THE STRUCTURE AND EXPRESSION OF *XENOPUS BOREALIS* ACTIN GENES

A thesis submitted for the degree of

Doctor of Philosophy

at the University of Warwick

February 1984

by

Gareth S. Cross M.A.

Developmental Biology Group

Department of Biological Sciences

University of Warwick

Coventry

U.K.

TABLE OF CONTENTS

FIGURES

Acknowledgements

I must begin by thanking my supervisor, Professor Hugh Woodland, for all his helpful advice during this project. A big thank you must also go to Dr. Phil Turner for his valuable advice (and materials) over these years; a special thank you for teaching me how to take off sequencing gels Phil (sorry, an obscure, Phil-like, joke). Thanks also to Dr. R. W. Old for advice and materials, Dr. William Bains, for letting me screen his library, Dr. Giorgio Valle, for his help with the rabbits (and his radioactive ink); Dr. Tim Mohun, for his sequence comparisons; Professor Roger Whittenbury, for getting me some more money; and Elizabeth Ballantine, for various materials and just generally being nice (will that do Elizabeth?). Finally, a big thank you to my parents and all my friends who've helped me keep on going.

Declaration

Apart from a very small section of DNA sequence (shown in Fig. 3.19), all of the results presented in this thesis were obtained by the author.

Abbreviations

ATP, dATP	Adenosine triphosphate, deoxyadenosine triphosphate
bp	Base pairs
cDNA	Complementary DNA
Ci	Curie
dCTP	Deoxycytidine triphosphate
DEAE	Diethylamino Ethyl
DMSO	Dimehtylsulphoxide
DNA	Deoxyribonucleic acid
EDTA	Sodium ethylenediaminetetraacetate
EGTA	Ethyleneglycol-bis-(beta-aminoethyl ether)-N,N,N',N-tetraacetic acid
GTP, dGTP	Guanosine triphosphet, deoxyguanosine triphosphate
g	Gramme
Kb	Kilobase
mg	Milligramme
mRNA	Messenger RNA
PIPES	1,4-Piperazinediethanesulfonic acid
PPO	2,5-Diphenyloxazole
RNA	Ribonucleic acid
rRNA	Ribosomal RNA
rpm	Revolutions per minute
SDS	Sodium dodecylsulphate
SSC	Standard saline citrate (see Section 2.2)
ss	Single stranded
tRNA	Transfer RNA
TEMED	N,N,N',N'-Tetramethyl-ethylenediamine
VLDL	very-low-density lipoprotein

Tris	Tris (hydroxymethyl) aminomethane, hydrochloride salt buffer
dTTP	Deoxythymidine triphosphate
w/v	Weight per volume
µg	Microgramme
5AP	The 2.6 kb Eco RI insert of the plasmid p5AP or λ 5AP
35A	The 3.6 kb Eco RI insert of the plasmid p35A or λ35A

SUMMARY

Two cloned Eco RI fragments have been isolated from a Xenopus borealis genomic library in phage lambda, using an actin cDNA clone as a probe. Both fragments have been restriction-mapped and partially sequenced. As a result of this, one fragment, 2.6 kb in size, has been found to encode an amino acid sequence identical to the first 149 amino acids of the mammalian cardiac muscle actin, while the other, 3.6 kb in size, encodes a protein which is almost identical to the first 267 amino acids of the mammalian beta-cytoskeletal actin. The putative muscle actin gene contains introns at different positions to those of the cytoskeletal actin gene. These intron positions are identical to those of the analogous genes in mammals. These data suggest that vertebrate muscle and cytoskeletal actin genes became separate before the divergence of amphibians from the rest of the vertebrates.

Other sequence studies performed include an examination of codon usage and sequence divergence in actin genes.

'Repetitive DNA' has been found in the 3.6 kb Eco RI fragment and has been roughly localized.

Primer extension has been used to show that the partial leader sequence of an oocyte mRNA species is identical to that found in the beta-like cytoskeletal actin clone. A large intron is present in the leader of this gene. Near the transcriptional start site, several small sequences have been found which exist in similar positions in the rat beta-cytoskeletal actin gene, and may possibly be involved in the control of gene expression. Another oocyte actin mRNA species encodes a protein similar to the N-terminal end of the mammalian gamma-cytoskeletal actin, and its leader sequence is surprisingly similar to that of the beta-like actin message.

Primer extension has also been used to study the expression of actin genes in several X. laevis non-muscle tissues, as well as in X. borealis skeletal muscle.

1. INTRODUCTION

1.1 General Aims

One of the most interesting biological enigmas is that of the development of multicellular organisms. A single fertilized egg can give rise to a wide range of different cell types, each characterized by a specific protein complement, yet all containing the same set of genes. The latter point was shown in Xenopus laevis, where the nuclei from skin and intestinal epithelium (i.e. differentiated cells) were injected into enucleated eggs, and shown to be capable of giving rise to swimming tadpoles, containing a wide range of differentiated tissues (Gurdon, 1974). Recently, swimming tadpoles have also been formed following the injection of adult erythrocyte nuclei of Rana pipiens into oocytes, the nuclei of which were later removed after maturation (Di Berardino and Hoffner, 1983). Therefore, it appears that differentiation does not involve the elimination or irreversible inactivation of parts of the genome in particular tissues. It must instead involve the controlled expression of genes in such a way that specific sets of proteins are expressed only in the appropriate place (in addition to the proteins expressed in all cells, the so-called 'house-keeping proteins'). How this occurs, is one of the major questions at which the research described in this thesis is aimed.

The molecular mechanisms involved in determining the types of cell differentiation that take place during the development of an organism are, of course, particularly interesting. They are, however, most easily investigated in organisms where developmental mutations can be mapped, enabling the genes involved to be cloned. The best known examples of these are the bithorax, antennapedia, and similar genes in

Drosophila (Bender et al., 1983; Akam, 1983; North, 1983). However, it should also be possible to gain some insight into the molecular mechanisms of differentiation by studying the structure and control of expression of those genes whose products characterize differentiated cells. Indeed, it will clearly be necessary to know this information to fully understand the action of mechanisms which bring about differentiation. The original aim of this project was, therefore, to study the structure, pattern of expression, and eventually, the mechanism of control of a family of such genes: the actins. (In fact, not all the actin isoforms are tissue-specific, but as explained later, the degree of expression of the cytoskeletal actin genes does appear to be modulated in different tissues.)

I have chosen to study the actin genes of the frog Xenopus borealis, partly because of the existence, already, of a certain amount of data concerning events at the molecular level during Xenopus development, including some on actin gene expression (Ballantine et al., 1979; Sturgess et al., 1980), and partly because of the manipulative usefulness of Xenopus oocytes. This usefulness arises from their enormous size (roughly 1 mm in diameter) which makes them easy to micromanipulate and microinject with exogenous DNAs and RNAs. The Xenopus oocyte has become a powerful system for introducing cloned genes into a living cell to study the sequences important in their expression. For example, the deletion or mutation of certain regions of cloned DNA, before recloning and microinjection, has been used to study the function of sequences 5' to certain genes (Grosschedl and Birnstiel, 1980, 1982; Grosschedl et al., 1981; Wasylyk and Chambon, 1981). When using such a system, it would seem reasonable to judge the results as more meaningful when the genes injected are from the same, or a closely related, species

(especially in the case of the cytoskeletal actins, where the genes are known to be expressed in the oocyte itself).

As it involved the isolation and sequencing of cloned actin genes, this study has also contributed towards our understanding of the evolution of actins, both at the protein level (e.g. the significance of various amino acid substitutions), and at the level of the genes themselves (e.g. the relationship between amphibian actin genes and those of other vertebrates), no information on amphibian actin genes having been published yet.

To appreciate the questions that a study such as this can attempt to answer, it is useful to describe our current knowledge of the family of actin proteins. The next section will describe the function of actin in general, the differences between the various actins, and their differing patterns of expression.

1.2 Actin: The Protein

The Function of Actin

Actin is an important element in the contractile mechanism of muscle cells, as well as performing an important role in all eukaryotic cells, both structurally and in motility. In its monomeric form ('G-actin') it has a molecular weight of approximately 42,000 daltons, but it also exists in muscle, and other cells, as a filamentous polymer ('F-actin'). In muscle the polymerization occurs only once, but in non-muscle cells the large pool of unpolymerized actin, and the extensive rearrangement of actin during motility and the cell cycle, suggest that the pools of actin monomer and polymer may be in a more dynamic state (reviewed by Pollard and Craig, 1982; Craig and Pollard, 1982). It is in the polymerized form that actin functions physiologically.

The mode of action of actin was first elucidated in striated muscle (Huxley, 1969), where contraction is accomplished by the sliding of overlapping actin filaments and myosin filaments, powered by ATP hydrolysis; the actin filaments are anchored at lattice-like structures, called Z bands, which contain the protein alpha-actinin. This sliding movement is mediated by the continual attachment of myosin 'heads' (myosin has a bipartite structure with two globular heads and a long, fibrous tail) to the actin filament, followed by a conformational change, a breakage of the actomyosin contact, the hydrolysis of ATP, and a return to the original conformation. Regulation of the contraction is mediated by reversible inhibition of the actin-myosin binding reaction, on the lowering of the intracellular calcium ion concentration. In molluscan muscle this regulation resides in the myosin molecule, but in vertebrate striated muscle the thin actin filaments contain a complex of

the control proteins troponin and tropomyosin, which undertake the interaction with the calcium ions.

A different regulatory mechanism exists in smooth muscle and non-muscle tissues, where it is believed that a similar type of actin-myosin interaction gives rise to contractility, but where no troponin has been found. Here, the protein caldesmon binds to actin and inhibits crossbridge formation until, in the presence of raised calcium ion levels, caldesmon prefers to bind to calmodulin, and the actin-myosin interaction can take place (Kakiuchi and Sobue, 1983). In non-muscle cells, actin exists in a variety of structural forms and is involved in a wide variety of motile processes, including cell locomotion, cytoplasmic streaming and transport, secretion, phagocytosis and cytokinesis (Buckley and Porter, 1967; Pollard and Weihing, 1967; Korn, 1978). Examples of the different structural forms that actin may take include the stress fibres and the filamentous network that may be found in moving cells (Lazarides and Weber, 1974; Lazarides, 1976; Goldman and Knipe, 1972; Heath and Dunn, 1978; Small et al., 1978). Stress fibres consist of tightly-packed bundles of actin-containing filaments in association with myosin, tropomyosin and alpha-actinin, and are associated at their termini with the cell membrane. The dense filament meshwork can be found in the leading lamellae and ruffling membranes, and undergoes dramatic changes in viscosity and stiffness. Many proteins that bind to actin have been found that might control this sort of versatility (although the in vivo action of few have been demonstrated; Craig and Pollard, 1983; Weeds, 1982). These include proteins that cross-link filaments into networks or bundles, proteins which cap the ends of filaments to regulate their length (and may also sever filaments), and proteins that inhibit polymerization by binding to

actin monomers.

Actin is a conserved protein

The large number of proteins with which actin must interact, as well as the necessity for the self-association of G-actin, probably explains why actin is so highly conserved in amino acid sequence. Thus, chicken and mammalian skeletal muscle actins are identical, as are chicken gizzard and mammalian gamma-smooth muscle actins, and these differ from cytoskeletal actins at only a small number of residues (Vandekerckhove and Weber, 1978b, 1978d, 1979). Physarum actin shows only 5% amino acid substitutions when compared with mammalian cytoskeletal actins (Vandekerckhove and Weber, 1978e), and a sea urchin actin, only 6% divergence from Drosophila actins (Cooper and Crain, 1982). The most diverged actin sequence so far detected is that of Oxytricha fallax (Kaine and Spear, 1982), but this may encode an actin-related fibre protein, since the encoded polypeptide is smaller than other actins (being only 356 amino acids long, whereas actins from other organisms vary between 374 and 376 amino acids in length). Amongst other actins the highest divergence is that of 19%, between Drosophila and soybean actins (Shah et al., 1982). This evolutionary conservation has been very useful when identifying clones, because it enables genes from phylogenetically distant organisms to be used as hybridization probes (see Section 3.1).

Mammalian actins: some are tissue-specific

As suggested in the previous paragraph, some actins are tissue-specific. At present, six different primary sequences have been described in different mammalian tissues (Vandekerckhove and Weber,

1979). Their nomenclature is based on the ability of isoelectric focussing techniques to resolve them into three groups, called alpha, beta and gamma, in order of decreasing negative charge (Whalen et al., 1976; Garrels and Gibson, 1976). There is an alpha-actin sequence specific to skeletal muscle, and a different alpha-actin in the heart (Vandekerckhove and Weber, 1978a). (In contrast to the situation for myosin, tropomyosin and troponin, there is no evidence for different alpha-actins in fast and slow skeletal muscle; Vandekerckhove and Weber, 1979.) In smooth muscle, two other actins have been identified. Their relative proportions vary in different smooth muscle tissues; in aorta, an alpha-type predominates, whereas in stomach (or chicken gizzard), a gamma-type is the major component (Vandekerckhove and Weber, 1981). Two cytoskeletal actins, beta- and gamma-types, have been found in all mammalian non-muscle cells examined (Garrels and Gibson, 1976), and it has been shown that they are both expressed in a single cell (Bravo et al., 1981). This information has been summarized in Figure 1.1, for easy reference. In addition to these six, a few additional isoforms have been postulated: in the nucleus of non-muscle cells (Bremer et al., 1981), and in brain synaptosomes (Marotta et al., 1978).

What are the differences between these various isoforms? The amino acid differences between the accepted six mammalian isoforms are small in number. Only 4 out of 375 amino acids differ between cardiac and skeletal muscle actins, 6 to 8 between smooth and skeletal muscle actins, 24 to 25 between striated muscle and cytoskeletal actins, and 22 to 23 between smooth muscle and cytoskeletal actins (Vandekerckhove and Weber, 1979). It is possible that these differences are the result of the different interactions of actin (with control proteins, etc.) in these tissues. The most variable part of the actin sequence is the 5 to

Fig.1.1 The 6 major mammalian actin isoforms

	Tissue	Position on 2D gel
1	heart (and fetal skeletal) muscle	α
2	skeletal muscle	α
3	aorta smooth muscle (minor in stomach)	α
4	stomach smooth muscle (minor in aorta)	γ
5	most tissues	γ
6	most tissues	β

Figure 1

The

together

actins ca

, in orde

specific

muscle, b

the latte

Figure 1.1: The 6 Mammalian Actin Isoforms

The location of the 6 major mammalian actin isoforms is summarized together with their running position on 2-dimensional gels. Mammalan actins can be separated into 3 spots by isoelectric focussing: α, β and γ, in order of decreasing negative charge. The two smooth muscle-specific isoforms exist in varying proportions in both aorta and stomach muscle, but the α-type is predominant in the former, and the γ-type in the latter.

Position on 2D gel

α
α
α
γ
γ
β

6 amino acid region at the N-terminal end of the protein, and this region has recently been implicated in the binding of actin to the myosin heavy chain (Sutoh, 1982).

There is evidence that, in mammals at least, the expression of actin genes is controlled, not just in a tissue-specific manner, but also temporally in the same tissue. Minty et al. (1982) have shown that cardiac muscle actin is coexpressed with skeletal muscle actin in mouse fetal skeletal muscle, but that in adult skeletal muscle, the mRNA for the cardiac muscle isoform is no longer present. This raised the interesting possibility that groups of muscle-specific genes might all be turned on together at myogenesis; the restricted isoform patterns evident in adult muscle might result from the selective turn-off of expression of inappropriate isoforms later in development. This certainly seems to be applicable to the expression of fast and slow isoforms of myosin light chains (Keller and Emerson, 1980; Stockdale et al., 1981), alpha-tropomyosin (Montarras and Fiszman, 1982) and troponin C (Hastings and Emerson, 1982), in cultured myoblasts from slow or fast avian skeletal muscles. Minty et al. (1982), however, could detect only extremely low levels of skeletal muscle actin mRNA in 17 to 20 day fetal and adult heart, although it is possible that larger amounts of this mRNA may be synthesized at earlier stages of fetal heart development. At this time, therefore, it seems unwise to attribute a single programme of control to all the contractile proteins that are expressed in any one tissue. Thus, although fetal and adult isoforms of myosin light chain 1 are also co-expressed in fetal skeletal muscle (Whalen et al., 1978), and in myotube cultures (Robert et al., 1952), the disappearance of fetal actin (cardiac muscle) mRNA during skeletal muscle development does not appear to be tightly co-ordinated with that of fetal myosin

light chain. The three developmentally regulated forms of myosin heavy chain have been shown to be expressed sequentially in muscle development; and not all together at any time (Whalen *et al.*, 1981). Additionally, none of the myosin heavy chains appear to be both activated and repressed coordinately with fetal actin. Such data indicate that there may be programmes of developmental regulation specific to each multigene family of muscle-specific proteins.

Non-mammalian actins

How many actins do non-mammalian organisms have? As mentioned above, the chicken skeletal muscle and gizzard muscle actins are identical to their mammalian counterparts, so it is possible that the whole set of actins in birds is identical to that in mammals. The only other vertebrate actins to be studied are the cytoskeletal actins of the amphibians *Xenopus laevis*, *Rana pipiens* (frogs), *Triturus cristatus* (newt) and *Pleurodeles waltlii* (salamander) (Vandekerckhove *et al.*, 1981). Some of these have been shown to possess differences in the 3 N-terminal acidic residues possessed by the mammalian cytoskeletal actins. This information is summarized in Figure 1.2. Thus, while the mammalian beta- and gamma- types possess the end sequences X-asp-asp-asp (designated type 1) and X-glu-glu-glu (type 8), respectively (where asp = aspartic acid, glu = glutamic acid, and X is a blocking group recognized as an acetyl group in other actins), *X. laevis* possesses 3 cytoskeletal types: X-glu-asp-asp (type 4), X-asp-glu-glu (type 5) and a type 8 (in kidney, brain, liver and oocytes). Type 1, type 8 and type 3 (X-asp-glu-asp) actins are found in cultured cells of *Triturus*. Brain and liver tissues of *Rana* only show two actin forms, type 4 and type 5, while liver tissue of *Pleurodeles* shows only a single major actin form

Fig. 1.2

Actin Type	X. laevis	R. pipiens	T. cristatus	mammals
1= X _asp_asp_asp_			+	+
3= X _asp_glu_asp_			+	
4= X _glu_asp_asp_	+	+		
5= X _asp_glu_glu _	+	+		
8= X _glu_glu_glu _	+		+	+

Figure 1.2

The
so far fr
which the
known to
terminal
recogniz
as sugge

Figure 1.2: 5 Vertebrate Cytoskeletal Actins

The 5 different cytoskeletal actin isoforms that have been isolated so far from vertebrates are presented, together with the species in which they have been found (marked +). As indicated, these isoforms are known to vary in the acidic amino acid residues present at their N-terminal end (asp = aspartate, glu = glutamate, X = a blocking group recognized as an acetyl group in other actins). The numbering system is as suggested by Vandekerckhove et al. (1981).

(as yet uncharacterized). Clearly, cytoskeletal actins are more complex and variable in amphibians than in warm-blooded vertebrates. The reason for this is unknown. The possible localization of the two mammalian cytoskeletal isoforms, in different regions of the muscle cell, is discussed in Section 4.1. It is possible that this apparent specialization of function in mammals has not occurred in amphibians and that these differences represent neutral mutations.

Two genes coding for different cytoskeletal actins have been found in the sea urchin Strongylocentrotus purpuratus (Cooper and Crain, 1982; Schuler et al., 1983). The total number of different actin isoforms in sea urchins has not been studied. Some regulation of their expression during development does take place, however. Of the two actin mRNA species observed in Northern blot experiments, a 1.8 Kb transcript is most abundant in embryos after the blastula stage, while a 2.2 Kb message is most abundant during oogenesis, early embryogenesis, and in adult cell types, such as coelomocytes or intestine (reviewed by Davidson et al., 1982).

Drosophila has 6 actin genes which code for at least 3 different polypeptides (Fyrberg et al., 1981). 3 different actins can also be separated by isoelectric focussing (Storti et al., 1978; Fyrberg and Donady, 1979). Very recently, Fyrberg et al. (1983) have surveyed the expression of the six genes during development, and found each gene to be transcribed to form a functional mRNA which accumulates with a distinct pattern. Two of the genes are expressed in undifferentiated cells, and probably encode cytoskeletal actins; another two are expressed predominantly in larval, pupal and adult intersegmental muscles. A fifth is expressed only in the muscles of the adult thorax, and the sixth in both the adult thorax and leg muscles. Clearly, there

is tissue-specific expression of Drosophila actin genes.

It is interesting that the amino acid sequence of each of the 6 Drosophila actin genes is more similar to that of the vertebrate cytoskeletal actins, than to vertebrate muscle actins (Fyrberg et al., 1981). Possibly, the differences found between vertebrate muscle and cytoskeletal actins arose after the divergence between protostomes and deuterostomes and are not necessary for the function of all muscles. In support of this, Davidson et al. (1982) have reported the results of Files et al. (1983) which show that the DNA sequence of an actin gene expressed in the body wall musculature of the nematode worm Caenorhabditis elegans indicates that this gene also codes for a protein similar to vertebrate cytoskeletal actins. One interesting prospect for this project, therefore, was to discover at what stage this specialization of muscle isoforms occurred in the development of vertebrates (i.e. do amphibian muscle-specific actins resemble mammalian muscle or cytoskeletal isoforms?).

So far, I have described the occurrence of tissue-specific expression of actin genes in certain Metazoa, the vertebrates, insects and nematodes, and pointed out that in sea urchins, different actins become predominant at different developmental stages. Can this be extended to other forms of life? Do other organisms possess multiple forms of actin, and are the genes for these expressed in different ways?

The exact number of actin isoforms in plants has not been investigated. One gene has been cloned from soybean, however, and found to resemble the vertebrate cytoskeletal actins in amino acid sequence (Shah et al., 1982). Only one form of actin can be found in the plasmodia of the slime mould Physarum polycephalum (Vandekerckhove and Weber, 1978e), and the yeast Saccharomyces cerevisiae only possesses one

actin gene (Ng and Abelson, 1980; Gallwitz and Seidel, 1980). This apparent simplicity of actin composition in lower eukaryotes cannot necessarily be extended to Acanthamoeba castellaani, however, as Jantzen (1981) has described the separation of at least three actin forms by 2-dimensional electrophoresis, although it was not demonstrated that these were not merely post-transcriptional modification of the products of the same gene, as originally found in the slime mould Dictyostelium (Kindle and Firtel, 1978; Firtel, 1981). The latter organism has 17 actin genes of which at least 8 are known to be expressed (McKeown et al., 1978; McKeown and Firtel, 1982), yet there appears to be only one type of actin primary sequence (similar to the vertebrate cytoplasmic type) expressed by all the genes (Vandekerckhove and Weber, 1980; McKeown and Firtel, 1982), although not all of the genes have been sequenced and minor protein forms representing less than 5% of the total actin would not have been detected. This phenomenon is even more remarkable, as it seems that several of these genes are differentially expressed during the Dictyostelium life cycle (McKeown and Firtel, 1981; Tsang et al., 1982). The reason for this is unknown.

The levels of cytoskeletal actins can vary

So far, the genes that are expressed in a tissue-specific manner and have been characterised have all been found to be those coding for muscle actins. Does this mean that the cytoskeletal actins are expressed constitutively in all cell types?

Although the cytoskeletal actins have been found in nearly all vertebrate cell types, their synthesis is not independent of the cell in which they are expressed. Thus, regulation appears to occur, both in total synthesis rate, according to cell morphology or differentiated

state of the cell, and in the relative proportions of the gamma- and beta-type mammalian isoforms. Garrels and Gibson (1976) reported that the ratio of these latter isoforms (as measured on 2-dimensional gels, when labelled with ^{35}S-methionine), although always less than 1, was not constant in various cell lines; these included nerve, glial cell, postfusion muscle precursor and myotube rat cell lines. A decrease in the rate of actin synthesis has been reported when certain sublines of 3T3 cells undergo adipose differentiation (Spiegelman and Farmer, 1982). The altered synthesis of actin (down by 90%), as well as tubulin and vimentin, was quite specific in that most preadipocyte proteins did not undergo large decreases during differentiation. In contrast to this, there is an increase in cytoskeletal actin synthesis during the differentiation of myeloid leukaemia cells (Hoffman-Lieberman and Sachs, 1978).

There appears to be a control of cytoskeletal actin gene expression during some types of differentiation. It is not clear, however, in the case of adipocyte differentiation, whether the change in actin synthesis is not merely due to the change in cell morphology at differentiation. Thus, the suspension of anchorage-dependent mouse fibroblasts has been reported to result in a decrease in total protein synthesis, with actin synthesis being preferentially inhibited (Farmer *et al.*, 1983). Reattachment initiates a preferential increase in actin synthesis. Thus, there appears to be a close relationship between the morphological state of the cell and the regulation of cytoskeletal gene expression. Growth may also be an important factor, as Riddle *et al.* (1979) have suggested that actin synthesis peaks during the transit from Go to S phase, during the serum-stimulated growth of mouse 3T3 cells.

Is there a change in cytoskeletal actin synthesis at muscle cell

differentiation? Several groups have reported the reduction in quantity of cytoskeletal actin messages following the fusion of myoblasts during myogenesis (Schwarz and Rothblum, 1981; Shani et al., 1981; Caravatti et al., 1982), and Garrels and Gibson (1976) were unable to detect cytoskeletal actins in adult muscle tissue using 2-dimensional gels and Coomassie-blue staining. Caravatti et al. (1982), however, did not observe any decrease in beta- and gamma-actin synthesis for at least 72 hours after the approximate time of myoblast fusion. Pardo et al. (1983) used affinity chromatography to isolate a polyclonal antibody to chicken gizzard actin which was specific to gamma-type actins (including the cytoskeletal type). Using immunofluorescence methods, they found that this antibody colocalizes with mitochondria in muscle fibres of the mouse diaphragm, indicating that cytoskeletal actins do exist in fully differentiated muscle cells. A similar conclusion was reached by Lubit and Schwartz (1980), using an antibody to Aplysia actin. The inability of Garrels and Gibson (1976) to find cytoskeletal actin in adult muscle is explicable because the staining method they used is comparatively insensitive.

Cytoskeletal actin synthesis also appears to be regulated during early development. Ballantine et al. (1979) and Sturgess et al. (1980) have found that the cytoskeletal actins of Xenopus laevis are among the most rapidly made proteins of oocytes, but that their synthesis, relative to other proteins, is low in the egg or cleaving embryo. In blastulae and later embryonic stages, however, cytoskeletal actins are once more among the most rapidly made proteins. Abreu and Brinster (1978) and Martin et al. (1978) have reported that actin synthesis in the developing mouse is reduced just after fertilization. The functional reasons for these changes is not known. In the case of

Xenopus, Sturgess et al. (1980) have pointed out that, because the oocyte contains a large store of actin, making only a small amount of actin for 5 to 6 hours of early development will have very little effect on the actin content of the early embryo.

Summary of main conclusions

Actin is a highly conserved protein and can be found in widely divergent organisms, including deuterostomes, protostomes, plants and protists. It is therefore suitable for evolutionary studies covering a long period of time. The muscle actins have been found to be tissue-specific, while the ratio of beta- to gamma-cytoskeletal actins can also vary from tissue to tissue. The total levels of cytoskeletal actins appear to be determined by many different factors, which may include cellular morphology, growth phase and differentiation, as well as the developmental stage of the organism.

How does the control of actin gene expression come about? This will be discussed in the next chapter.

1.3 The Control of Actin Gene Expression

Various actin isoforms are expressed in a tissue-specific manner while the extent of expression of others appears to be finely controlled. How are these controls brought about? To begin this section, the possible control sites in the pathway of gene expression will be discussed, where possible, with reference to previous work on actin expression. The second part will discuss the types of control elements that may be involved in this regulation and will also consider the role of gene clustering in the control of those genes' expression; the spatial relationship between co-expressed contractile protein genes will be discussed.

1.3.1 At what level is actin gene expression controlled?

The production of a functional protein from a eukaryotic gene is a process that involves many steps. When investigating the control of gene expression, it is necessary to consider which of these steps is most likely to be involved in that control. The protein-coding sequence of most eukaryotic polymerase II genes, with the exception of such genes as histones and alpha-interferon (Kedes, 1979; Nagata et al., 1980), contain inserts of non-coding sequence called introns (Breathnach and Chambon, 1981). All of the actin genes so far cloned, with the possible exception of the Dictyostelium (McKeown and Firtel, 1981b) and Oxytricha (Kaine and Spear, 1982) genes, have been found to contain introns (Hamada et al., 1982; Nudel et al., 1983; Zakut et al., 1982; Fornwald et al., 1982, Davidson et al., 1982; Fyrberg et al., 1981; Sanchez et al., 1982; Files and Hirsch, 1982; Nellen and Gallwitz, 1982; Gallwitz and Sures, 1980; Ng and Abelson, 1980; Shah et al., 1982). After

transcription, such introns must be spliced out of the primary transcript. Not only must this RNA be removed, but transcription in some genes, such as the mouse and chicken beta-globin genes (Hofer and Darnell, 1981; Groudine and Weintraub, 1981), has been shown to proceed past the polyadenylation site, and hence this RNA must be removed before polyadenylation of these messages can occur. Preceeding these steps, the primary transcript is capped with an m7G residue (Shatkin, 1976), and once all such modifications are made, the message must still be transported from the nucleus. Once in the cytoplasm, the production of protein depends on the translation of that message by ribosomes, and the amount produced per mRNA molecule depends on its stability. Once the protein is made, it may still have to undergo modification before becoming active and, of course, the amount of protein present at any time will depend on its stability.

Which of these steps is most likely to be involved in controlling the tissue-specific expression of muscle-specific actin genes or regulating the level of cytoskeletal actins? As I will show later, the answer to the second part of this question may vary depending on the particular phenomenon being considered (e.g. cell detachment and reattachment to surfaces, heatshock, etc.), although for most of such cases no firm conclusion can be reached as to the site of control. The first part of the question relates to differentiation in general, and examples of other tissue-specific abundant proteins will be used in the argument. I will consider each stage of gene expression successively (in reverse order), and try to evaluate the importance of each in the controlled expression of the various actin genes.

There is little information available on the stability of actin proteins, and there is no evidence that the regulation of stability of

any other protein is responsible for its appearance or non-appearance in differentiated cells. It is possible that stability may have an effect on the regulation of protein levels, however, as studies have shown that the stability of liver proteins may change (Schimke and Doyle, 1970).

Comparisons of cloned vertebrate actin gene sequences with protein amino acid sequences, have revealed that post-translational modification does take place in actins, to replace the first one or two amino acids with an acetyl blocking group (Gaetjens and Barany, 1966; Alving and Laky, 1966; Vandekerckhove and Weber, 1978a; Nudel et al., 1983; Zakut et al., 1982; Hamada et al., 1982; Gunning et al., 1983). It is possible to envisage such a step being controlled, but there is no evidence for this. Post-translational modification also occurs in Drosophila, although the facts are, at present, somewhat confusing. The Drosophila actins can be separated into 3 types by isoelectric focussing. Type III is an unstable species except in adult thorax (Horovitch et al., 1979). Fyrberg et al. (1983) translated poly A(+) message from adult thorax in reticulocyte lysate and were unable to detect this actin type, suggesting that a post-translational modification is involved in its formation. This might appear to be an example of tissue-specific post-translational modification. This may be true; however, it seems that there are two Drosophila actin genes whose messages are found only in adult thorax and leg muscle (suggesting pre-translational control to be operating). The unstable type III actin found in non-thorax tissues may possibly be a different gene product to the type III protein found in the thorax.

The tissue-specific expression of vertebrate muscle actin (or myosin heavy and light chains) is not a result of translational control. This conclusion is partly based on experiments using myoblasts from

developing skeletal muscles, which proliferate in culture and then enter a phase of cell fusion and formation of multinucleated muscle fibres. A close correlation has been found between the levels of the muscle-specific contractile proteins, and their mRNAs. This has been found by in vitro translation studies (Paterson et al., 1974; Daubas et al., 1981) on the RNA extracted from various stages. It has also been found using cloned probes in hybridization experiments with primary chick cultures (Schwartz and Rothblum, 1981), in rat L8 cultures (Shani et al., 1981), mouse cultures (Caravatti et al., 1982) and primary quail cultures (Hastings and Emerson, 1982). Ballantine et al. (1979) showed that alpha-actin (muscle-specific) first appeared at gastrulation, during the development of Xenopus embryos, and in vitro translation studies showed that the mRNA for this protein also first became detectable at around this time. The changes in beta- and gamma-actin levels during the formation of chicken gizzard (smooth muscle) have been found, by in vitro translation studies, to correlate with the changes in mRNA levels (Saborio et al., 1979).

If translational control were to play a major role in tissue differentiation, this might mean that the ability to translate particular kinds of mRNAs would only appear during the differentiation of particular types of cell. Sturgess et al. (1980) injected mRNA from tadpole tail into Xenopus eggs and found that alpha-actin was synthesised. A similar result has been obtained using mRNAs specific to other types of differentiated cell. Thus, globin is synthesized on injection into oocytes or eggs, and is still made in swimming tadpoles; even in regions differentiated into non-erythroid cells (Woodland et al., 1974; Gurdon et al., 1974).

Translational control may not play a role in the tissue-specific

expression of muscle actin, but seems to be used by some cells to control the rate of synthesis of the cytoskeletal actins. Cytoskeletal actins are among the most rapidly made proteins of oocytes, blastulae and later embryonic stages of Xenopus laevis, but the rate of synthesis is thought to be lower in the egg or cleaving embryo (Ballantine et al., 1979; Sturgess et al., 1980). However, by extracting RNA and translating it in vitro, it was shown that the cytoskeletal actin mRNA is present in equal amounts at each of these stages; demonstrating that translational regulation is acting here.

Another example of a change in translational utilization takes place during muscle cell differentiation: Caravatti et al. (1982) found a sharp decrease in cytoskeletal actin message levels after the fusion of myogenic cells, but could not detect any change in the amount of beta- and gamma-actin synthesized at this time, or up to 72 hours later. They have suggested that there may be an increase in actin mRNA utilization by the translational machinery of these cells.

Translational control may also play a role in controlling the rate of cytoskeletal actin synthesis in response to changes in cell configuration. Farmer et al. (1983) found that when a 3T3 line of anchorage-dependent fibroblasts were suspended, total protein synthesis was strongly inhibited. Actin synthesis was preferentially inhibited, decreasing from 12% of total protein synthesis to 6% in suspended cells. The level of actin mRNA, as determined by Northern blot hybridization with an actin cDNA clone, fell only slightly on suspension of the cells, so it was suggested that the phenomenon was due to some form of translational control. Farmer et al. suggested that the mRNAs of the suspended cells were rendered untranslatable by some kind of modification of the RNAs. It is also possible, however, that the

reduction in actin synthesis occurs when there is an overall decrease in the efficiency of initiation of protein synthesis by ribosomes, and that actin messages merely lose out in the competition with other mRNAs. This type of phenomenon has been described by Lodish (1974) and Alton and Lodish (1977), and is a different form of control from that where specific messages are bound to proteins that inhibit translation, as probably occurs, for example, in Spisula development (Rosenthal et al., 1980). The latter process could well account for the control of cytoskeletal actin in early Xenopus development as well, since the decrease of the actin synthesis occurs at the same time as a 50-fold increase in nucleosomal core histone synthesis and a 2 to 5 fold increase in overall protein synthetic rate (Woodland et al., 1980; Ballantine et al., 1979).

Although translational control appears to have an effect on actin expression, it is not responsible for the tissue-specific expression of muscle actins. Is gene expression in differentiated tissues controlled by selectively stabilizing certain messages while degrading others? The accumulation of alpha-actin mRNA at myogenesis (Caravatti et al., 1982; Schwartz and Rothblum, 1981; Shani et al., 1981) could be explained by an increase in its stability as well as by an increase in transcription or primary transcript processing.

Although there is no evidence for such a form of control with muscle-specific actins, there is a substantial amount supporting its influence on other tissue-specific abundant protein levels. Thus, when prolactin is given to breast tissue explants, there is a 20-fold increase in casein mRNA halflife (Guyette et al., 1979). The halflives of vitellogenin and VLDL mRNAs, in rooster liver, drop from 24 hours to 3 hours on withdrawal of oestrogen (Wiskocil et al., 1980). There are

also two established cases where preservation of a specific group of mRNAs occurs at the same time as the destruction of a large fraction of the total cellular RNA. In mammalian reticulocyte differentiation, the final stage, where 90% of the synthesized protein is globin, appears to be reached with the conservation of globin (and a few other erythrocytic mRNAs) and the specific destruction of all other mRNAs (Aviv et al., 1976; Bastos et al., 1977; Volloch et al., 1981). Dictyostelium amoebae, on aggregation, synthesize an extra set of mRNAs in addition to the normal set. The new set of RNA species is specifically destroyed if the cells are artificially disaggregated (Margiarotti et al., 1982).

The stability of mRNA species does seem to be an important factor in development and cannot be ignored. It is conceivable that the stability of various actin messages may be altered in some situations. It is interesting, in this respect, that there is an immediate decrease in cytoskeletal actin mRNA content on the fusion of myogenic cells, concomitant with the increase in muscle-specific protein RNAs (Caravatti et al., 1982; Shani et al., 1981). However, evidence is presented below that the appearance of muscle-specific actin mRNAs at myogenesis is the result of transcriptional activation, just as the actual tissue-specific appearance of globin RNA in erythroid tissues is a result, primarily, of transcriptional activation (see below).

Before moving on to transcriptional control, the possibility of control within the nucleus should be considered. There is little evidence that nuclear transport is important, but there is evidence, at least in certain systems, that control may be exerted during processing of the primary transcript. Nuclear turnover of RNA transcripts in HeLa cells has been found to be substantial (Soeiro et al., 1968), and nuclear RNA complexity has been found to be 4 to 10 times greater than

the complexity of mRNA (Lewin, 1980d). This is perhaps greater than the difference expected on the basis of intron loss during processing, suggesting that complete transcripts fail to appear in the cytoplasm. Analysis of nuclear RNA in cultured CHO cells has shown that the primary transcripts can be divided into two classes: those that give rise to mRNA sequences and those that do not. Thus, greater than 90% of non-histone mRNAs entering polysomes contain polyA, while polyA is added in the nucleus to about one quarter of the total primary transcripts (Harpold et al., 1979; Salditt-Giorgieff, 1981). That the polyA(-) species were different from those that were polyadenylated was shown by hybridizing cDNA clones made from polyA(+) mRNA to nuclear RNA. They hybridized mainly to the polyA(+) fraction (Harpold et al., 1979). To show that there is tissue-specific control at the nuclear level, however, it would be necessary to demonstrate that some of these unexpressed primary transcripts can appear in the polysomes of a different cell type. The only positive evidence for this type of control has been found in sea urchins where Wold et al. (1978) found, for example, that while 85% of blastula mRNA sequences were not present in the adult intestine mRNA population, virtually all were present in the intestine nucleus.

There is no evidence that this type of control is important for the tissue-specific abundant proteins of vertebrates. Instead, transcriptional control is believed to be exerted over such genes. The experiments performed to demonstrate this, have involved the isolation of nuclei from various tissues, followed by the synthesis of radioactively-labelled RNA by bound RNA polymerases for a short length of time (1 to 2 minutes). Such a protocol minimizes the chance that primary transcripts will not be seen due to degradation (as might occur

if they were destined not to be processed in that cell type). Such a protocol has been used by Groudine and Weintraub (1981) to demonstrate the synthesis of globin primary transcripts in red blood cell nuclei, but not in primitive erythroblasts, and also by Dermann et al. (1981) to show that liver-specific mRNAs were transcribed only in the liver, and not other tissues.

Are vertebrate muscle-specific actin genes also under transcriptional control? There is some indirect evidence that this is the case. Studies using a variety of cell types have shown that genes which are being expressed in a cell type are organized in chromatin in a conformation which renders them more susceptible to mild digestion with pancreatic DNAase I than genes which are not active in the same tissue (Weintraub and Groudine, 1976; Garel and Axel, 1976; Panet and Cedar, 1977; Wu et al., 1979; Storb et al, 1981). Thus, for example, the globin, but not the ovalbumin gene, is digested by DNAase I in avian erythrocytes (Weintraub and Groudine, 1976), but the ovalbumin gene and not the globin gene is preferentially digested in the chick oviduct (Garrel and Axel, 1976). This might seem to be a convenient method of showing that a gene is transcriptionally active at any given time. However, other experiments have shown that this change in chromatin structure may not be sufficient for transcription, since a number of apparently inactive genes have been reported to be DNAase I sensitive; for example, the chick adult beta-globin gene in primitive erythroblats, and the foetal gamma-globin genes of adult sheep bone marrow (Stalder et al., 1980; Young et al., 1978). Nevertheless, it appears that this change in DNA conformation is necessary before transcription may occur, as there have been no instances, so far, of genes which are apparently active in a tissue (in that their mRNA product is abundant there), but

are not DNAase I sensitive. This fact is important in the argument for transcriptional control over the skeletal muscle actin gene. Thus, Carmon et al. (1982) have found tht the genes for skeletal muscle actin and myosin light chain 2 only become DNAase I sensitive when rat myogenic cells fuse into multinucleated muscle fibres. This increase in sensitivity occurs close to the time that the mRNAs from these genes were shown to appear (Shani et al., 1981; Caravatti et al., 1982). The coincident increase in production of messages and the change in chromatin conformation is, I think, convincing evidence that transcriptional activation does occur during myogenesis. This conclusion would not go undisputed, however, as Schwartz and Rothblum (1981), although finding a greatly increased content of muscle-specific actin mRNA at myogenesis, were able to detect a significant amount (150 molecules/cell) of this message in replicating prefusion myoblasts. These workers however, were not able to rule out the possibility that such messages were being produced from a few precociously differentiating cells within the cultures.

Even if transcriptional control is responsible for the non-expression of muscle-specific actins in non-muscle cells, once activated, other factors may increase the expression of such genes. Zimmer and Schwartz (1982) have produced evidence that the chicken skeletal muscle gene is amplified approximately 90-fold during the fusion of myogenic cells, both in cell culture, and in intact embryonic muscle. The amplification takes the form of two new EcoRI bands in Southern blots of genomic DNA, the first of which appears prior to the expression of alpha-actin mRNA, and has an actin sequence dosage of about 8 to 10-fold greater than the alpha-actin structural gene. The second species appears later in myogenesis and has a gene dosage of about 85-fold greater than the

alpha-actin gene, during the maximal induction of mRNA. These amplified DNA species both disappear in concert with the drop in alpha-actin mRNA levels in late postfusion myotubes. The rough correlation between the appearance and disappearance of the amplified species and the change in levels of the alpha-actin mRNA would suggest that these phenomena are linked, although there is, as yet, no evidence that the amplified DNA is transcribed (and these results need confirmation).

Only one other type of protein-coding gene has been shown to be amplified during cellular differentiation: that of the chorion proteins in Drosophila (Spradling and Mahowald, 1980; Thireos et al., 1982; Spradling, 1982). Unlike the chorion genes, which are endoreplicated in the macrochromosomes (probably the result of multiple DNA replication initiation events at a replication origin within the cluster of 4 chorion genes, giving rise to a set of nested replication forks), and where the amplified DNA eventually disappears via cell death when the fly pupates, the amplified actin genes disappear very rapidly (within 24 hours of their appearance). Because of this, Zimmer and Schwartz (1982) have suggested that the actin genes are amplified extrachromosomally (probably by a rolling circle type of mechanism; the mechanism by which the first extrachromosomal copy is produced from the chromosomal gene is unknown), in the same way as the ribosomal RNA genes in Xenopus oocytes (Brown and Dawid, 1968).

