscle myoblasts differentiating *in vitro* synthesise both slow and fast muscle light chains. Dev. Biol. 82: 168-171.

STROHMAN, R.C., MOSS, P.S., MICOU-EASTWOOD, J. and SPECTOR, D. (1977). Messenger RNA from myosin polypeptides: Isolation from single myogenic cell cultures. Cell 10: 265-273.

STRUHL, K., CAMERON, J.R. and DAVIS, R.W. (1976). Functional genetic expression of eukaryotic DNA in *E. coli*. Proc. Natl Acad. Sci. USA 73: 1471-1475.

SUGINO, A., GOODMAN, H.M., HEYNEKER, H.L., SHINE, J., BOYER, H.W. and COZZARELLI, N.R. (1977). Interaction of bacteriophage T4 RNA and DNA ligase in joining of duplex DNA at base-paired ends. J. Biol. Chem. 252: 3987-3995.

SULLIVAN, D., PALACOIS, R., STAVNEZER, J., TAYLOR, J.M., FARAS, A.J., KIELY, M.L., SUMMERS, N.M., BISHOP, J.M. and SCHIMKE, R.T. (1973). Synthesis of a deoxyribonucleic acid sequence complementary to ovalbumin messenger ribonucleic acid and quantitation of ovalbumin genes. J. Biol. Chem. 248: 7530-7539.

SUN, S.M., SLIGHTOM, J.L. and HALL, T.C. (1981). Intervening sequences in a plant gene-comparison of the partial sequence of cDNA and genomic DNA of French bean phaseolin. Nature 289: 37-41.

SUZUKI, Y. and SUZUKI, E. (1974). Quantitative measurements of fibroin messenger RNA synthesis in the posterior silk gland of normal and mutant *Bombyx mori*. J. Molec. Biol. 88: 393-407.

SYROVÝ, I. (1977). Myosin - characteristics of the semi-tendinosus and extensor digitorum longus muscle of new born rabbits and adult rabbits. Int. J. Biochemistry 8: 847-848.

SYVANAN, M. (1974). *In vitro* genetic recombination of bacteriophage λ. Proc. Natl Acad. Sci. USA 71: 2496-2499.

SYVANAN, M. and YIN, J. (1978). Studies of DNA packaging into the heads of bacteriophage lambda. J. Molec. Biol. 126: 333-346.

THOMAS, M., CAMERON, J.R. and DAVIS, R.W. (1974). Viable molecular hybrids of bacteriophage lambda and eukaryotic DNA. Proc. Natl Acad. Sci. USA 71: 4579.

TIEMEIER, D., ENQUIST, L. and LEDER, P. (1976). Improved derivative of a phage λ EK2 vector for cloning recombinant DNA. Nature 263: 526-529.

TIEMEIER, D.C., TILGHMAN, S.M., POLSKY, F.I., STEIDMAN, J.G., LEDER, A., EDGELL, M.H. and LEDER, P. (1978). A comparison of two cloned mouse beta globin genes and their intervening sequences. Cell 14: 237-245.

TILGHMAN, S.M., TIEMEIER, D.C., SEIDMAN, J.G., PETERLIN, B.M., SULLIVAN, M., MAIZEL, J.V. and LEDER, P. (1978a). Intervening sequence of DNA identified in the structural portion of the mouse β-globin gene. Proc. Natl. Acad. Sci USA 75: 725-729.

TILGHMAN, S.M., CURTIS, P.J., TIEMEIER, D.C., LEDER, P. and WEISSMANN, C. (1978b). The intervening sequence of a mouse β-globin gene is transcribed within the 15S globin mRNA precursor. Proc. Natl Acad. Sci. USA 75: 1309-1313.

TRAYER, I.P. and PERRY, S.V. (1966). The myosin of developing skeletal muscle. Biochem. Z. 345: 87-100.

TSAI, S.Y., ROOP, D.R., TSAI, M.J., STEIN, J.P., MEANS, A.R. and O'MALLEY, B.W. (1978). Effect of oestrogen on gene expression in the chick oviduct. Regulation of the ovomucoid gene. Biochemistry 17: 5773-5780.