Thus, the tissue-specific expression of skeletal muscle actin genes is probably under transcriptional control, although gene amplification may increase the amount of mRNA produced. It would seem likely that the control of expression of the other muscle-specific actin genes is also transcriptional. Does the same apply to cytoskeletal actins? The question here is a little different, as the latter genes are expressed

to some extent in most cell types. The ratio of the beta- and gamma-cytoskeletal actins varies from one non-muscle tissue to the next (Garrels and Gibson, 1976). There has been no work designed to find out what causes this phenomenon, and it seems, at present, that post-transcriptional mechanisms are just as likely to be responsible as those acting at transcription. There are also some situations where the total levels of cytoskeletal actin mRNA change. I have already explained that some of these are likely to be the result of translational control. In some cases, however, other levels of control may be responsible. Thus, on re-attachment of anchorage-dependent mouse 3T3 fibroblasts, there is a dramatic increase in actin synthesis (more so than other proteins) which coincides with an increase in actin mRNA production (Farmer et al., 1983). On fusion of cultured myogenic cells, during myogenesis, there is an immediate decrease in the levels of the cytoskeletal actin mRNAs (Caravatti et al. 1982). The former phenomenon may be the result of control at any pre-translational level. The latter appears to be the result of specific RNA degradation, although the lack of quantitative information given on the speed of this decrease and the prefusion halflives of these messages leaves this conclusion uncertain; transcriptional control of cytoskeletal actin genes at myogenesis cannot be ruled out yet. The rat beta-actin gene is DNAase I sensitive in both differentiated and undifferentiated myogenic cultures (Carmon et al., 1982). DNAase I sensitivity, however, appears to be unrelated to the rate of transcription of a gene (Garel et al., 1977) and, therefore, this cannot be taken to show that the cytoskeletal actin genes of muscle cells are not under transcriptional control (especially as several DNAase I-sensitive genes have been found which do not appear to be transcribed at all; Stalder et al., 1980; Young et al., 1978).

There is little evidence indicating the level at which the expression of most cytoskeletal actin genes is controlled. However, evidence does exist that the transcriptional control of cytoskeletal actins can take place in Drosophila. Findly and Pederson (1981) found that the normally high level of 5 minute-pulsed actin nuclear RNA transcripts decreased within 10 minutes of subjecting Drosophila cells to a 35°C heat-shock, and that it returned to normal within 2 hours of returning the cells to 25°C. It seems likely that this is a demonstration of a rapid and flexible transcriptional control. It is, of course, possible to argue that this is a special situation and that control at other levels accounts for the other changes in cytoskeletal actin synthesis that I have described.

Summary of main conclusions

It seems likely that muscle-specific actin genes are transcriptionally inactive in non-muscle cells. In muscle cells, however, once these genes are being transcribed, further factors (such as gene amplification and changes in mRNA utilization) may effect the final muscle actin levels. There is no evidence relating to the tissue-specific regulation of beta/gamma cytoskeletal actin ratios. Other changes in the rate of synthesis of cytoskeletal actins may be brought about through control at the translational, transcriptional or message stability levels.

1.3.2 Sequences involved in the control of gene expression and the possible role of gene clustering

For the precise regulation of gene expression, it seems necessary

There is little evidence indicating the level at which the expression of most cytoskeletal actin genes is controlled. However, evidence does exist that the transcriptional control of cytoskeletal actins can take place in Drosophila. Findly and Pederson (1981) found that the normally high level of 5 minute-pulsed actin nuclear RNA transcripts decreased within 10 minutes of subjecting Drosophila cells to a 35°C heat-shock, and that it returned to normal within 2 hours of returning the cells to 25°C. It seems likely that this is a demonstration of a rapid and flexible transcriptional control. It is, of course, possible to argue that this is a special situation and that control at other levels accounts for the other changes in cytoskeletal actin synthesis that I have described.

Summary of main conclusions

It seems likely that muscle-specific actin genes are transcriptionally inactive in non-muscle cells. In muscle cells, however, once these genes are being transcribed, further factors (such as gene amplification and changes in mRNA utilization) may effect the final muscle actin levels. There is no evidence relating to the tissue-specific regulation of beta/gamma cytoskeletal actin ratios. Other changes in the rate of synthesis of cytoskeletal actins may be brought about through control at the translational, transcriptional or message stability levels.

1.3.2 Sequences involved in the control of gene expression and the possible role of gene clustering

For the precise regulation of gene expression, it seems necessary

that the regulatory molecules concerned should recognize specific DNA (or RNA) sequences. The alternative would require gene specific differences in chromatin conformation that are replicated independently of DNA sequence; a seemingly unlikely occurrence. Differences in chromatin conformation do exist in the eukaryotic genome (see below), but many of these are the results of, not the cause of, regulatory events. One type of such phenomena, DNAase I hypersensitive sites (Elgin, 1981), can apparently be transmitted from one cell generation to the next (Groudine and Weintraub, 1982), but it is thought that such structures are DNA sequence dependent (Weintraub, 1983) and are, in any case, often only found associated with active genes (Elgin, 1981; Weisbrod, 1982; Igo-Kemenes et al., 1982).

The models that have been proposed for gene regulation in eukaryotes all represent a reformulation of the concept of the bacterial operon proposed by Jacob and Monod (1961). The most detailed model was proposed by Britten and Davidson (1969). This involved the control of the protein-producing 'structural genes' by a closely-linked 'receptor' element. They proposed that the activation of the structural genes requires the interaction of an activator protein, or RNA, with the receptor sequence. The genes coding for the activator molecules, termed 'integrator' genes, could in turn be activated by a nearby 'sensor' element, which might respond, for example, to an external stimulus. In order for genes to be transcribed together, in 'batteries', at the appropriate time in the appropriate tissue, several structural genes might be associated with the same type of receptor sequence or, alternatively, a single sensor element might activate several different integrator genes; the activator molecules produced reacting with

different receptor sequences. There are many possible variations on such a model, and Davidson and Britten (1979) later proposed that the interaction of activator with receptor elements might take place post-transcriptionally (activator RNA interacting with intron-containing structural gene primary transcripts; in fact, it is possible to extend the model to postnuclear control mechanisms, involving the interaction of proteins with specific sequences in mature cytoplasmic messages). The theory is useful in that it predicts that, where genes are coordinately expressed, it might be possible to find a common sequence associated with each of the genes. It might, therefore, be possible to find such a sequence associated with the alpha-actin genes and other muscle-specific protein genes. Of course, as explained by the theory, such genes may possess different 'receptor' elements, but it may be possible to identify these sequences by examining one gene type from several organisms, in the hope that sequences with an important controlling function may be more conserved than the surrounding DNA. As this was one of the starting aims of my project, it is worth examining the evidence that this type of control mechanism might exist.

The Britten and Davidson hypothesis was initially attractive because of its potential ability to explain the function of the large amount of repeated sequences that make up a considerable proportion of plant and animal genomes (Lewin, 1980b). For example, DNA reannealing studies on Xenopus laevis, a frog closely related to the species studied in this project, have shown 10% of its genome to be composed of sequences repeated 110 times, 31% of sequences repeated 2000 times, and 6% of sequences repeated 300,000 times; 54% being single copy DNA (Davidson et al., 1973). The major part of this repetitive DNA does not code for protein; so, what does it do? A high fraction of repetitive

sequences in most eukaryotes are present as small blocks of sequence interspersed with non-repetitive sequences (Lewin, 1980c). Repetitive sequences have been found, at least in some mammalian and echinoderm species, in a large fraction of nuclear transcription units (Davidson and Posakony, 1982). Furthermore, the representation of sea urchin repetitive sequences in nuclear RNA has been found to be tissue-specific (Scheller et al., 1978). Several sets of Dictyostelium polyA(+) RNAs have been reported which are developmentally regulated; a set being defined by the particular repeat sequences covalently attached to its members (Kimmel and Firtel, 1979; Zuker and Lodish, 1981). Repetitive sequences have been found around, and sometimes even within, many cloned structural genes (Davidson et al., 1975; Cochet, 1979; Bell et al., 1980a; Proudfoot et al., 1980; Shen and Maniatis, 1980a; Page et al., 1981; Pearson et al., 1981; Ryffel et al., 1981; Young et al., 1982; Schibler et al., 1982). Consequently, much attention has been paid to the possibility that such repetitive sequences might be the control elements suggested by the hypothesis (e.g. the members of one repeat family might constitute the receptor elements adjacent to a set of genes which are coordinately expressed). No evidence, however, has yet been presented that identifies such sequences as being necessary for the control of gene expression. The 'evidence' described above can be explained in different ways. Thus, the proximity of repetitive sequences near and within transcription units may serve no useful function, yet this arrangement might, for example, lead to such sequences being transcribed (Davidson and Posakony, 1982).

The identity of repetitive DNA as control elements remains unproven. Attention, however, has now focussed on the possibility that such DNA may have no function at all and may be merely 'selfish' in

nature (Orgel and Crick, 1980; Doolittle and Sapienza, 1980; Doolittle, 1982), its repeated nature merely being a result of its ability to increase its copy number in the genome, and for the copies to be integrated elsewhere in the genome by transposition. Thus the transposable elements of yeast and Drosophila make up most or all of the middle-repetitive DNA components of these organisms (Finnegan et al., 1982; Doolittle, 1982); physically similar DNAs being found in other eukaryotes (Adams et al., 1980; Anderson et al., 1981). Such sequences are several kilobases in length, however; longer than the interspersed middle repetitive elements in species like Xenopus (300 to 500 basepairs). The best studied of such sequences are the members of the AluI family, which represent more than half of all human middle-repetitive DNA, with relatives in other mammals (Jelinek et al., 1980; Jelinek and Schmidt, 1982). The transcription of many of these elements by RNA polymerase III (Elder et al., 1981), together with the observation that most of the elements are flanked by direct repeats 7 to 20 nucleotides long (specific to each element), suggests that many of these repeats are formed via an RNA intermediate which is possibly reverse transcribed, and the resulting DNA integrated into the genome (Sharp, 1983). AluI-like sequences have been found to exist as closed circular extrachromosomal DNA, just like the copia transposable elements of Drosophila (Calabretta et al., 1982; Flavell and Ish-Horowicz, 1981).

Although these sequences may be replicated and inserted elsewhere in the genome this does not necessarily mean they have no function. However, it is certainly not necessary to propose that they all have a function as control elements.

If such sequences are not necessarily all control elements, then it would seem that some sequences must be. But are the control elements

necessarily located adjacent to the structural genes, as proposed by the hypothesis? Might they not be located many tens of kilobases away, making their identification difficult? It is possible to envisage the transcription of genes resulting from, say, a change in conformation of a large chromatin domain, in which case the controlling element need not be positioned adjacent to the structural gene. Indeed, it is possible to envisage many genes located together in one such region so that their coordinate expression might not require control elements for each individual gene. That large sections of chromatin might be controlled in this way, was suggested by work on Drosophila, where transposition of a DNA segment to a different part of the chromosome affects the level of gene expression (Spofford, 1976).

Although there is no evidence to suggest that their transcription is controlled in this way, there are several gene families whose members are expressed at the same time, and are clustered in the genome. These include the histone genes (Hentschel and Birnstiel, 1981), the heat-shock genes in Drosophila (Mirault et al., 1979; Artavanis-Tsakonis et al., 1979), the ribosomal RNA genes of most eukaryotes (Long and Dawid, 1980), the silkworm and Drosophila chorion genes (Eickbush and Kafatos, 1982; Jones and Kafatos, 1980; Spradling, 1981), the chick ovalbumin family (Royal et al., 1979) and the antennapedia gene complex related to the control of segmental pattern of Drosophila (Lewis et al., 1980). There is, however, little real evidence that the clustering of these genes has any functional role. It is possible that the clustering of related genes may only reflect their evolutionary origins. They presumably arose by gene duplication and, for some unknown reason, the products have not yet been separated. There is a growing body of evidence which suggests that, at least for some genes, the sequences

responsible for controlling their expression are located very close to the structural genes.

Recently, Chao et al. (1983) have shown that sequences important in the induction of mouse beta-globin genes exist within the 1.2 kilobases of 5' flanking DNA. Similar results were obtained with a human beta-globin gene. Thus, for example, a human beta-globin gene was introduced into the genome of murine erythroleukemia cells. The latter cells are arrested in erythroid maturation, but can be induced to follow an erythroid developmental programme with chemical inducers. Such inducers were found to increase the level of transcription from the exogenous globin gene, but not of a non-globin exogenous gene. Similarly, cloned chick delta-crystallin genes have been injected into the nuclei of various mouse tissues, and were found to be expressed only in lens epithelial cells; the correct tissue for this tissue-specific gene (Kondoh et al., 1983). Experiments have shown that the sequence necessary for the inducibility of the metallothionein gene by cadmium are located within 148 base-pairs of the transcriptional start site (Mayo and Palmiter, 1982). Other instances where a cloned gene is inducible when introduced into cells include the interferons (Mantei and Weissman, 1982; Ohno and Taniguchi, 1982; Pitha et al., 1982), and the heat shock proteins (Corces et al., 1981; Burke and Ish-Horowicz, 1982; Pelham, 1982).

If such control sequences exist adjacent to genes, can they be identified? One approach to this problem has been to change regions of cloned genes 5' to the start of transcription, and test the effect of such deletions, or mutations, on transcription. Several sequences specific to one type of gene, and which are necessary for maximal transcription have been reported, including the HSV TK gene (McKnight

and Kingsbury, 1982), the sea urchin H2A gene (Grosschedl and Birnstiel, 1980, 1982), the SV40 early region (Benoist and Chambon, 1981; Fromm and Berg, 1982), and the beta-globin gene (Dierks et al., 1983). The latter sequence was also found to be conserved in all but one of the beta-globin genes of mouse, human, goat and rabbit (Efstratiadis, 1980) as well as present, but in a slightly different position, in the chicken (McGhee et al., 1981). The physiological significance of such sequences has not been proven, however. For example, in the case of the beta-globin experiments, the genes were introduced into non-erythroid cells, where the endogenous globin genes would not be expected to be expressed. It has been proposed that the SV40 and sea urchin H2A sequences are 'enhancer' sequences. The latter are promoters (i.e. their presence increases the transcription of genes), capable of acting over long distances (reviewed by Khoury and Gruss, 1983), and in either direction.

Other sequences have been found which do seem to function in a way which can be related to a possible physiological role. These have recently been reviewed by Davidson et al. (1983) who suggest that their small size (9 to 15 bases) and relatively high degree of mismatch between members of one family may have made them invisible to the traditional, DNA reannealing, methods of identifying repeated sequences. Thus, four yeast genes for amino acid synthesis, that are activated when the organism is starved for amino acids, possess sequences in their 5' flanking regions which have a high percentage homology with a 9 base-pair consensus sequence. In transformation experiments, deletion mutants of the gene, which lack the repeated sequence element, cannot be induced (Donahue et al, 1983). A short repeat is also shared by the Drosophila heat-shock genes, and Pelham (1982) has shown that deletions in the region of this sequence destroy the transcriptional response of

the hsp70 gene to heat-shock. Furthermore, when the hsp70 5' flanking sequence was ligated to the herpes thymidine kinase gene, the latter became inducible by heat-shock, and this effect was not seen if the hsp70 DNA induced lacked the shared repeated sequence element. Other shared repeats have been found in the conalbumin, lysozyme, and ovalbumin genes of chicken (steroid-induced egg white proteins) (Grez et al., 1981), the late silkmoth chorion genes and the middle silkmoth chorion genes (Jones and Kafatos, 1980). There has been no functional assay for the repeated element in these last 3 groups, however. The putative control elements in all of these groups are different, as is necessary if such sequences are to control the expression of these groups in response to different stimuli. This difference is also necessary if such elements are to be distinguished from sequences such as the 'TATA box' and the 'CAAT box' which appear in many gene types; perhaps functioning as promoter elements (Breathnach and Chambon, 1981). Also, where no functional test has been made for such elements, it is important that the non-coding DNA sequence surrounding the elements have much less homology among the members of a group. This is important, for example, in the case of the chorion genes, where the structural genes themselves are ancestrally related, and are located in a cluster so that sequence homology in long stretches of non-coding sequences might merely indicate the occurrence of concerted evolution, as has been documented in a number of gene families (Dover and Coen, 1981; Jeffreys, 1981).

So, there are several instances of small sequences adjacent to genes playing a role in the expression of those genes. In a few cases (heat-shock genes, chorion genes) these sequences have been found in clustered genes. Does this mean that clustering definitely has no function?

Not necessarily; there may be other levels of control involved. For example, various members of the beta-globin cluster of mammals are not expressed at the same time during the life of the animal, yet the structure of the 50 kilobase cluster seems to be important for their expression. The arrangement of the cluster has been found to be indistinguishable in man, gorilla and baboon, indicating that it has been faithfully preserved for 20 to 40 million years (Barrie et al., 1981). Moreover, the intergenic sequences are much more highly conserved than if they were functionless (Barrie et al., 1981), and mutations in one part of the cluster have been found to have a significant effect on the expression of distant genes expressed at a later stage (Van der Ploeg et al., 1980). Jeffreys (1982) has suggested that the preservation of the overall structure may result from this region being controlled as one chromatin domain. I have already described how beta-globin genes have their own control elements adjacent to them. This does not rule out a preliminary regulatory event, however, which, for example, might activate the beta-globin domain only in erythroid cells. It is possible that chromatin conformation may have to change so that regulatory molecules can interact with the DNA. This might explain why the deletion of the gamma and delta genes influences beta-globin expression (Van der Ploeg et al., 1980).

That several separate events may be involved in the activation of eukaryotic genes, was suggested by experiments by Burch and Weintraub (1983), which showed the appearance of a specific set of DNAase I hypersensitive sites (Elgin, 1981), on committment of the chicken liver vitellogenin gene to vitellogenin synthesis during embryogenesis. Only the liver genes underwent this chromatin conformational change and only these genes became inducible by oestrogen. (DNAase I hypersensitive

sites have been shown to be propagated from one cell generation to the next; Groudine and Weintraub, 1982.) The other changes that appear to be necessary, but not sufficient for gene activation, are, of course, also suggestive that more than one regulatory step occurs. These include the general sensitivity to DNAase I (Weintraub and Groudine, 1976; Stalder et al., 1980; Young et al., 1978) and hypomethylation (Razin and Riggs, 1980; Chisholm, 1982) (although not all active genes are hypomethylated; Gerber-Huber et al., 1983). Such findings suggest that several levels of regulation, and hence several regulatory sequences, may exist.

Is there linkage of genes coding for coordinately expressed actins and other contractile proteins?

One of the main questions posed at the outset of this project was whether the expression of actins was also controlled by adjacent regulatory sequences. It is perhaps worth examining the evidence obtained from other species which addresses the question of whether clustering is important in the coordinate expression of actin and other contractile protein genes.

Where coordinately expressed genes can be shown to be unlinked, then it seems that trans-acting mechanisms must be responsible for that control. This is the case with several groups of actin genes which are coordinately expressed. Drosophila actin genes are, in most tissues, expressed together in pairs (Fyrberg et al., 1983). Vertebrate beta- and gamma-actins are expressed in the same cell (Bravo et al., 1981), and the two striated muscle actin mRNAs accumulate together in foetal skeletal muscle and during the differentiation of a cloned mouse myoblast cell line (Minty et al., 1982). In the corresponding genomes,

the six actin genes of Drosophila are dispersed at different chromosomal loci (Fyrberg et al., 1980), the beta- and gamma-cytoskeletal genes are found in different chromosomal fractions in the chick (Cleveland et al., 1981), and the two striated muscle genes do not cosegregate in recombinant inbred mouse lines and are, therefore, not genetically linked (Minty et al., 1983). These genes, then, are coordinated by a trans-acting system.

Are all actin genes unlinked? In several species, linked actin genes have been isolated on cloned DNA. Two linked genes were found in Dictyostelium (McKeown et al., 1978), but have since been found not to be expressed and are thought to be pseudogenes (McKeown and Firtel, 1982). Three of the four actin genes of the nematode C. elegans are clustered (Files and Hirsch, 1983). The times at which these genes are expressed have not yet been established, however. Several actin genes are believed to be linked at one locus in Physarum, although nothing is known of their expression (Schedl and Dove, 1982). The only other organism in which clustering has been observed is the sea urchin (Scheller et al., 1981; Schuler and Keller, 1981; Overbeek et al., 1981; Davidson et al., 1982), where 3 genes have been found on one 30 kilobase fragment, 2 genes on a 12 kb fragment (Scheller et al., 1981), and 2 others are also linked to each other (Davidson et al., 1982). The sea urchin actin genes can be classed into subgroups according to the homology of their 3' non-coding regions. Linkage occurs, both of genes in the same subgroup, and of genes in different subgroups (Davidson et al., 1982). Various members of each of the 3 subgroups have been shown to be transcribed. The transcripts from the different subgroup types become prevalent at different times during development. Thus the linkage of genes of different subgroups is possibly an example of the

linkage of two genes which are not coordinately expresed; a situation similar to the globin genes (Jeffreys, 1982). Caution must be applied here, however, as the individual genes present in this cluster have not yet been shown to be transcribed (Scheller et al., 1981; Davidson et al., 1982). The linkage of members of the same subgroup may only reflect the process by which they were formed: gene duplication (Jeffreys and Harris, 1982). It is conceivable, however, that these genes may not have separated because they are coordinately expressed; both possibly being situated in a functionally important domain.

Actin genes, then, are found both in clusters, and separately. There is little evidence to support the view that coordinately expressed actins must be clustered in order for that form of regulation to take place. What about the linkage of actin genes with other types of genes with which they are expressed?

There is much less information available on contractile protein genes other than actin. However, recent results show that three linked tropomyosin genes (two muscle, one non-muscle) map in the same chromosomal region, 88F, 2-5, on the right arm of chromosome 3 (Bautch et al., 1982), as the actin gene DmA1 (presumably a muscle actin), which is expressed in the thorax of Drosophila (Fyrberg et al., 1983). This mapping, however, was based on in situ hybridization, which cannot localize sequences to a region less than 50 to 100 kilobases in size, and so does not necessarily imply that these genes are immediately adjacent on the DNA. In addition to this, Bernstein et al. (1981) have reported that a Drosophila myosin heavy chain gene maps in a different region from any of the actins. Czosnek et al., (1982) have recently shown that the genes coding for the skeletal muscle-specific myosin heavy chain, myosin light chain 2 and actin proteins are all located on

separate chromosomes of the mouse.

Generally, these observations suggest that trans-acting mechanisms are responsible for the coordinate expression of contractile protein genes. This, and the discovery of sequences important in the activation of several types of gene, immediately 5' to such genes, suggests that similar sequences may be involved in the control of actin gene expression. It should also be remembered, however, that data discussed in Section 1.3.1 suggest that post-transcriptional control of some types of actin gene may sometimes occur. In such instances, control sequences might be located within the region of the gene that will give rise to the mature mRNA molecule.

1.4 Aims

The primary aim of this project was to obtain clones of *Xenopus borealis* actin genes in order to find out if regulatory sequences were present adjacent to these genes. The main genes of interest were those expressed in a tissue-specific manner: the muscle actin genes. Initially, the search for these sequences was to be made by sequencing the genes, finding out where transcription starts, and comparing the sequence in this region with sequences of coordinately expressed genes, or the same gene-type in another organism. It was also decided to sequence as much of the genes as possible (i.e. not just the 5' regions), to gain more knowledge of actin evolution.

These goals have not all been achieved. In particular, the transcriptional studies on a muscle-specific gene have not yet been made. However, the 5' portions of a cytoskeletal and a putative muscle-specific actin gene have been almost completely sequenced, yielding some interesting information of an evolutionary nature. The former gene has been transcriptionally mapped by primer-extension, and several conserved sequences, which possibly may be involved in the control of gene expression, have been discovered. The existence of quite a large amount of sequence upstream from the putative muscle-specific actin gene may aid the future transcriptional mapping of this gene.

2 MATERIALS AND METHODS

2.1 Materials

Restriction enzymes were obtained from Boehringer-Mannheim (London) Ltd., Bethesda Research Laboratory (U.K.), New England Biolabs, and Miles Labs Ltd. *E. coli* DNA polymerase I and *E. coli* DNA polymerase Klenow fragment was obtained from Boehringer Mannheim. T4 DNA polymerase I, bacterial alkaline phosphatase, T4 DNA kinase, and T4 DNA ligase were all obtained from Bethesda Research Laboratory.

All radioisotopes were obtained from Amersham International Ltd. at the following specific activities: alpha-^{32}P-dNTP at 2000-3000 Ci/mM at 10 mCi/ml in aqueous solution, gamma-^{32}P-dNTP at >5000 Ci/mM at 10 mCi/ml in aqueous solution, and ^{35}S-methionine at 1200 Ci/mM at 12 mCi/ml in 20 mM potassium acetate containing 0.1% mercaptoethanol.

X-ray film was obtained from Fuji Photo Film Co. Ltd. and Eastman Kodak Company. Nitrocellulose sheets were obtained from Schleicher and Schull (W. Germany). Type II Agarose for electrophoresis was bought from Sigma Chemical Company. Acrylamide for sequencing gels was obtained from BDH Chemicals Ltd. (Anala R grade). Materials for bacteriological media were obtained from Difco Laboratories and Oxoid Ltd. Other chemicals were obtained from BDH Chemicals, Sigma, and Fisons Scientific Apparatus.

Xenopus laevis frogs were obtained from the South African Snake Farm, Fishhoek, South Africa, while *Xenopus borealis* frogs (originally found in Kenya) were laboratory bred. *X. laevis* and *X. borealis* genomic DNA, used for experiments in section 7, were generous gifts from Dr. Phil Turner, as was the *X. borealis* ovary RNA used for experiments in Sections 3 and 5. The *X. laevis* kidney, ovary, and XTC cell RNAs used

in Section 5.3 were kindly supplied by Ms. Elizabeth Ballantine.

2.2 Some Commonly Used Solutions

These will be frequently referred to throughout this section and are listed here for convenience.

T.E.: 1 mM EDTA, 10 mM Tris/ HCl pH 7.4

1 x SSC: 0.15 M NaCl, 0.015 M sodium citrate

L Broth: 10 g NaCl, 10 g Bactotryptone, 5 g Yeast Extract in 1 litre

2 x TY: 5 g NaCl, 16 g Bactotryptone, 10 g Yeast Extract in 1 litre

TBE buffer: 10.9 g Tris base, 5.5 g boric acid, 0.95 g EDTA in 1 litre, pH 8.3

2.3 Preparation of Plasmid DNA

This was the 'cleared lysate' method adapted from Clewell and Helinski (1970). A 10 ml inoculum of plasmid-containing *E. coli* was grown overnight in L Broth supplemented with antibiotic (e.g. 0.1 mg/ml ampicillin for the pBR325 recombinants). This was added to 1 litre of L. broth (+ antibiotic) in a 2 litre flask and shaken vigorously at 37°C, until the optical density at 590 nm was about 0.6 to 0.7. Solid chloramphenicol was then added to 0.1 mg/ml, and the flasks shaken overnight at 37°C. The bacteria were harvested by centrifugation at 10,000 rpm in an MSE 6 x 250 ml rotor, for 15 minutes. The centrifuge pots were drained thoroughly, and then the bacteria were resuspended in precooled 25% sucrose, 50 mM Tris/HCl pH 8.0, and kept on ice. For each ml of this suspension, 0.2 mls of a freshly made lysozyme solution were added (20 mg/ml in 10 mM Tris/HCl pH 8.0). This was left on ice for 5

minutes before 0.4 mls of 25 mM EDTA/Tris/HCl pH 8.0 was added. This was mixed gently and, after another 5 minutes on ice, 1.6 mls of Lysis Mix (0.1% Triton X-100, 0.0625 M EDTA, 0.05 M Tris/HCl pH 8.0) were added and, again, mixed gently. This was left on ice for 20 minutes, inverting every 5 minutes. The mix was then centrifuged for 1 hour in an MSE 8 x 50 rotor at 18,000 rpm, at 4°C. The supernatant, containing the plasmid DNA, was carefully decanted off.

The plasid DNA was purified by banding on a CsCl equilibrium gradient. 26 g of CsCl was added to the supernatant of the above centrifugation, the volume of which had been made up to 25 ml with 0.1 x SSC (i.e. final CsCl concentration = 0.91 g/ml). 1.5 ml of 10 mg/ml ethidium bromide was then added to this, in a 30 ml, self-sealing, Beckmann centrifuge tube. The tubes were centrifuged for 16 hours at 50,000 rpm, 23°C, in a Beckmann VTi50 vertical rotor. The resulting DNA bands were visualized with ultra violet light; the upper band being genomic DNA, the lower being the required plasmid in supercoiled form. The latter was removed with a metal tube attached to a syringe, and the ethidium bromide removed by extraction with isoamyl alcohol (previously equilibrated with CsCl-saturated water). The plasmid preparation was then dialyzed overnight against 1 mM EDTA, 10 mM Tris/ HCl pH 7.4 (TE), at 4°C. To the resulting DNA solution, was added a tenth volume of 3 M sodium acetate, and 2 volumes of ethanol. After precipitation, overnight at -20°C, the DNA was pelleted by centrifugation (e.g. 10,000 rpm in an MSE 8 x 50 rotor, at 4°C, for 30 minutes), rinsed with 75% ethanol and vacuum dried for a short period of time. DNA was usually then redissolved in TE, at 3 to 4 mg/ml, and stored at 4°C. Stocks of important plasmids were kept as a pellet, under ethanol, at -20°C.

2.4 DNA Electrophoresis on Agarose Flatbed Gels

100 ml, 15 x 20 cm agarose slab gels were made up, and run, in 1 x TAE buffer (4.84 g Tris base, 2.72 g sodium acetate, 0.74 g EDTA in 1 litre, adjusted to pH 8.3 with glacial acetic acid) (Sharpe et al., 1973). The percentage of the gels was varied according to the size of the DNA fragments to be separated. Gels of DNA samples, to be visualized by ethidium bromide fluorescence, were made up containing 0.001 mg/ml ethidium bromide; as was the running buffer. Gels of radioactive DNA samples were usually dried down onto DEAE-cellulose paper. Samples were diluted with 1/5 volume of loading buffer (20 mM Tris pH 8.0, 100 mM EDTA pH 8.0, 50% glycerol, orange G as a dye) before running. Electrophoresis was usually at 20 volts overnight, or at 120 volts for several hours.

2.5 Extraction of DNA from Agarose Gels

Many methods of extracting DNA from agarose gels were tried. Several methods usually gave reasonable (at least 50%) yields. One of these simply involved the placing of the excised agarose gel fragment into a large (1.5 ml) Eppendorf tube, followed by its partial disaggregation with a blunt instrument. 0.3 ml of TE was then added and the DNA allowed to diffuse out, overnight at 4°C. This was then spun through glass wool, to remove the agarose, extracted with phenol, and precipitated with 1/10th volume of 3 M sodium acetate, 2 volumes of ethanol, in dry ice for 10 minutes. The DNA was then pelleted by centrifugation for 10 minutes in an Eppendorf centrifuge, washed in 75% ethanol, dried (but not over-dried, as it was then difficult to

redissolve), and redissolved in TE.

A second method was used more often for large quantities of DNA. This involved using a gel made with 'Low Melting Point Agarose'. The buffer system for these gels was the same as for normal gels. Agarose slices were added to twice their volume of TE in Eppendorf tubes, and heated at 67°C for 30 minutes, before being transferred to a 37°C water bath. An equal volume of prewarmed phenol was then added, after which, the mixture was vortexed for 30 seconds, and centrifuged for 3 minutes. The aqueous upper phase was removed and re-extracted twice with phenol, and then once with chloroform. The DNA was then precipitated as above. The yield from this method was high, but the DNA was sometimes undigestable with restriction enzymes, and had to be re-extracted.

Probably the best method, involved the insertion of a strip of Whatman No. 1 (or 3MM) paper immediately in front of a required DNA band, in a small slit cut in a normal agarose gel. The DNA was run into this paper (and stopped with a small piece of dialysis membrane inserted next to the paper). The paper was removed and inserted into a small Eppendorf tube with a small hole in the bottom. This tube was then placed in a large Eppendorf tube and centrifuged. The DNA solution appeared in the bottom tube and was extracted with phenol, and precipitated as above. High yields can be attained in this way.

The second and third methods were used for the preparation of large amounts of DNA. Only the first method was practicable when small amounts of radioactivity-labelled DNA were involved. For fragments of the latter kind, under 0.7 kb in length, an acrylamide gel system was used, to achieve the maximum yield.

2.6 DNA electrophoresis in Polyacrylamide Gels (Non-Denaturing)

These were as in Maniatis et al. (1975) and Southern (1979). They were used to obtain high yields of radioactively-labelled small fragments (<0.7 kb) of ds DNA. 25 cm vertical gels were run, with 0.4 mm spacers. The gels were usually 5% acrylamide, 0.17% bis-acrylamide and 1 x TBE (10.9 g Tris base, 5.5 g boric acid, 0.95 g EDTA in 1 litre and pH altered to 8.3 with solid boric acid or solid Tris), and were polymerized with 6.6% ammonium persulphate and TEMED (e.g. 0.15 ml and 0.05 ml, respectively, for a 25 ml gel mix). The running buffer used was 1 x TBE. The sample buffer used was 1 x TBE, 20% glycerol and Bromophenol Blue. Radioactive DNA samples were usually precipitated, the pellets washed and dried, and redissolved in this sample buffer. These gels were usually electrophoresed at 600 volts until the bromophenol blue had run 2/3 the length of the gel. The glass plates were then separated and the gel covered with cling film. Radioactive ink was used to asymmetrically mark the gel. A strip of X-ray film was placed next to the gel, the glass plates clipped together, and the whole thing was wrapped in foil. This was then placed at 4°C to expose (usually for 1 to 2 hours).

2.7 The Electrophoresis of Single-Stranded DNA on Polyacrylamide Gels.

The same gel system was used for the separation of sequencing reaction products, the preparation of primer and the separation of primer extension products. Most of the gels contained 8% acrylamide: 3.6 g acryalmide, 0.18 g bis-acrylamide, 18.9 g urea, 4.5 ml of 10 x TBE made up to 45 ml with water. They were polymerized with 0.3 ml of 6.6% ammonium persulphte and 0.03 ml of TEMED. Gels were formed between two 40 x 20 cm glass plates, taped together with 0.4 mm spacers. Samples

were dissolved in sample buffer (98% deionised formamide, 4% (w/v) Ficoll, 0.01% bromophenol blue, 0.01% xylene cyanol), and heated to 90°C for 3 minutes, before being rapidly cooled in ice, and loaded on to the polymerized gel (at least 1 hour allowed for polymerization) with a drawn-out capillary tube. The gels were run vertically, in fresh 1 x TBE buffer, at 26 mA. For sequencing gels, it was generally found that the bands became fuzzy if the gel was run for more than 5 to 7 hours (so that it was difficult to read sequence more than 250 to 300 bases from the labelled end). After running, the gels were treated as in section 2.6 except that the sequencing and primer extension gels were autoradiographed frozen, at -70°C.

2.8 Extraction of DNA from Acrylamide Gels

This was adapted from Maxam and Gilbert (1977). The DNA-containing slice of gel was transferred to a 1.5 ml Eppendorf tube, to which was added 0.3 ml of 0.5 M ammonium acetate, 0.01 M magnesium acetate, 0.1% sodium dodecyl sulphate, 0.1 mM EDTA. This was left overnight at 37°C. The gel slice was then removed from the tubes and 0.7 ml of ethanol added to the remaining solution. The DNA was precipitated by placing the tube in dry ice for 10 minutes. The precipitate was pelleted in an Eppendorf centrfuge. When double stranded DNA was being extracted from non-denaturing gels, this pellet was redissolved in TE and reprecipitated without the addition of any salt. With single-stranded DNA from urea-containing gels, however, this was not done (as the DNA would not precipitate without additional salt). *E. coli* tRNA carrier was added to most precipitations involving small amounts of DNA (<1 µg), as it was found to be necessary to reduce the amount of the latter that

'stuck' to the tube. It was also crucial to avoid over-drying of the DNA pellet, to avoid the same problem.

2.9 E. coli Genomic DNA Preparation

This was similar to the plasmid preparation method (see 2.3), except that, after addition of the lysozyme, the tube was kept on ice for 30 minutes. Also, the Lysis Mix consisted of 0.5% Triton X-100, 25 mM EDTA, 50 mM Tris/ HCl pH 8.0. After addition of the latter, the tube was kept on ice for 30 minutes, with vigorous vortexing. After the clearing spin, the DNA solution was extracted several times with phenol/chloroform and then ethanol-precipitated.

2.10 Preparation of Muscle RNA

Several *X. borealis* froglets were taken and anaesthetized in 0.1% 'MS222' (Ethyl m-amino benzoate), before being decapitated. They were then skinned and the leg and body-wall muscle cut out, These were chopped up with scissors and homogenized in 10 mM EDTA, 2% sodium dodecyl sulphate, 1% LiCl, 0.1 M Tris/ HCl pH 8.6 (with added octan-1-ol to reduce frothing). This was then treated directly with proteinase K. The latter was preincubated at 37°C for 30 minutes, to remove any RNAase contamination of the enzyme, and then added to the RNA preparation at 0.1 mg/ml and incubated for 30 minutes. The preparation was extracted 3 to 4 times with phenol/ chloroform (1:1), once with chloroform/ isoamyl alcohol (24:1), and ethanol precipitated, overnight at -20°C. After pelleting the RNA, by centrifugation for 30 minutes at 10,000 rpm, 4°C in an MSE 8 x 50 ml rotor, it was redissolved in 50 mM Tris/ HCl pH 7.6,

3 mM magnesium chloride and then incubated at room temperature for 1 hour with RNAase-free DNAase (at 0.1 mg/ml). This was re-extracted with phenol/ chloroform and ethanol precipitated. The RNA was pelleted as before, rinsed with 75% ethanol, dried, and dissolved in 100 µl of sterile distilled water.

2.11 Bacteriophage Plating

This was essentially as described by Murray and Murray (1974) and Enquist et al. (1974). Bacteria and phage were plated on 9 cm diameter circular plates containing approximately 30 mls of autoclaved L. Agar (10 g NaCl, 10 g Bacto Tryptone, 5 g Yeast Extract, 15 g Difco agar in 1 litre). For the initial screening, 10 x 10 cm square plates were used. The plates were dried well, before addition of the Top Agar. Phage were diluted to a suitable concentration in Phage Buffer (1.5 g potassium dihydrogen phosphate, 3.5 g disodium hydrogen phosphate, 2.5 g Nacl, 0.125 g magnesium sulphate, 0.1 mM calcium chloride and 50 mg gelatin in 500 mls). The E. coli strain LE392 were grown overnight at 37°C in 10 mls of L. Broth containing 0.4% maltose and 10 mM magnesium sulphate, until a dense culture was obtained. The bacteria were centrifuged for 3 minuts in an MSE benchtop centrifuge (5,000 rpm) and resuspended in 1 ml of 10 mM magnesium sulphate. 0.1 ml of this suspension was then added to 0.1 ml of diluted phage and left at room temperature for 10 minutes (to allow the phage to adsorb onto the bacterial cells). Then, 3.5 mls (or 10 mls for the square plates) of Top Agar (as above for L. agar, but containing only 8g/litre agar and supplemented with 0.4% maltose and 10 mM magnesium sulphate), autoclaved and cooled to 45°C, was added to the phage and bacteria, and this mixture poured onto an L. Agar plate. The

plates were incubated overnight at 37°C.

2.12 Bacteriophage Storage

Plaques of potential interest were transferred with the sterile large end of a Pasteur pipette into 1 ml of Phage Buffer. After mixing, the supernatant was removed into another large Eppendorf tube, which was centrifuged to pellet any bacteria. The phage-containing supernatant was transferred into a sterile bijoux bottle and 1 drop of chloroform added from a Pasteur pipette. This was stored at 4°C.

2.13 Benton and Davis Plaque-Lift Procedure

This is the phage-screening procedure of Benton and Davis (1977). Phage were plated out so as to produce separated plaques (e.g. 3000 to 4000 plaques per 10 cm square plate). These plates were then cooled at 4°C (to prevent the Top Agar sticking to the nitrocellulose). Nitrocellulose filters, previously autoclaved, were layed on the cooled Top Agar for 3 minutes. During this time, the position and orientation of the nitrocellulose was marked in the Agar with a toothpick. The nitrocellulose was removed and washed in 0.5 M NaOH, 1.5 M NaCl for 20 seconds, 0.5 M Tris/HCl pH 7.0, 3 M NaCl for 20 seconds, and finally rinsed n 2 x SSC (1 x SSC = 0.15 M NaCl, 0.015 M sodium citrate), before blotting dry. The filters were then baked at 80°C for 2 hours. They were then ready for hybridisation with a radioactive probe.

2.14 Bulk Preparation of Phage DNA

This was derived from Murray and Murray (1974). A phage plaque was first plated out onto a fresh plate. Then varying numbers of plaques (the optimum number appeared to depend on the individual recombinant concerned) were removed by scraping off the Top Agar and mixing with Phage Buffer (see 'phage plating'). The bacterial and Agar debris was spun down in an MSE bench centrifuge at 8000 rpm, for 15 minutes. This was the starting inoculum. E. coli (LE392) bacteria were grown in 10 mls of L Broth supplemented with 0.4% maltose and 10 mM magnesium sulphate. While the bacteria were still in log phase, 0.2 mls was added to 0.9 mls of phage inoculum and left at room temperature for 5 to 10 minutes. This was added to 1 litre of L. Broth, containing 10 mM magnesium sulphate, which had been autoclaved in a 2 litre flask, and then prewarmed to 37°C. This was incubated at 37°C overnight, with vigorous shaking. Successfully completed growth was indicated by the presence of bacterial debris at the bottom of the flask.

0.5 mls of chloroform were then added to each litre culture. The cultures were centrifuged at 5000 rpm in an MSE 6 x 250 ml rotor at 4°C, to pellet the bacterial debris. The supernatant was made 0.5 M in NaCl, 10% in polyethylene glycol (PEG). This was stirred to thoroughly dissolve all the PEG, and left on ice for at least 2 hours. The phage precipitate was pelleted by centrifugation at 5000 rpm for 15 minutes in an MSE, 6 x 250 ml rotor, at 4°C. The supernatant was carefully poured off, and as much residual liquid removed, as possible, with a Pasteur pipette. The pellet from 1 litre of bacteria was resuspended in <7 ml of Phage Buffer.

To each ml of this phage suspension was added 0.71 g CsCl. This was dissolved and the volume made up to a convenient one (for the centrifuge tubes being used) with Phage Buffer (with 0.71 g CsCl added

per ml). This was centrifuged in an MSE 10 x 10 ml rotor at 40,000 rpm, 20°C, for 36 to 48 hours. The phage band in the resulting density gradient was removed, made up to a convenient volume with more Phage Buffer (with 0.71 g CsCl added per ml), and recentrifuged as before. The phage band was removed, once again, and dialyzed for 2 hours at room temperature against TE (1 mM EDTA, 10 mM Tris/HCl pH 7.4). This was twice extracted with an equal volume of phenol and dialyzed overnight against TE, at 4°C, to remove the phenol. The DNA was then ready for use. At no stage was the phage DNA precipitated from ethanol as it was found to be extremely hard to subsequently redissolve.

2.15 Nitrocellulose Filter Hybridization

The conditions used were after Denhart (1966). This method was used for Southern blots and screening the Benton and Davis plaque lifts. Usually, the filter was prehybridized overnight with Denhart's additives (0.2% Ficoll, 0.2% polyvinylpyrolidone, 0.2% bovine serum albumin) and *E. coli* DNA at 10 µg/ml, in 3 or 4 x SSC. 5 to 10 ml of this solution was added to the nitrocellulose and sealed into a plastic bag, which was placed at the bottom of a water bath at the required temperature. For the hybridization, a fresh solution of the same composition was used. The nick-translated probe was first boiled for 10 minutes, and rapidly cooled on ice, before being made up to 5 ml with the hybridization solution (usually made in concentrated form so as to avoid dilution by the probe). The hybridization was performed overnight or over the weekend. The nitrocellulose was then washed at the appropriate temperature and salt conditions for 1 hour (usually several changes of 500 ml) dried, and autoradiographed with an intensifying screen.