TSE, T.P.H. and TAYLOR, J.H. (1977). Translation of albumin messenger RNA in a cell free protein synthesising system derived from wheat germ. J. Biol. Chem. 252: 1272-1278.

ULLRICH, A., SHINE, J., CHIRGWIN, J., DICTET, R., TISCHER, E., ROLTES, W.J. and GOODMAN, H.J. (1977). Rat insulin genes: construction of plasmids containing the coding regions. Science 196: 1313-1319.

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.

UMEDA, P.K., KAVINSKY, C., SINHA, A.M., HSU, H.J., JAKOVCIC, S. and RABINOWITZ, M. (1982). Molecular cloning of myosin heavy chain cDNA: From skeletal and cardiac muscle. In, "Proceedings of the CHS Meeting of Molecular and Cellular Control of Muscle Development" (Ed. M. Pearson). Cold Spring Harbour, New York.

VANDEKERCKHOVE, J. and WEBER, K. (1978). The amino acid sequence of Physarum actin. Nature 276: 720-721.

VAN DEN BERG, J., VAN OOYEN, A., MANTEI, N., SCHAMBOCK, A., GROSVELD, G.C., FLAVELL, R.A. and WEISSMANN, C. (1978). Comparison of cloned rabbit and mouse β-globin genes showing strong evolutionary divergence of two homologous pairs of introns. Nature 276: 37-44.

VAN DER PLOEG, L.H.T., GROFFEN, J. and FLAVELL, R.A. (1980). A novel type of secondary modification of two CCGG residues in the human γ δ β -globin gene locus. Nucl. Acids Res. 8: 4563-4574.

VENDRELY, R. and VENDRELY, C. (1949). La teneur due noyan cellulaire en DNA a travers les organes, les individus et les especes animals. Experimentia 5: 327-329.

VERMA, I.M., TEMPLE, G.F., FAN, H. and BALTIMORE, D. (1972). *In vitro* synthesis of DNA complementary to rabbit reticulocyte 10S RNA. Nature New Biol. 235: 163-167.

VERTEL, B.M. and FISCHMAN, D.A. (1976). Myosin accumulation in mononucleated cells of chick muscle cultures. Dev. Biol. 48: 438-446.

VLAD, M. (1977). Quantitative studies of rDNA in amphibians. J. Cell. Science 24: 109-119.

WAHLI, W., DAWID, I.B., WYLER, T., WEBER, R. and RYFFEL, G.U. (1980). Comparative analysis of the structural organisation of two closely related vitellogenin genes in *Xenopus laevis*. Cell 20: 107-117.

WAHRMAN, J.P., WINAND, R. and LUZZATI, D. (1973). Effects of cyclic AMP on growth and morphological differentiation of an established myogenic cell line. Nature New Biol. 245: 112-113.

WAHRMAN, J.P., DRUGEON, G., DELAIN, E. and DELAIN, D. (1976). Gene expression during the differentiation of myogenic cells of the L$_6$ line. Biochimie 58: 551-562.

WALKER, C.R. and WILSON, B.W. (1975). Control of acetylcholinesterase by contractile activity of cultured muscle cells. Nature 256: 215-216.

WARBURG, O. and CHRISTIAN, W. (1941). Biochemische Zeitschrift 310: 384.

WEBER, K. and OSBORN, M. (1969). The reliability of molecular weight determinations by dodecyl sulphate polyacrylamide gel electrophoresis. J. Biol. Chem. 244: 4406-4410.

WEEDS, A.G., TRENTHAM, D.R., KEAN, C.J.C. and BULLER, A.J. (1974). Myosin from cross re-innervated cat muscles. Nature 247: 135-139.

WEIGEL, P.H., ENGLUND, P.T., MURRAY, K. and OLD, R.W. (1973). The 3'-terminal nucleotide sequences of bacteriophage λ DNA. Proc. Natl Acad. Sci. USA 70: 1151-1155.

WENSINK, P.C., FINNEGAN, D.J., DONELSON, J.E. and HOGNESS, D.S. (1974). A system for mapping DNA sequences in the chromosomes of *Drosophila melanogaster*. Cell 3: 315-325.