For the experiment described in section 7.1, however, the prehybridization and hybridization mixes contained 50% deionized formamide pH 7.0, 3 x SSC, 10 µg/ml *E. coli* DNA, and 50 mM Hepes pH 7.0. The prehybridization mix contained 0.1% bovine serum albumin, Ficoll and polyvinyl pyrolidone, while the hybridization mix contained these ingredients at 0.02% plus 10% (w/v) dextran sulphate. Both stages were at 42°C, in a 3 ml volume, overnight in plastic bags. Afterwards, the filter was washed 4 times in 2 x SSC (15 minutes each), at room temperature, before being washed twice in 0.1 x SSC (15 minutes each). The temperature at which the last wash was done was varied from room temperature (low stringency), to 60°C (high stringency).

2.16 Hybrid Release Translation

The protocol used was a modified version of that used by Cleveland *et al*. (1980). 1 cm square nitrocellulose filters were autoclaved and soaked in 6 x SSC. 50 µg of recombinant DNA was then made 6 x SSC, boiled briefly, and quenched on ice (to denature the DNA). This was loaded equally onto 2 nitrocellulose filters, which were air-dried, and baked at 80°C for 2 hours. Each filter was broken up and placed in a 1.5 ml Eppendorf tube. To each was added either 100 µg of *X. borealis* skeletal muscle total RNA or 200 µg of *X. borealis* total ovary RNA. The RNA was dissolved in 10 mM PIPES (pH 6.4), 0.4 M NaCl, 0.2% SDS. The tubes were sealed and incubated at 50°C for 2 hours in a shaking waterbath. The RNA was then removed and the filters washed 6 times in 500 µl of 1 x SSC at 60°C for 1 minute each. This was followed by two 1 minute-washes in 500 µl of 0.2 x SSC at 50°C. The bound RNA was then eluted from the filters by boiling in two successive 200 µl aliquots of

water. 10 µg of E. coli tRNA carrier was added to the pooled eluent, as well as 40 µl of 3 M sodium acetate and 1 ml of ethanol. This was left overnight at -20°C to precipitate the RNA. The tubes were spun to pellet the RNA, the latter then being washed, dried and redissolved in 10 µl water.

2.17 Wheat Germ in vitro Translation

The wheat germ extract used for this procedure was a gift from Drs. R. W. Old and A. Colman. This extract was first treated in the manner described by Pelham and Jackson (1976) to remove the endogenous messages. 1 µg of micrococcal nuclease was added to 100 µl of wheat germ extract as well as 2 µl of 50 mM $CaCl_2$. This was incubated at room temperature for 15 minutes, after which the reaction was stopped with 2 µl 100 mM EGTA (EGTA specifically chelates Ca^{++} ions, which are essential to the activity of the nuclease). A reaction mix was made up consisting of the following: 10 µl of Energy Mix (20 mM dithiothreitol, 10 mM ATP, 0.2 mM GTP, 80 mM creatine phosphate, 0.4 mg/ml creatine phosphokinase), 10 µl of 0.4 M HEPES/KOH pH 7.6, 10 µl of 1.1 M KCl, 10 µl of a mixture of all 20 amino acids (minus methionine) at 0.5 mM and pH 7.8 each, and 60 µl of ^{35}S-labelled methionine (360 µCi). The final incubation mixture consisted of 5 µl of this reaction mix + 5 µl of nuclease-treated wheat-germ extract + 2.5 µl of RNA in water. This was incubated for 1 hour at 25°C.

2.18 Translation in Reticulocyte Lysate

The lysate used was made by myself (the preparation is described

separately) and was used after removing the endogenous RNA, as described by Pelham and Jackson (1976). To 1 ml of untreated lysate was added 30 ml of 1 mM haemin, 5 µl of creatine kinase (10 mg/ml), 10 µl of 100 mM $CaCl_2$ (20 mM Tris HCl pH 7.2) and 10 µl of micrococcal nuclease (10 mg/ml). This was then incubated at 20°C for 17 minutes, when the reaction was stopped by addition of 10 µl 0.25 M EGTA pH 7.8. This treated lysate can be stored for several months in liquid nitrogen.

To make 100 µl of reaction mix, 50 µl of nuclease-treated lysate was taken and to it was added 8 µl of a mixture of all the amino acids (except methionine) at 0.5 mM pH 7.8, 5 µl of 2 M KCl, 5 µl of ^{35}S-labelled methionine (60 µCi), 4 µl of mouse liver tRNA (2 mg/ml), 3 µl of 400 mM creatine phosphate, and 25 µl of 20 mM Tris HCl pH 7.2.

30 µl of such a mix would be incubated with 3 µl of RNA (in water) at 30°C for 1 hour.

2.19 Preparation of Rabbit Reticulocyte Lysate

The starting material was 6 rabbits. These were first injected with vitamins (mainly vit.B). On each of the next 5 days each rabbit was injected subcutaneously with 0.6 ml of 2.5% phenylhydrazine. They were then left one day before performing heart puncture on the 4 surviving rabbits.

The rabbits were anaesthetized with 1 ml of sodium pentabarbitone (60 mg/ml) by injection into an ear vein. Rapidly, a long syringe needle was inserted through the body wall into the heart. On piercing the latter, a 50 ml syringe, previously rinsed in heparin (1000 units/ml), was attached and 50-75 ml of blood withdrawn (I would like, at this point, to thank Dr. Giorgio Valle for his manual help and moral

support). The blood was kept on ice before being centrifuged for 15 minutes at 2000 rpm in the Mistral 6L 4 x 1000 rotor in heparin-treated 250 ml glass pots. The pellet obtained was washed 3-4 times in washing buffer (140 mM KCl, 50 mM NaCl, 5 mM $MgCl_2$) until the supernatant became clear. The volume of the pellet was measured and an equal volume of cold sterile water added to lyse the cells. This was centrifuged at 8000 rpm in an MSE 8 x 50 rotor for 10 minutes in oak ridge tubes. The supernatant was removed carefully, so as to include no nuclear pellet, and dispensed into 1 ml aliquots. These were rapidly frozen in liquid nitrogen. From 4 rabbits, 22 mls of lysate were finally obtained. This can be kept for over 18 months in liquid nitrogen.

2.20 Protein Gel Electrophoresis

For the separation of the *in vitro* translation products the SDS-polyacrylamide gel electrophoresis system of Laemlli (1970) was used. The gels used for reticulocyte lysate products were 12.5% acrylamide, those for wheat germ products, 18%. To make an 18% separating gel the following components were combined: 15 ml of 60% acrylamide, 0.3% NN bis-acrylamide, 6.25 ml 36.6% Tris HCl pH 8.8, 0.5 ml 10% SDS, 28 ml water, 0.25 ml ammonium persulphate and 10 µl TEMED. To make a 12.5% gel, the amount of acrylamide/ bis-acrylamide solution was reduced accordingly. After the separating gel had set, a 10 ml stacking gel was poured on top: 3 ml of 10% acrylamide, 0.5% NN bis-acrylamide, 4.4 ml water, 2.4 ml 6% Tris HCl pH 6.8, 0.1 ml 10% SDS, 100 µl 10% ammonium persulphate, 5 µl TEMED. Slots were formed with a slot former on pouring the stacking gel. Gels were run in Tris/glycine running buffer (50 mM Trizma base, 400 mM glycine, 0.5% SDS). 5 µl of reticulocyte

translation mix were loaded with 10 µl of sample buffer: 1 mM phenylmethylsulfonyl fluoride, 20 mM Tris pH 7.6, 20% glycerol, 1% mercaptoethanol, 0.01% Bromophenol blue, 2% SDS. 10 µl of wheat germ mix were loaded with 4 µl of 40% glycerol, 8% SDS, 2% mercaptoethanol, 18.25% Tris HCl pH 6.8, 0.004% Bromophenol blue. These mixes were boiled before loading. Gels were usually run overnight at 50 volts and then dried down.

The ^{35}S label was detected in gels by fluorography (Laskey and Mills; 1975). Gels were stained in Coomassie blue (1 g coomassie blue in 227 mls methanol, 227 mls water, 46 mls acetic acid for 1 hour, followed by destaining in 875 mls water, 50 mls methanol, 75 mls acetic acid for 1 hour) to locate the unlabelled markers. The gel was then soaked in two changes of dimethylsulphoxide (DMSO) for 1 hour each, with gentle shaking, then in 22% PPO in DMSO for 3 hours. The gel was then rinsed under running water for 1 hour, dried onto Whatman 3 MM paper and autoradiographed with flashed film.

2.21 Annealing of 35A-Derived Primer to RNA

The conditions for this were similar to those used by Tsang *et al.* (1982). The hybridization mixes which were used have been described in section 5.1.2. When up to 20 µg of RNA were involved, the volume used was 10 µl, the RNA and primer being precipitated together and the pellet redissolved in the hybridization buffer. It was then transferred to a small Eppendorf tube, covered with paraffin and hybridized overnight at the appropriate temperature. The sample was then diluted into 0.1 ml of TE and ethanol-precipitated twice.

2.22 Primer Extension Reaction

The pellet obtained from 2.21 was redissolved in 5 µl of 10 mM Pipes pH 6.4, 0.4 M NaCl. 0.04 ml of premade Primer Extension Mix (50 mM Tris/HCl pH 8.2, 10 mM dithiothreitol, 6 mM magnesium chloride, 25 µg/ml actinomycin D and 0.5 mM dATP, dCTP, dGTP, dTTP) was added to this, as well as 5 to 10 units of AMV reverse transcriptase. This was incubated at 42°C for 60 minutes. Sometimes, further enzyme was added and the incubation continued for another 30 minutes. This was then extracted with phenol, ethanol-precipitated twice, and the pellet redissolved in sequencing gel sample buffer. This was run on an 8% sequencing gel.

(When 750 µg of ovary RNA were annealed to primer, for sequencing, the hybridization was performed in 0.35 ml volume, and the primer extension used 0.48 ml of Primer Extension Mix plus 0.06 ml of 10 mM Pipes pH 6.4, 0.4 M NaCl.)

2.23 Phosphatasing and Kinasing DNA

This was only successful using 5' overhangs. The DNA was dissolved in 45 µl of TE and 5 µl of 0.5 M Tris/HCl pH 8.0. 1 µl of bacterial alkaline phosphatase (250 units) was added to this and incubated for 1 hour at 37°C. 0.5 µl of 500 mM EDTA, pH 8.0 was added to this and heated to 60°C for 10 minutes. It was then extracted 3 times with phenol/chloroform (1:1) and ethanol-precipitated.

The kinase reaction was performed in 20 µl volume. This included 2 µl of 10 x Kinase Buffer (0.7 M Tris/HCl pH 7.6, 100 mM magnesium chloride, 50 mM dithiothreitol) as well as the gamma-labelled ^{32}P-ATP

and 12 units of T4 polynucleotide kinase. The reaction was performed at 37°C for 15 minutes. This was extracted with phenol (with 10 µg of E. coli tRNA added) and passed down a Sephadex G50 column to check for labelling of the DNA.

2.24 Nick Translation

This was adapted from Rigby et al. (1977). Usually the DNA was made as hot as possible, for the probing of nitrocellulose filters. For restriction mapping, however, the DNA was only lightly labelled, so that it was still possible to visualise discrete restriction digest products on gels.

The latter type of reaction was performed in 20 µl volume, which included 4 µl of 5 x NT Buffer (0.25 M Tris/HCl pH 7.9, 50 mM magnesium chloride, 50 mM dithiothreitol, 250 µg/ml bovine serum albumin), 1 µl each of three unlabelled dNTPs (1 mM each), 1 to 2 µl of alpha-^{32}P dNTP (10 mCi/ml in aqueous buffer) and water. 1 µl of E. coli DNA polymerase I (2-5u/µl) was added to start the reaction, which was run for 1 hour at 24°C.

For heavily labelled material, the labelled dNTP in the reaction was increased, 50 pg of DNAase I wre added (to 20 µl), and 1 µl of the three unlabelled dNTPs at 10 µM was added. The reaction was at 37°C for 3 hours.

In both cases the reaction mix was subsequently extracted with phenol, and passed down a Sephadex G50 column (with 10 µl of E. coli tRNA as carrier) to check for successful labelling of the DNA.

2.25 Labelling DNA with T4 DNA Polymerase I

This technique was used for the labelling of DNA markers as well as for the end-labelling of DNA for sequencing. It was taken from O'Farrell *et al*. (1980). The reaction mix in the latter type of reaction consisted of 1 µl of 10 x TA (0.33 M Tris acetate pH 7.9, 0.66 M potassium acetate, 0.1 M magnesium acetate, 1 mg/ml nuclease-free bovine serum albumin, 5 mM dithiothreitol) to 2 µg of DNA, 20 µCi of alpha-^{32}P-dNTP, 1 unit of T4 DNA polymerase made up to 10 µl. This was incubated at 20^{o}C for 30 minutes, and then extracted with phenol. The low temperature was used to minimize the amount of 3' to 5' exonuclease activity, so that only one labelled base would be incorporated. Consequently, only 5' overhangs could be labelled in this way.

For the labelling of DNA markers the reaction was set up in 20 µl with 2 µl of 10 x TA buffer, 1 µg DNA and 1 unit of T4 polymerase. This was incubated at 37^{o}C for 3 to 5 minutes (at 1.25 units/µg DNA, 20 nucleotides/min should be excised from each 3' end). After this time, 2 µl of alpha-^{32}P-dNTP (10 mCi/ml in aqueous solution) was added directly to the reaction mix, together wsith 1 µl of a mixture of the other 3 unlabelled dNTPs (at 2 mM), and incubated for 30 minutes at 37^{o}C.

After each type of reaction, the mix was extracted with phenol, and then passed down a Sephadex G50 column.

2.26 Ligation of Restriction Fragments

Ligation was usually performed in a volume of 10 to 20 µl, and in a solution of 50 mM Tris HCl pH 7.6, 10 mM magnesium chloride, 20 mM dithiothreitol, 1 mM ATP (adenosine triphosphate) and 1 unit of T4 DNA ligase. The reactions were performed at 4^{o}C overnight (or occasionally

at 15°C for 7 hours). Typically, such a ligation used 20 ng of vector DNA and 40 ng of donor DNA. The ligation mix was then added directly to competent cells in the transformation procedure.

2.27 Transformation of E. coli with Double-Stranded M13 and Plating

A colony of *E. coli* (JM103) was grown up in 20 mls of 2 x TY medium (10g Yeast Extract, 5 g NaCl, 16 g Bactotryptone in 1 litre) until its optical density, at 590 nm, was 0.3. The cells were pelleted at full speed in an MSE bench centrifuge (10 minute spin). The cells were resuspended in 10 mls of precooled 50 mM calcium chloride and left on ice for 40 minutes. They were then spun down again and resuspended in 2 ml of 50 mM calcium chloride. 0.3 ml aliquots of this were added to each ligation mix and then left on ice for 40 minutes (meanwhile a fresh culture of JM103 was prepared so that, on plating, it was still in exponential growth phase). The cells were heatshocked at 42°C for 2 minutes. At this time 30 µl of BCIG (20 mg/ml in dimethylformamide), 25 µl of IPTG (24 mg/ml in water) and 0.2 ml of exponentially growing JM103 were added to 3 mls of H-Top Agar (10 g Bactotryptone, 8 g Difco agar, 8 g NaCl in 1 litre), at 45°C. The transformed cells were then added to this H-Top Agar and the mixture poured onto a dried 9 cm diameter minimal agar plate.

Minimal Agar plates were made by autoclaving 390 ml of water with 7.5 g of Difco agar. When cool, 100 mls of A salts (5.25 g of dipotassium hydrogen phosphate, 2.25 g of potassium dihydrogen phosphate, 0.5 g of ammonium sulphate, 0.25 g of trisodium citrate in 100 mls), 0.5 ml of 1 M magnesium sulphate, 0.125 ml of 1% thiamine hydrochloride (vitamin B1) and 10 ml of 20% glucose, were added, and 30

at 15°C for 7 hours). Typically, such a ligation used 20 ng of vector DNA and 40 ng of donor DNA. The ligation mix was then added directly to competent cells in the transformation procedure.

2.27 Transformation of E. coli with Double-Stranded M13 and Plating

A colony of E. coli (JM103) was grown up in 20 mls of 2 x TY medium (10g Yeast Extract, 5 g NaCl, 16 g Bactotryptone in 1 litre) until its optical density, at 590 nm, was 0.3. The cells were pelleted at full speed in an MSE bench centrifuge (10 minute spin). The cells were resuspended in 10 mls of precooled 50 mM calcium chloride and left on ice for 40 minutes. They were then spun down again and resuspended in 2 ml of 50 mM calcium chloride. 0.3 ml aliquots of this were added to each ligation mix and then left on ice for 40 minutes (meanwhile a fresh culture of JM103 was prepared so that, on plating, it was still in exponential growth phase). The cells were heatshocked at 42°C for 2 minutes. At this time 30 µl of BCIG (20 mg/ml in dimethylformamide), 25 µl of IPTG (24 mg/ml in water) and 0.2 ml of exponentially growing JM103 were added to 3 mls of H-Top Agar (10 g Bactotryptone, 8 g Difco agar, 8 g NaCl in 1 litre), at 45°C. The transformed cells were then added to this H-Top Agar and the mixture poured onto a dried 9 cm diameter minimal agar plate.

Minimal Agar plates were made by autoclaving 390 ml of water with 7.5 g of Difco agar. When cool, 100 mls of A salts (5.25 g of dipotassium hydrogen phosphate, 2.25 g of potassium dihydrogen phosphate, 0.5 g of ammonium sulphate, 0.25 g of trisodium citrate in 100 mls), 0.5 ml of 1 M magnesium sulphate, 0.125 ml of 1% thiamine hydrochloride (vitamin B1) and 10 ml of 20% glucose, were added, and 30

mls poured onto each plate.

Once the transformed cell-containing Top Agar had set, the plates were inverted and incubated overnight at 37°C. The presence of blue plaques indicated that the transformation had worked because only when the defective forms of beta-galactosidase of the M13 and JM103 complement each other (on induction by IPTG = isopropyl-beta-D-thio-galactopyranoside), is the lactose analogue BCIG (5-bromo-4-chloro-3-indoyl-beta-galactoside) broken down to give a blue derivative. Recombinant phages, in which foreign DNA has been inserted into one of the cloning sites within the beta-galactosidase gene, usually yielded white plaques.

2.28 Preparation of Single-Stranded M13 (for sequencing)

A 10 ml culture of JM103 was grown overnight at 37°C. A drop of this was added to 25 ml of 2 x TY and 1 ml aliquots of this were dispensed into sterile Bijoux bottles. Material from white plaques was transferred by toothpick into these cultures and then they were incubated for 5 to 6 hours at 37°C, with shaking. After this, the bacteria were pelleted by centrifugation in 1.5 ml Eppendorf tubes for 5 minutes. The supernatant (containing free phage) was transferred to another Eppendorf tube (when it could be stored for several days at 4°C, but was then recentrifuged before proceeding with the next step) where 0.2 ml of 2.5 M NaCl, 20% polyethylene glycol was added. After mixing, this was left at room temperature for 15 minutes. The phage precipitate was pelleted with another 5 minute centrifugation, and the supernatant discarded. The tubes were respun and all the supernatant removed with a Pasteur pipette. The pellet should be visible. 0.1 ml of 10 mM

Tris/HCl pH 8.0, 0.1 mM EDTA was added and this was extracted with 50 µl of neutralized phenol (vortexed 10 seconds, stood for 5 minutes, revortexed and centrifuged for 1 minute). The upper aqueous layer was extracted with 0.5 ml diethyl ether and the M13 ssDNA precipitated with 10 µl of 3 mM sodium acetate, 0.25 ml of ethanol, in dry ice for 10 minutes. After centrifugation and drying, the DNA was redissolved in 10 µl of 10 mM Tris/HCl pH 8.0, 0.1 mM EDTA. If successful, 1 µl of this DNA solution should show up well on an agarose gel, stained with ethidium bromide. The DNA was stored at 4°C.

2.29 Testing the Orientation of Cloned Inserts in M13

This was used, when a single fragment had been cloned, to test which DNAs to sequence. The basic principle involved, was to hybridize all possible combinations of pairs of the ssDNAs together, and to then run these on an agarose gel, next to unhybridized DNA. DNAs resulting from the cloning of the original fragment in opposite directions will, in single stranded form, be complementary. These will thus hybridize and run as double stranded DNA on the gel, i.e. they are retarded. The hybridization reaction was performed in a 9 µl volume, and included 0.5 µl of each DNA, 3 µl of water, 1 µl of agarose gel loading buffer (see section 2.4) and 4 µl of 'fresh' 10 x Hin buffer (see section 2.30). This mixture was hybridized for 30 minutes at 67°C, and then loaded directly onto the gel.

2.30 M13 Sequencing

This was based on the chain termination method of Sanger *et al*.

(1977), where single stranded DNA is used as a template for the synthesis of a radioactively labelled second strand. In each of 4 separate reactions, a small amount of one dideoxynucleoside triphospahte is included, so that the newly synthesized strand tends to terminate at that type of base. The quantities of the nucleotides are varied so that a ladder of such terminations is achieved for each base. When the products from the 4 reactions are electrophoresed, side by side, on an 8% urea-polyacrylamide sequencing gel, the sequence can be read by following the order of appearance of bands in the 4 tracks.

The DNA to be sequenced was cloned into the replicative form of M13mp7 (Messing et al., 1981), M13 mp8, or M13 mp9 (Messing, 1981; Messing and Vieira, 1982). Single stranded DNA was prepared (see section 2.28) and the radioactively labelled second strand primed from a universal primer, as in Sanger et al. (1980).

2 µl of single stranded M13 recombinant was hybridized with a 15 base-long universal primer (1 µl of 5 µg/ml) in 10 µl of volume, including 6 µl of water and 1 µl of 'fresh' 10 x Hin buffer (1 vol of 100 mM dithiothreitol + 9 vol of storage 10 x Hin buffer = 0.1 M Tris/HCl pH 7.4, 0.1 M magnesium chloride, 0.5 M NaCl). The small Eppendorf tube, containing this mixture, was heated to 90-100°C for 3 minutes, and then allowed to cool slowly, at room temperature, for 20 minutes. 0.5 µl of ^{32}P dGTP was added to this, followed by 1 µl of diluted Klenow fragment solution (1 unit of the Klenow fragment of E. coli DNA polymerase I, diluted in 50 mM potassium dihydrogen phosphate, 50% glycerol) 2 µl of this mixture was added to each of 4 large Eppendorf tubes, containing varying proportions of the different dideoxy- and deoxynucleoside triphosphates:

1. The T mix: A volume of 0.5 mM dideoxy TTP was added to an equal volume of T^{o} solution (1 µl of 0.5 mM dTTP, 20 µl of 0.5 mM dCTP, 20 µl of 0.5 mM dATP, 1 %l of 0.066 mM dGTP pluse 5 µl of 'low TE' = 5 mM Tris/HCl pH 8.0, 0.1 mM EDTA).

2. The C mix: 0.2 mM dideoxy CTP was added to an equal volume of C^{o} solution (20 µl of 0.5 mM dTTP, 1 µl of 0.5 mM dCTP, 20 µl of 0.5 mM d ATP, 1 µl of 0.0966 mM dGTP plus 5 µl of 'low TE').

3. The A mix: 0.2 mM dideoxy ATP was added to an equal volume of A^{o} solution (20 µl of 0.5 mM dTTP, 20 µl of 0.5 mM dCTP, 1 µl of 0.5 mM dATP, 1 µl of 0.066 mM dGTP plus 5 µl of 'low TE').

4. The G mix: 0.016 mM dideoxy dGTP was added to an equal volume of G^{o} solution (20 µl of 0.5 mM dTTP, 20 µl of 0.5 mM dCTP, 20 µl of 0.5 mM dATP, 1 µl of 0.066 mM dGTP + 5 µl of 'low TE').

2 µl of each mix was added to a separate tube with the primer-annealed M13 mix. The mixture was incubated at room temperature for 15 minutes, before the addition of 2 µl of 0.5 mM dGTp, and a further 15 minute incubation. 5 µl of sequencing-gel sample buffer (section 2.21) was then added, to stop the reaction. This mixture was then heated at 100°C, with the tube tops open, for 3 minutes, then rapidly cooled on ice, and loaded onto an 8% sequencing gel.

2.31 Southern Blotting

This is the procedure of Southern (1975), to transfer DNA from

agarose gels to nitrocellulose. After running the gel, the DNA was denatured *in situ* in the gel by soaking it for 15 minutes in 0.5 M NaOH, 1.5 M NaCl, followed by neutralisation for 15 minutes in 0.5 M Tris/HCl pH 7.0, 3 M NaCl. Typically, 500 ml of each buffer was used. The gel was laid on 4 sheets of 3 MM chromatographic paper, previously moistened with 20 x SSC (175 g NaCl, 88.25 g trisodium citrate, in 1 litre), on a glass platform. The latter was raised above the surface of a tray so that the gel was above the surface of the 20 x SSC, but the 3 MM paper hung over the edge of the plate, to form a wick. A section of nitrocellulose was laid over the DNA-containing region of the gel. Lengths of polythene were placed so that a few mm protruded between the nitrocellulose and the gel, but otherwise entirely covered the rest of the apparatus. Two sheets of 3 MM were laid over the nitrocellulose, and a pile of Kleenex tissues, several inches thick, over this. The apparatus was now assembled so that the liquid in the tray should be drawn up into the tissues via the window created with polythene, i.e. through the gel and nitrocellulose. The next day, the nitrocellulose was removed, rinsed in 2 x SSC, and blotted dry. It was then baked at 80°C for 2 hours.

2.32 Maxam and Gilbert Sequencing

This was essentially as described in Maxam and Gilbert (1977). The starting material consisted of double-stranded DNA, labelled at one end with 32-phosphorus. This was divided into 5 siliconized 1.5 ml Eppendorf tubes and ethanol-precipitated, together with 1 µg of *E. coli* tRNA 'carrier'. Then, 5 reactions were routinely performed, each resulting in the cleavage of the DNA at one or two of the 4 types of

base:

1. The G reaction: The DNA pellet was dissolved in 0.2 ml of 50 mM sodium cacodylate pH 8.0, 1 mM EDTA. To this was added 1 µl of dimethyl sulphate, and after mixing, this was left at room temperature for 2.5 minutes. The reaction was stopped with 0.05 ml of 1.5 M sodium acetate pH 7.0, 1 M beta-mercaptoethanol, 0.1 mg/ml *E. coli* tRNA.

2. The A>C reaction: The DNA pellet was redissolved in 0.1 ml of 1.2 M NaOH, 1 mM EDTA, and incubated at 90°C for 9 minutes. The reaction was stopped with 0.15 ml of acetic acid and 5 µl of 1 mg/ml *E. coli* tRNA added.

3. The C reaction: The DNA pellet was redissolved in 15 µl of 5 M NaCl and 5 µl of water. 30 µl of hydrazine was added to this and mixed. After incubation at room temperature for 7.5 minutes, the reaction was stopped with 0.2 ml of 'C-stop' (0.3 M sodium acetate, 0.1 mM EDTA, 25 µg/ml *E. coli* tRNA).

4. The C + T reaction: The reaction was the same as the C reaction except that the DNA pellet was dissolved in 20 µl of water (with no NaCl). Again, 0.2 ml of 'C-stop' was used to stop the reaction.

5. The A + G reaction: The DNA pellet was redissolved in 6 µ of water. 25 µl of 95% formic acid was added to this, and the mixture incubated at room temperature for 4 minutes. The reaction was stopped with 0.2 ml of 1 M sodium acetate, 5 µg/ml *E. coli* tRNA.

Immediately after each of these reactions, 0.75 ml of ethanol was added to each tube, and the latter placed on dry ice for 10 minutes. The reprecipitated DNA was then pelleted, redissolved in 0.2 ml of 0.3 M sodium acetate and precipitated once more with 0.5 ml of ethanol. The DNA pellet, this time, was then washed with 75% ethanol, and vacuum-dried. It was resuspended in 0.1 ml of 1 M piperidine and incubated at 90°C for 30 minutes. The piperidine was then removed by freeze-drying. The DNA was resuspended in 0.1 ml of water and freeze-dried. Another 20 µl of water was then added and also freeze-dried. The DNA was resuspended in formamide-containing sample buffer and run on urea-containing polyacrylamide gels (i.e. the 5 reactions in 5 slots, side by side).

2.33 Transformation of E. coli with Plasmid DNA (pBR325 recombinants)

This method was that used by Mandel and Hiya (1970). A loop-full of E. coli (HB101) was transferred from a freshly streaked plate into 10 ml of L. broth, and shaken overnight at 37°C. A 1 ml aliquot of this was diluted into 50 ml of L. broth. The culture was grown to an optical density of 0.6 (at 590 nm) and centrifuged at 3,000 rpm for 10 minutes in an MSE bench centrifuge (highest setting). The bacteria were resuspended in 25 ml of precooled 50 mM calcium chloride and left on ice for 30 minutes. The cells were then centrifuged again, as before, and resuspended in 5 ml of 50 mM calcium chloride. A small amount of the DNA to be transformed (up to 10 µl) was added to an equal volume of 100 mM calcium chloride in a sterile Eppendorf tube, and then 0.05 ml of TCM buffer (10 mM Tris HCl pH 7.5, 10 mM calcium chloride, 10 mM magnesium chloride) was added. 0.1 ml of calcium-treated bacteria was added to

this and the mixture left on ice for 60 minutes before the bacteria were heat-shocked at 42°C for 3 minutes. 0.85 ml of sterile L. Broth was added to the tube and this was incubated at 37°C for 60 minutes. The bacteria were spread onto selective agar-plates (usually approximately 0.1 ml per plate). The latter were made by autoclaving a mixture of L. broth with Difco Agar at 15 g/litre and, when this had cooled to 45°C, adding the antibiotic and pouring the plaets (30 ml per 9 cm diameter plate). In this case, the antibiotic was ampicillin at 0.05 mg/ml. After spreading the bacteria on the dried plates, the latter were incubated overnight at 37°C. Only transformed bacteria should appear as colonies. Such colonies were then picked onto plates containing chloramphenicol at 1 µg/ml, to test for insertion of DNA into the pBR325 plasmid (see section 3.5).

3 ISOLATION AND SEQUENCING OF ACTIN GENES

3.1 Screening the Lambda Library

3.1.1 Introduction

A gene library made by Dr. William Bains (Bains, 1982) was used to probe for actin genes. This library had been constructed using the cloning vector, lambda-gtWES (Tiemeier *et al.*, 1976). This involved the removal of a central EcoRI fragment from the phage DNA, followed by the ligation to the remaining 'arms', of a partial EcoRI digest of *Xenopus borealis* genomic DNA (made from the blood of 3 females). The subsequent packaging of the ligated DNA into phage particles, and infection of the *E. coli* strain LE392 with these, produced a primary library of 425,000 phage. This was subsequently amplified 10^3 to 10^4 times by growth on plates.

One of the reasons for using this library for the screening was because it was a 'partial library'. EcoRI is a 'six-cutter' restriction enzyme; that is, it cuts at the six bases GAATTC. The random chance of these 6 bases arising in any given stretch of DNA sequence is 1 in 4^6; i.e. on average, there should be such a site every 4096 bases. When looking for an actin-type gene, where the coding sequence alone will occupy [374 x 3 =] 1122 bases, and it is suspected that there will be a considerable length of DNA present in intervening sequences, there is a reasonable chance that any single EcoRI fragment isolated will contain only part of the desired sequence. Using a partial digest for the cloning will increase the chance of obtaining the whole gene in one phage isolate.

William Bains found the mean insert size of isolates obtained from this library to be 6.1 kb. Taking the haploid genome size of *X.*

borealis as 3.2×10^9 bases (Theibaud and Fischberg, 1977), and using the formula of Clarke and Carbon (1976), $N = \ln(1-P)/\ln(1-F)$ (where N = the number of recombinants, P = the required probability, and F = the fraction of the genome represented by the average recombinant), 10^6 phage would have to be screened to have a 90% chance of finding any one actin gene. This was too large a number to screen initially, and as studies on mammalian actins have shown that there are at least 6 different primary sequences, and so at least 6 separate genes, only 200,000 phage were screened. This should have been enough to find at least one actin gene, assuming that the actin genes in X. borealis are not closely linked.

The probe used in this experiment was the Dictyostelium actin cDNA clone pcDd actin B1 (Bender et al., 1978). This consisted of a 1.1 kb insert, cloned by GC-tailing into the Pst I site of pBR322. The actins are known to be very highly conserved between different species. For example, it was known that there is a region between amino acids 7 and 159 where there is only 1 amino acid difference between Dictyostelium and mammalian cytoskeletal actins (Vandekerckhove and Weber, 1980). Because of this, it was expected that cross-hybridization between the Dictyostelium and Xenopus actin sequences would occur in low stringency hybridization conditions.

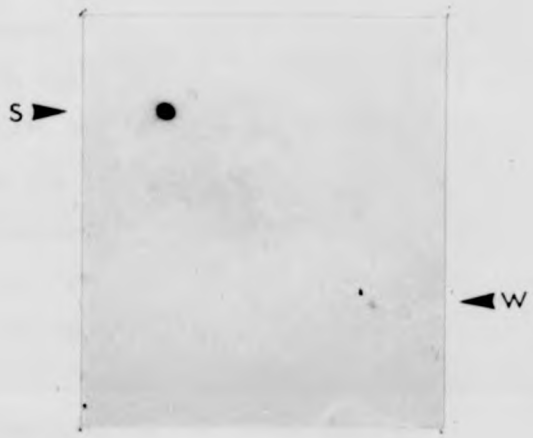
3.1.2 Results

(i) The first screen

40 square L Agar plates (9 cm x 9 cm) were plated with approximately 5,000 plaques each. Nitrocellulose filteres were prepared from these plates according to Benton and Davis (1977). These were hybridised overnight with nick-translated insert from pcDd actin B1 in 4

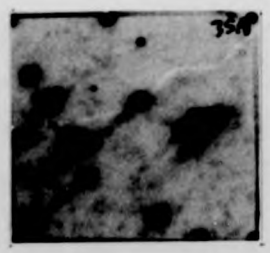
Fig. 3.1

a)



b)

S



W

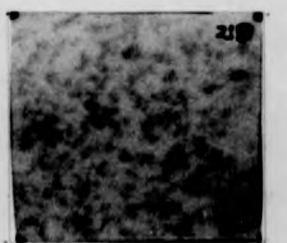
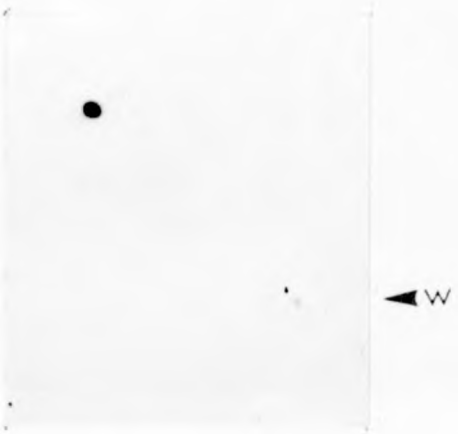


Fig. 3.1

a)



b)

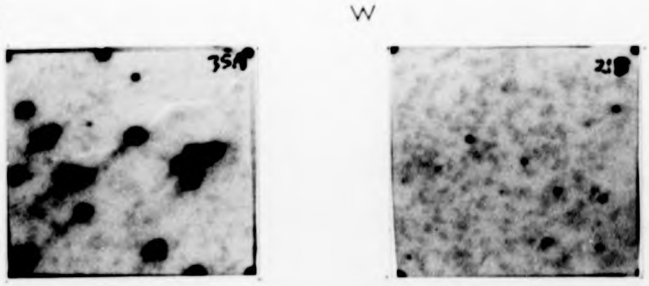
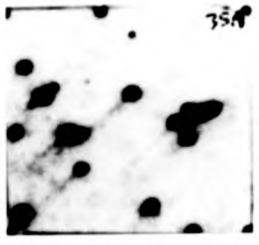


Fig. 3.1

a)

b)

S



W

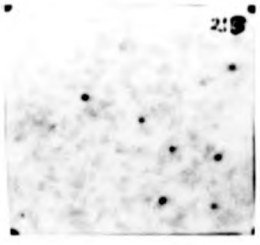


Figure 3.1: First and Second Screens of the Lambda Library

(a): An example of an autoradiograph produced from screening one nitrocellulose filter with nick-translated insert from the Dictyostelium pcDd actin B1 clone. S is an example of a strong positive, and W, a weak positive.

(b): Examples of autoradiographs from a second screening; i.e. the plaque-containing agar portions corresponding to the positives from the first screen, were isolated and the phage from these replated. The plaques from these were transferred to nitrocellulose (see Section 2.13) and rescreened in exactly the same way as for the first screening (the hybridization in 4 x SSC at 60°C, the wash in 4 x SSC at 55°C). Both filters in this example were exposed to X-ray film for the same length of time. The plate giving rise to S was produced from a strong positive, and that giving rise to W, from a weak positive.

x SSC, 1 x Denhardt's, 60 µg/ml sonicated *E. coli* DNA, 100 µg/ml poly(A), at 60°C and washed in 4 x SSC at 55°C, before exposure. An example of the result obtained is shown in Figure 3.1(a). 7 strongly hybridizing plaques were found, as well as roughly 20 weakly hybridizing plaques (the inaccuracy in estimation is due to the difficulty in distinguishing between the latter plaques and artefacts).

(ii) The second screen

All of the strong positives and most of the weak positives were picked in plugs of agar (with the blunt end of a Pasteur pipette), and replated at 100 to 200 plaques per plate. These were then rescreened using the same hybridization and washing conditions as before. Figure 3.1(b) shows examples of filters taken from a replated strong positive and replated weak positive. Clearly, the weakly hybridizing plaque retains this characteristic in the second screening (i.e. the weak positives are not all artefactual). Unfortunately, only four of the strong positives were picked properly and observed in the second screen. I shall refer to these as λ 5AP, λ 13A, λ 35A, and λ 37BP.

Individual plaques corresponding to positives were picked from the second screening, and replated. To ensure clonal purity, single plaques were picked once again, and stored in phage buffer at 4°C. These were the stocks used for subsequent experiments.

3.2 Hybrid-Release Translation (HRT)

3.2.1 Introduction

This section will describe the method used to ensure that the recombinants isolated by the above screen contained genuine actin

sequences, and were not the result of hybridization of non-actin sequences in the probe; such as the stretches of poly dC and poly dG which were left attached to the cDNA insert after cleavage with PstI.

The method of hybrid-release translation is now in wide use (e.g. Cleveland et al., 1980). The technique involves the binding of recombinant DNA to a small piece of nitrocellulose paper, followed by the use of this in a hybridization reaction with an RNA preparation. Only those RNA species complementary to some portion of the recombinant DNA will be retained on the filter at the end of the reaction after the non-hybridizable RNA has been washed off. These bound RNA species are then eluted off the DNA and used as a template for protein synthesis in a cell-free translation system. The protein products of a cloned gene can then be identified by electrophoresing these in vitro translation products on the same polyacrylamide gel as proteins of known size and identity. Of course, to identify a specific cloned gene, the RNA preparation used in the hybridization reaction must contain RNA species homologous to that sequence. In this instance, the cloned DNA was thought to contain actin sequences, but the type of actin was unknown. However, because of the relatively few amino acid differences between the vertebrate muscle isoforms (see Section 1.2), it was hoped that the hybridization of an actin message to a filter-bound actin clone could be achieved, no matter what type of actin gene was involved, by using a skeletal muscle RNA preparation and an ovary RNA preparation (both from X. borealis). The latter cell-type has been shown to contain a considerable amount of beta- and gamma-actin mRNA (Ballantine et al., 1979; Sturgess et al., 1980). The skeletal muscle actin of mammals has been shown to differ by only 4 amino acids from cardiac muscle actin, and only 6 to 8 amino acids from the smooth muscle actins

sequences, and were not the result of hybridization of non-actin sequences in the probe; such as the stretches of poly dC and poly dG which were left attached to the cDNA insert after cleavage with PstI.

The method of hybrid-release translation is now in wide use (e.g. Cleveland et al., 1980). The technique involves the binding of recombinant DNA to a small piece of nitrocellulose paper, followed by the use of this in a hybridization reaction with an RNA preparation. Only those RNA species complementary to some portion of the recombinant DNA will be retained on the filter at the end of the reaction after the non-hybridizable RNA has been washed off. These bound RNA species are then eluted off the DNA and used as a template for protein synthesis in a cell-free translation system. The protein products of a cloned gene can then be identified by electrophoresing these in vitro translation products on the same polyacrylamide gel as proteins of known size and identity. Of course, to identify a specific cloned gene, the RNA preparation used in the hybridization reaction must contain RNA species homologous to that sequence. In this instance, the cloned DNA was thought to contain actin sequences, but the type of actin was unknown. However, because of the relatively few amino acid differences between the vertebrate muscle isoforms (see Section 1.2), it was hoped that the hybridization of an actin message to a filter-bound actin clone could be achieved, no matter what type of actin gene was involved, by using a skeletal muscle RNA preparation and an ovary RNA preparation (both from X. borealis). The latter cell-type has been shown to contain a considerable amount of beta- and gamma-actin mRNA (Ballantine et al., 1979; Sturgess et al., 1980). The skeletal muscle actin of mammals has been shown to differ by only 4 amino acids from cardiac muscle actin, and only 6 to 8 amino acids from the smooth muscle actins

(Vandekerckhove and Weber, 1979).

3.2.2 Results

The isolate, designated λ13A (a strong positive), was grown up in a litre culture to produce several hundred µg of DNA. This was then used for the HRT experiment. Two control types of DNA were also used; these were λ bh2, a recombinant isolated from the same library by Dr. William Bains (Bains, 1982), which was known to hybridize to several histone genes; and pcDd actin B1, the Dictyostelium actin cDNA cloned into pBR322 (the insert from this had been used to screen the library). The latter acted as a positive control.

DNA of each type was loaded onto a pair of filters, 25 µg on each. One filter from each pair was hybridized to 100 µg of unfractionated X. borealis skeletal muscle RNA, and the other to 200 µl of unfractionated X. borealis ovary RNA. The eluted RNA was translated in both the reticulocyte lysate and wheat germ cell-free translation systems, and the products size-separated by SDS-polyacrylamide gel electrophoresis. The wheat germ system was used to detect histones translated from λ bh2 hybrid-released RNA. This is because, in reticulocyte lysates, the large amount of globin synthesized from the remaining endogenous message, as well as the non-radioactive globin, runs in the same position as the histones, and obscures their presence. However, the wheat germ system gave a high background due to endogenous message, and a clearer result (in the region where actin runs) was obtained with reticulocyte lysate.