WHALEN, R.G., BUTLER-BROWN, G.S. and GROS, F. (1978). Identification of a novel form of myosin light chain present in embryonic muscle tissue and cultured cells. J. Molec. Biol. 126: 415-431.

WHALEN, R.G., SCHWARTZ, K., BOUVERET, P., SELL, S.M. and GROS, F. (1979). Contractile protein isozymes in muscle development: Identification of an embryonic form of myosin heavy chain. Proc. Natl Acad. Sci. USA 76: 5197-5201.

WHALEN, R.G. and SELL, S.M. (1980). Myosin from foetal hearts contains the skeletal muscle embryonic light chain. Nature 286: 731-733.

WHALEN, R.G., SELL, S.M., BUTLER-BROWN, G.S., SCHWARTZ, K., BOUVERET, P. and PINSET, I. (1981). Three myosin heavy chain isozymes appear sequentially in developing rat muscle. Nature 292: 805-809.

WILKIE, D.R. (1968). In, "Muscle". Edward Arnold Ltd.

WILLIAMS, B.G. and BLATTNER, F.R. (1979). Construction and characterisation of the hybrid bacteriophage lambda charon vectors for DNA cloning. J. Virol. 29: 555-575.

WILLIAMS, J.G. and LLOYD, M.M. (1979). Changes in the abundance of polyadenylated RNA during slime mould development measured using cloned molecular hybridisation probes. J. Molec. Biol. 129: 19-35.

WOO, S.L.C., MONAHAN, J.J. and O'MALLEY, B.W. (1977). The ovalbumin gene: Purification of the anticoding strand. J. Biol. Chem. 252: 5789-5797.

WU, R. and TAYLOR, E. (1971). Nucleotide sequence analysis of DNA: II. Complete nucleotide sequence of the cohesive ends of bacteriophage λ DNA. J. Molec. Biol. 57: 591-511.

WYDRO, R., NGUYEN, H.T., GUBITS, K. and NADAL-GINARD, B. (1983). Characterisation of sarcomeric myosin heavy chain genes. J. Biol. Chem. 258: 670-678.

YAFFE, D. and DYM, H. (1972). Gene expression during differentiation of contractile muscle fibres. Cold Spring Harbour Symp. Quant. Biol. 37: 543-547.

YOUNG, E. and SINSHEIMER, R. (1967). Vegetative phage λ DNA: II. Physical characterisation and replication. J. Molec. Biol. 30: 165-200.

YOUNG, E. and SINSHEIMER, R. (1968). Vegetative phage λ DNA: III. Pulse labelled components. J. Molec. Biol. 33: 49-59.

ZALIN, R.J. and MONTAGUE, W. (1974). Changes in adenylate cyclase, cyclic AMP, and protein kinase levels in chick myoblasts, and their relationship to differentiation. Cell 2: 103-108.

ZIMMER, W. and SCHWARTZ, R.J. (1982). Amplification of chicken actin genes during myogenesis. In, "Gene Amplification" (Ed. R.T. Schimke). Cold Spring Harbour Laboratories, New York.

APPENDIX

Solutions and Formulae

Acetate-EDTA	3M	sodium acetate
	1mM	EDTA pH 7.0
Binding Buffer	0.5M	potassium chloride
(autoclave)	0.1M	Tris-HCl pH 7.6
Concentration Buffer	40mM	Tris-HCl pH 8.0
(Filter sterise)	10mM	sodium azide
	10mM	magnesium chloride
	10mM	putrescine
	0.1%	β-mercaptoethanol
	7%	dimethylsulphoxide
	10mM	spermidine
Denhardt Solution	3x	SSC
(10x concentrated)	0.2%	polyvinylpyrolidine
	0.2%	BSA fraction V
	0.2%	ficoll
	50μg/ml	denatured *E. coli* DNA
Dilute Saline Citrate	15mM	sodium chloride
(DSC)	15mM	trisodium citrate pH 7.1
Dissociation Buffer	12M	urea
	1.5%	SDS
	1.5%	β-mercaptoethanol
	15mM	sodium phosphate pH 7.0
EcoR1 Buffer	0.1M	Tris-HCl pH 7.5
	10mM	magnesium chloride
	10mM	β-mercaptoethanol
	50mM	sodium chloride
Eluting Buffer	10mM	Tris-HCl pH 7.6
(autoclave)		
Grinding Buffer	20mM	Hepes pH 7.8
(wheat germ translation	0.12M	potassium acetate } autoclave
system)	5mM	magnesium acetate
	1mM	dithiothreitol

L-Agar*
(autoclave)

1%   bactotryptone (Difco)
1%   sodium chloride
0.5% yeast extract (Difco)
1%   bacto-agar (Difco)
pH 7.2 with sodium hydroxide

L-Agarose*
(autoclave)

As for L-agar with 1% agarose in place
of the bacto-agar

L-Broth*
(autoclave)

1%   bactotryptone (Difco)
1%   sodium chloride
0.5% yeast extract (Difco)
pH 7.2 with sodium hydroxide

Medium 9
(filter sterilise)

10%   (v/v) salt mixture
0.4% glucose
0.1mM calcium chloride
1mM   magnesium sulphate

Muscle Buffer (M)

0.25M potassium chloride
10mM magnesium chloride
10mM Tris-HCl pH 7.4   autoclave
50µg/ml dextran sulphate

Muscle Buffer (MIB)
(autoclave)

0.15M potassium chloride
10mM magnesium chloride
20mM Tris-HCl pH 7.6
6mM   β-mercaptoethanol
10% or 50% glycerol

Parish-Kirby Solution

1%   sodium tri-isopropyl naphtholene sulphonate
6%   sodium-4-amino salicilate
1%   sodium chloride
6%   butanol
0.5% SDS

Phage Buffer
(filter sterilise)

0.3% potassium dihydrogen orthophosphate
0.7% di-sodium hydrogen orthophosphate
0.5% sodium chloride
1mM   magnesium sulphate
0.1mM calcium chloride
0.001% gelatin

---

*Growth media were supplemented with 40 µg/ml thymidine and 10 µg/ml diaminopimelic acid for the growth of DP50supF cells.

Phenol Solution	275 ml	phenol, water saturated, freshly distilled
	275 ml	chloroform
	35 ml	m-cresol
	0.25 g	8-hydroxyquinoline
Phosphate-EDTA (PEB)		Equimolar amounts of sodium dihydrogen orthophosphate and di-sodium hydrogen orthophosphate
	1mM	EDTA pH 7.0
Polysome Buffer (P)	10mM	Tris-HCl pH 9.0
	0.5%	SDS
RNA Gel Buffer	50mM	boric acid
	5mM	sodium borate
	10mM	sodium sulphate
	1mM	EDTA
S1 Buffer	0.3M	PEB
	0.2M	sodium chloride
	0.3M	sodium acetate pH 4.5
	2mM	EDTA
	6mM	zinc sulphate
	0.52N	acetic acid
Saline-EDTA	0.15M	sodium chloride
	0.1M	EDTA pH 8.0
Salt Mixture	7%	di-sodium hydrogen orthophosphate
	3%	potassium di-hydrogen orthophosphate
	0.5%	sodium chloride
	1%	ammonium chloride

Sample Buffer

(a) Laemmli gels	10%	glycerol
	5%	β-mercaptoethanol
	2.3%	SDS
	62.5mM	Tris-HCl pH 6.8
(b) Weber and Osborn gels	8M	urea
	1%	SDS
	1%	β-mercaptoethanol
	10mM	sodium phosphate pH 7.0

Soft Top Agar*   As for L-agar with only 0.65% bacto-agar

Soft Top Agarose*   As for L-agarose with only 0.65% agarose

Standard Saline Citrate   0.15M   sodium chloride
(SSC)   15mM tri-sodium citrate pH 7.1

Stop Buffer   60mM PEB
                                   120µg/ml denatured, sheared *E. coli* DNA

---

*Growth media were supplemented with 40 µg/ml thymidine and 10 µg/ml
diaminopimelic acid for the growth of DP50SupF cells.