The fluorographs produced from these gels can be seen in Figure 3.2. Both translation systems show that, when ovary RNA is used, λ 13A hybrid-releases a protein that runs at the same position as the

Fig.3.2a

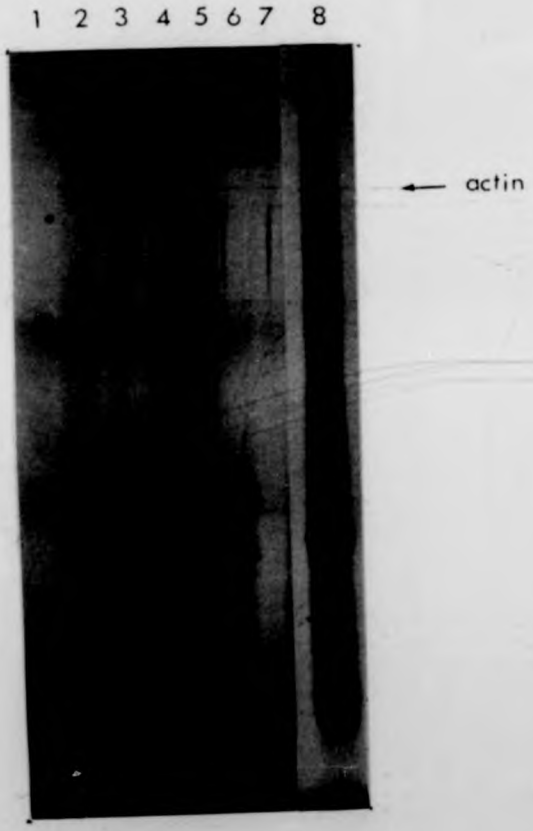


Fig.3.2a

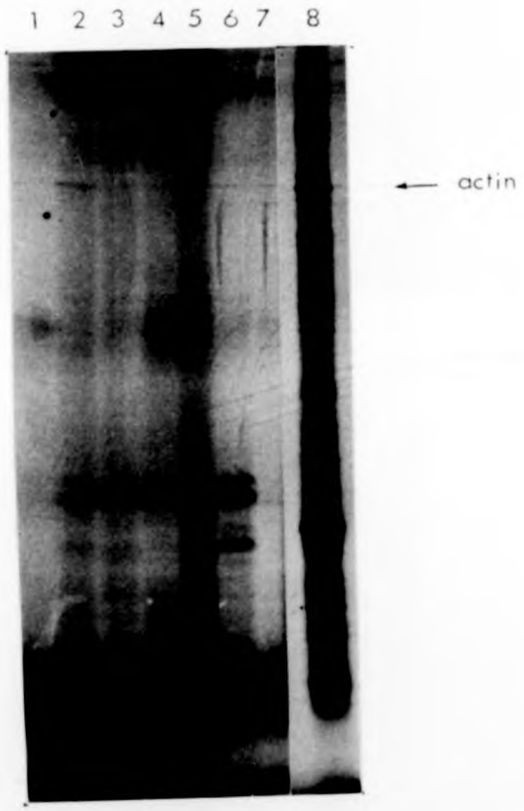


Fig. 3.2a

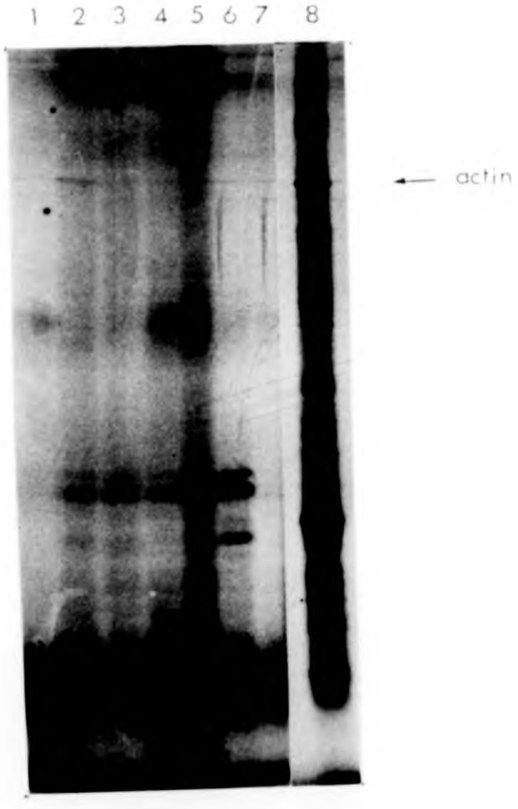


Fig. 3.2b

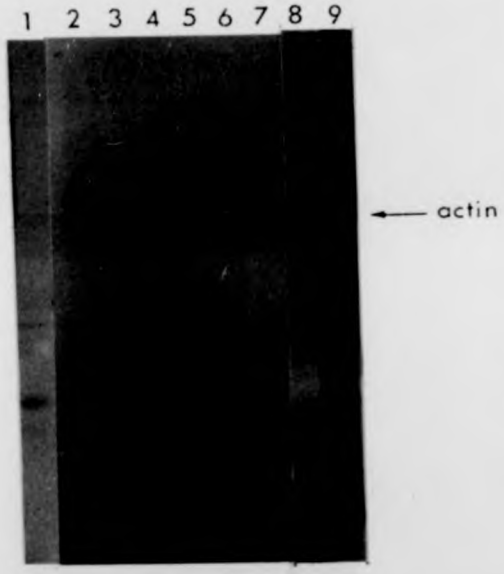


Fig. 3.2b

Fig. 3 2b

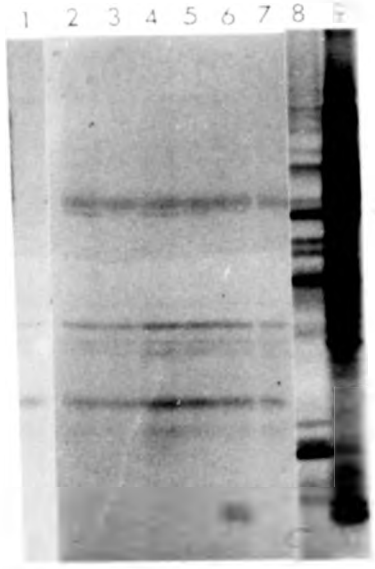


Figure 3.2: Hybrid-Release-Translation Products

Hybrid-released RNA was translated in (a) a message-dependent wheat germ system and (b) a message-dependent reticulocyte lysate system. In the former case the products were electrophoresed on an 18% SDS-polyacrylamide gel, and in the latter case, a 12.5% SDS-polyacraylamide gel. The incorporated ^{35}S-methionine in each case was detected by fluorography.

The products of each translation system without the addition of RNA were run in track 1. Track 8 contains the translation products of the skeletal muscle RNA. Tracks 2 to 7 contain the translation products of RNA eluted from filters: filters 2 and 3 were each loaded with 25 µg of λ 13A, 4 and 5 with 25 µg Dictyostelium actin cDNA clone B1, 6 and 7 with 25 µg of λbh2. Filters 2, 4 and 6 were each hybridized with 200 µg of unfractionated X. borealis ovary RNA, and filters 3, 5 and 7 with 100 µg of unfractionated X. borealis skeletal muscle RNA. Track 9 in (b) contains the translation products of X. borealis ovary poly(A)+ RNA.

The arrow shows the position where an unlabelled rabbit actin marker ran.

unlabelled rabbit actin marker. This was also the size of one of the most abundant translation products of the skeletal muscle RNA and the ovary poly(A)+ RNA. The Dictyostelium actin clone also hybrid-released this protein, while λ bh2 did not. The latter, as predicted, only hybridized to histone RNAs (see track 6 of Figure 3.2a). Neither translation system produced the actin-like protein when no message was added, confirming that this was a genuine product of hybrid-release RNA. It was concluded that λ 13A contained actin-like sequence.

The significance of the stronger actin band produced when ovary RNA was used in the hybridization with λ 13A is not clear. The proportion of actin message contained in each preparation could have been contributory, as well as differing degrees of homology between each actin RNA type and the clone.

The conclusion from this experiment was that the recombinant λ13A probably contained genuine actin sequence, and that it was worth proceeding with the characterization of this, as well as the other three 'strong positives'.

3.3 Comparison of the Four 'Strong Positives'

3.3.1 Introduction

The purpose of the experiments described in this section was to establish how many different actin sequences were present in these clones. The experiments performed with this aim were as follows: the sizing of the cloned EcoRI fragments, the identification of those EcoRI fragments that hybridize to the Dictyostelium actin clone, and the comparison of the 4-cutter restriction di gest patterns of 3 of these fragments. In this way, it was hoped that genes coding for different

isoforms, or highly diverged genes coding for the same isoform, would be detected. This analysis would not necessarily detect allelic differences; a desirable attribute, as the man-hours involved in gene-sequencing would make the sequencing of two alleles extremely wasteful.

3.3.2 Results

What size inserts do the clones have?

DNA produced from clones λ5AP, λ13A, λ35A and λ37BP (all strong positives) and λ 22A (a weak positive) was digested with EcoRl and run on a 0.6% agarose gel next to a Hind III digestion of 'wild type' DNA. Figure 3.3(a) shows the resulting bands. A plot of the marker sizes against the reciprocal of the distance travelled can be used to obtain size estimates for the EcoRl inserts. The two largest bands in each of the EcoRl digested tracks are due to the lambda arms.

λ5AP has a 2.6 kb insert not shared by the others. λ 35A, λ13A and λ 37BP all have a 3.6 kb insert while λ13A also contains a 1.7 kb insert; λ 22A has a 6.7 kb insert. The similarity in the insert sizes of λ 35A and λ 37BP suggested that they probably represent the same DNA sequence (which has been multiplied in the library amplification step). It was decided, therefore, to concentrate on only one of these inserts.

λ 13A contains an extra 1.7 kb fragment, and so presumably arose from a separate cloning event in the formation of the primary library. Although it is obvious that the 3.6 kb fragment must contain sequences homologous to the Dictyostelium probe (because 35A and 37BP contain this fragment alone), it was necessary to decide whether the 1.7 kb fragment would be of interest in this respect. Therefore the EcoRl fragments were transferred to nitrocellulose from the gel shown in Figure 3.3(a) and probed with the Dictyostelium actin clone once more. The result is

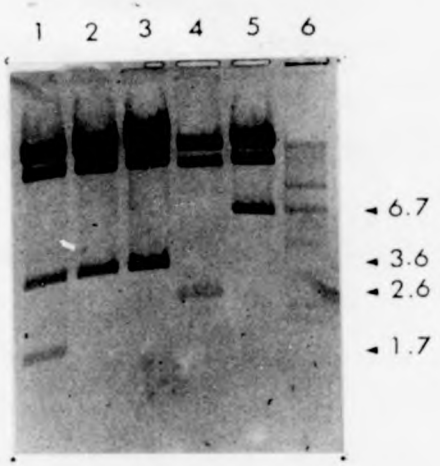
shown in Figure 3.3(b), which has been over-exposed so that the background in track 1 can be seen. Clearly, the probe has not hybridized to the 1.7 kb fragment, even though the same low-stringency conditions were used as in the original screening. The 3.6 and 2.6 kb 'strong positive' fragments have once more hybridized. The 6.7 kb 'weak positive' fragment can also be seen to hybridize, to a lesser extent. Despite this encouraging sign, all further work was confined to 'strong positives'.

Because they were cloned separately, the 3.6 kb inserts of λ13A and λ 35A were investigated to see how similar they were. This was done by isolating the two fragments, mildly nick-translating them (see p63), digesting the labelled DNA with various 4-cutter enzymes, and running the products of the 2 fragments side-by-side on the same 3% agarose gels. These were then dried down onto DEAE cellulose paper and autoradiographed. The products of λ5AP's 2.6 kb insert were also included on the same gels, for comparison. These digests are shown in Figure 3.4, partly for this comparison and also because reference will be made to these fragments in the next section, on restriction mapping.

4-cutter restriction enzymes recognize and cut at sites consisting of 4 bases; e.g. Hpa II cuts at CCGG. They will therefore tend to cut, on average, once every 4^4 (= 256) base pairs, and, so, are useful in comparing the sequences of two large lengths of DNA. With all six 4-cutters used in this experiment, there were no detectable differences between the 3.6 kb inserts of λ13A and λ35A. Also there were few, if any, 4-cutter fragments that were common to both λ 5AP and λ 35A (or λ 13A). Therefore, from this point on, all work was concentrated on the 3.6 kb fragment of λ 35A, and the 2.6 kb fragment of λ 5AP.

Fig. 3.3

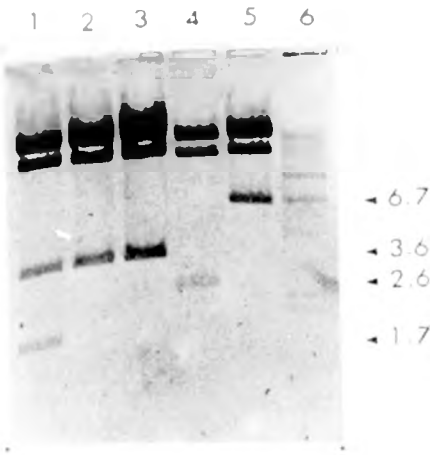
a



b

Fig. 3.3

a



b

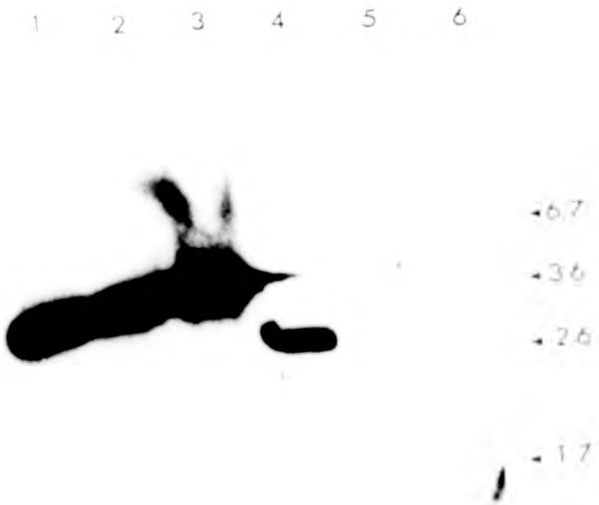


Figure 3.3: Probing the Lambda Inserts with the Dictyostelium Actin Clone

Approximately 1 µg of various recombinant lambda DNAs were digested with EcoRI and the products electrophoresed on a 0.6% agarose gel, containing 1 µg/ml ethidium bromide. The gel was then exposed to ultraviolet light and the resulting DNA bands are shown in (a). The DNA was then transferred to nitrocellulose paper by the Southern blot method and hybridized with nick-translated insert from the Dictyostelium pcDd actin B1 clone. The hybridization and washing conditions were the same as those originally used to screen the library (see Section 3.1.2 and Fig. 3.1). After drying, the nitrocellulose was placed next to X-ray film with an intensifying screen. The resulting autoradiograph is shown in (b).

The tracks contain the EcoRI digests of λ13A (1), λ 35A (2), λ 37BP (3), λ5AP (4) and λ22A (5). Track 6 contained a HindIII digest of wild type lambda as markers. The sizes of the EcoRI inserts are given in kilobases.

Fig. 3.4

a

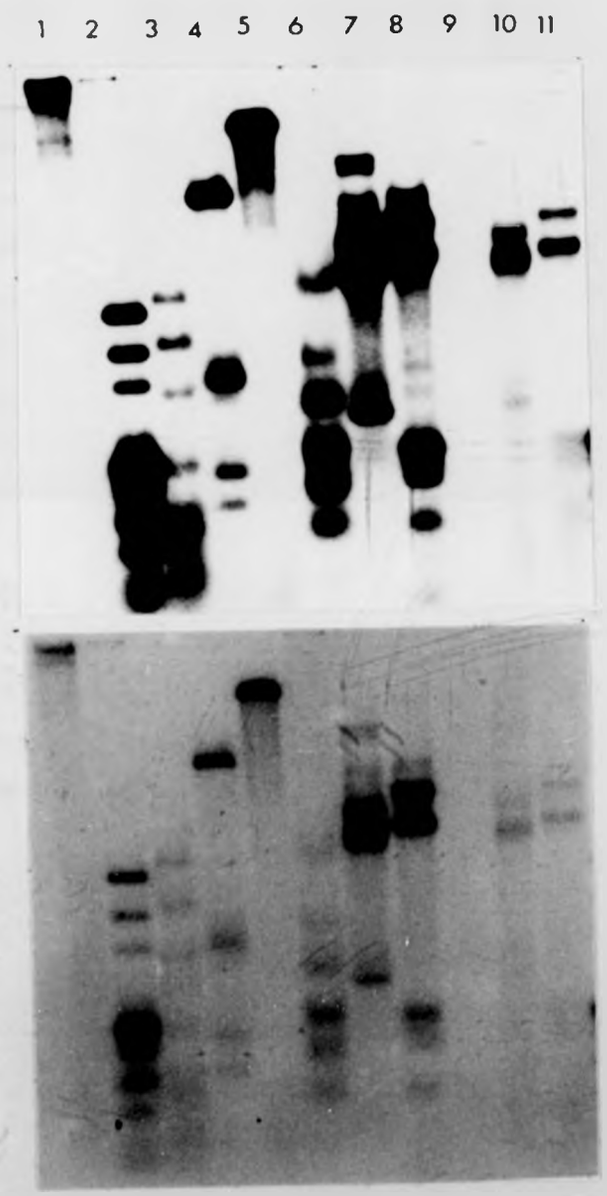


Fig. 3.4

a

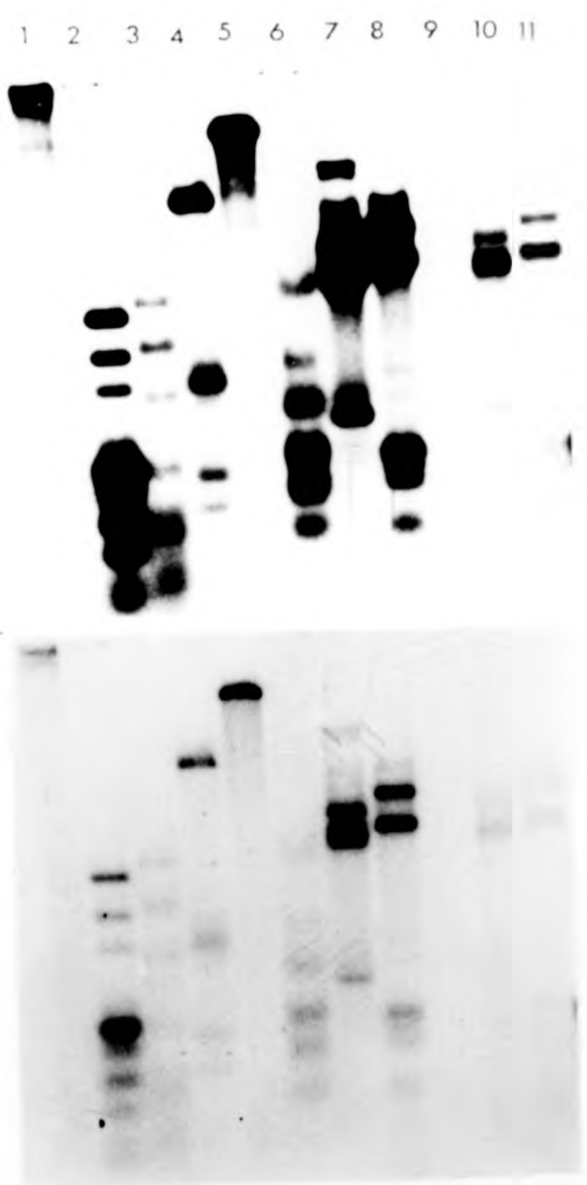


Fig. 3.4

b

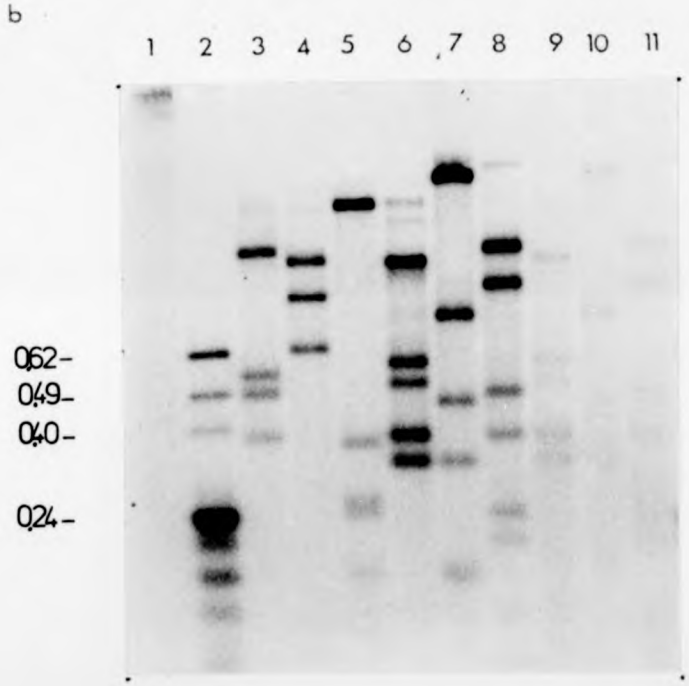


Fig. 3.4

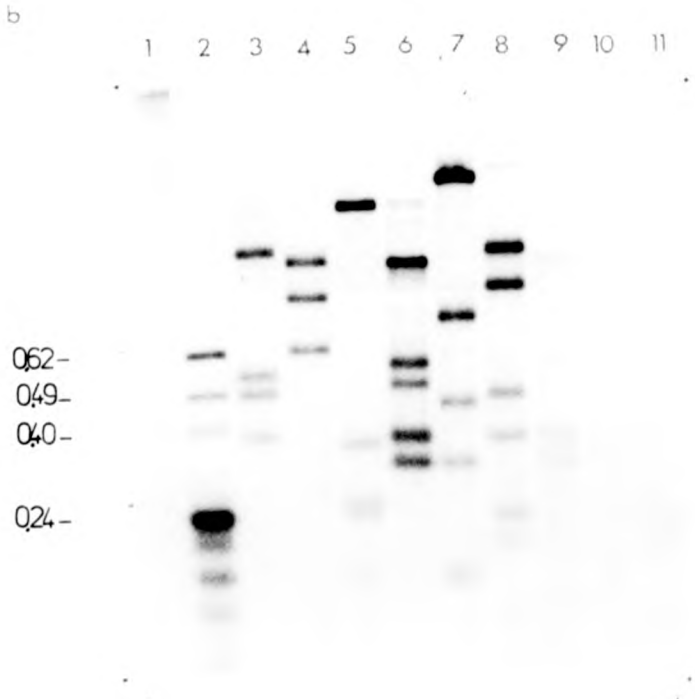


Figure 3.4: Comparing 4-Cutter Digests of 3 Lambda Inserts

λ13A, λ35A and λ5AP were digested with EcoRI, and the 3.6 kb and 2.6 kb inserts gel-purified. These were then mildly nick-translated (see p63) and restricted with various 4-cutter restriction enzymes. The products were electrophoresed on a 3% agarose gel, dried down onto DEAE cellulose paper, and autoradiographed with an intensifying screen. Included on the gels were radioactive markers consisting of a T4 polymerase end-labelled Hpa II digest of pAT153 (track 2) and a T4 polymerase end-labelled HindIII digest of wild type lambda (track 1). In (a), the enzymes used were Alu I (3, 6 and 9), Hae III (4, 7 and 10), and Hha I (5, 8 and 11). Two different lengths of autoradiograph exposure are shown. In (b) the enzymes used were Hinf I (3, 6 and 9), Hpa II (4, 7 and 10), and Rsa I (5, 8 and 11). In (a) and (b) the digestions were done with the λ5AP insert (tracks 3 to 5), the λ13A 3.6 Kb insert (tracks 6 to 8), and the λ35A insert (tracks 9 to 11).

The sizes of these fragments are given in Figures 3.6 and 3.14. The marker sizes given are in kilobases.

3.4 Restriction Mapping and Localization of the Genes

3.4.1 Introduction

Depending on the approach adopted for DNA sequencing, a restriction enzyme map may or may not be an essential prerequisite. If the method of M13 sequencing is adopted initially, there may be no need for such a map. There are several variations in the overall strategy of this technique, but most involve the generation of restriction fragments which are then cloned, either separately or mixed, into the M13 vector. After production of single stranded DNA, the inserted DNA is sequenced by the dideoxy method (see Methods). The objective is to obtain the sequence of all the restriction fragments generated by one enzyme, and then to do exactly the same for a second enzyme. It should then be possible to overlap the sequences obtained using the two different enzymes (in practice, more may have to be used). The essential point is that one need not know the relationship of the restriction fragments to each other before the sequence is obtained.

Such a strategy would have been possible with the 2 clones described here if the technique had been working efficiently in the laboratory at the time. One of its failings is that large amounts of uninteresting sequence may have to be produced before the investigator obtains the sequence of the region he is interested in. Without any preliminary mapping, he will not even know how much of the gene is present on the section of DNA he has cloned. I decided, therefore, to make a detailed restriction map of these clones, to find the position of the actin sequences on these, and then sequence the relevant regions using the method of Gilbert and Maxam. The latter typically involves the end-labelling of a specific restriction fragment, followed by the

cleavage of that fragment with a second enzyme, to produce material labelled at only one end (which is then used in the sequencing reaction; for details, see Methods). There is obviously a need with such a method to have as detailed a map as possible. This is the reason for the work to be described in this section.

There are two basic approaches to restriction mapping in current use. In the first of these, the cloned DNA is digested with an enzyme, and the fragments sized by electrophoresis. The DNA may then be doubly digested with this and a second enzyme, and the products analysed. The second enzyme may cut one or more fragments produced by the first, leaving others intact. By repeating this with many other pairs of enzymes, it may be possible, eventually, to deduce the positions on the whole molecule where each enzyme cuts. There is a drawback with such a procedure, in that, to map the restriction sites of many 4-cutters (which tend to cut very often; once every 256 bases, on average) from several kilobases of DNA, a large mass of data would have to be analysed. Such an analysis is very difficult, unless performed on smaller portions of the DNA.

In the second procedure, the DNA is prepared labelled at one end. It is then partially digested with the test enzyme to generate a mixture of products, which are then separated by electrophoresis and detected by autoradiography (Smith and Birnstiel, 1976). Only those fragments with the labelled end attached will be detected. The smallest band on the autoradiograph will thus correspond to the terminal fragment, the second smallest being the terminal fragment plus its neighbour, and so on. In this way, the distances of the various sites, from the point of labelling, can be worked out. Provided pure end-labelled material can be made, this is a more rapid technique than that of double-digests.

Both of these methods were used in assembling the maps in this study; as well as several variations on these. As the positioning of the actin coding sequence was an integral part of the mapping process, this will be described in this section as well.

3.4.2 Results

Restriction map of λ 35A insert

The 6-cutter map

To provide a framework around which the more numerous 4-cutter sites could be placed, I initially digested nick-translated insert with numerous 6-cutters. Ava I, Bam HI, Sal I, Hind III, Pst I, Hinc II all failed to cut it. Kpn I, Sst I and Xba I each cut the insert once, as shown in Figure 3.5(a). To position these sites, the 3 possible combinations of double digests were performed. The results of these are shown in Figure 3.5(b).

If we fix the Kpn I site at 0.75 kb from one end, the question this experiment is asking becomes: in what order does one find the Xba I and Sst I sites, on proceeding towards the centre of the insert? Examination of Figure 3.5(b) shows that in the double digests it is the 2.1 kb Sst I fragment and the 1.7 kb Xba I fragment that is digested by Kpn I. This gives the order of sites as being Kpn-Xba-Sst, as well as giving the distance between the sites. Figure 3.5(c) illustrates the site positions that are consistent with the double digests. The only inconsistency in Figure 3.5(b) is the lack of 0.4 kb fragment in track 2. This may be attributed to the breakdown of the nick-translated DNA (with the passing of time) which tends to obscure smaller fragments. It should be explained that the largest band in each of tracks 1 to 3 has

been interpreted as being due to a partial digestion, i.e., if the digestion had gone to completion, these bands would have disappeard. In such nick-translated (or unlabelled) material such bands can often be identified because of the greater intensity of smaller bands (the intensity should be roughly proportional to the size of the fragment).

35A: The 4-cutter map

In this section, reference will be made to the various fragments produced by the enzymes on digestion of λ35A insert. The autoradiographs showing these products have been shown in Figures 3.5, but the fragment sizes are summarized, for convenience, in Figure 3.6.

Figure 3.7 shows the results of partial digestions carried out on the two Sst I fragments. The whole Eco RI fragment was end-labelled with T4 DNA polymerase, digested with Sst I, and the two fragments separated on a 1% agarose gel. Figure 3.7(a) shows the partial digestions on the 1.5 kb fragment, and Figure 3.7(b), those on the 2.1 kb fragment. Because DNA extracted from agarose gels tends to be contaminated with DNA smaller in size, there are bands in partials of the 2.1 kb fragment that can be attributed to the 1.5 kb fragment. These were eliminated from the analysis, although it is possible that some may represent genuine sites in the larger fragment.

Some information was obtained by the complete digestion of end-labelled λ 35A insert which had not been cut into two (shown in Figure 3.7c). This gave the end fragments of some enzymes. Fortuitously, the two ends of the EcoRI fragment labelled to differing extents with T4 DNA polymerase, and the end fragments could each be assigned to one end. The interpretation of all of these experiments is summarized in Figure 3.9.

Confirmation of some of this information was achieved with the double digests shown in Figure 3.8. The position of the 1.05 kb Hinf I fragment was confirmed because it was cut by Sst I, and not by Xba I (or Kpn I); see Figure 3.8(a). Similarly, the 0.9 kb Rsa I fragment and the 1.85 kb Hpa II fragment were cut only by Sst I; see Figure 3.8(b).

Some extra information was generated by the double digests shown in Figure 3.8(b). The largest (0.9 kb) Hae III fragment was cut by Sst I but not by the other 6-cutters. One of the two 0.85 kb Hha I bands is cut by each of the 6-cutters (it is known from the end-labelling experiments that the 1.05 kb fragment is at one end and cannot be cut by these enzymes). As it is known that one 0.85 kb fragment is an end fragment, the other 0.85 kb Hha I fragment can now be positioned. These additions are shown as dotted lines in Figure 3.9.

Where, on this map, is the actin sequence?

Having constructed this map, the region containing the actin sequence was located. Insert from λ 35A was digested with various 6- and 4-cutter enzymes, the products separated on a 3% agarose gel, and blotted onto nitrocellulose. This was then probed with insert from the Dictyostelium actin cDNA clone under the same, low stringency hybridization and washing conditions as used in the original screening. The resulting autoradiograph is shown in Figure 3.10. The hybridizing region is clearly confined to the small 1.5 kb Sst I fragment. Neither the 2.1 kb Sst I nor the 1.7 kb Xba I fragment has 'lit up'. All the hybridizing bands in Figure 3.10 are consistent with this localisation. Clearly this was the region in which to start sequencing.

Fig.3.5

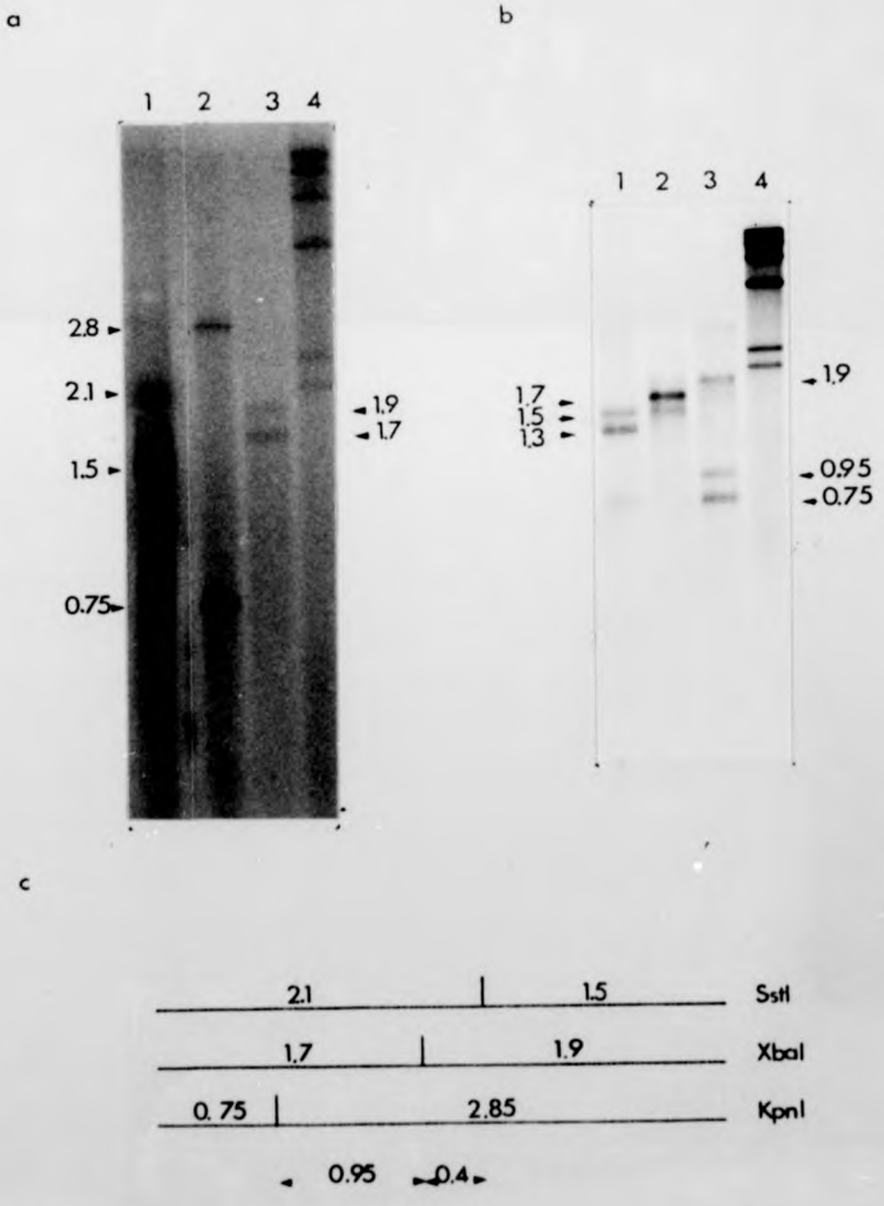
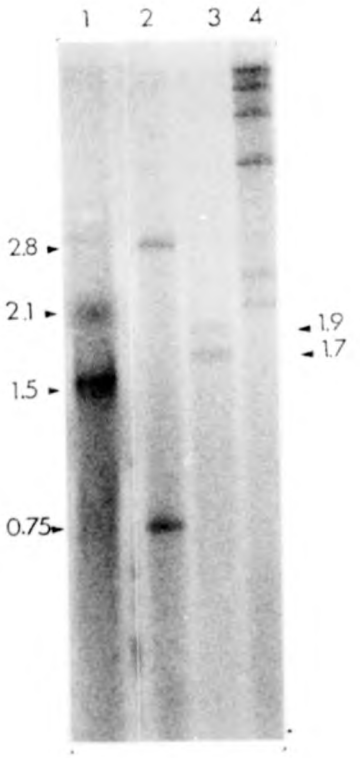
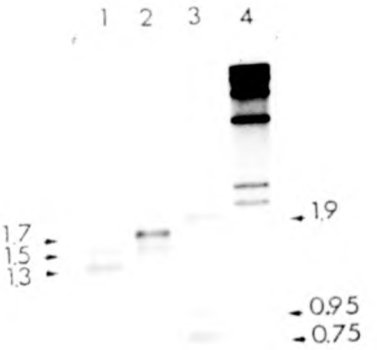

a
b
1 2 3 4
2.8
2.1
1.5
0.75
1.9
1.7
1 2 3 4
1.7
1.5
1.3
1.9
0.95
0.75
c
2.1
1.5
SstI
1.7
1.9
XbaI
0.75
2.85
KpnI
0.95
0.4

Fig. 3.5

a


1
2
3
4
2.8
2.1
1.5
0.75
1.9
1.7

b


1
2
3
4
1.7
1.5
1.3
1.9
0.95
0.75

c

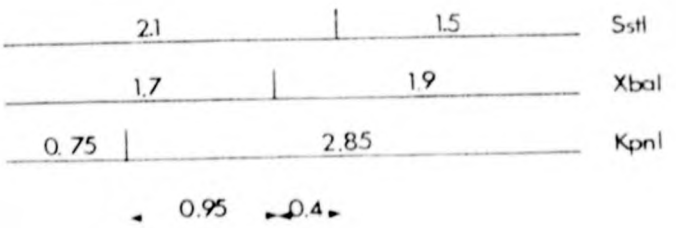

2.1
1.5
SstI
1.7
1.9
XbaI
0.75
2.85
KpnI
0.95
0.4

Fig. 3.5

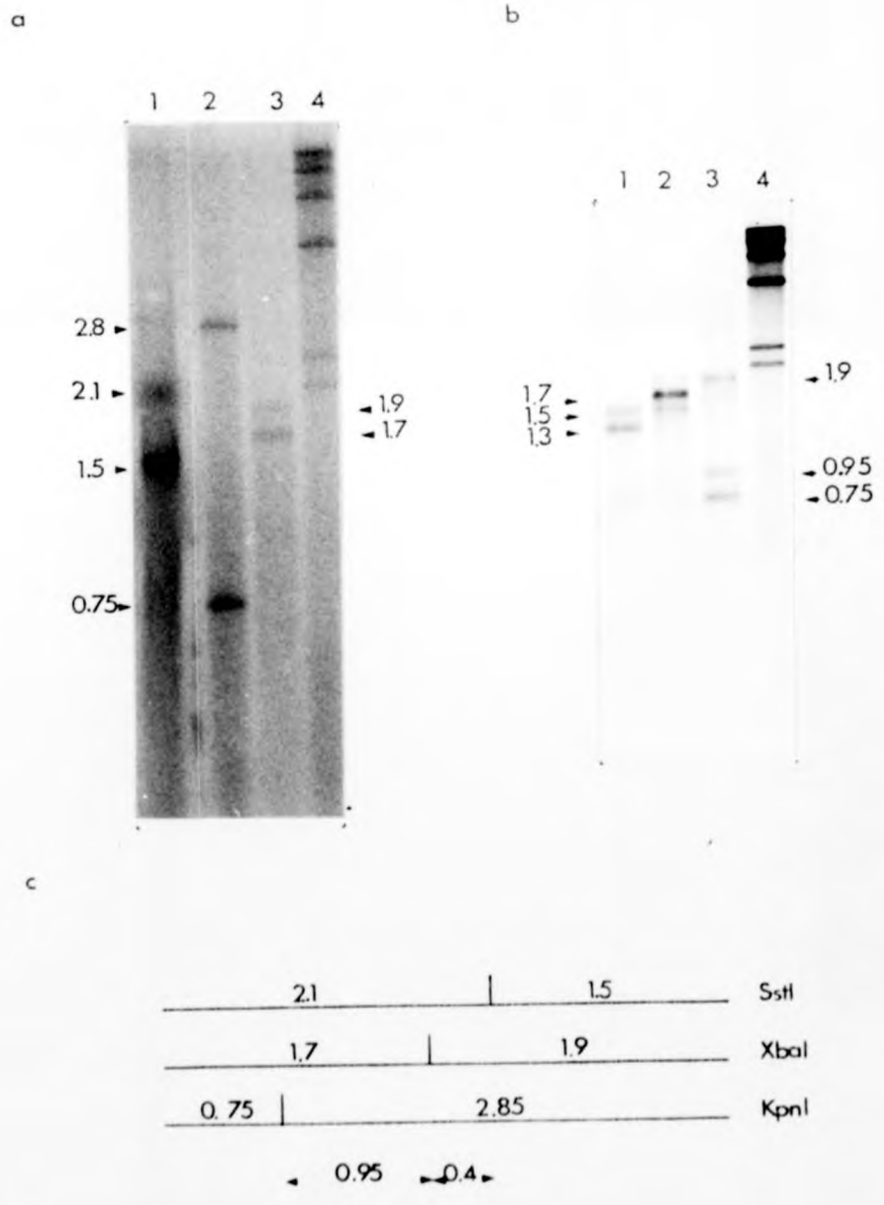
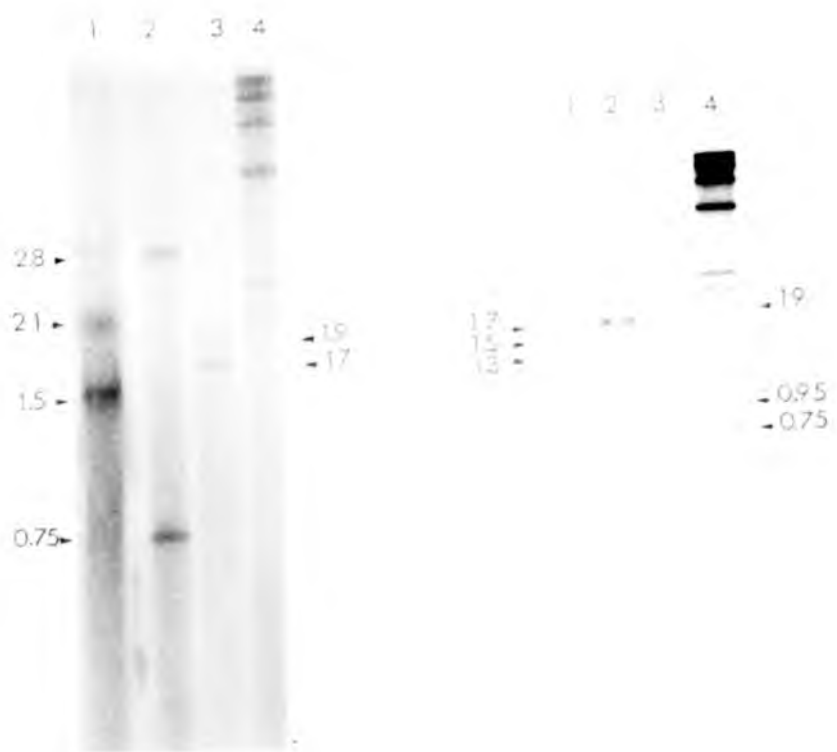

a
1 2 3 4
2.8
2.1
1.5
0.75
1.9
1.7
b
1 2 3 4
1.7
1.5
1.3
1.9
0.95
0.75
c
2.1
1.5
SstI
1.7
1.9
XbaI
0.75
2.85
KpnI
0.95
0.4

Fig.3.5



c

2.1 | 1.5 — SstI

1.7 | 1.9 — XbaI

0.75 | 2.85 — KpnI

0.95 0.4

Figure 3.5: 6-Cutter Digests and Double-Digests on 35A

λ35A DNA was digested with EcoRI and the 3.6 kb insert gel-purified. This was digested with various enzymes and the products electrophoresed on 1.5% agarose gels. These were dried onto DEAE cellulose paper and autoradiographed with screens. The enzymes used in (a) were Sst I (1), Kpn I (2) and Xba I (3). The enzymes used in (b) were Sst I and Kpn (1), Sst I and Xba I (2), and Xba I and Kpn I (3). Track 4, in each case, contained the radioactive λHind III markers (see Appendix I for the sizes of these). The numbers refer to the sizes, in kilobases, of the bands.

The positions of the enzyme sites, that are consistent with these digests, are shown in (c). The numbers represent the distance, in kilobases, betwen various sites.

Fig. 3.6 35A restriction fragments

HinfI	HpaII	HhaI	HaeIII	RsaI
1.05	1.85	1.05	0.9	1.1
0.6	0.75	(0.85) × 2	0.85	0.9
0.5	0.5	0.25	(0.8) × 2	0.5
(0.4) × 2	0.3	0.2	0.2	0.4
(0.35) × 2	0.15	0.18		0.25
		0.15		0.2
				?

AluI	SstI	XbaI	KpnI
0.7	2.1	1.9	2.85
0.5	1.5	1.7	0.75
0.37			
0.23			
0.15			
?			

Figure 3.6: Restriction Fragment Sizes of 35A

Restriction fragment sizes (in kilbases) of the λ 35A 3.6 kb Eco RI insert. In the case of Rsa I and Alu I there are definitely many small fragments that cannot be separated and sized with the gel system used.

Rsa I
1.1
0.9
0.5
0.4
0.25
0.2
?

Fig. 37

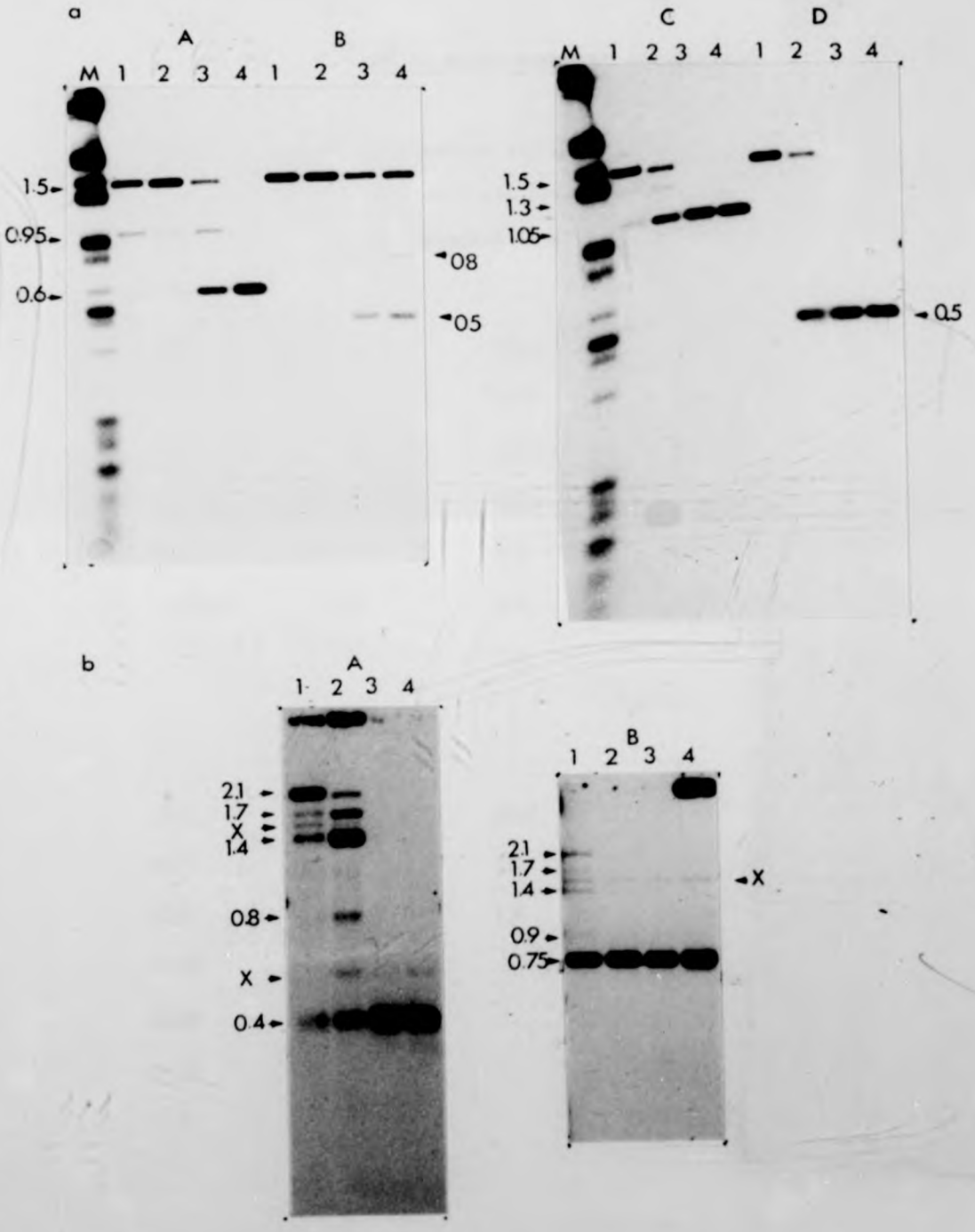

a
A
B
C
D
M 1 2 3 4 1 2 3 4
M 1 2 3 4 1 2 3 4
1.5
0.95
0.6
0.8
0.5
1.5
1.3
1.05
0.5
b
A
1 2 3 4
2.1
1.7
X
1.4
0.8
X
0.4
B
1 2 3 4
2.1
1.7
1.4
X
0.9
0.75

Fig. 3.7

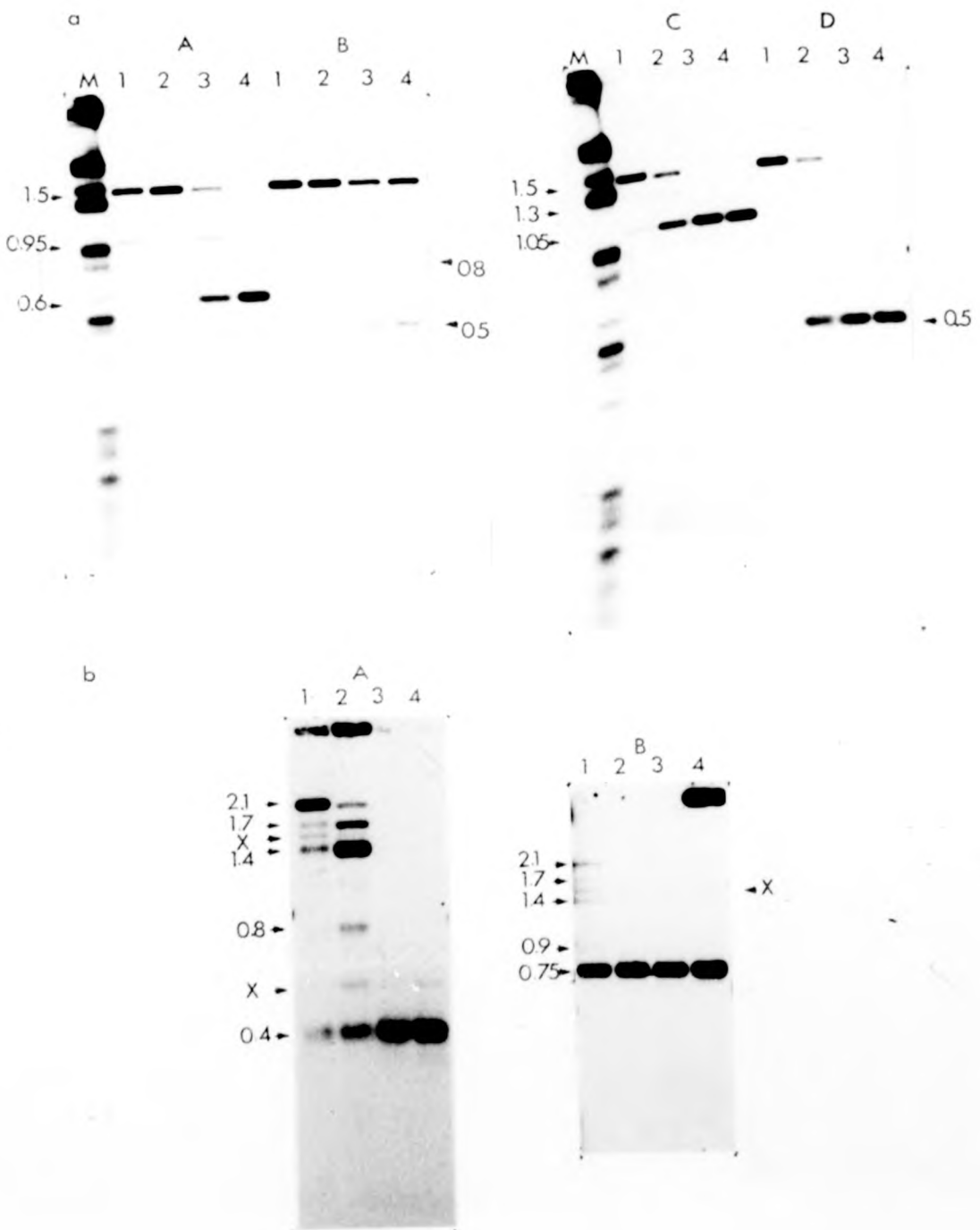

a
A
B
C
D
M 1 2 3 4 1 2 3 4
M 1 2 3 4 1 2 3 4
1.5
0.95
0.6
0.8
0.5
1.5
1.3
1.05
0.5
b
A
1 2 3 4
2.1
1.7
X
1.4
0.8
X
0.4
B
1 2 3 4
2.1
1.7
1.4
X
0.9
0.75

Figure 3.7(a) and (b): Partial Digests on End-Labelled Sst I Fragments of 35A

Partial digests on end-labelled Sst I fragments of 35A. The 3.6 kb EcoRI fragment ofλ35A was end-labelled using T4 DNA polymerase I and then digested with Sst I. The 2.1 kb and 1.5 kb fragments were separated by electrophoresis on a 1% agarose gel and extracted. (a) shows the results of partial digestion of the end-labelled 1.5 kb Sst I fragment with Sau 3AI (A), Rsa I (B), Hha I (C) and Hinf I (D). (b) shows the results of partial digestion of the (impure) 2.1 K fragment with Hinf I (A) and Hpa II (B). For each enzyme, 1 µg of unlabelled DNA was added to each of 4 tubes, together with a small amount of the relevant labelled fragment. The tubes were incubated with the relevant enzyme for 0 minutes (1), 5 minutes (2), 30 minutes (3), or 120 minutes (4), before the reactions were stopped by the addition of Tris-acetate sample buffer, followed by heating to 70°C for 10 minutes. The radiolabelled markers (a mixture of a Hinf I digest of pAT153 and an EcoRI/Hind III digest of wild-type lambda DNA - see Appendix I) were run in the tracks labelled M. In (b), X marks the bands thought to be contaminants from the 1.5 Kb Sst I fragment. The band sizes are in kilobases.

Fig. 3.7

c

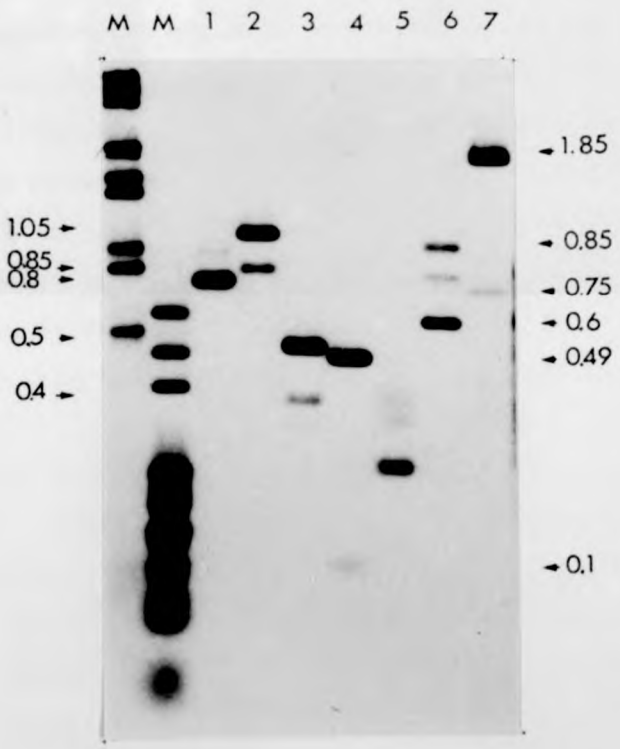


Fig. 3.7

c

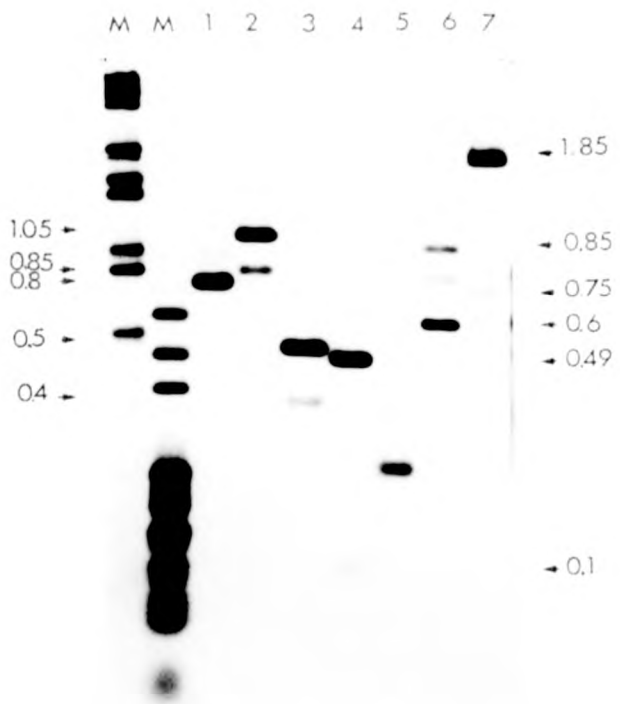
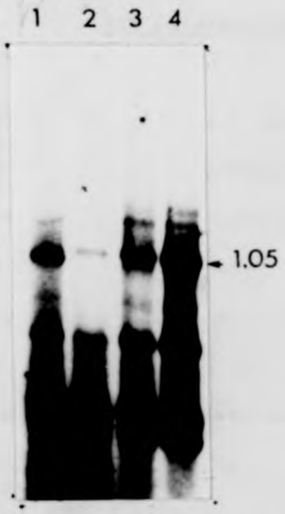


Figure 3.7(c): Total Digests on End-Labelled 35A

Total digests on the end-labelled 3.6 kb EcoRI insert of λ35A. The 3.6 kb EcoRI insert of λ35A was end-labelled with T4 DNA polymerase I, and after digestion with various enzymes, the products were separated by electrophoresis on a 3% agarose gel. This was then dried onto DEAE cellulose paper and autoradiographed with an intensifying screen. The enzymes used were HaeIII (1), Hha I (2), Hinf I (3), Rsa I (4), Alu I (5), Sau 3Al (6) and Hpa II (7). The markers (in the tracks labelled M) were as in (a) and (b) except that the pAT153 and lambda digests were run in separate tracks.

Fig. 38

a



b

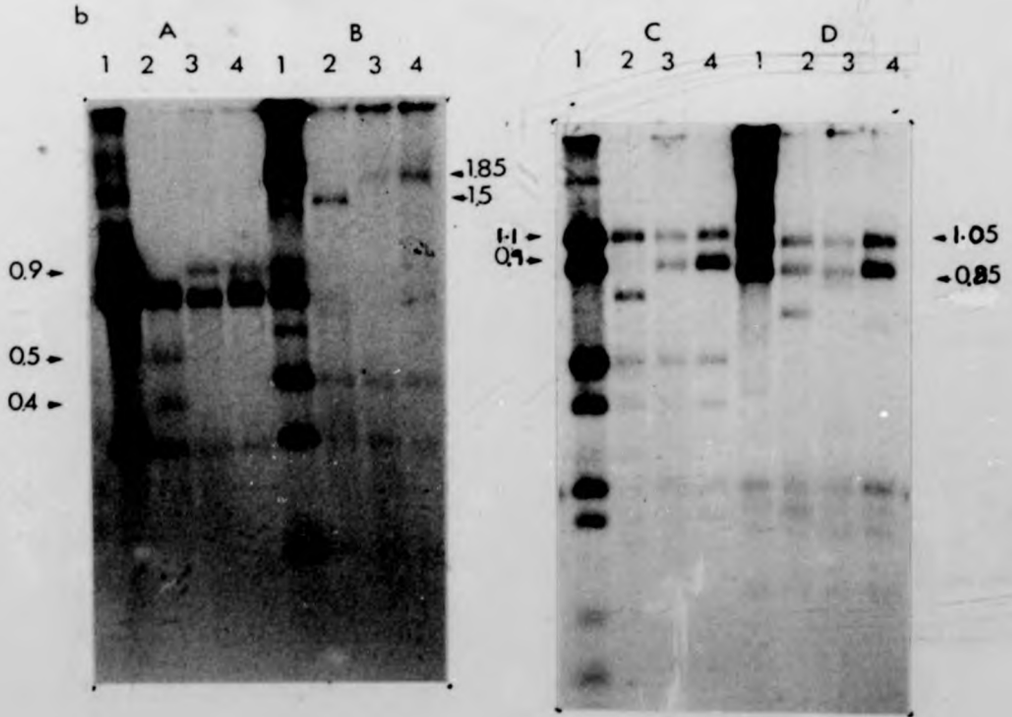
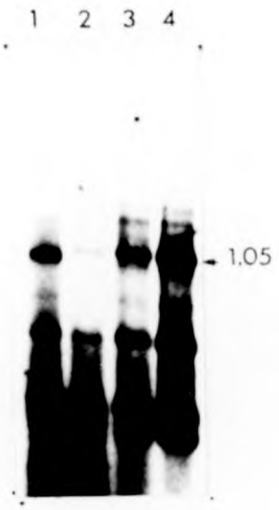


Fig. 38

a



b

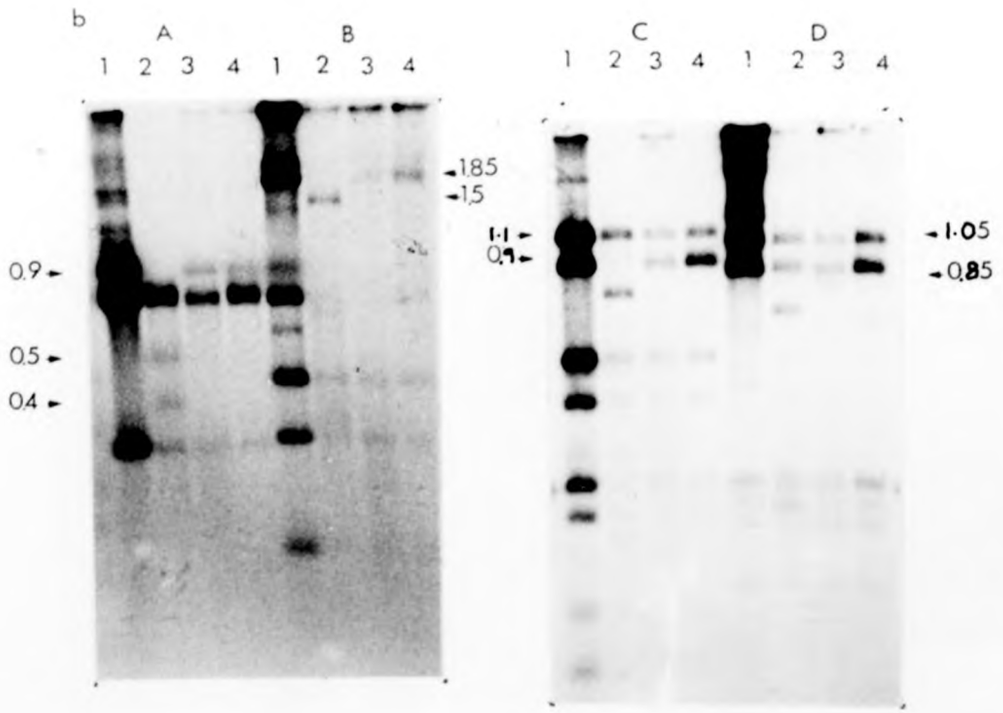


Figure 3.8: Double Digests on 35A

(see p63)

The 3.6 kb Eco RI fragment of λ 35A was mildly nick-translated and digested with various combinations of enzymes. The products were electrophoresed on 3% agarose gels, dried onto DEAE cellulose paper, and autoradiographed with intensifying screens. In (a), the enzymes used were: Hinf I + Sst I (1), Hinf I + Kpn I (2), Hinf I + Xba I (3), and Hinf I alone (4). In (b), the labelled Eco RI insert was first digested with either Sst I (2), Kpn I (3) or Xba I (4). Each of these was then digested with Hae III (A), Hpa II (B), Rsa I (C), and Hha I (D). Track 1 in each case was digested with the 4-cutter enzyme alone. The sizes are in kilobases.

-1.05
-0.85

Fig. 3.9 35A : 4_cutter map

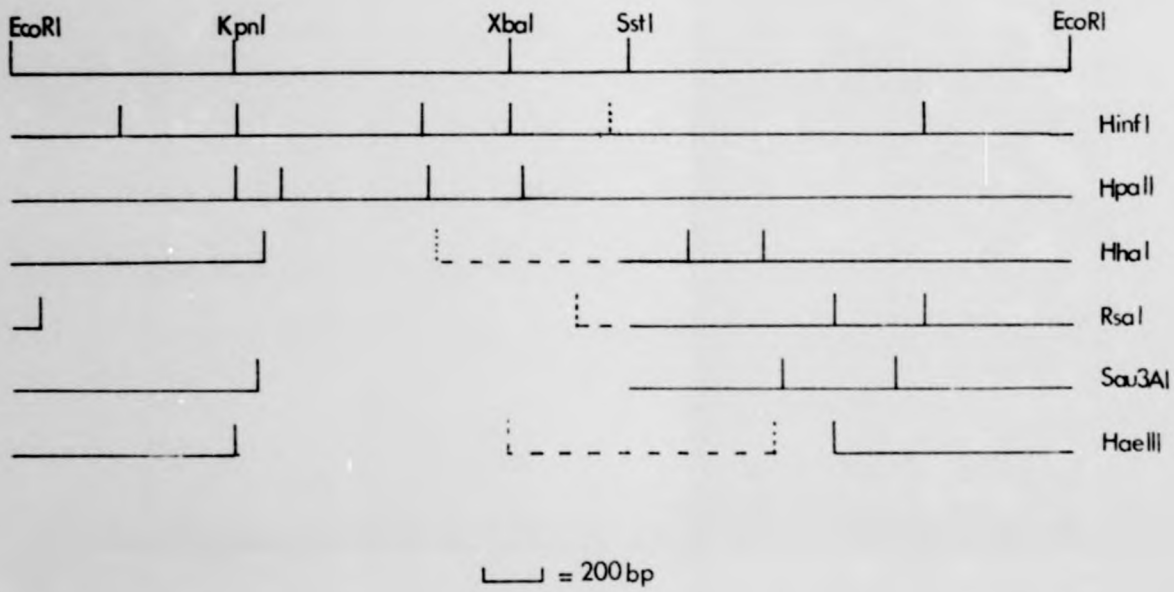


Figure 3.9: Restriction Map of 35A

The interpretation of the gels shown in Figures 3.7 and 3.8. Most of the sites are derived from the end-labelling work,but the positions of restriction fragments drawn with dotted lines were determined by double digests. Additions to this map, as a result of sequencing, are shown in Fig. 3.20.

Fig. 3.10

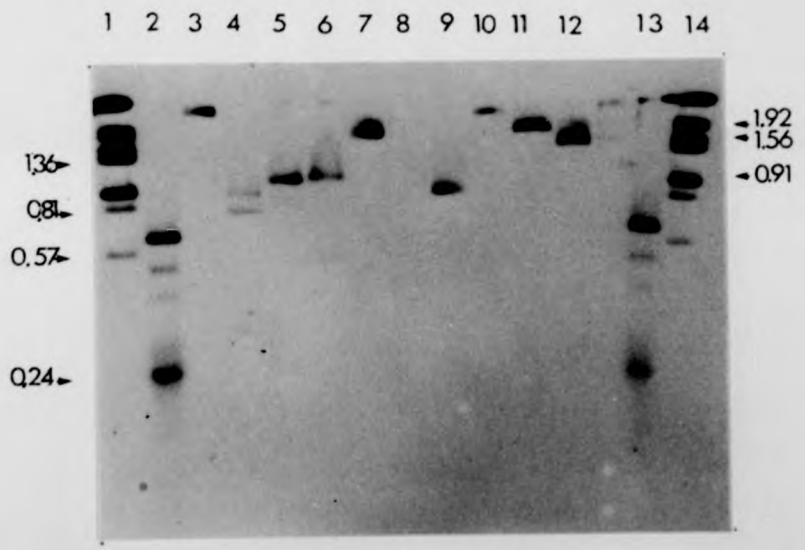


Fig. 3.10

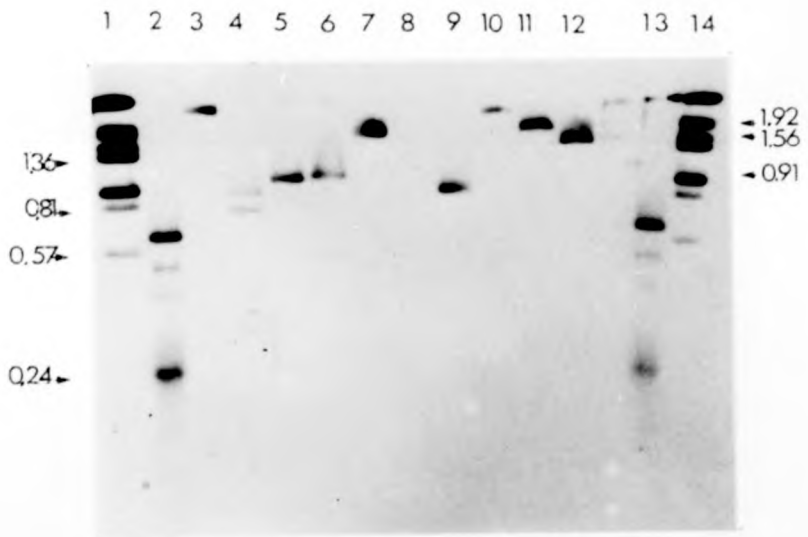


Fig. 3.10

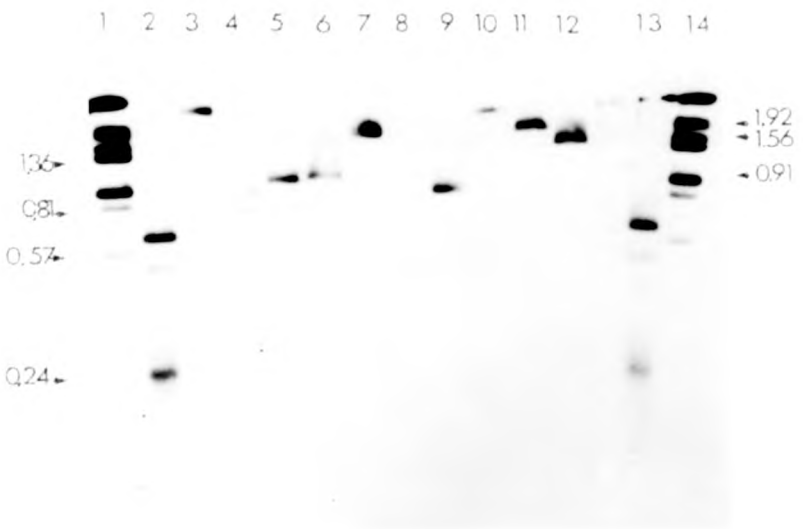


Figure 3.10: Locating Actin Sequence on the 35A Map

The 3.6 kb Eco RI insert from λ35A was gel-purified and digested with Hae III (4), Hha I (5), Hinf I (6), Hpa II (7), Alu I (8), Rsa I (9), Kpn I (10), Xba I (11), and Sst I (12). The products were separated on a 3% agarose gel and Southern-blotted onto nitrocellulose. The latter was then probed with ^{32}P-nick-translated insert from the Dictyostelium pcDd actin B1 clone under the same hybridization conditions as the original screening (see Fig. 3.1). The dried filter was then autoradiographed with an intensifying screen. Tracks 1 and 14 contained an end-labelled Eco RI/Hind III digest of wild-type lambda, 2 and 13, an end-labelled Hpa II digest of pAT153, and 3, undigested EcoRI fragment. The marker sizes are in kilobases (see Appendix 1 also).

Restriction Map of λ5AP Insert

The 6-cutter map

The insert of λ5AP was mildly nick-translated, and incubated with various 6-cutters. Ava I, Bam HI, Sal I, Sst I, Kpn I, Xba I, and Acc I, all failed to cut it. Figure 3.11(a) shows the products produced by digestion with Pst I, Hind III and Hinc II and Figure 3.11(b), the products of the three double digestions possible with these enzymes. As there is one Pst I site, it is easiest to orientate everything around this. Clearly, the largest Hind III fragment (1.05 kb) is the one cut by Pst I. Reference to the total digestion of end-labelled insert, shown in Figure 3.12, shows that the 1.25 kb and 0.8 kb Hind III fragments are on the outside; while the 0.45 kb fragment is unlabelled and so must be in the interior. Similarly, Figure 3.12 shows the 0.95 kb and 0.19 kb Hinc II fragments to be labelled, and therefore are on the ends. Figure 3.11 shows that it is the large (1.2 kb) Hinc II fragment that is cut by Pst I. The order of six-cutter sites is shown in Figure 3.13(a). The position of the 0.21 kb Hinc II fragment, adjacent to the 0.175 kb, can be deduced from the fact that the 1.2 kb Hinc II fragment is cut by Hind III into 0.4 kb and 0.8 kb fragments (and not into 0.2 kb and 1 kb fragments).

The 4-cutter map

As with 35A, the total digests on nick-translated insert with various 4-cutters have already been shown in Figure 3.4. The fragment sizes are summarized, for convenience, in Figure 3.14. Figure 3.12 shows the results of T4 polymerase I end-labelling the 2.6 kb Eco RI insert of λ5AP, and then digesting with various 4 cutters. Only the outermost fragments should remain labelled after a total digestion. In

the case of Hpa II, these are 1.1 kb and 0.85 kb in size, and in the case of Hinf I, 0.54 kb and 0.39 kb.

It was at this point in making the map of 5AP that the localization of the actin sequence was used in constructing the map itself, and it is therefore more convenient to describe the two processes in an integrated fashion.

The 5AP insert was digested with various enzymes, and the products ran out on a 3% agarose gel. These were then blotted onto nitrocellulose and probed with insert from the Dictyostelium clone (exactly the same experiment as with 35A). The results are shown in Figure 3.15. The interpretation of this experiment is more complicated than the equivalent one performed with 35A.
Nevertheless, the actin sequence is clearly localized on the larger Pst I fragment, and hybridizes most strongly to the 0.95 kb Hinc II fragment. If it is accepted that the right hand end of the insert (see Fig. 3.13) hybridizes the most strongly, then one can place the 0.54 kb Hinf I, and the 0.85 kb Hpa II, fragments at this end as well. This fixes the Hpa II sites with respect to Figure 3.13(a), but still leaves the orientation of the central two Hinf I fragments in doubt. This was clarified by the experiments shown in Figure 3.16. If the 1.15 Hinf I fragment is adjacent to the 0.54 kb fragment, then the 0.64 kb Hpa II fragment would not be cleaved in a Hpa II/ Hinf I double digest (nor would the 0.49 kb Hinf I fragment). However, Figure 3.16(a) shows that the 0.64 kb fragment is cut in this digest. Therefore, it is the 0.49 kb Hinf I fragment that is adjacent to the 0.54 kb Hinf I fragment. This orientation is confirmed by the fact that it is the 1.15 kb Hinf I fragment that is cut by Pst I (see Figure 3.16(b)). The resulting map and hybridising regions are shown in Figure 3.13(b). In Figure 3.15,

Fig. 3.11

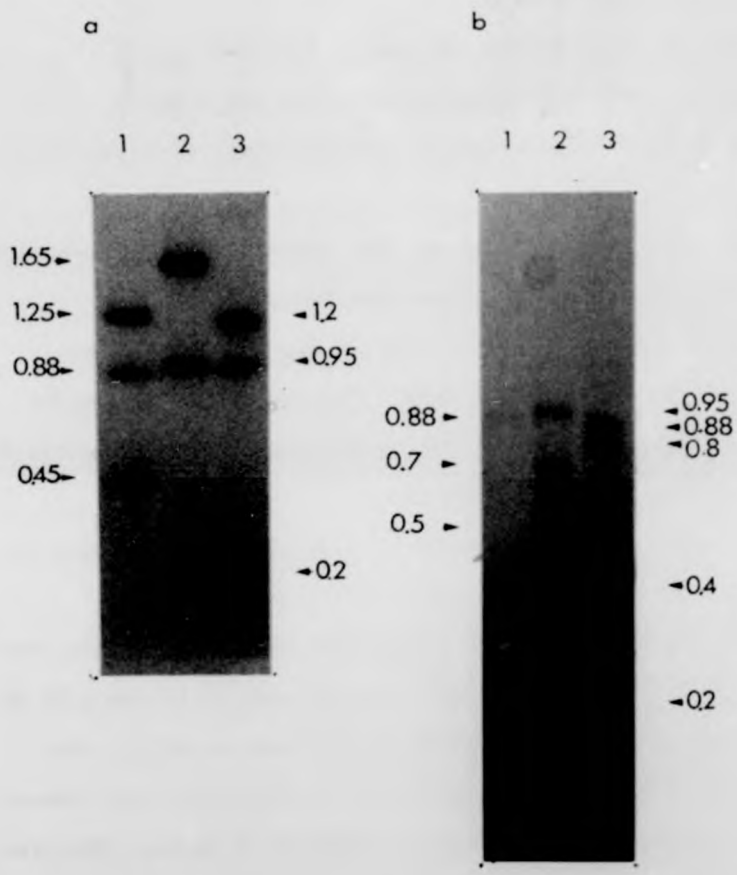


Fig. 3.11

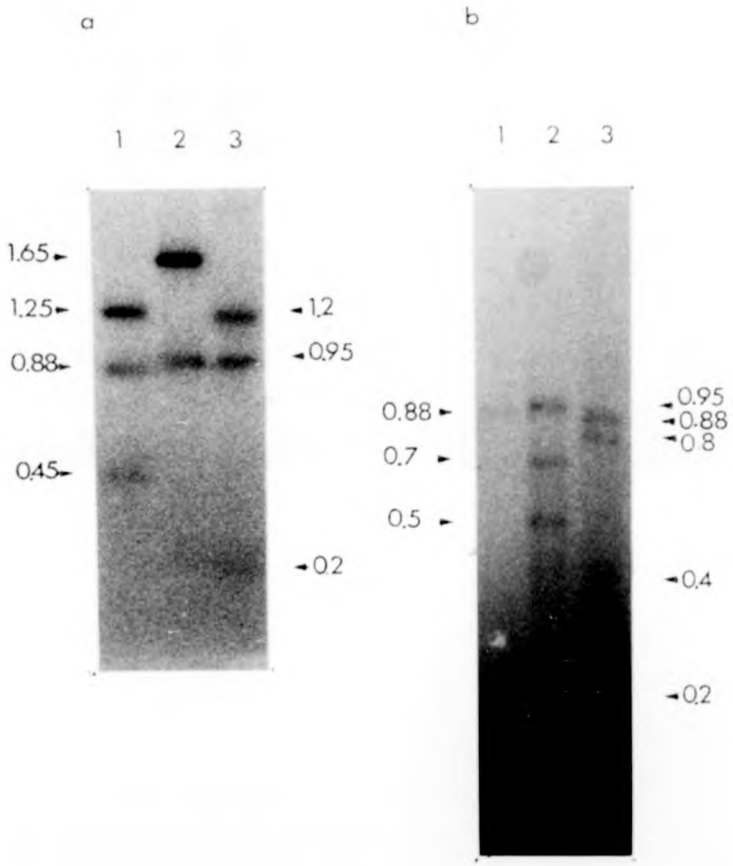


Fig. 3.11

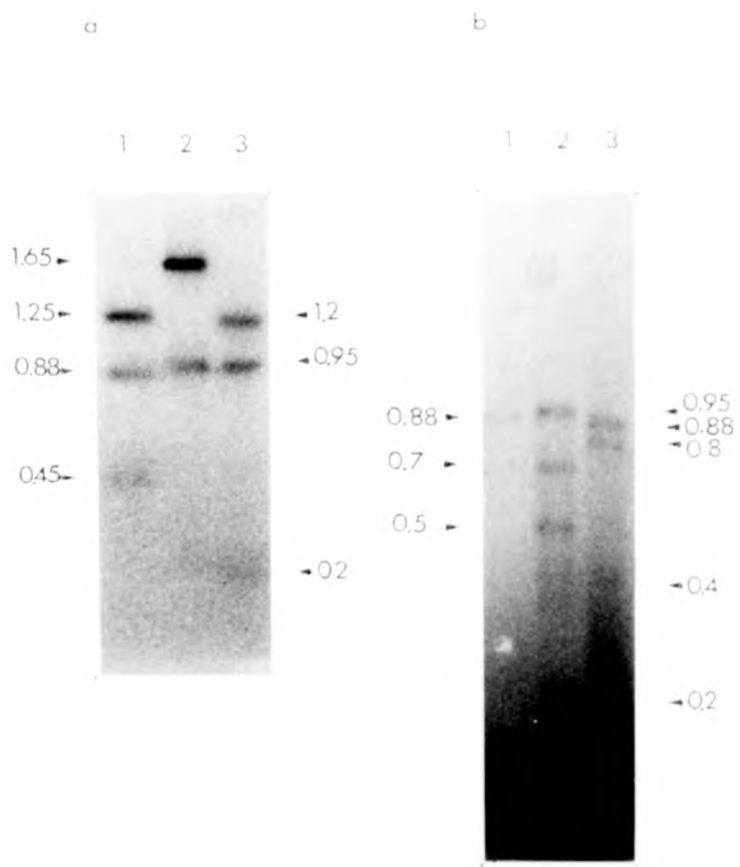


Figure 3.11: 6-Cutter Digests and Double-Digests on 5AP

The 2.6 kb insert of λ 5AP was mildly nick-translated and digested with various 6-cutter restriction enzymes. The products were electrophoresed on (a) a 1.5% agarose gel, and (b) a 3% gel. These were then dried onto DEAE-cellulose paper and autoradiographed with intensifying screens. For (a), the enzymes used were: Hind III (1), Pst I (2), and Hinc II (3). For (b), the enzymes used were: Pst I + Hind III (1), Pst I + Hinc II (2), and Hind III + Hinc II (3). The band sizes are given in kilobases.

Fig.3.12

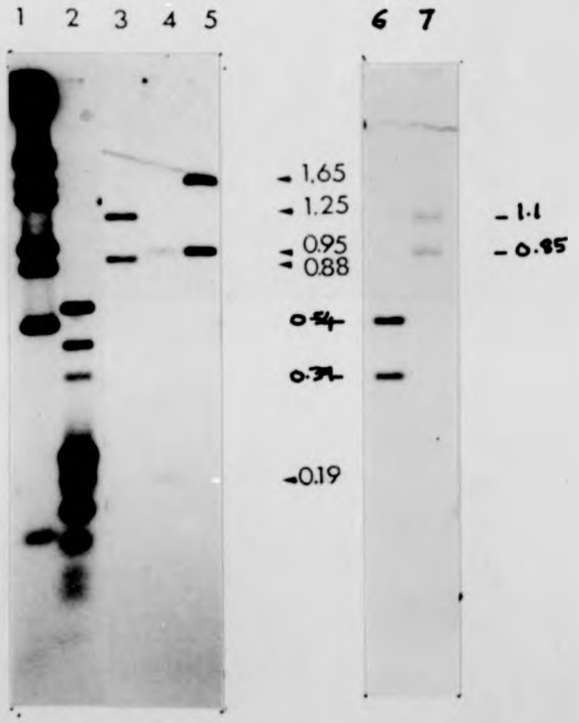


Fig. 3.12

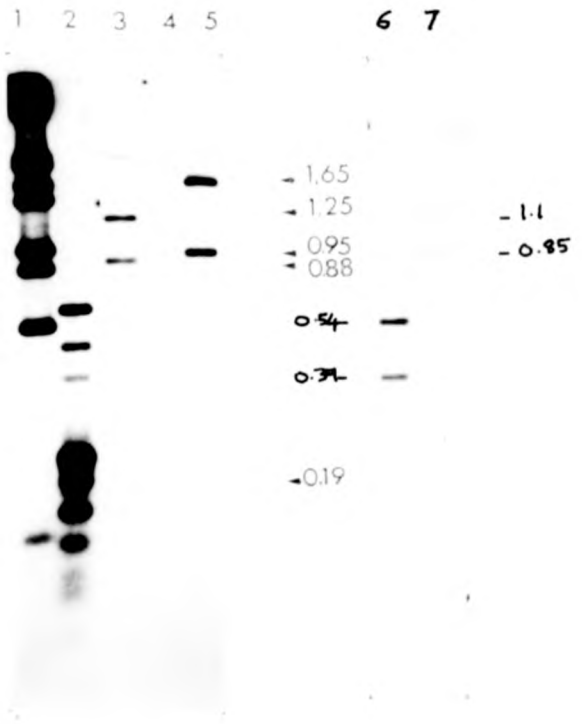


Figure 3.12: Total Digests on End-Labelled 5AP

The 2.6 kb insert of λ 5A P was end-labelled with T4 DNA polymerase I and then digested with Hind III (3), Hinc II (4), and Pst I (5), Hinf I (6) and Hpa II (7). The products were electrophoresed on a 3% agarose gel, dried onto DEAE-cellulose paper and autoradiographed with an intensifying screen. Track 1 contained a radiolabelled Eco RI/Hind III wild-type lambda digest, and track 2, a radiolabelled Hpa II pAT153 digest (see Appendix 1 for the sizes of these markers). The band sizes are given in kilobases.

Fig.3.13 SAP: 6_cutter and 4_cutter maps

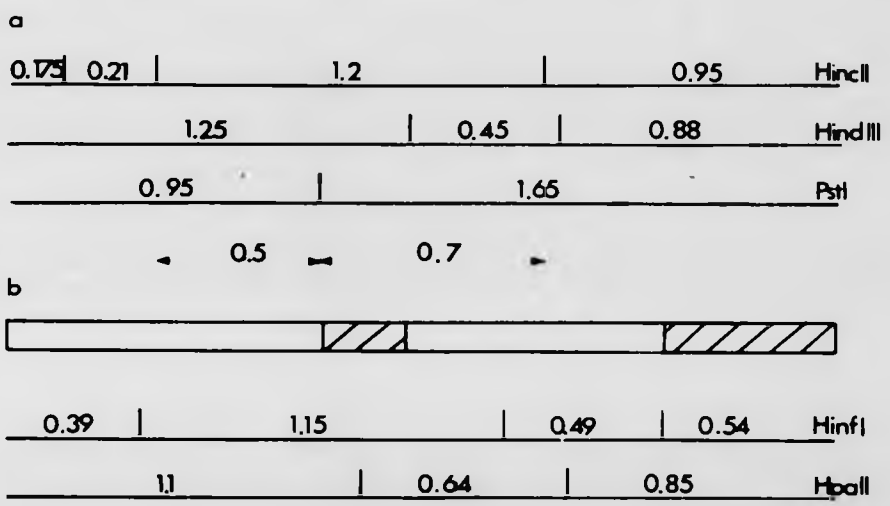


Figure 3.13: Restriction Map of 5AP

The 6-cutter restriction map of 5AP (a) together with the 4-cutter map and the regions that apparently hybridized to the Dictyostelium pcDd actin B1 insert (b). The 6-cutter map results from the digests shown in Figures 3.11 and 3.12, and the 4-cutter map, from digests shown in Figures 3.15 and 3.16.

Fig. 3.14 5AP restriction fragments

HinfI	HpaII	HaeIII	AluI
1.15	1.1	1.35	0.71
0.54	0.85	0.4	0.51
0.49	0.64	0.35	0.37
0.39		0.2	0.23
		0.17	0.15
		?	?

RsaI	HindIII	HincII	PstI
1.55	1.25	1.2	1.65
0.38	0.88	0.95	0.95
0.27	0.45	0.21	
0.25		0.19	
0.17			

Figure 3.14: Restriction Fragment Sizes of 5AP

The sizes (in kilobases) of the fragments generated on digestion of the λ5AP Eco RI insert with various restriction enzymes. There are no Hha I sites.

Fig. 3.15

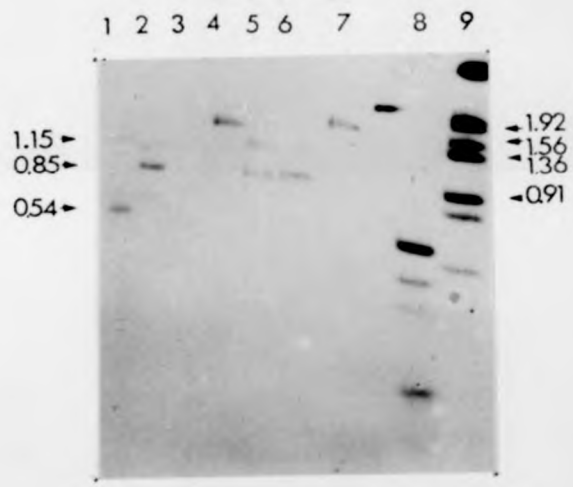


Fig. 3.15

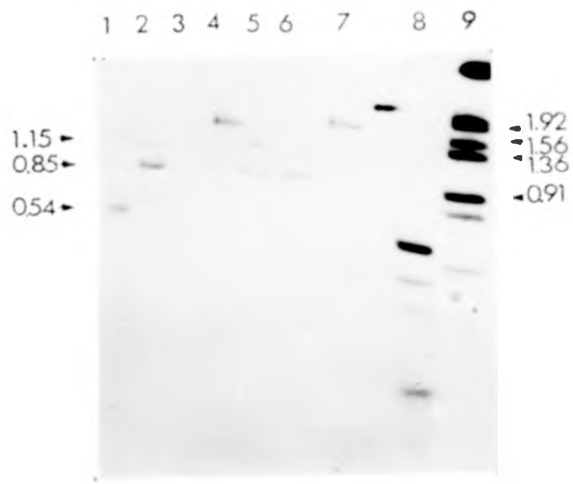


Fig. 3.15

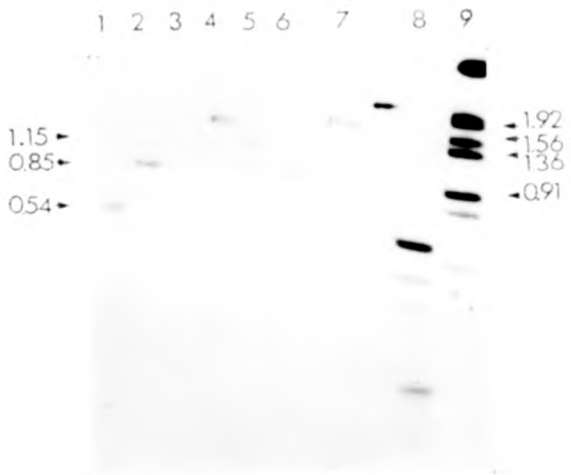


Figure 3.15: Localization of Actin Sequence on the 5AP Map

The Eco RI insert of λ5AP was digested with Hinf I (1), Hpa II (2), Alu I (3), Rsa I (4), Hind III (5), Hinc II (6), and Pst I (7). The products were electrophoresed on a 3% agarose gel and Southern-blotted. The nitrocellulose filter was then probed with nick-translated insert from the Dictyostelium actin pcDd actin B1 clone (in the same hybridization conditions as used to screen the library - see Section 3.1.2 and Figure 3.1). The dried filter was autoradiographed with an intensifying screen. Track 9 contained a radiolabelled EcoRI/Hind III digest of wild type lambda, and track 8, a labelled Hpa II pAT 153 digest (see Appendix 1 for the marker sizes).

Fig 3.16

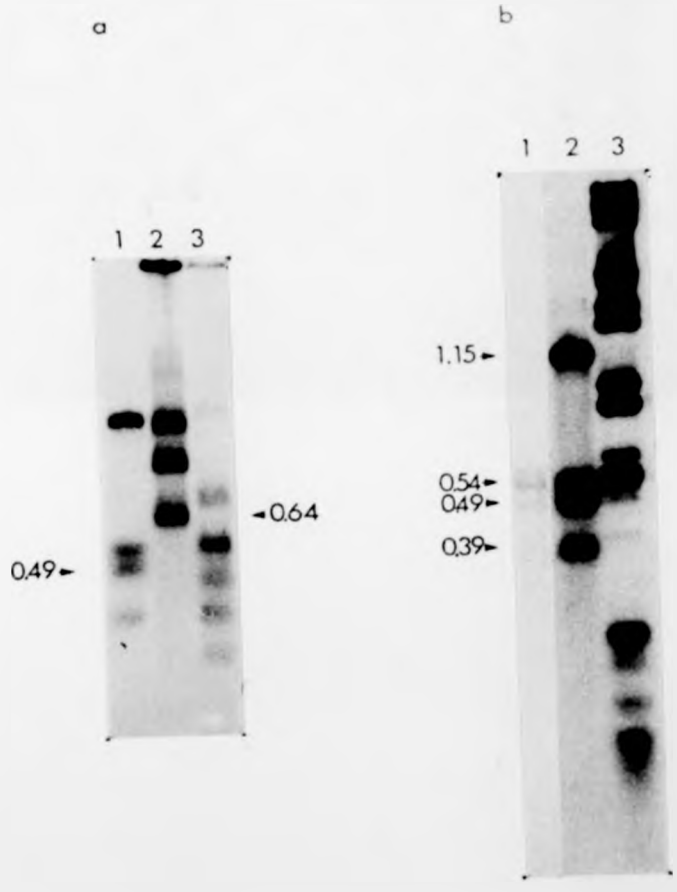


Fig 3.16

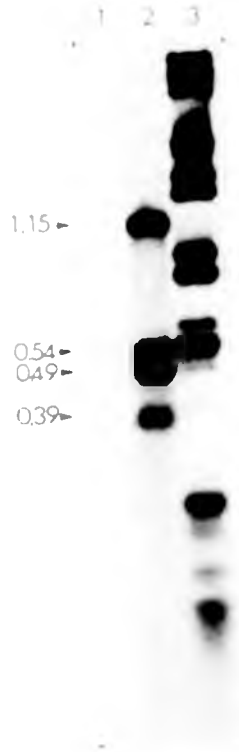


Figure 3.16: Double-Digests on 5AP

The 2.6 kb Eco RI insert of λ5AP was mildly nick-translated and digested with (a) Hinf I (1), Hpa II (2), and Hpa II + Hinf I (3); and (b) Hinf I + Pst I (1), and Hinf I (2). Track 3 in (b) is a marker track consisting of a mixture of a radiolabelled EcoRI/Hind III digest of wild type lambda and a radiolabelled Hpa II digest of pAT 153 (see Appendix 1 for the marker band sizes). The DNA was electrophoresed on 3% agarose gels and dried onto DEAE-cellulose paper. This was autoradiographed with an intensifying screen. The sizes of several bands discussed in the text are shown (in kilobases).

the probe does not seem to have hybridized to the 0.95 kb Pst I fragment, yet has hybridized to the 1.25 kb Hind III fragment, leaving the 0.45 kb Hind III fragment unhybridized. This suggested two separate actin coding regions on the clone, and the sequencing that followed was concentrated, initially, on these areas.

3.5 Recloning into pBR325

For greater convenience of DNA preparation, the 3.6 kb Eco RI insert of λ35A and the 2.6 kb Eco RI insert of λ5AP were recloned into the Col E1-derived vector pBR325 (Bolivar, 1978; Prentki et al., 1981). In each case, gel-purified insert (approx. 0.2 μg) was ligated with Eco RI-linearized pBR325 (approx. 0.5 μg) and transformed into the E. coli strain HB101. The plasmid pBR325 contains a gene that confers resistance to the drug ampicillin. Consequently, the bacteria were spread onto ampicillin-containing plates to select for those containing plasmids. These colonies were then picked onto chloramphenicol- and ampicillin-containing plates. The unique Eco RI site in pBR325 is in the middle of the chloramphenicol-resistance gene, and insertion into this site inactivates the gene (Bolivar, 1978). Therefore, those colonies that grew on ampicillin plates, but not on chloramphenicol plates, contained recombinant plasmids. Such colonies were picked from the ampicillin plates and grown up on 10 ml cultures. DNAs prepared from these were restricted with Eco RI and the products run on agarose gels. Those clones that gave rise to the same size Eco RI band as the λ 35A or λ5Ap inserts, used as markers, were subsequently grown up (and shown to give the same restriction products as the inserts of λ35A and λ 5AP; data not shown). These clones will be referred to as p5AP and

p35A. For simplicity, I will often refer to the 2.6 kb and 3.6 kb Eco RI inserts from these clones as 5AP and 35A, respectively.

3.6 Sequencing the Clones

Having obtained a rudimentary map and the rough positions of the actin sequences, I began to sequence the clones. This usually involved the Maxam and Gilbert (1975) method, but on several occasions a restriction fragment was cloned into the appropriate M13 vector and sequenced by the dideoxy method developed by Sanger *et al* (1977). Both of these methods are described in detail in Section 2.

When using Maxam and Gilbert sequencing one must be able to generate end-labelled material. Three methods of doing this were used in this study. Two different DNA polymerase enzymes were used. One of these, the Klenow fragment of *E. coli* DNA Polymerase I, was used only rarely because it was frequently found to produce internal labelling of the fragment. (This may have been a property of the particular batch of enzyme in use at the time.) Far more frequently used, was T4 DNA polymerase I, which was free from this problem. Both of these enzymes can add labelled nucleotides, in a 5' to 3' direction, onto an end produced by a restriction enzyme leaving a 5' overhang; the latter acting as the template. I have also used the 3' exonuclease, or proof-reading, activity of the T4 enzyme, on several occasions, to produce labelled ends for sequencing (e.g. labelling Eco RI and Hinf I sites at the G position). The third method used here exploits the ability of T4 polynucleotide kinase to add a radioactive phosphate group onto the 5' end of a 5' overhang, once the original phosphate has been removed by a phosphatase (see Methods).

Whatever the method used, the end-labelled fragment was usually cut with another restriction enzyme so as to produce fragments unequal in size that could easily be separated on gels. (Although, on several occasions, the initial fragment was such that it was possible to label one end only; e.g. Hpa II - Eco RI, Hind III - Eco RI, and some Hinf I - Eco RI fragments.) The fragments were then subjected to the sequencing chemistry, and the products run on sequencing gels, as described in Methods. Figures 3.17 and 3.19 give the detailed sequencing strategies for both clones.

The regions of DNA sequenced, together with the resulting restriction maps and gene structures are shown in Figures 3.18 and 3.20.

Fig.3.17 The sequencing strategy for SAP

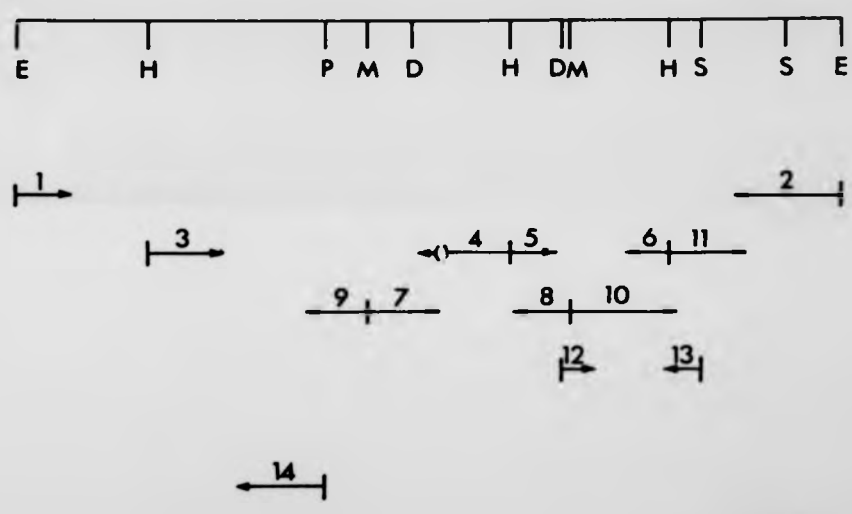


Figure 3.17: Sequencing Strategy for 5AP

A map of 5AP showing the sequencing strategy. Most of the regions were sequenced by the Maxam and Gilbert method. The exception, labelled M13, was sequenced using the method of Sanger et al. (1978) - see Section 2.13. The restriction enzyme sites for Eco RI (E), Hinf I (H), Pst I (P), Hind III (D), Hpa II (M) and Sau 3Al (S) are shown.

Sections 1 to 12 were end-labelled using T4 DNA polymerase. For sections 1 and 2 the 2.6 kb Eco RI fragment was labelled with ^{32}P dGTP, and then cut with Hinc II. For sections 3 and 4 the 1.15 kb Hinf I fragment was labelled with ^{32}PdGTP, and similarly for 5 and 6 using the 0.49 kb Hinf I fragment, both being cut with Hpa II. The 0.64 kb Hpa II fragment was used for sections 7 and 8 and the 1.1 kb Hpa II/Eco RI fragment for section 9. Both were labelled with ^{32}P dCTP and cut with Hinf I. For section 10 the 0.85 kb Hpa II-Eco RI fragment was labelled only at the Hpa II end, using ^{32}P dCTP and unlabelled dATP. The 0.54 kb Hinf I-Eco RI fragment was used for section 11, being labelled only at the Hinf I end using ^{32}P dGTP and unlabelled dATP. The same labelling method was used to label only the Hind III end of the 0.34 kb Hind III/Hinf I fragment, to obtain section 12. Section 13 used the 2.15 kb Sau 3AI fragment which was labelled with Klenow fragment and ^{32}P dGTP before being cut with Hinf I.

Section 14 was sequenced by cloning the 0.95 kb Pst I- Eco RI fragment into an M13 mp8 vector, which had previously been restricted with Eco RI and Pst I, and sequencing by the Sanger method.

Fig. 3.18 5AP

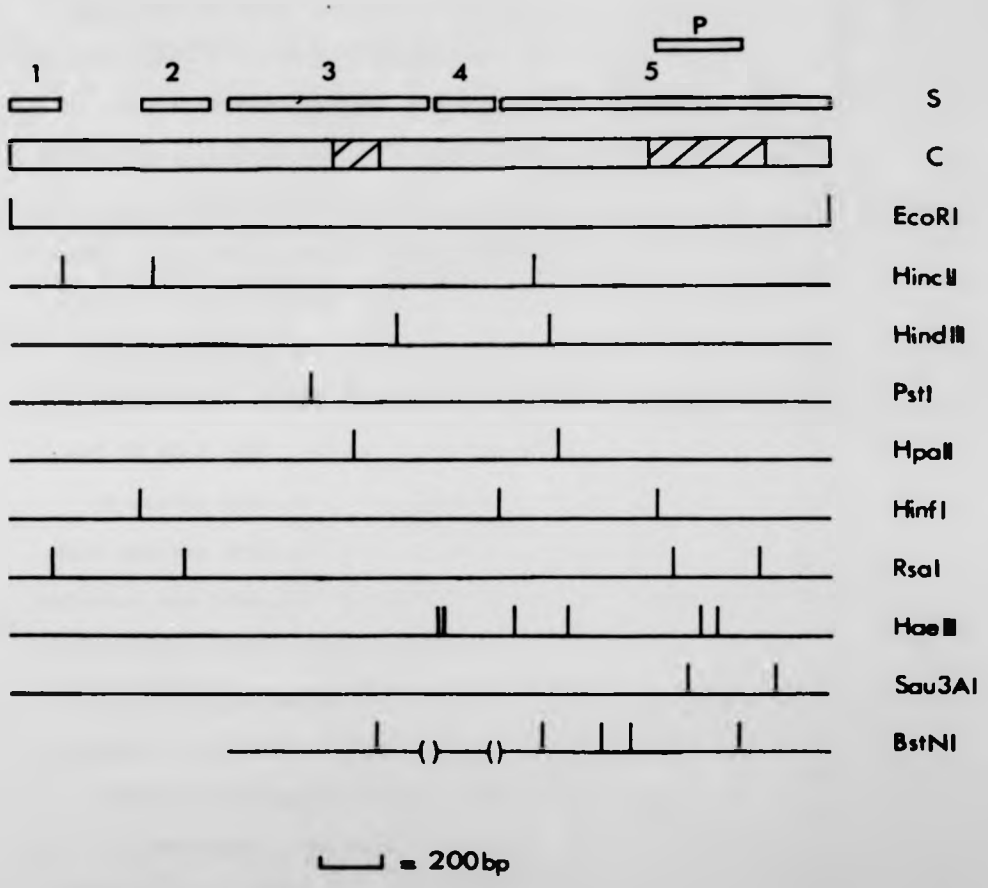

P
1
2
3
4
5
S
C
EcoRI
HincII
HindIII
PstI
HpaII
HinfI
RsaI
HaeIII
Sau3AI
BstNI
= 200bp

Figure 3.18: Detailed Restriction Map of 5AP

A more detailed restriction map of the p5AP insert showing those regions that were sequenced (S), and those regions (hatched) that code for actin (C). The region labelled P is the Hinf I - Bst NI fragment used as a probe for genomic blots (see Section 7.1).

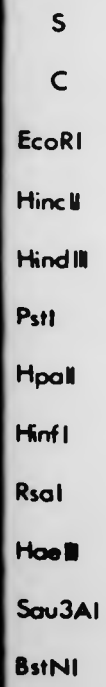


Fig.3.19 The sequencing strategy for 35A

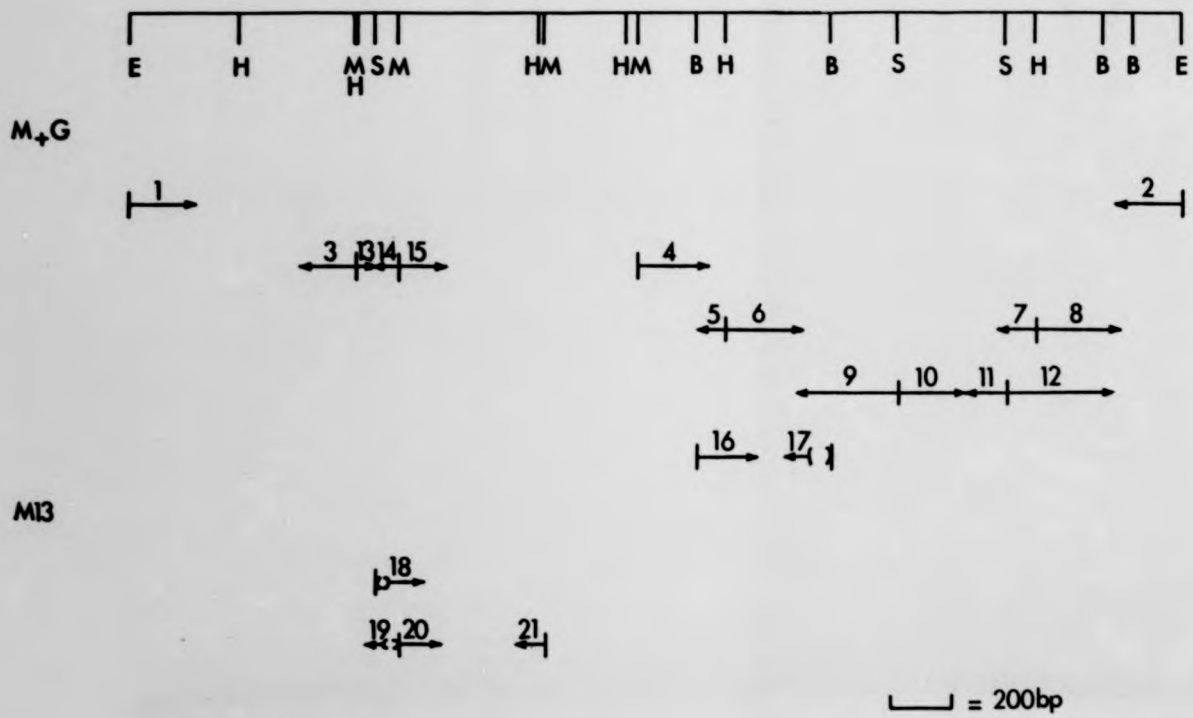


Figure 3.19: Sequencing Strategy for 35A

A map of 35A showing the sequencing strategy. Those regions sequenced by the Sanger method are shown under the 'M13', and those by the Maxam and Gilbert method, between the latter and the 'M + G'. The restriction sites of Eco RI (E), Hinf I (H), Hpa II (M), Bst NI (B) and Sau 3A1 (S) are shown (NB: only part of 35A has been mapped for Bst NI - see Fig. 3.20).

Sections 1 to 12 were end-labelled using T4 DNA polymerase. For sections 1 and 2 the 3.6 kb Eco RI fragment was labelled with ^{32}P dGTP, and cut with Hinf I. For sections 3 and 4 the 0.75 kb and 1.85 kb Hpa II-Eco RI fragments were labelled at the Hpa II end only with ^{32}P dCTP and unlabelled dATP. For section 5 a 0.35 kb Hinf I fragment was labelled with ^{32}P dGTP, and cut with Hpa II. For 6 and 7, the 1.05 kb Hinf I fragment was labelled with ^{32}P dGTP and cut with Hha I (see Figure 3.20). Section 8 used the 0.5 kb Hinf I-Eco RI fragment, labelled at the Hinf I end, only, with ^{32}P dGTP and unlabelled dATP. For section 9, the 0.94 kb Sau 3A1-Xba I fragment was labelled at the Sau 3AI end only, with ^{32}P dGTP. For sections 10 and 11, the 0.38 kb Sau 3AI fragment was labelled with ^{32}P dGTP and cut with Hae III. For section 12, the 0.6 kb Sau 3AI-Eco RI fragment was labelled at the Sau 3AI end, only, with ^{32}P dGTP and unlabelled dATP.

Sections 13 to 17 were end-labelled, using T4 polynucleotide kinase, with gamma-labelled ^{32}P ATP. Sections 13 and 14 used the 0.15 kb Hpa II fragment, which was cut with Sau 3A1 after labelling. Similarly, the 0.75 kb Hpa II fragment was used for section 15, and cut with Hinf I. The 0.46 kb Bst NI fragment (from the 1.9 kb Xba I-Eco RI fragment) was used for sections 16 and 17, being cut with Sst I.

For section 18, an M13 mp8 vector was cleaved with Bam HI and a Sau 3A1 digest of the whole 3.6 kb Eco RI fragment was ligated with this preparation. A clone with an insert over 2 kb in size was partially sequenced by Mr. A. Mohammed and was found to contain the region shown (i.e. the sequence spanning the Hpa II site).

For sections 19 to 21, an M13 mp7 vector was cleaved with Acc I and gel- purified 0.15 kb and 0.75 kb Hpa II fragments were ligated, separately, with this. The insert orientation test (see Section 2.29) was used to identify recombinants with inserts in opposite orientations. The clones of the 0.75 kb fragment were found to include some in both orientations, and so, sequence was obtained from both ends. The sequence obtained from the 0.15 kb clone was used only for spanning the Sau 3A1 site.

Fig. 3.20 35A

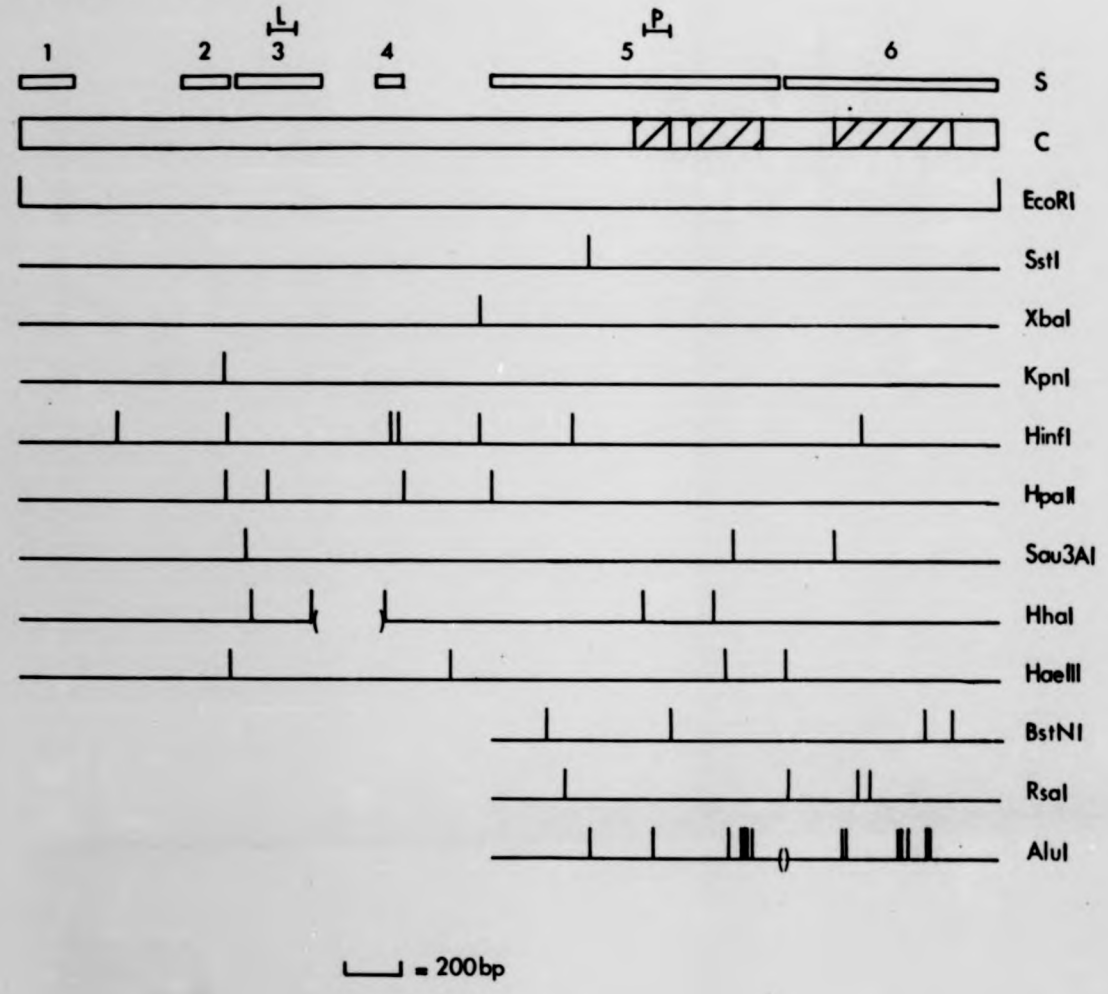

L
P
1
2
3
4
5
6
S
C
EcoRI
SstI
XbaI
KpnI
HinfI
HpaII
Sau3AI
HhaI
HaeIII
BstNI
RsaI
AluI
= 200bp

Figure 3.20: Detailed Restriction Map of 35A

A more detailed restriction map of the p35A insert showing those regions that were sequenced (S) and those regions (hatched) that code for actin (C). The position of the majority of the leader sequence (L) is shown (see Section 5.1.2; 8 bases of the 5' untranslated region also lie adjacent to the first coding exon). The region marked P corresponds to the Hha I-Bst NI fragment used in Section 5 for primer extension studies.

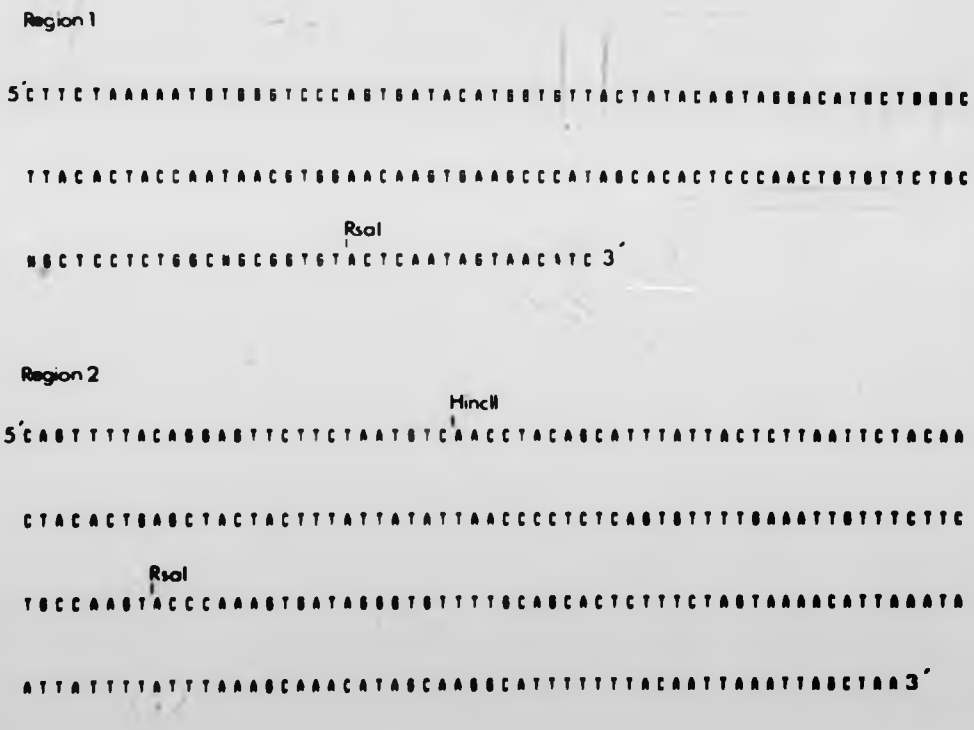
4 SEQUENCE ANALYSIS

4.1 The Coding Sequence

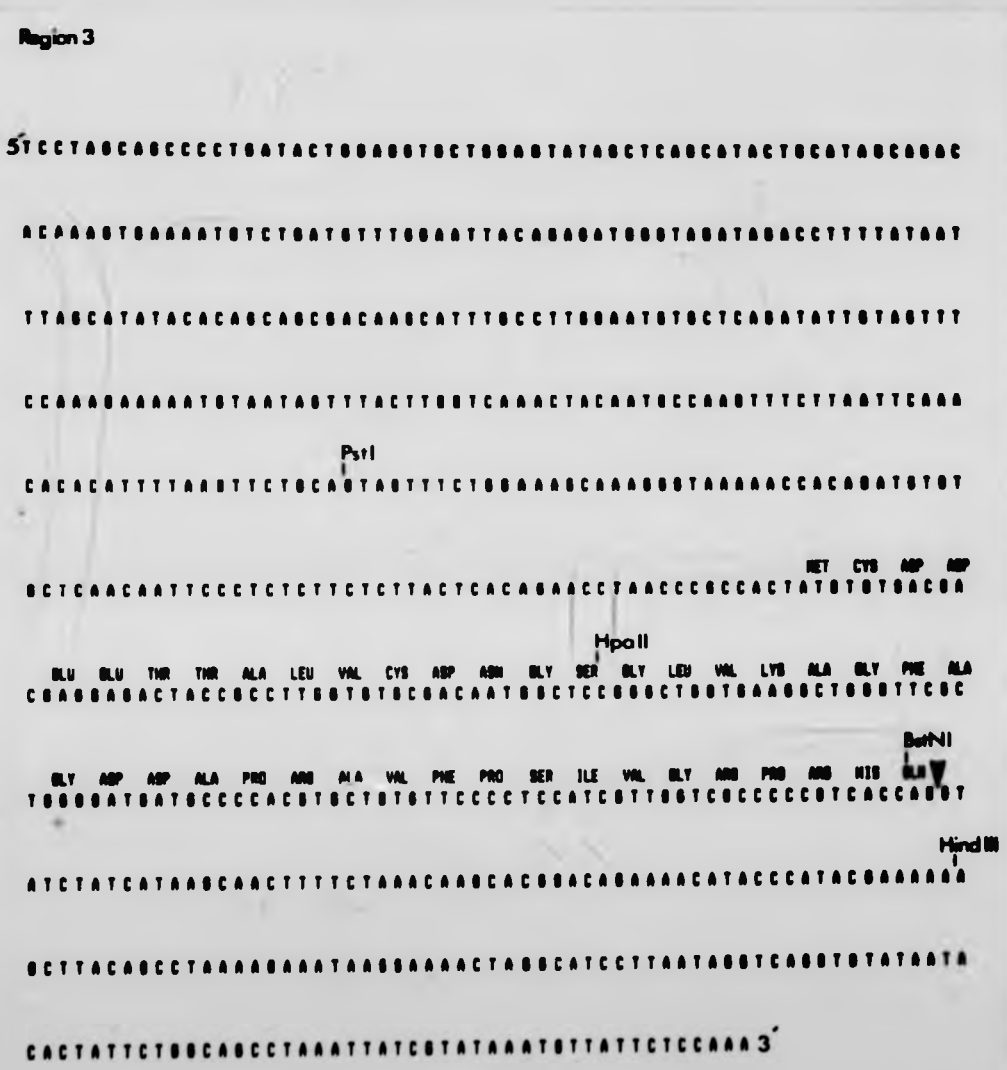
Do actin-coding regions exist in 35A and 5AP? Figures 4.1 and 4.2 show all of the sequence obtained from 5AP and the sequence of regions 4, 5 and 6 from 35A (see Fig. 3.20). In the sequence of 5AP, two blocks of sequence can be found that, when translated, differ from the revised rabbit skeletal muscle actin sequence (Elizinga and Lu, 1976; Collins and Elzinga, 1975; Vandekerckhove and Weber, 1978b) in only 2 out of 149 amino acid positions. Similarly, in 35A there are 3 regions that translate to give 267 amino acids, in which there are 18 differences from the rabbit sequence. There can be little doubt, therefore, that these regions do represent the exons of actin genes. It comes as no surprise that these coding regions are separated by regions that presumably serve no coding function. Such introns have been found in the majority of eukaryotic polymerase II genes to date (Breathnach and Chambon, 1981).

What does the amino acid sequence tell us? Figure 4.3 shows a comparison made between the amino acid sequences of the 5AP and 35A-encoded proteins, rabbit skeletal muscle actin and mammalian cytoskeletal actin. Vandekerckhove and Weber (1978b) have determined the amino acid sequences of a number of vertebrate actins, comparisons between which have revealed a number of amino acid positions at which muscle actins display one characteristic residue, and cytoskeletal actins another. A comparison of positions 5 to 266 in Figure 4.3 shows that, in each, the 5AP-encoded protein has the amino acid that is present in rabbit skeletal muscle actin while the 35A-encoded protein has the substitution present in the mammalian cytoskeletal types. There

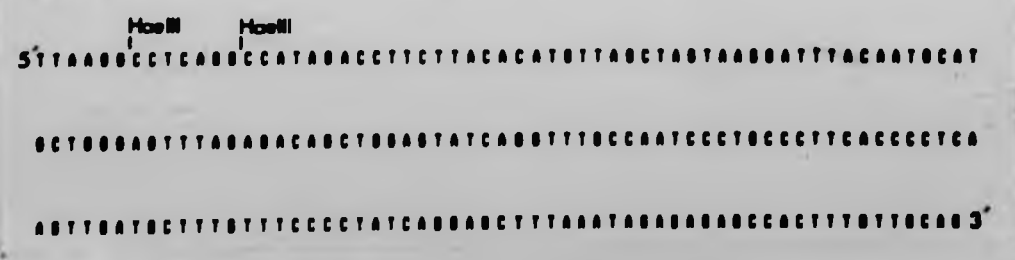
Fig. 4.1 : SAP


Region 1
5'CTTCTAAAAATGTGGGTCCCAGTGATACATGGTGTTACTATACAGTAGGACATGCTGGGC
TTACACTACCAATAACGTGGAACAAGTGAAGCCCATAGCACACTCCCAACTGTGTTCTGC
NGCTCCTCTGGCNGCGGTGTACTCAATAGTAACATC 3'
RsaI
Region 2
5'CAGTTTTACAGGAGTTCTTCTAATGTCAACCTACAGCATTTATTACTCTTAATTCTACAA
CTACACTGAGCTACTACTTTATTATATTAACCCCTCTCAGTGTTTTGAAATTGTTTCTTC
TGCCAAGTACCCAAAGTGATAGGGTGTTTTGCAGCACTCTTTCTAGTAAAACATTAAATA
ATTATTTTATTTAAAGCAAACATAGCAAGGCATTTTTTTACAATTAAATTAGCTAA 3'
HincII
RsaI

Region 3



Region 4



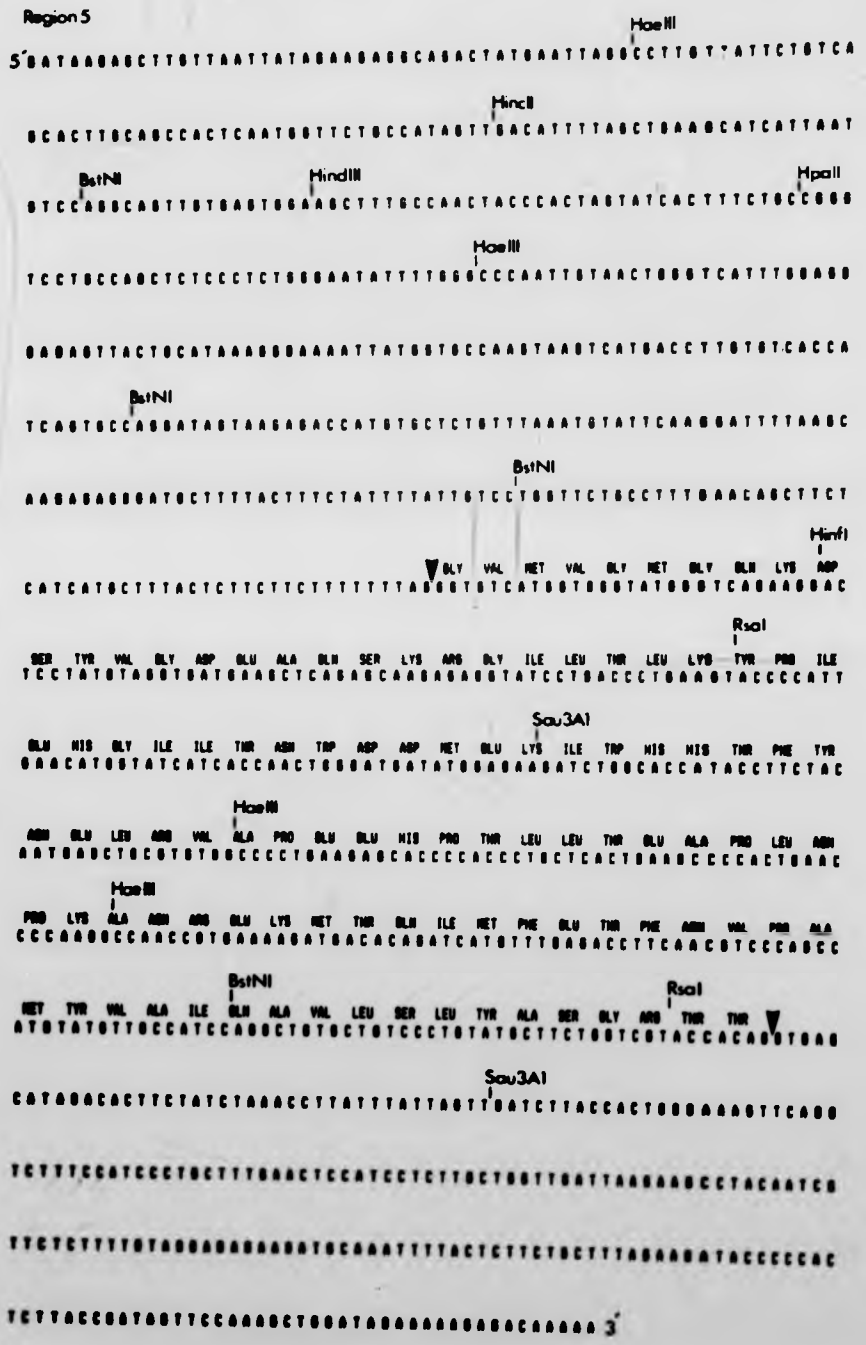
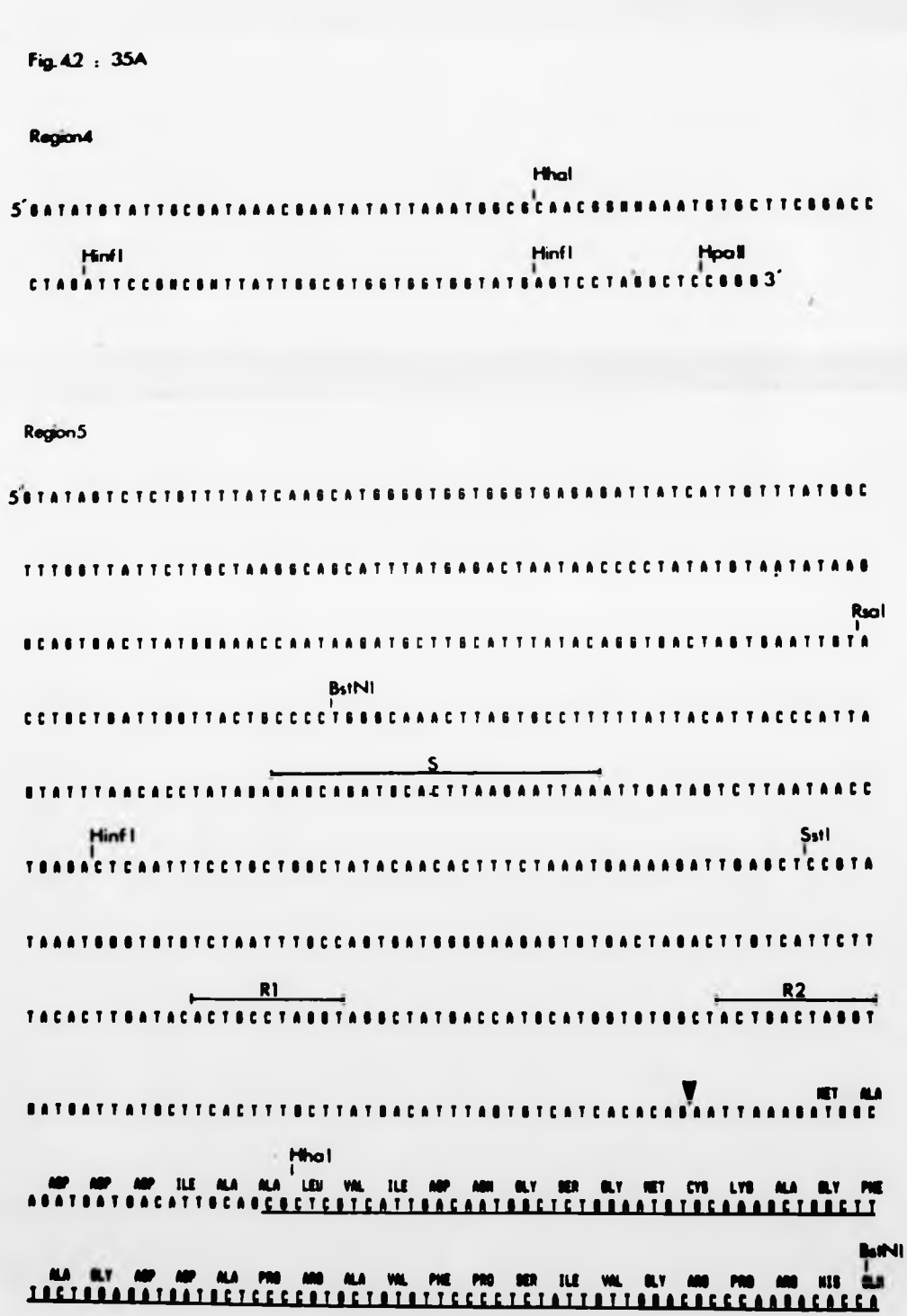

Region 5
HaeIII
HincII
BstNI
HindIII
HpaII
HaeIII
BstNI
BstNI
HinfI
RsaI
Sau3AI
HaeIII
HaeIII
BstNI
RsaI
Sau3AI
3'

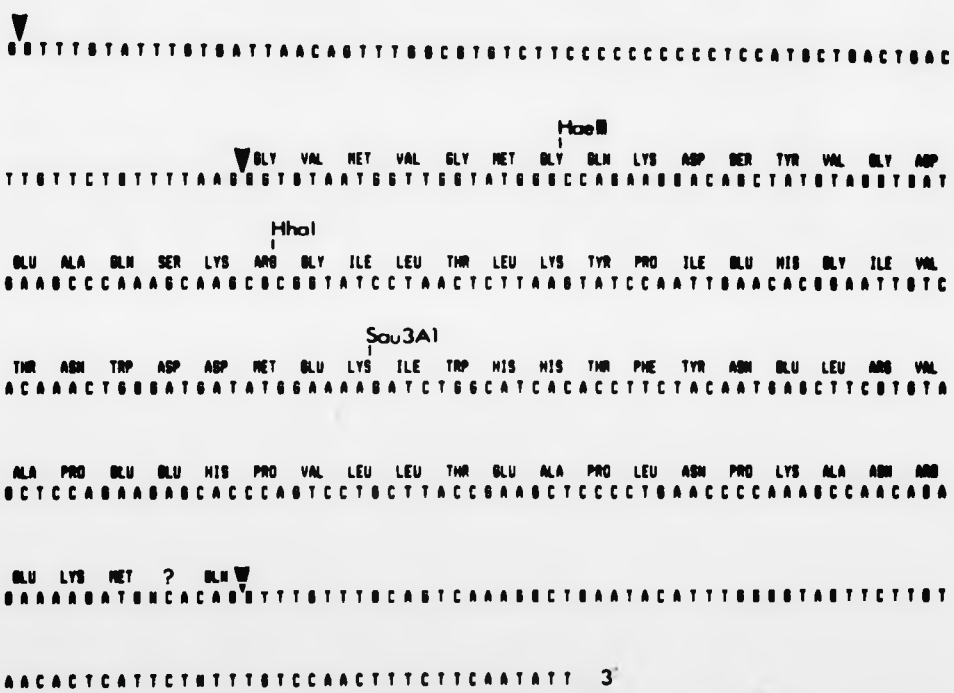
Figure 4.1: Sequence of 5AP: Regions 1 to 5

The antisense strand sequence of regions 1 to 5 of 5AP (see Figure 3.18) is shown. Each full line contains 60 bases. In regions 3 and 5, two blocks of sequence are found which encode an amino acid sequence similar to a mammalian muscle-specific actin (the encoded amino acids are shown above the DNA sequence). Putative splice sites that obey Breathnach and Chambon's rule (see Section 4.4) are indicated by large arrow heads. The positions of the restriction enzyme cutting sites shown in Figure 3.18 are indicated.

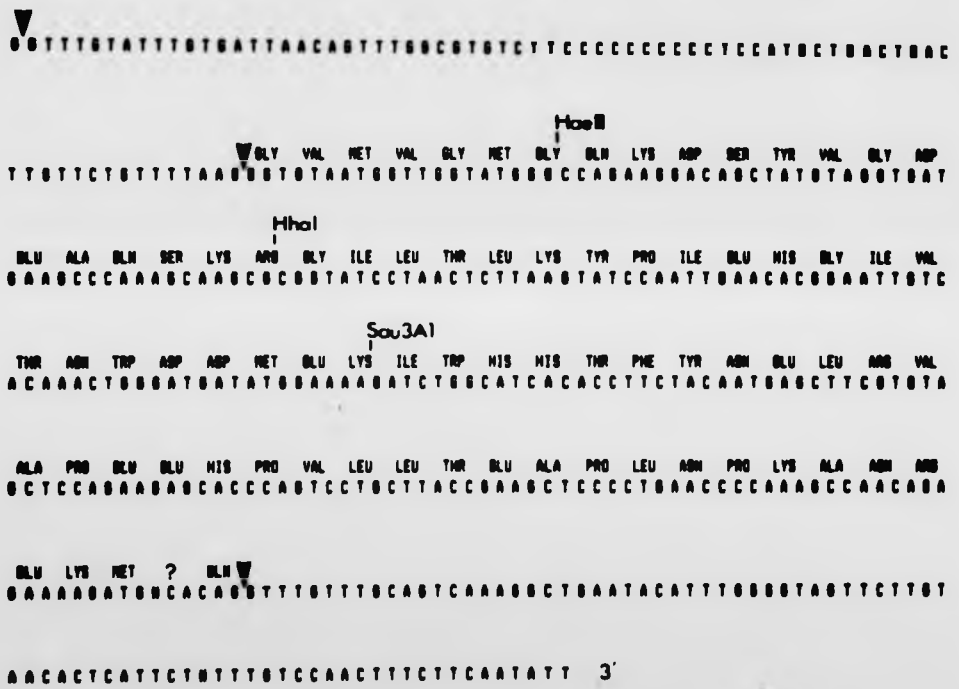
Fig. 42 : 35A


Region4
HhaI
HinfI
HinfI
HpaII
Region5
RsaI
BstNI
S
HinfI
SstI
R1
R2
MET ALA
HhaI
BstNI

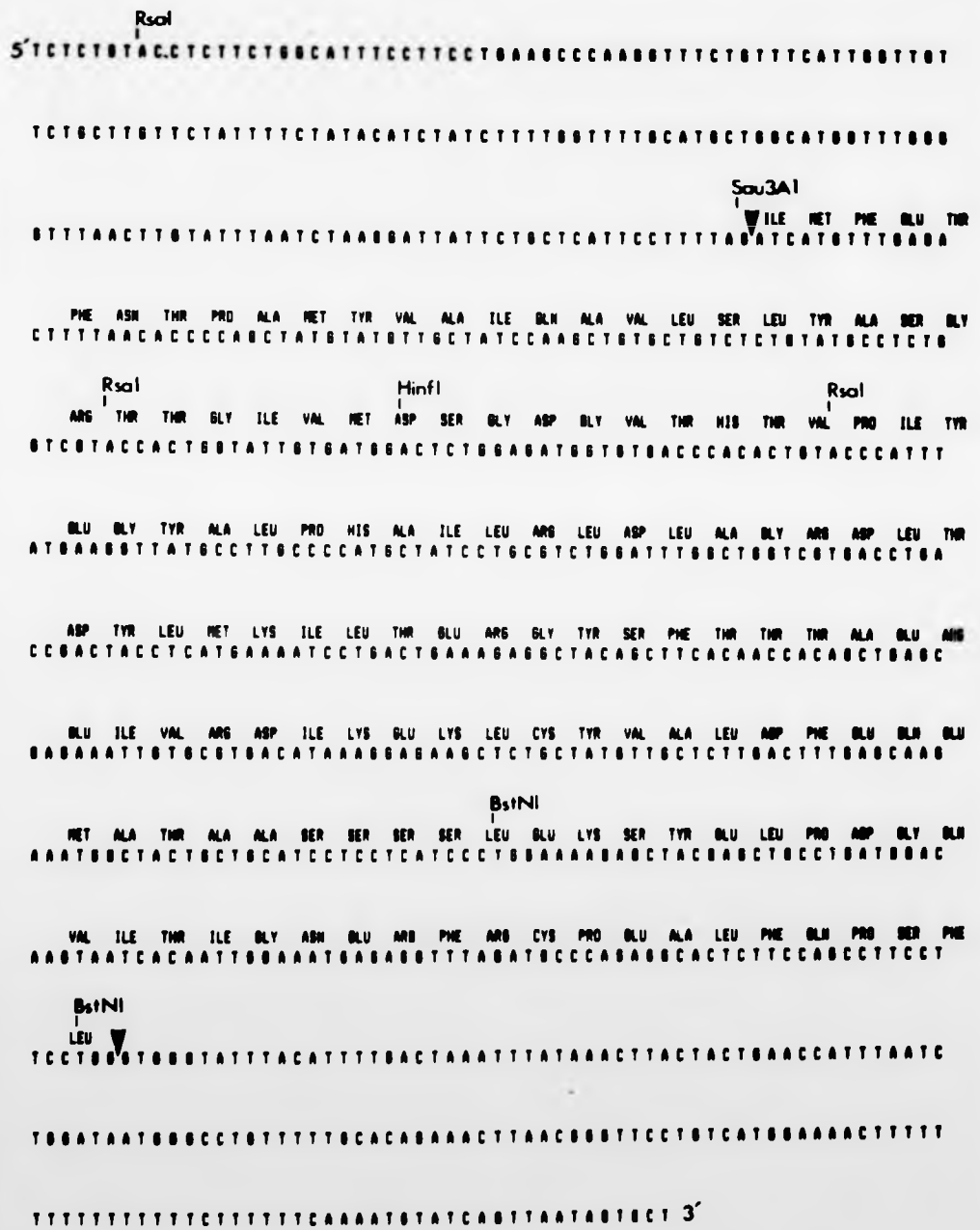
__Region 5 contd.



__Region 5 contd.



Region 6


RsaI
Sau3AI
ILE MET PHE GLU THR
PHE ASN THR PRO ALA MET TYR VAL ALA ILE GLN ALA VAL LEU SER LEU TYR ALA SER GLY
RsaI
HinfI
RsaI
ARG THR THR GLY ILE VAL MET ASP SER GLY ASP GLY VAL THR HIS THR VAL PRO ILE TYR
GLU GLY TYR ALA LEU PRO HIS ALA ILE LEU ARG LEU ASP LEU ALA GLY ARG ASP LEU THR
ASP TYR LEU MET LYS ILE LEU THR GLU ARG GLY TYR SER PHE THR THR THR ALA GLU ARG
GLU ILE VAL ARG ASP ILE LYS GLU LYS LEU CYS TYR VAL ALA LEU ASP PHE GLU GLN GLU
BstNI
MET ALA THR ALA ALA SER SER SER SER LEU GLU LYS SER TYR GLU LEU PRO ASP GLY GLN
VAL ILE THR ILE GLY ASN GLU ARG PHE ARG CYS PRO GLU ALA LEU PHE GLN PRO SER PHE
BstNI
LEU
5′
3′

Region 6

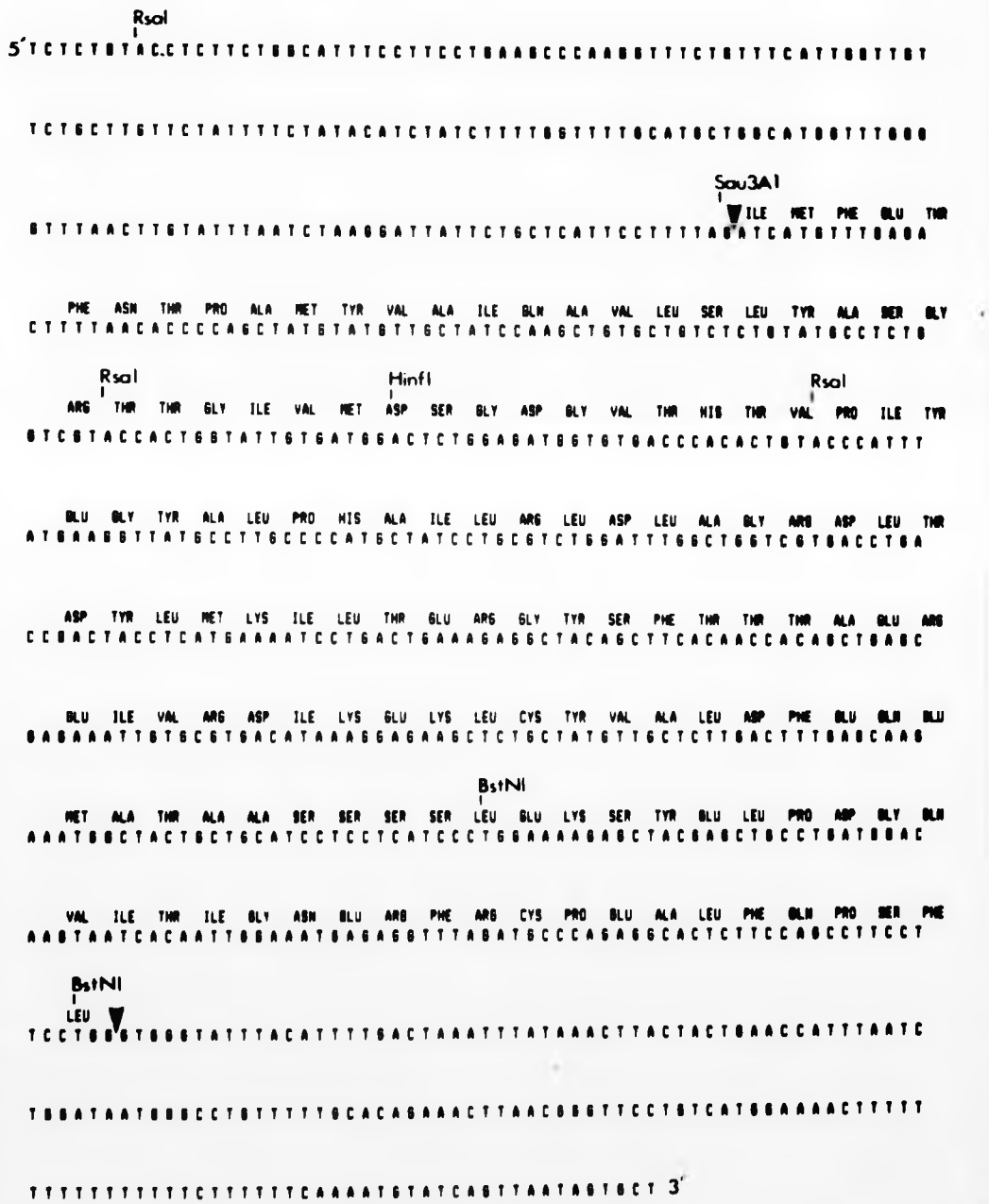

RsaI
5'TCTCTGTACCTCTTCTGGCATTTCCTTCCTGAAGCCCAAGGTTTCTGTTTCATTGGTTGT
TCTGCTTGTTCTATTTTCTATACATCTATCTTTTGGTTTTGCATGCTGGCATGGTTTGGG
Sau3AI
ILE MET PHE GLU THR
GTTTAACTTGTATTTAATCTAAGGATTATTCTGCTCATTCCTTTTAGATCATGTTTGAGA
PHE ASN THR PRO ALA MET TYR VAL ALA ILE GLN ALA VAL LEU SER LEU TYR ALA SER GLY
CTTTTAACACCCCAGCTATGTATGTTGCTATCCAAGCTGTGCTGTCTCTGTATGCCTCTG
RsaI
HinfI
RsaI
ARG THR THR GLY ILE VAL MET ASP SER GLY ASP GLY VAL THR HIS THR VAL PRO ILE TYR
GTCGTACCACTGGTATTGTGATGGACTCTGGAGATGGTGTGACCCACACTGTACCCATTT
GLU GLY TYR ALA LEU PRO HIS ALA ILE LEU ARG LEU ASP LEU ALA GLY ARG ASP LEU THR
ATGAAGGTTATGCCTTGCCCCATGCTATCCTGCGTCTGGATTTGGCTGGTCGTGACCTGA
ASP TYR LEU MET LYS ILE LEU THR GLU ARG GLY TYR SER PHE THR THR THR ALA GLU ARG
CCGACTACCTCATGAAAATCCTGACTGAAAGAGGCTACAGCTTCACAACCACAGCTGAGC
GLU ILE VAL ARG ASP ILE LYS GLU LYS LEU CYS TYR VAL ALA LEU ASP PHE GLU GLN GLU
GAGAAATTGTGCGTGACATAAAGGAGAAGCTCTGCTATGTTGCTCTTGACTTTGAGCAAG
BstNI
MET ALA THR ALA ALA SER SER SER SER LEU GLU LYS SER TYR GLU LEU PRO ASP GLY GLN
AAATGGCTACTGCTGCATCCTCCTCATCCCTGGAAAAGAGCTACGAGCTGCCTGATGGAC
VAL ILE THR ILE GLY ASN GLU ARG PHE ARG CYS PRO GLU ALA LEU PHE GLN PRO SER PHE
AAGTAATCACAATTGGAAATGAGAGGTTTAGATGCCCAGAGGCACTCTTCCAGCCTTCCT
BstNI
LEU
TCCTGGGTGGGTATTTACATTTTGACTAAATTTATAAACTTACTACTGAACCATTTAATC
TGGATAATGGGCCTGTTTTTGCACAGAAACTTAACGGGTTCCTGTCATGGAAAACTTTTT
TTTTTTTTTTCTTTTTCAAAATGTATCAGTTAATAGTGCT 3'

Figure 4.2: Sequence of 35A: Regions 4 to 6

The antisense strand sequence of regions 4 to 6 of 35A (see Figure 3.20) is shown. Each full line contains 60 bases. In regions 5 and 6, three blocks of sequence were found which encode an amino acid squence identical (except in one position) to the mammalian beta actin (the encoded amino acids are shown above the DNA sequence). Putative splice sites that obey Breathnach and Chambon's rule (see Section 4.4) are indicated by large arrow heads. The positions of most of the restriction enzyme cutting sites shown in Figure 3.20 are indicated (Alu I sites, AGCT, have been omitted to avoid congestion). In region 5 the sequence which has been overlined and labelled S resembles part of a repetitive element found by Spohr et al. (1981), while 10 out of 11 bases of the overlined sequence R1 are repeated in R2 (see Section 6). Also in region 5, I have underlined the section of this gene's transcript to which the primer described in Section 5 hybridizes (i.e. the primer is the sense strand of this Hha I-Bst NI restriction fragment - see Figure 5.1).

Figure 4.3

Residue	M		C		5AP	35A
-2					met	
-1					cys	met
1	asp				asp	ala
2	glu	asp(β)		glu(γ)	asp	asp
3	asp	asp(β)		glu(γ)	glu	asp
4	glu	asp(β)		glu(γ)	glu	asp
5	thr		ile		thr	ile
6	thr		ala		thr	ala
10	cys	val(β)		ile(γ)	cys	ile
16	leu		met		leu	met
17	val		cys		val	cys
76	ile		val		ile	val
103	thr		val		thr	val
129	val		thr		val	thr
153	leu		met			met
162	asp		thr			thr
176	met		leu			leu
201	val		thr			thr
225	asn		gln			gln
259	thr		ala			ala
266	ile		leu			leu

Figure 4.3: Comparisons Between 5AP, 35A, Cytoskeletal and Muscle Actins

The amino acid differences between mammalian cytoskeletal actins (C), rabbit skeletal muscle actin (M), and the proteins encoded by 5AP and 35A are shown. Only the first 267 amino acids have been compared, as neither Xenopus clone contains sequence encoding residues beyond this part of the protein (there are another 7 positions, between residues 267 and 375, where mammalian skeletal muscle and cytoskeletal actins differ). The amino acids encoded by the 5AP and 35A genes have been numbered to match the mammalian sequences (Vandekerckhove and Weber, 1978b), and this numbering system is used in the text. As explained in the text, the initial met-cys of the 5AP gene and the met-ala of the 35A gene, may be post-translationally removed. The 5AP gene only encodes amino acids up to residue 149.

Figure 4.4

R	SkM	CM	SmM(ν)	SmM(ϑ)	SAP
-2					met
-1					cys
1	asp	asp		glu	asp
2	glu	asp	glu	glu	asp
3	asp	glu	glu	glu	glu
4	glu	glu	glu	asp	glu
5	thr	thr	thr	ser	thr
6	thr	thr	thr	thr	thr
17	val	val	cys	cys	val
89	thr	thr	ser	ser	thr

Figure 4.4: Comparison Between 5AP and the Muscle Actins

The differences between the mammalian muscle actin isoforms are tabulated for comparison with the protein encoded by the 5AP gene. The numbering system is as in Figure 4.3. The mammalian actin sequences are from Vandekerckhove and Weber (1978d), and are as follows: SkM = rabbit, bovine, and chicken skeletal muscle actin, CM = bovine cardiac muscle actin; S mM (γ) = chicken gizzard smooth muscle actin; S mM (α) = bovine aorta smooth muscle actin.

Fig 4.5a

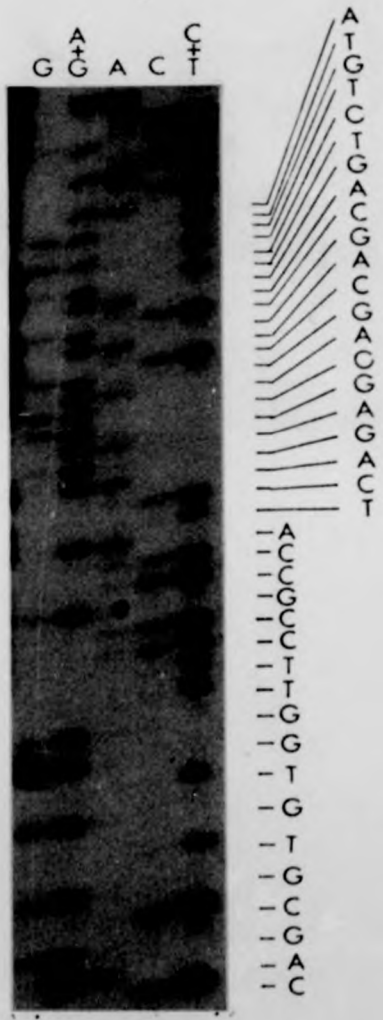


Fig 4.5a

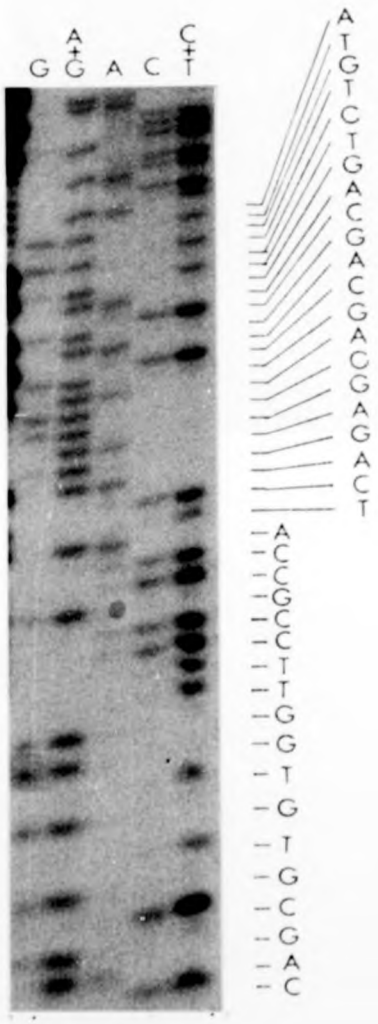


Fig 4.5a

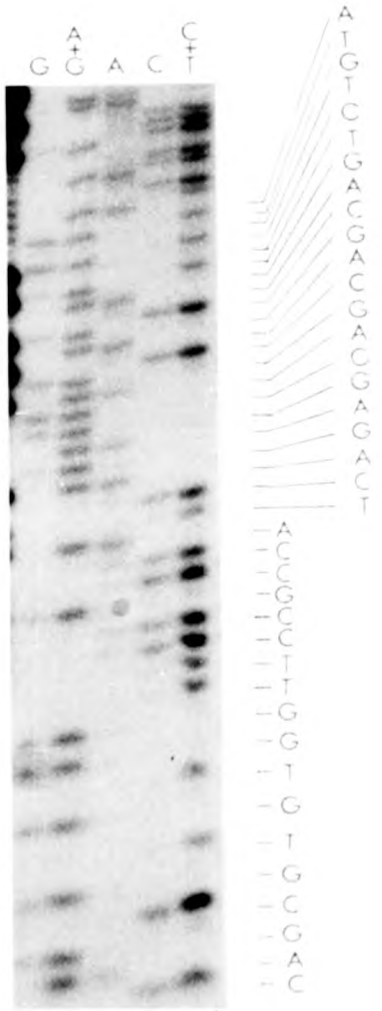


Fig. 4. 5 b

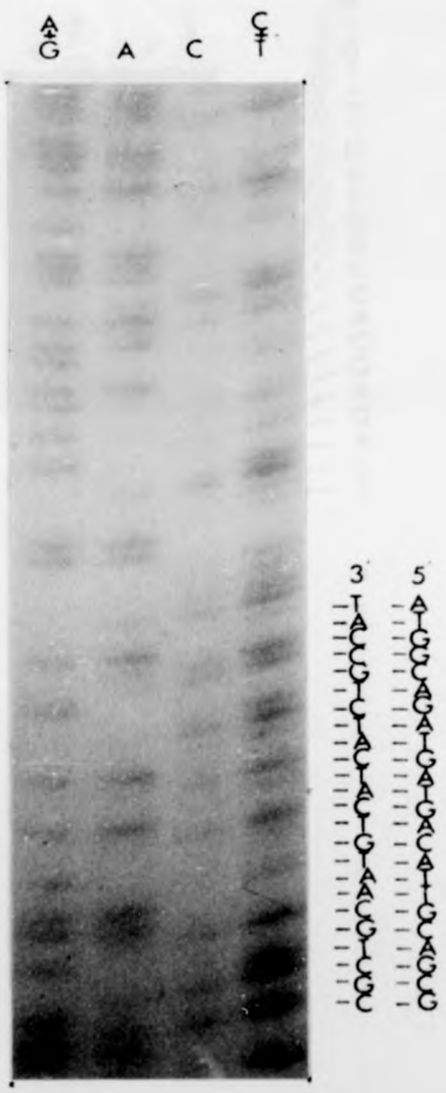


Fig. 4. 5 b

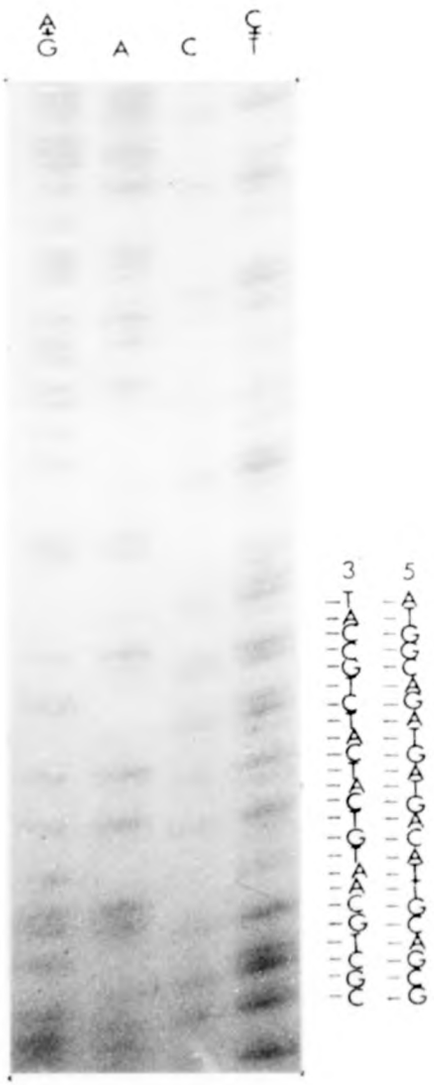


Figure 4.5: Sequencing Gels: N-Terminals of the 5AP and 35A Actins

Sequencing gels showing the first 39 bases of the coding region of the 5AP gene (a), and the first 24 bases of the coding region of the 35A gene (b). In the former case the antisense strand was 3' end-labelled for sequencing (see section 9 of Figure 3.17), and so the sequence reads 5' to 3' from top to bottom. In the latter case the sense strand was 5' end-labelled (section 17, Figure 3.19) and reads 3' to 5' from top to bottom; the complementary strand has also been supplied, for easy comparison with the non-sense strand sequence shown in Figure 4.2.

are no other amino acid differences in this region, between 5AP and 35A, and between either of these two and the mammalian actins. This is highly suggestive that the 5AP gene encodes a muscle-specific actin and the 35A gene, a cytoskeletal actin (although I have yet to show that the 5AP gene is actually expressed in muscle, and there is no amino acid sequence available on Xenopus muscle actins).

Vandekerckhove and Weber (1978a) have shown that at least 6 different actin types are expressed in mammals and that 4 of these are muscle-specific. Figure 4.4 shows the differences between the muscle types (Vandekerckhove and Weber, 1978d). Fig. 4.5(a) shows the sequencing gel from which the N-terminal end of the 5AP gene product was determined. The 5AP-encoded protein is the same as a mammalian striated muscle actin at positions 17 and 89. It is also dissimilar to the aorta smooth muscle actin at positions 4 and 5. The gizzard smooth muscle actin is one amino acid shorter at the N-terminal end than both the striated muscle actin and the 5AP protein. In fact, the latter is not only dissimilar to the mammalian smooth muscle actin but is identical to the bovine cardiac muscle sequence (except for the initial methionine and cysteine residues, but - see below). Because it is not yet known to what extent the differences between the actin isoforms are the same in amphibians as in mammals it is not possible, from this data alone, to conclude that the 5AP gene codes for a heart muscle actin. However, there is high degree of silent site conservation between 5AP and a putative X. laevis cardiac muscle actin cDNA clone (Tim Mohun; personal communication). The latter is possibly a cardiac actin because of its similarity to the mammalian cardiac actin at positions 298 and 357, and because of the similarity of part of its 3' untranslated region to the same region of a mouse cardiac actin cDNA clone (T. Mohun, per.

comm.). Out of 390 bases compared between the Xenopus clones there were only 7 silent base changes (and no replacement changes). This gives a figure of 25% divergence of silent sites, when using the method of Perler et al. (1980) to correct for multiple substitutions at each site. This compares with a value of 112% obtained for a comparison between human cardiac and skeletal muscle actin genes (Hanauer et al., 1983). The difference between the X. laevis and X. borealis clones is, therefore, much less than the difference between the two striated muscle actin types of one species of mammal. Therefore, it seems likely that the divergence of the two Xenopus genes represents that of the same gene in the time since the 2 species diverged. However, even if the gene, of which 5AP is part, is identical to the mammalian cardiac actin gene, it is possible that the amphibian cardiac and skeletal muscle actins have not diverged in the same way as have the mammalian actins; there is, as yet, no evidence in Xenopus of a different actin isoform in cardiac muscle to that in skeletal muscle. It is impossible to date, accurately, the divergence of the mammalian cardiac and skeletal muscle genes from the silent site changes, because of the saturation of such changes over long periods of time (Hanauer et al., 1983; Perler et al., 1980). Furthermore, a great range of such divergence rates has been found in various globin genes (Perler et al., 1980). However a corrected silent site divergence of only 86%, between human and chicken, and 70% between rat and chicken skeletal muscle actins (Hanauer et al., 1983), is suggestive that the cardiac and skeletal muscle gene divergence may have occurred before the separation of the mammalian and avian ancestral lines (cf. 112% between human cardiac and skeletal genes). It is not known, however, if this divergence occurred before, or after, the separation of these lines from the amphibians.

The main difference between the amino acid sequences encoded by 5AP and 35A and the vertebrate protein sequences published to date lies in the first two amino acids. All the latter sequences begin with an aspartate or glutamate residue which is acetylated. The 5AP-encoded protein, however, begins with a methionine and cysteine before the first acidic residue, and the 35A-encoded protein, a methionine followed by an alanine. Cysteine has not been found at the amino terminus of any known actin isoform (Vandekerckhove and Weber, 1979). There is good reason to believe, however, that this does not represent a significant difference between amphibian actin proteins and others. One of those actins that has been sequenced is the chicken skeletal muscle gene (Vanderkerckhove and Weber, 1978d). A chicken gene has been sequenced by Fornwald et al. (1982) and found to code for a protein identical to this except for the presence of an additional methionine-cysteine at the N-terminal end. A met-cys has also been found at the start of the human cardiac and skeletal muscle, a rat skeletal muscle, a sea urchin, all 4 nematode, and all 6 Drosophila actin genes (Hamada et al., 1982; Zakut et al., 1982; Fyrberg et al., 1981; Files et al., 1983; Cooper and Crain, 1982; Schuler et al., 1983; Hanauer et al., 1983; Gunning et al., 1983). Therefore, it seems likely that these amino acids are cleaved off after translation.

The same explanation may apply to the methionine-alanine at the N-terminus of the 35A gene product, although there are no amino acid data for Xenopus borealis actins. (The sequencing gel from which the N-terminus of the 35A gene product was determined is shown in Fig. 4.5b.) The three cytoskeletal actin isoforms found in Xenopus laevis oocytes (a species closely related to X. borealis) have all been found to begin with 3 acidic residues. Therefore, it is likely that this X. borealis

gene (which is expressed in the oocyte; see Section 5.1) expresses a protein which also starts with 3 acidic residues; the met-ala being post-translationally removed.

There are also invertebrate actins which possess amino acids between the initiating methionine and the acidic residues. There is, however, little evidence to support the notion that all such amino acids must be cleaved off. The Acanthamoeba gene codes for methionine-glycine at the N-terminus (Nellen and Gallwitz, 1982), while a soybean gene has been found to begin with methionine-alanine (Shah et al., 1982). In the former case, the glycine appears to be a substitution for an acidic residue, it being followed by only two acidic residues. This glycine may be present in the mature protein (the authors suggest that this contributes to the protein having a more alkaline isoeletric point than the mammalian actins). In the latter case there is no amino acid sequence information and consequently no evidence that the alanine is cleaved off.

It is interesting that the only other vertebrate cytoskeletal actin genes for which there is any sequence information, the rat beta-gene (Nudel et al., 1983), and the human beta- and gamma-genes (Gunning et al., 1983), also do not possess the cysteine found at the N-terminal of all vertebrate muscle, sea urchin, and protostome actin genes that have been sequenced. The methionine of these genes, however, is immediately followed by the acidic residues found in the mature proteins, unlike the 35A-encoded protein with its intervening alanine residue. It is tempting to speculate that the post-translationally removed cysteine of these genes is conserved for a functional purpose; one which is not needed by vertebrate cytoskeletal actins.

When did this difference arise? It seems unlikely that this

cysteine arose separately in vertebrate muscle, sea urchin and protostome genes, by convergent evolution. It also seems unlikely that these cys(+) and cys(-) genes arose from two separate lines of genes which were present before the divergence of the protostomes and deuterostomes. This is because all of the Drosophila and nematode genes (all Cys+) resemble the vertebrate cytoskeletal genes (Cys-) more than the vertebrate muscle-specific (Cys+) genes (see section 1.2). Also, the sea urchin cytoskeletal actin genes 1 and 2 are of the Cys(+) type (Schuler et al., 1983). Consequently, it seems likely that at some stage of deuterostome evolution, the muscle-specific genes diverged from the cytoskeletal genes, and that after the divergence of the echinoderms and the vertebrates, the cytoskeletal genes lost their N-terminal cysteine codon. In mammals, the codon was removed entirely, while in amphibians, it mutated to an alanine codon.

In the above sequence of evolutionary events, I have said that the muscle-specific genes diverged from cytoskeletal genes 'at some time during deuterostome evolution'. The presence of an amphibian actin sequence identical to a mammalian muscle type now suggests that this occurred before the separation of the amphibian, bird and mammalian lineages. The actin genes of sea urchins (Davidson et al., 1982) have not all been sequenced. The further dating of this gene divergence must await such work.

What type of cytoskeletal actin does the 35A gene encode? Immediately after the met-ala, 35A codes for 3 aspartate residues (see Fig. 4.5b). It is, in this sense, similar to the mammalian beta-actin (see Fig. 4.3), and unlike the mammalian cytoskeletal gamma-actin (the latter having 3 glutamate residues in these positions). The 35A gene, however, possesses a residue at position 10 which, as Fig. 4.3 shows, is

the same as the mammalian gamma-actin. The presence of 3 aspartates at positions 2 to 4 (numbered from the rabbit skeletal muscle actin) of this X. borealis actin was surprising, as none of the 3 cytoskeletal actins found by Vandekerckhove et al. (1981) in X. laevis were found to have this sequence. No actin of this type was found in oocytes, kidney epithelial cells, brain or liver. Several other amphibians had been found to possess actins different at their N-terminus to those of X. laevis, as well as mammals (Vandekerckhove et al., 1981; described in Section 1.2). The question, raised by these results is whether these differences have evolved because they were functionally important, or whether they show that, in amphibians at least, there is no functional difference between the cytoskeletal actin isoforms; i.e. that the differences represent neutral mutations. The difference in the isoforms found in 2 such closely related species as X. laevis and X. borealis seems to support the latter explanation (immunological data on the albumins would suggest that these species diverged from a common ancestor about 8 million years ago; Bisbee et al., 1977), but, of course, does not prove it. This apparent difference does depend on the 35A gene being expressed; evidence is produced in support of this in Section 5.1.

In mammals, it seems likely that the two cytoskeletal isoforms do perform different functions. It is thought likely that all mammals possess the beta- and gamma-isoforms shown in Fig. 4.3, because sequences corresponding to these actins have been found in humans (Vandekerckhove et al., 1980b), cows, rats, and mice (Vandekerckhove and Weber, 1978b, 1978c). This conservation of the mammalian cytoskeletal isoforms, together with the tissue-specific variation in their ratios (Garrels and Gibson, 1976), would suggest that they perform different

functions. There is also some more direct evidence for this proposal. This stems from the work of Pardo et al. (1983) on the distribution of cytoskeletal actins in muscle cells. A polyclonal antibody preparation, which recognizes gamma-actins, colocalized with the mitochondria in these cells. Lubit and Schwartz (1980) and Hall et al. (1981) made an antibody to Aplysia (a mollusc) actin, which does not recognize the vertebrate skeletal muscle actin. This antibody localized in a completely different region of muscle cells to that of Pardo et al. (1983): to the sarcolemma (especially at the neuromuscular junction), the sarcoplasmic reticulum, and transverse tubules. Pardo et al. have therefore postulated that this second antibody was recognizing a different isoactin to the gamma-type: presumably the beta-actin. They pointed out, however, that they had produced no evidence that these proteins were the same as those present in non-muscle cells. Anti-actin sera commonly bind to many actin types; for example, the antibody preparation of Pardo et al. (1983) binds to gamma-actins of both the smooth muscle and cytoskeletal type, and could possibly be binding to a third protein in skeletal muscle, although there is no evidence for this.

Cacares et al. (1983) have recently used monoclonal antibodies against quail muscle actin in studies on the dendrites of brain neurones. One such antibody clone was found to stain dendrites as well as the dendritic spines ('mushroom-like' projections on dendrites), while another only stained dendrites. ⟶

→ This would be explained by isoforms of actin having a differential distribution in the spines and parent dendrite. However, another interpretation of this result is that the difference of labelling observed with the second monoclonal antibody was due merely to the different accessibility of this particular antigenic site in these two locations (this is a problem which is presumably not encountered when polyclonal antibodies are used, as with Pardo et al. (1983) and Lubit and Schwartz, 1980).

Summary

5AP and 35A each contain the 5'-portion of an actin gene. Comparison with mammalian protein sequences makes it seem likely that the 5AP gene encodes a muscle-specific actin, while the 35A gene encodes a cytoskeletal actin. If this is correct, then there are almost no amino acid sequence differences between the proteins encoded by these Xenopus clones, and their mammalian counterparts. Evidence from other vertebrate muscle-specific genes suggests that the initial met-cys residues encoded by the 5Ap gene are probably post-translationally removed. Comparison with amino acid sequences of Xenopus laevis cytoskeletal actins suggests that a similar fate might befall the initial met-ala encoded by the 35A gene. The amino acid sequence available for the 5AP gene is identical to that of a mammalian cardiac muscle actin, but there is no evidence that amphibian heart and skeletal muscle contain different actin isoforms. The amino acid sequence of 35A is identical to the mammalian beta-isoform (ignoring positions -1 and -

2) except for position 10, where it is the same as the gamma-cytoskeletal isoform. It is, however, different in positions 2 to 4 (at least) from the 3 X. laevis cytoskeletal actins. Together with data presented in Section 5.1 which indicates that the 35A gene is expressed, this suggests that, in amphibians, there may be little functional difference between the cytoskeletal isoforms. However there is some evidence for the separate localization of cytoskeletal actins in mammals.

4.2 Actin Gene Sequence Comparison

The tissue-specific nature of vertebrate actins has already been described in Section 1.2, where it was also pointed out that the amino acid sequence of a given isoform appears to be completely conserved in different mammalian and bird species. This, of course, makes it impossible to establish the divergence of two vertebrate actin genes from the replacement site substitutions of one relative to the other, as was done with globin genes (Perler et al., 1980; Efstratiadis et al., 1980). In the latter case, the rate of divergence was calibrated by comparing known globin types between species where the time since their divergence was known. Using this calibration, the approximate date of various gene duplications could be estimated.

The only way in which the divergence time of actin genes could possibly be estimated is by studying the silent site substitutions. The latter are base changes that give rise to a synonymous codon rather than one specifying a different amino acid. The rate of silent site substitution has also been studied in globin genes (Efstratiadis et al., 1980; Perler et al., 1980), where it has been found to occur at an initial rate of about 0.7% per million years. This accumulation of changes appears to saturate after 85 million years, after which, they accumulate at a much reduced rate.

Does the same phenomenon occur in actin genes? A recent study of two linked sea urchin actin genes appears to show a strong conservation of silent site bases (Schuler et al., 1983). The marked divergence of most of the non-coding DNA, including the introns, suggests that these genes diverged about 90 million years ago. The silent site divergence, however, was only 12% after correction for multiple mutations (Perler et al., 1980). This compares with an approximately 60% corrected

divergence found for globins after a similar divergence time. This silent site conservation could be the result of the need for particular features of secondary structure by these transcripts or, possibly, the result of some mechanism of concerted evolution, such as gene conversion or unequal crossover (Slightom et al., 1980; Jeffreys, 1981). If the former is true then the necessity for such silent site conservation would appear to have been lost by other actin genes. I have compared the rat and rabbit skeletal muscle actin genes (Zakut et al., 1982; Putney et al., 1982; Minty et al., 1982), which I have assumed to have diverged at approximately the time of the mammalian radiation, 85 million years ago (Romero-Herrara et al., 1973). I found the corrected silent site divergences (corrected exactly as described in Perler et al., 1980; see Appendix 2) in both comparisons to be about 50%; considerably greater than that of the two sea urchin genes. Hanauer et al. (1983) have recently done a similar calculation and found a 61% corrected divergence between a human and rat skeletal muscle actin genes.

Could the sea urchin genes be similar because of some form of concerted evolution? This kind of process appears to have been responsible for the similarity of the silent sites and introns of the human alpha-globin genes (Zimmer et al., 1980) and the 5' halves of the human gamma-globin genes (Slightom et al., 1980), as well as the lack of divergence of the first intervening sequence in the mouse major and minor beta-globin genes (Konkel et al., 1979). Concerted evolution has also been described inthe heat-shock genes of Drosophila (Torok et al., 1982; Leigh Brown and Ish-Horowicz, 1981), and for the human and great ape ribosomal genes (Arnheim et al., 1980). There is, therefore, a reasonable body of data which suggests that portions of related genes

can be corrected against each other in some way. That often only part of such genes are involved in this process, suggests that it may be responsible for the silent site similarity of the two sea urchin actin genes (Schuler et al., 1983). Most of the non-coding sequence, including most of the introns, are highly diverged. 55 bases of one intron, however, also show 12% corrected divergence between the two genes; this is evidence in favour of concerted evolution acting here.

I have also compared the rat and chicken skeletal muscle genes (Zakut et al., 1982; Fornwald et al., 1982), which presumably diverged approximately 270 million years ago (Dickerson, 1971; Moore et al., 1976; Wilson et al., 1977), and found a corrected divergence of only 68%. This is similar to the average value obtained for a similar divergence time with globin genes (Efstratiadis et al., 1980), and suggests that the accumulation of silent changes also saturates in actin genes (i.e. this value is similar to the corrected divergence of 61% between human and rat alpha-actin genes).

This saturation presumably occurs when all the neutral silent sites have been fully randomized. The slow change after saturation may represent the change of silent sites which were previously conserved for, perhaps, structural reasons. That all possible silent replacements are not necessarily neutral was suggested by Kafatos et al. (1977), who found that silent changes were not scattered at random along the mRNA, but were clustered into regions that tended also to be rich in replacement site substitutions. The slow change may be because of the gradual appearance of adaptive changes in gene expression, or the slow shift in codon utilization pattern of the genome. The occurrence of certain 'invariant codons' (see Section 4.3.1), where the codon used only changes after a long period of divergence, may be an example of

such a conserved silent site. Perler et al. (1980) suggested that such conserved silent sites in the globin genes were those silent changes in the same codon with one or two replacement substitutions. This cannot be the case here, though, as there are no replacement substitutions between rat and chicken actins. Jeffreys (1981) has queeried whether the apparent change in silent site substitution rate is real, or an artefact caused by analysing highly divergent sequences, or by the methods used for correcting these divergences for multiple substitutions. For example, these methods will be inaccurate at high levels of substitution since they assume that the 3 possible base changes per substitution occur with equal probability; this has been found not to be the case (Van Ooyen et al., 1979; Jukes, 1980).

It might appear that the silent site substitutions of actin are similar to those in globin. Hanauer et al. (1983) have also pointed out that the divergence of 61% between rat and human skeletal muscle genes is similar to data obtained for other rat/human gene comparisons using the same calculation: insulin, 64%, growth hormone, 71%; prolactin, 76% (Perler et al., 1980; Cooke et al., 1981). All these values may be misleading, however. Silent site substitutions may not represent a very good evolutionary clock. Thus Perler et al. (1980) found that the individual divergence values for various pairs of globin genes over approximately the same time period differed considerably. For example, the alpha-globin comparison between human and rabbit, and human and mouse, gave corrected divergences of 31% and 83%, respectively. This phenomenon must place considerable doubt on any conclusions based on silent substitution data.

This may be further exemplified by several strange values obtained with various actin gene comparisons. Thus, I have compared the rat

beta-gene (Nudel et al., 1983) with the 35A gene, and the first 267 amino acids of a sea urchin gene (Cooper and Crain, 1982), and found a higher corrected silent site divergence for the former pairing than the latter; 149% and 115%, respectively. Two vertebrate cytoskeletal actin genes might have been expected to be more closely related than vertebrate and echinoderm genes. A comparison betweenthe rat alpha- and beta-genes, also over the first 267 amino acids, (Zakut et al., 1982; Nudel et al., 1983) gives a value of only 82%. This also is a strange result, when compared to the rat beta/35A pairing, as I have shown that Xenopus possesses an actin like the vertebrate muscle-type, and therefore the divergence between this and the cytoskeletal actins is likely to have occurred before the separation of the amphibian and mammalian ancestral lines.

Summary

The high degree of silent site conservation found between two sea urchin actin genes is more likely to be the result of concerted evolution than of a conservation of silent site bases. This conclusion is drawn because of the much higher divergence values found between other actin genes that diverged at approximately the same time as the sea urchin genes. Such figures are roughly similar to the average silent site divergences between globin genes that diverged at the same time. However, because individual divergence values between different pairs of globin genes that have all diverged at a similar time have been found to be quite different, it has previously been concluded that silent site divergence cannot be used as an accurate molecular clock. Comparisons between several actin genes support this conclusion.

4.3 Invariant Codons and Codon Usage

4.3.1 Invariant codons

Turner and Woodland (1982) and Bains (1982) found positions in the histone genes coding for H4 and H3, respectively, where the same codon was used in a large number of genes, in preference to the synonymous codons. I shall refer to these as 'invariant codons'.

The authors of these studies used the codon usage data obtained for their genes, to produce an estimate of the significance of these invariant codons. Thus, Turner and Woodland multiplied together the probabilities of finding a given codon encoding a particular amino acid in each of the genes they had available. In two instances, the probability was low enough to suggest that the occurrence of these codons, in these positions, had been conserved for some function that would not be fulfilled by the synonymous codon.

In view of such claims, I decided to perform a similar survey using sequence of actin genes available at the time of writing. Some of the data obtained from this comparison have been shown in Figure 4.6. In (a), the first 149 amino acids of 14 genes are compared, including the 5AP and 35A sequences; in (b), amino acids 150 to 266 of 13 genes, including 35A, and in (c) amino acids 267 to 375 of 14 genes. As might be expected, the number of positions using the same codon tends to decrease when genes from less related species are compared. Column A compares all the deuterostome and protostome genes, while the other columns each introduce a gene from the lower eukaryotes, and a plant, into the comparison. The information is presented in this way because it is conceivable that some codons may be conserved, but that the function these codons serve may diverge over long evolutionary periods,

Figure 4.6(a)

A		B	C	D	E
asp 11	GAC	+			
phe 31	TTC	+	+	+	
gln 49	CAG	+			
lys 50	AAG	+			+
ser 60	AGC				
lys 61	AAG	+	+	+	
ile 64	ATC	+	+		
lys 68	AAG	+			+
asn 78	AAC	+	+		+
lys 84	AAG	+	+	+	+
phe 90	TTC	+	+	+	
tyr 91	TAC	+	+	+	
glu 100	GAG	+		+	+
lys 118	AAG	+	+		+
ile 122	ATC				
val 139	GTG	+			
thr 148	ACC	+			

Figure 4.6(b)

A		B	C	D	E
pro 104	CCC	+		+	
tyr 188	TAC	+			
ile 192	ATC	+			+
asp 211	GAC	+	+		+
glu 214	GAG	+			
lys 215	AAG	+		+	
cys 217	TGC	+		+	
lys 238	AAG	+		+	
ile 248	ATC	+	+	+	+
glu 253	GAG	+		+	+
phe 262	TTC	+	+	+	+
gln 263	CAG	+			

Figure 4.6(c)

A		B	C	D	E
ile 274	ATC	+			
ile 282	ATC	+	+	+	+
ile 289	ATC	+			+
ala 295	GCC				
asn 296	AAC	+	+	+	+
lys 315	AAG	+	+	+	+
lys 326	AAG	+	+		+
lys 328	AAG	+	+		+
tyr 331	TAC	+	+		
ile 345	ATC	+	+		+
ser 350	TTC	+			
thr 351	ACC	+	+	+	+
lys 359	AAG	+			
his 371	CAC	+	+	+	+
cys 374	TGC				

Figure 4.6: Invariant Codons

E

The sequences of all of the published actin genes were compared at each amino acid position in order to identify those positions where the same codon is used by many genes. The data has been divided into 3 sections: in (a), the first 149 amino acids are compared, including sequence from the 5AP and 35A genes, in (b), amino acids 150 to 267 are compared, including 35A sequence, and (c) where amino acids 268 to 375 are taken, with no Xenopus sequence. The amino acid positions and sequences of those codons invariant in deuterostomes and protostomes are given in column A: rat skeletal muscle and beta-cytoskeletal genes (Zakut et al., 1982; Nudel et al., 1983), a chick skeletal muscle gene (Fornwald et al., 1982), a human cardiac muscle gene (Hamada et al., 1982), two Drosophila genes, 79B and 88F (Sanchez et al., 1982) and two sea urchin genes (Cooper and Crain, 1982; Schuler et al., 1983) (except for part (c), where only 1 sea urchin gene was used). In each of the other columns, an extra gene was compaerd to this set: the Acanthamoeba gene in B (Nellen and Gallwitz, 1982), the yeast gene in C (Gallwitz and Sures, 1980; Ng and Abelson, 1980), the Oxytricha gene in D (Kaine and Spear, 1982), and the soybean gene in E (Shah et al., 1982). Where the same codon was used in these genes, as well, a (+) has been marked.

so that a different codon may be found in a gene from a very distantly related organism. It would be possible, of course, to present comparisons of genes separated by even shorter periods of divergence, but I have been unable to demonstrate the significance of any such codons.

In order to attempt a numerical estimate of the significance of any invariant codons, the frequency of occurrence of all the codons in each gene has been counted and reproduced in Figure 4.7. The frequency of occurrence of each codon in each gene can also be expressed as a fraction of the total occurrence of that amino acid, and used as the probability of finding that codon in any given position where that amino acid occurs. It is clear that, as with other genes in various organisms (Grantham et al., 1980), the use of synonymous codons is not random (see below). This is important when considering the significance of the invariant codons. It can also be seen that the pattern of codon usage varies between the organisms. In particular, soybean, yeast and Oxytrycha differ greatly. For this reason, the numbers have not been expressed as an average of all the genes available, as was done by Turner and Woodland (1982).

Consideration of Figure 4.6 reveals that there are very few codons that are invariant in all the genes for which sequence is available. Comparison between this and Figure 4.7 will show that almost all the invariant codons in Figure 4.6 involve the most widely used of the synonymous codons. If it were assumed (for the moment) that all of the genes used here are totally unrelated to each other, then it is possible to calculate the probability of a given codon occuring at a specified position in a given number of genes. This was done by calculating, for each gene, the probability of a given codon being used for a given amino

acid, and then multiplying together these numbers. Thus, the probability of finding the codon AAG used for lysine 84 in the 14 genes of Figure 4.6(a) is approximately 0.08. But what is really required is the probability of finding a given invariant codon, a given number of times, anywhere in the gene. Thus, as there are 4 positions in the 14 actin genes shown in Figure 4.6(a) where lysine is present in all the genes, the probability required is that of obtaining an invariant codon in one amino acid position with no invariant codons in the other three. In this case, there are 4 possible combinations of 1 invariant codon and 3 non-invariant codons that can be achieved (the formula is, in fact, $n! \div [(n-r)! \times r!]$ where ! = factorial, and, in this case, n = 4, and r = 1). Thus, in this example, the probability is $0.08 \times (1-0.08)^3 \times 4 = 0.25$. This does not appear to be very significant. Similarly high (>0.1) probabilities were obtained for a selection of several other invariant codons where the codons involved were preferred to the synonymous codons. Of course, this analysis does not show that such invariant codons are <u>not</u> the result of conservation for functional reason; the overall preference of a gene for certain codons might be the result of such conservation. However, there may be other explanations for this preference (e.g. conservation or mutability of meCpG, preferred codon-anticodon binding energies, isoacceptor tRNA abundance; see Section 4.3.2). Only when the codon used is one that is infrequently used in the rest of the gene, is it possible to demonstrate that an invariant codon is not a chance observation.

I have found it possible to generate smaller probability values by varying the genes considered in the comparison. For example, if only 11 genes are compared (Fig. 4.6a, column B), then AAG occurs for lysine in all the genes at 5 positions (out of a possible number of 7), with a

probability of 0.0115. If 8 genes are compared (Fig. 4.6a,b, and c, column B), 11 lysines out of 18 use AAG in all the genes, with a probability of 0.026. However, the only examples which are entirely convincing are those that use the least-preferred codons. Serine 60 is encoded by AGC in 10 genes (Fig. 4.6a) and Valine 139 by GTG in 11 genes (Fig. 4.6a). There are 6 positions where this could occur for the serine and 9 for the valine. Consequently, the probabilities of these occurrences are 2.48×10^{-8} and 1.46×10^{-4}, respectively. Serine 60 is also AGC in the partial sequences for the mouse skeletal and cardiac muscle actin genes (Minty et al., 1982), as well as the human skeletal muscle actin gene (Hanauer et al., 1983). It is not, however, the codon used at this position in the recently published sequences of the 4 nematode genes (Files et al., 1983). Valine 139 has also been found to be GTG in a partial rabbit skeletal muscle actin sequence (Putney et al., 1983) and the human skeletal muscle actin gene (Hanauer et al., 1983).

Another example of a significant-looking invariant codon is that of the TGT codon used for the cysteine following the first methionine in 8 of the genes. The probability of this occurring at this position by chance, separately, in each of these genes, is 4.67×10^{-5}. In fact, the same codon is used in this position in the other 4 *Drosophila* genes (not fully sequenced and so not included in the survey, Fyrberg et al., 1981), as well as the 4 nematode genes (Files et al., 1983). The significance of this finding is diminished, however, by the recent publication of the human skeletal muscle actin gene sequence (Hanauer et al., 1983; Gunning et al., 1983), where the TGC codon is used in this position.

The probability values calculated for such invariant codons may

appear quite significant. It must be remembered, however, that the probabilities calculated in this way are likely to be underestimates. This is because the method of multiplying together the probabilities of codon usage from several genes (to give the probability of that codon being used by all the genes) relies on each gene having been formed as a completely separate event. This, of course, is not true and, although the neutral silent sites of genes from distantly related species may have been randomised, this may not be true for other genes, such as the two closely related sea urchin genes, and perhaps the two Drosophila genes. It is not known how long it takes for the silent site substitutions of actin genes to saturate; the reasons for this have been discussed in Section 4.2. Despite these reservations, I believe that the invariant codons used for Serine 60, Valine 139, and Cysteine-1 are not the result of chance, but are the result of their conservation for some unknown function.

4.3.2 Codon usage

Figure 4.7 shows the frequency of use of the various codons in 14 partial and complete actin genes (including the 5AP and 35A genes). It is clear that, for each amino acid, there are codons that are preferred over others. Figure 4.8 shows the frequency of occurrence of each base at the third codon position for each of these genes. The first and second positions have not been included as these will obviously be influenced by the amino acid composition of the actins. For the same reason the methionine (ATG) and trytophan (TGG) codons have not been included in the third position analysis. Most of the other amino acids are encoded by a set of synonymous codons where the third position is

Figure 4.7

		A	B	C	D	E	F	G	H	I	J	K	L	M	N
Arg	CGT	5	6	8	7	6	8	8	9	6	6	3	5	1	6
	CGC	1	2	6	3	6	6	1	1	12	10	14	0	1	1
	CGA	0	1	0	1	2	0	1	1	0	1	0	0	0	3
	CGG	0	0	3	3	2	1	0	0	0	0	0	0	0	0
	AGA	1	4	1	1	0	1	2	1	0	0	0	13	21	6
	AGG	0	1	0	0	2	2	5	4	0	1	1	0	0	2
Leu	TTA	0	0	1	0	0	0	0	1	0	2	0	2	1	0
	TTG	1	2	3	1	2	3	2	0	2	1	0	19	7	4
	CTT	0	4	0	0	2	1	10	9	2	0	3	2	0	11
	CTC	1	4	6	6	5	5	11	1	3	2	18	1	11	9
	CTA	0	1	1	1	0	1	0	1	3	0	0	2	3	3
	CTG	8	1	16	19	17	16	4	4	18	24	8	0	1	0
Ser	TCT	1	5	5	5	5	3	5	5	2	0	1	14	4	7
	TCC	4	4	11	14	11	16	13	7	16	11	6	12	5	5
	TCA	0	1	0	1	2	0	3	4	1	1	0	3	11	5
	TCG	0	0	0	0	1	0	0	0	4	11	17	0	0	0
	AGT	0	0	0	0	0	0	1	0	0	0	0	3	1	7
	AGC	1	4	5	5	4	4	4	3	2	1	0	0	7	7
Thr	ACT	2	6	9	7	4	4	3	4	2	1	3	13	9	9
	ACC	7	7	12	15	20	16	15	13	20	22	22	8	11	6
	ACA	2	4	5	4	2	7	2	0	0	0	0	0	1	3
	ACG	0	0	0	0	1	0	1	0	1	3	1	0	0	0
Pro	CCT	1	2	6	11	5	5	1	1	0	3	1	6	0	5
	CCC	5	6	9	6	10	10	9	8	12	11	15	0	5	6
	CCA	3	6	1	2	2	4	9	6	3	4	0	13	9	7
	CCG	0	0	3	0	2	0	0	0	4	1	3	0	1	0
Ala	GCT	6	15	12	9	12	10	10	10	4	7	4	15	9	11
	GCC	7	4	13	18	15	18	19	16	21	18	23	11	14	8
	GCA	0	4	2	1	0	0	1	0	2	1	0	0	0	9
	GCG	0	1	2	3	2	1	0	0	3	3	2	0	0	0
Gly	GGT	8	9	8	9	10	10	10	11	10	16	12	28	15	14
	GGC	1	4	15	14	13	12	6	4	17	12	16	0	1	4
	GGA	0	7	2	1	0	1	12	10	2	1	2	0	9	8
	GGG	3	0	3	2	5	5	0	0	0	0	0	0	0	3
Val	GTT	2	4	3	2	0	0	9	9	1	1	2	14	14	13
	GTC	2	3	10	5	5	8	10	10	8	5	15	12	9	7
	GTA	1	5	2	2	0	1	1	0	4	2	0	0	3	3
	GTG	6	5	6	13	16	12	3	3	10	13	6	0	0	5

		A	B	C	D	E	F	G	H	I	J	K	L	M	N
Ile	ATA	0	1	0	0	0	0	0	0	0	0	0	0	0	1
	ATC	7	7	18	20	25	22	26	16	25	24	24	15	19	11
	ATT	1	9	12	8	5	8	2	1	2	5	3	14	7	16
Lys	AAA	0	3	3	3	5	5	1	1	0	2	0	7	4	7
	AAG	7	8	16	17	14	16	18	11	19	17	19	11	15	13
Asn	AAC	4	4	4	6	8	8	9	9	8	8	9	9	13	4
	AAT	2	3	7	3	3	3	0	0	1	1	0	0	1	4
Gln	CAA	0	4	2	1	1	4	1	2	3	1	1	14	13	7
	CAG	5	4	9	11	10	7	12	7	10	10	10	0	0	2
His	CAC	3	5	7	7	8	7	7	6	9	7	10	7	5	3
	CAT	2	2	2	2	1	2	2	2	0	2	0	3	1	7
Glu	GAA	5	11	7	4	3	7	6	4	4	0	0	26	14	12
	GAG	6	9	21	21	25	21	20	18	24	26	28	1	13	17
Asp	GAC	4	8	13	18	18	16	15	13	16	16	18	9	10	6
	GAT	5	10	9	5	4	6	8	8	5	8	3	11	10	17
Tyr	TAC	2	4	10	10	10	8	15	14	10	12	14	14	10	2
	TAT	3	7	6	5	6	8	0	0	6	3	1	0	3	11
Cys	TGC	1	3	3	4	4	5	5	3	6	6	4	0	5	3
	TCT	1	0	3	2	2	1	1	1	1	1	0	4	1	1
Phe	TTT	1	5	4	1	3	4	1	1	1	1	2	2	0	5
	TTC	4	5	8	12	9	8	13	10	11	12	11	12	14	6

M	N
0	1
19	11
7	16
4	7
15	13
13	4
1	4
13	7
0	2
5	3
1	7
14	12
13	17
10	6
10	17
10	2
3	11
5	3
1	1
0	5
14	6

Figure 4.7: Codon Usage

Codon usage in 11 complete and 3 incomplete (5AP, 35A, and sea urchin 1) actin genes.

A = 5AP gene

B = 35A gene

C = human cardiac actin gene (Hamada et al., 1982)

D = rat beta-actin gene (Nudel et al., 1983)

E = rat skeletal muscle actin gene (Zakut et al., 1982)

F = chick skeletal muscle actin gene (Fornwald et al., 1982)

G = sea urchin pSp G17 gene (Cooper and Crain et al., 1982)

H = sea urchin actin gene 1 (Schuler et al., 1983)

I = Drosophila 79B gene (Sanchez et al., 1983)

J = Drosophila 88F gene (Sanchez et al., 1983)

K = Acanthamoeba gene (Nellen and Gallwitz, 1982)

L = yeast gene (Gallwitz and Sures, 1980)

M = Oxytricha gene (Kaine and Spear, 1982)

N = soybean gene (Shah et al., 1982)

Figure 4.8

	T	C	C/T	A	G	G/A
5AP	40	54	1.35	12	36	3.00
35A	87	74	0.85	49	31	0.63
Human cardiac muscle	97	150	1.55	27	82	3.04
Rat beta cytoskeletal	76	163	2.14	22	93	4.23
Rat skeletal muscle	68	171	2.51	17	99	5.82
Chick skeletal muscle	73	169	2.32	29	84	2.90
Sea urchin pSpG17	71	178	2.51	39	65	1.67
Sea urchin 1	80	125	1.56	31	47	1.52
Drosophila 79B	43	196	4.56	22	95	4.32
Drosophila 88F	55	177	3.22	15	110	7.33
Acanthamoeba	78	211	2.71	3	95	31.67
Yeast	134	110	0.82	80	31	0.39
Oxytricha	102	141	1.38	89	37	0.42
Soybean	144	90	0.63	74	46	0.62

Figure 4.8: Third Base Frequences

The frequence of occurrence of each base in the third position of the codons of 14 actin genes (3 being incomplete, see Figure 4.7 for the references). To avoid the influence of amino acid composition only the occurrence of the two pyrimidines or the two purines should be compared. For the same reason, the single codons for methionine (ATG) and tryptophan (TGG) have been excluded.

G/A
3.00
0.63
3.04
4.23
5.82
2.90
1.67
1.52
4.32
7.33
31.67
0.39
0.42
0.62

occupied just by pyrimidines, just by purines, or both; the C to T ratio or A to G ratio, should not, therefore, be affected by the protein's amino acid composition. The only exception is isoleucine with the codons ATA, ATC and ATT. Here, however, the ATA codon is used only twice for 354 isoleucines and so will not appreciably affect the A:G ratio in this analysis.

From Fig. 4.8 it is clear that, in most of the genes, there is a marked under-usage of A in the third position, relative to G. The exceptions to this are the genes of 35A, yeast, Oxytricha and soybean, where A is preferred to G. Similarly, there is preference for C over T (but not so marked), except in 35A, yeast, and soybean. A preference for G over A in the third position has been noted by Grantham et al. (1980, 1981) in several different mammalian genes (but not, however, in the case of immunoglobulins and certain hormones). These authors also demonstrated that different genome types use distinct strategies of codon usage (e.g.: mammalian, bacterial, virus, mitochondrial, yeast and slime mould genes fall in different classes). These results appear to concur with this view. For example Figure 4.7 shows the CTG codon being preferred for leucine in three mammalian, one chicken and two Drosophila genes, while CTT or CTC is preferred for the sea urchin, Acanthamoeba, Oxytricha and soybean genes, and TTG for the yeast gene.

There also appear to be differences in usage of particular codons in different gene types within a species. Figure 4.7 shows that the Xenopus borealis 5AP gene, which probably codes for a muscle actin, often shows a different codon preference to that of the 35A gene, a Xenopus borealis cytoskeletal actin gene. This is observed with the codons for serine, alanine, glycine, glutamine, isoleucine, and phenylalanine. This phenomenon has been observed before. In Xenopus

laevis histone H4 genes, TTC is always used for phenylalanine (Turner and Woodland, 1982), while both codons occur equally in Xenopus laevis globin (Richardson et al., 1980). A distinct difference in isoleucine codon usage can also be found in these two types of gene. In mammalian immunoglobulin genes, almost twice as much degenerate A can be found as G (Grantham et al., 1981). This is in complete contrast to most other mammalian genes (see above).

What is the cause of non-random codon usage and the phylogenetic and gene-specific differences? A possible influence on the bias for certain codons may be related to the ease with which some codons can be mutated. For example, in animal cells, 50 to 70% of the dinucleotide 5'-CpG-3' occurs in a form with the cytosine methylated. MeC residues are thought to undergo spontaneous deamination to T residues, which are not recognized by any correction mechanism, and so are left in the DNA. Thus, where the meC of a meCpG dinucleotide has occurred in the third position of a codon, this might be expected to change to a T (without changing the amino acid encoded). Also, where the meC of a meCpG dinucleotide in the other strand has changed to a T, a meCG may also change to a meCA (where A is the third base of a codon). The soybean, yeast, and 35A genes prefer T and A to C and G, in the third position of their codons. Is this a result of meC deamination? Figure 4.9(a) shows the numbers of various dinucleotides occurring in the third and first base positions of these genes as well as the 5AP gene. The dinucleotides are arranged in pairs so that the ratio within a pair is not affected by the amino acid composition of the protein; dinucleotides using the third base of ATG and TGG codons have been excluded. The soybean and 35A genes both show a high ratio of TpG to CpG, which is higher than the ratio in any other pairs. Figure 4.9(b) shows the

Fig. 4.9

a

	soybean	35A	5AP	yeast
AA	27	20	1	32
GA	16	14	13	16
CA	40	23	18	27
TA	22	14	7	31
AC	11	7	4	13
GC	8	14	6	1
CC	25	24	18	14
TC	16	12	5	22

	soybean	35A	5AP	yeast
AG	20	23	6	22
GG	19	11	11	16
CG	6	7	8	40
TG	86	55	27	47
AT	15	3	0	14
GT	5	4	6	7
CT	18	22	10	27
TT	11	5	1	34

b

	soybean	35A
AG	32	20
AA	26	18
AC	16	22
AT	40	22
GG	5	1
GA	17	12
GC	16	13
GT	28	15

	soybean	35A
CG	0	1
CA	24	15
CC	23	21
CT	34	28
TG	12	18
TA	7	7
TC	33	20
TT	45	21

Figure 4.9: Dinucleotide Frequencies

Dinucleotide frequencies in the 5AP and 35A genes, the yeast actin gene (Gallwitz and Sures, 1980), and a soybean actin gene (Shah et al., 1982). The dinucleotides have been grouped into pairs, between which, direct comparisons can be made with no distortion occurring due to the amino acid composition of the protein. In (a) the dinucleotides are made up from the third base of each codon plus the first base of the next 3' codon. In (b) the dinucleotides are made up from the second and third bases of each codon. Dinucleotides including the third base of methionine (ATG) or tryptophan (TGG) codons have not been included.

ratios of the dinucleotides occurring in the second and third positions of codons in these two genes; there is clearly a predominance of CA over CG in both genes (although there is also a preference for GA over GG). Thus meC mutability appears to have affected the codon usage of these two genes. Figure 4.9(a) indicates, however, that this explanation cannot be applied so easily to the yeast actin gene.

The 5AP gene was included for comparison with the 35A gene, and it is interesting that the ratios of TpG to CpG differ between the two. This apparent difference may result from the small number of bases available in the 5AP gene, but similar differences in CpG content between related genes from the same species have been observed by McClelland and Ivarie (1982), comparing rabbit alpha- and beta-globin genes, and human alpha-2- and beta-globin genes.

Codon usage in the soybean and 35A genes may be affected by meC mutability, but most of genes shown in Figure 4.8 actually have preference for C and G over T and A. McClelland and Ivarie (1982) found mammalian genes to be, on average, enriched in CpG, relative to the rest of the genome. This difference was not confined to the coding regions, where selection of amino acids takes place, and they suggested that the genes may have maintained CpG levels by selection, because of the possible role of meC in the regulation of gene expression. An inverse correlation between gene expression and gene methylation has been found in the case of the Xenopus ribosomal genes (Bird et al., 1981), in globin genes of various species (McGhee and Ginder, 1979; van der Ploeg and Flavell, 1980; Shen and Maniatis, 1980b; Gerber-Huber et al., 1983), in the chicken ovalbumin gene (Mandel and Chambon, 1979), in a Xenopus albumin gene (Gerber-Huber et al., 1983), and various integrated viral genes (Groudine et al., 1981; Pollack et al., 1980; Sutter and Doerfler,

1980). Conservation of CpG pairs for gene regulation cannot, however, explain the third position's preference of C to T in the case of Drosophila and the sea urchin, as methylation is rare in the former, and most of the latter's genes appear to be unmethylated (Smith and Thomas, 1981; Bird et al., 1979).

What other factors are responsible for the nonrandom usage of codons? It is possible that certain codons may be required at specific sites to produce the correct secondary structure for the mRNA. Clearly, to influence the overall codon usage of a gene, a large number of such codons would be required. This type of phenomenon may indeed be what has been observed when 11 lysines out of 18 use AAG in 8 of the actin genes (see Section 4.3.1).

Another type of influence may be that of the differing codon-anticodon binding energies found with various codons. In bacteria, highly expressed genes tend to avoid codons of strongest binding energy (Grantham et al., 1981). Various authors have suggested that the codon usage of abundantly expressed genes in yeast has also evolved to produce optimum and uniform codon-anticodon binding energies (Bennetzen et al., 1982; Ikemura, 1981a, 1981b).

It has also been suggested that a major factor in determining the usage of synonymous codons is the abundance of the various isoacceptor tRNAs (Bennetzen and Hall, 1982; Ikemura, 1982; Gouy and Gautier, 1982). In yeast, the codon usage of the genes for the abundant proteins glyceraldehyde-3-phosphate dehydrogenase and alcohol dehydrogenase-1 is heavily biassed in favour of those codons whose tRNAs are most abundant (Bennetzen and Hall, 1982). However, in cells of certain highly specialised eukaryotic tissues, tRNA abundance is adjusted to the codon frequency in the mRNAs for the proteins predominantly produced in those

tissues, i.e. the tRNA profiles are tissue-specific. This would seem to indicate that it is not necessarily always the tRNA profile that determines codon usage; the codon usage can presumably also affect the levels of the various tRNAs. Thus, it is not a sufficient explanation for the non-random usage of codons to say that this is related to tRNA abundance. Although changes in the latter (random?) may certainly have been contributory, it also seems likely that other factors, such as those discussed above, may have affected codon usage.

Summary

Like other proteins, synonymous codons in actin genes are not used in a random fashion; there is a bias towards certain codons. Codon usage in actin genes, like other genes, varies over phylogenetic distances. The codon usage for the 5AP gene is different to that of the 35A gene; this has been found for other comparisons of genes expressed in different tissues. The biassed use of codons in the 35A and soybean genes can be partly explained by the mutation of a meC residue in the third position to T, or the mutation of a complementary meC residue giving rise to an A in the third position, replacing a G. This explanation, however, cannot be so easily applied to yeast, the only other organism that shows a preference for T and A over C and G, in the third position. All the other genes (except that of Oxytricha) show a preference for C and G over T and A in the third position. This might be explained, in some cases, as a result of CpG conservation for the purpose of control of gene expression by methylation. Drosophila and sea urchin genes, however, are unmethylated. Various other factors may effect codon usage, including codon-anticodon binding energies, mRNA secondary structure, and tRNA abundance.

4.4 The Introns: Their Location and Evolution

As in many other genes (Breathnach and Chambon, 1981), the boundaries of the introns are all ambiguous unless the GT-AG rule is employed. This rule is based on the observation that in all the polymerase II split genes so far studied, the intron-exon junctions can always be drawn so that the intron begins with the dinucleotide GT at its 5' end, and ends with the dinucleotide AG at its 3' end (Breathnach and Chambon, 1981; Breathnach et al., 1978). As an example of the ambiguities found in these genes, I will take the intron between amino acids 41 and 42 in 35A. I have assigned the splice sites as CAC,CAG/GTT--AAG/GTT,GTA (where / = the splice site, and the commas separate the codons). In this way, the amino acid sequence expected (by comparison with other vertebrate actins) is obtained. However, the same amino acids could also be obtained by drawing in the following splice sites: CAC,CAG,G/TT--AAGG/GT,GTA or CAC,CA/GGTT--AA/G,GGT,GTA or CAC,C/AGGTT--A/AG,GGT,GTA. These, however, do not follow the GT-AG rule. Similar alternatives can be drawn for all the other intron/exon boundaries in these clones. The common occurrence of this phenomenon can be attributed to the direct repeat nature of the 5' and 3' junctions. Thus AG/GT--AG/GT is often found at splice sites (Breathnach and Chambon, 1981). The splice site sequences in these genes will be discussed later, in Section 4.5. For the moment, I will deal with the most interesting aspect of these introns: their positions within the coding sequence.

The intron positions in 5AP and 35A are different. As Figure 4.1 shows, the introns in 5AP occur between amino acids 41 and 42 and in the middle of the codon for amino acid 150. Figure 4.2 shows that the introns in 35A occur between amino acids 41 and 42, between amino acids

121 and 122, and in the middle of the codon for amino acid 267.

This result was surprising because several other gene families had been found to consist of members whose introns were present at exactly the same location; for example, the globin genes (Efstratiadis et al., 1980), the ovalbumin genes and their homologues (Royal et al., 1979; Heilig et al., 1980), the vitellogenin genes (Wahli et al., 1980) and the immunoglobulin genes (Yamawaki-Katoaka et al., 1981). It had been the accepted view that intervening sequence positions are evolutionarily stable. Thus, the intron positions of the globin family have been found to remain constant throughout vertebrate evolution (Patient et al., 1980; Barrie et al., 1981).

A similar phenomenon has recently been found with other actin genes. Thus, chicken skeletal muscle, rat skeletal muscle and human cardiac muscle actin genes all have introns at positions 41/42, 150, 204, 267 and 327/328 (Fornwald et al., 1982; Zakut et al., 1982; Hamada et al., 1982), whereas the rat beta-cytoskeletal gene has introns at positions 41/42, 121/122, 267, and 327/328. The 5AP and 35A intron positions can be seen to fit in exactly with this difference in intron patterns, which appears to differ between the striated muscle types and the cytoskeletal types (whether this pattern differs further in the smooth muscle and gamma-cytoskeletal genes must await further cloning work). Although 5AP and 35A do not contain complete genes, and therefore possibly differ from the mammalian and chicken intron patterns, this data would seem to support the proposal put forward earlier that the muscle-type and cytoskeletal actin genes started to evolve separately before the divergence of amphibians from the rest of the vertebrates.

The intron positions of all the actin genes studied to date are

shown in Figure 4.10. It can be seen that Drosophila, a nematode, and a sea urchin also each contain genes with different positions. Intron positions also vary between the organisms compared; in fact, introns have now been found in 14 different positions in the actin genes (including one in the 5' untranslated region). The variability of sites, shown in Figure 4.10, has been discussed by several authors (Zakut et al., 1982; Davidson et al., 1982). Two extreme models can be envisaged to explain this variability: firstly, that the original actin gene contained many introns which have since been gradually eliminated, or secondly, that the original actin gene contained no introns, and that these have been introduced during evolution.

Intron elimination has been demonstrated in other genes (Lomedico et al., 1979; Hyldig-Nielsen et al., 1982). Thus, for example, the insulin genes of chicken (Perler et al., 1980) and man (Bell et al., 1980b) have 2 intervening sequences. The rat has 2 insulin genes; one of these has a similar structure with 2 introns, but the other gene has lost the second intron (Lomedico et al., 1979). Intron elimination would seem to explain the evolution of the deuterostome actin genes (vertebrate and sea urchin) from a primordial gene containing at least 7 introns. Each type of sea urchin gene contains intron positions which are found separately in the vertebrate muscle and cytoskeletal genes.

The events giving rise to the completely different intron positions in the Protista and lower eukaryotes are more difficult to imagine. These organisms contain far fewer introns; in the case of Dictyostelium and Oxytricha, probably none at all (although none of the former genes have been completely sequenced and, in the latter case, the gene was cloned from macronuclear DNA, so that introns may have been eliminated during the development of the macronucleus from the micronucleus; Kaine

Fig. 4.10

	L	4	13	18 – 19	41 – 42	63	105 – 106	121 – 122	150	204	267	307	327 – 328	353 – 354
human cardiac muscle		—	—	—	O	—	—	—	O	O	O	—	O	—
rat skeletal muscle	O	—	—	—	O	—	—	—	O	O	O	—	O	—
chick skeletal muscle	O	—	—	—	O	—	—	—	O	O	O	—	O	—
rat beta	O	—	—	—	O	—	—	O	—	—	O	—	O	—
SAP		—	—	—	O	—	—	—	O					
35A	O	—	—	—	O	—	—	O	—	—	O			
sea urchin C	—	—	—	—	—	—	—	O	—	O	—	—	—	—
sea urchin J		—	—	—	O	—	—	O	—	O	O	—	—	—
Drosophila 5C	O	—	—	—	—	—						—		
Dros. 57B	—	—	O	—	—							—		
Dros. 79B	—	—	—	—	—	—	—	—	—	—	—	O	—	—
Dros. 88F	—	—	—	—	—	—	—	—	—	—	—	O	—	—
nematode I II III		—	—	—	—	O								
nematode IV		—	—	O	—	—								
Oxytricha	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Acanthamoeba	—	—	—	—	—	—	O	—	—	—	—	—	—	—
yeast	—	O	—	—	—	—	—	—	—	—	—	—	—	—
Dictyostelium	—	—	—	—	—	—	—	—	—	—	—	—	—	—
soybean	—	—	—	O	—	—	—	—	O	—	—	—	—	O
corn	—	—	—	O	—	—	—	—	O	—	—	—	—	O

Figure 4.10: Intron Positions in Actin Genes

The 14 positions where introns have been found in various actin genes are shown along the top of the grid. The numbering system is that originally used for rabbit skeletal muscle actin (Collins and Elzinga, 1975) and, since the N-terminal sequences of actin proteins show different numbers of amino acids, and internal additions and deletions may occur in the actin protein sequence, the intron position given here is not necessarily equivalent to the actual numbering system for the actin gene in question. Where two numbers are given this signifies that the intron is found inbetween the triplets coding for those amino acids, while where there is only one, the intron occurs within that codon. 'L' represents the 5' untranslated (but transcribed) region of the gene. 'O' indicates the presence of an intron, '-' the absence of an intron, and a blank indicates that there is no information available concerning introns in that region. The genes are divided into four groups according to the type of organism they were isolated from: deuterostomes, protostomes, protists, and plants (from top to bottom). The actin genes examined include a human cardiac muscle gene (Hamada et al., 1982), a rat skeletal muscle gene (Zakut et al., 1982), a chicken skeletal muscle gene (Fornwald et al., 1982), a rat beta-cytoskeletal gene (Nudel et al., 1983), the 5AP and 35A genes (this work), the sea urchin C-type and J genes (Cooper and Crain, 1982; Davidson et al., 1982), the Drosophila 5C, 57B, 79B, and 88F genes (Fyrberg et al., 1981; Sanchez et al., 1982), all four nematode genes (Files et al., 1983), an Oxytricha macronucleus gene (Kaine and Spear, 1982), an Acanthamoeba gene (Mellen and Gallwitz, 1982), the yeast gene (Gallwitz and Sures, 1980; Ng and Abelson, 1980), the Dictyostelium genes (McKeown and

Firtel, 1981b), a soybean gene (Shah et al., 1982), and a corn (Shah, unpublished information) gene.

and Spear, 1982). None of the intron positions of these organisms has been found in deuterostomes. It might be argued, form this, that the second model (intron addition) might be applicable to the earlier evolution of actin genes. The first model is also possible, however, as protists may represent the more advanced evolutionary forms in that their smaller genomes may result from 'streamlining' introduced during the much higher number of divisions they have undergone (Doolittle, 1978). The original actin gene may have started with 14 or more introns, in that case, and most of these may have been eliminated in protists. Evidence that this might be the case comes from recent results in soybean and corn actin genes. One of the 3 introns of these genes occupies the same position as the vertebrate muscle genes. A second intron is in an identical position to an intron found in one of the nematode genes (see Figure 4.10). It seems unlikely that this could be due to coincidence; rather, this suggests that these introns were present in the ancestral gene leading to all of these lines (unless one proposed that there exist preferred sites for intron insertion).

It seems that some introns were present in very early ancestral genes and that these have been removed in some organisms. It is not possible, however, to exclude the possibility that these or any other introns, are inserted into the actin genes at some time. It seems that if such insertions were to occur, they would have to be accompanied by the ability of the gene to splice them out. This is because actin is such a conserved protein it seems unlikely that the insertion of extra coding sequence would be tolerated functionally, even for a small number of generations.

In a number of proteins different structural and functional domains have been shown to be represented by different exons (Stein et al.,

and Spear, 1982). None of the intron positions of these organisms has been found in deuterostomes. It might be argued, form this, that the second model (intron addition) might be applicable to the earlier evolution of actin genes. The first model is also possible, however, as protists may represent the more advanced evolutionary forms in that their smaller genomes may result from 'streamlining' introduced during the much higher number of divisions they have undergone (Doolittle, 1978). The original actin gene may have started with 14 or more introns, in that case, and most of these may have been eliminated in protists. Evidence that this might be the case comes from recent results in soybean and corn actin genes. One of the 3 introns of these genes occupies the same position as the vertebrate muscle genes. A second intron is in an identical position to an intron found in one of the nematode genes (see Figure 4.10). It seems unlikely that this could be due to coincidence; rather, this suggests that these introns were present in the ancestral gene leading to all of these lines (unless one proposed that there exist preferred sites for intron insertion).

It seems that some introns were present in very early ancestral genes and that these have been removed in some organisms. It is not possible, however, to exclude the possibility that these or any other introns, are inserted into the actin genes at some time. It seems that if such insertions were to occur, they would have to be accompanied by the ability of the gene to splice them out. This is because actin is such a conserved protein it seems unlikely that the insertion of extra coding sequence would be tolerated functionally, even for a small number of generations.

In a number of proteins different structural and functional domains have been shown to be represented by different exons (Stein et al.,

1980; Artymuik et al., 1981; Hyldig-Nielsen et al., 1982; Tonegawa et al., 1978; Craik et al., 1980; Sargent et al., 1981). The actin molecule has so far been shown to contain only two structural domains (Suck et al., 1981). However, because of the large number of interactions of actin with other proteins (Weeds, 1982), and its extraordinary evolutionary conservation, it seems likely tht it is composed of a large number of small functional domains. If the original actin gene contained many introns, it would be possible to invoke the idea of Gilbert (1978) that 'exon shuffling' may have served to construct this gene from a series of small exons, each having evolved to encode a polypeptide with a specific interaction. This might be the explanation for the presence of some of the introns now found in the genes, even if it doesn't account for all of them; some introns may have been added after the construction of the gene in this way.

Summary

The intron positions of the actin genes in 5AP and 35A are different. The positions of the former gene's introns are the same as those of the striated muscle actin genes of other vertebrates, and the position of the latter genes, the same as those of the vertebrate cytoskeletal genes. This suggests that the vertebrate muscle-specific and cytoskeletal actin genes became separate before the divergence of amphibians from the other vertebrates.

The introns of actin genes from various deuterostomes, protostomes, and protists occupy differing positions; in total, 14 different ones. It seems likely that some of these introns existed in a 'primordial actin gene', which existed before the separation of these lines, and have since been eliminated in certain organisms. It is not known,

however, whether some of these introns could have been introduced later on in the evolution of certain actin genes.

4.5 Introns: Splice Site Sequences and Splicing Mechanisms

Figure 4.11 compiles the sequences at the intron splice sites of 5AP and 35A, and compares these with the consensus sequence of Breathnach and Chambon (1981). The latter was based on a study of 90 5' splice junctions and 85 3' splice junctions of various protein-coding genes. Both that study, and the sequences presented in this thesis, show that only the GT at the 5' end of the intron, and the AG at the 3' end, are present at all the sites. The other positions in the consensus sequence represent bases that were present in a greater number of sites than the alternatives. Some of the splice sites of 5AP and 35A have sequences that differ from this consensus. For example, at position 1 of the intron at amino acid 267 of 35A there is a thymine which is found in only 4 out of the 90 sites examined by Breathnach and Chambon. Similarly the adenine at position 17 of the 35A (41/42) intron was found in only 3 of the 85 3' splice sites, and the cytosine at position 8 of the 5AP (41/42) intron was found in only 1 out of 90 5' splice sites. The number of differences of each sequence from the consensus has been shown in Figure 4.11(a), and it is clear that for most of the sites only a few positions differ.

It has been suggested that one or more of the small nuclear RNAs found in the nuclei of eukaryotic cells (Busch *et al.*, 1982) may be involved in the mechanism of splicing. One of these, designated U1, has a sequence which displays complementarity to the consensus sequence for both 5' and 3' splice sites. Consequently, Lerner *et al.*, (1980) and Rogers and Wall (1980) have proposed that the U1 RNA might hydrogen bond with both 5' and 3' sequences in such a way as to hold the two exon ends together for splicing to occur. Mount *et al.* (1983) have recently produced evidence that U1snRNPs will selectively bind 5' splice sites *in*

4.5 Introns: Splice Site Sequences and Splicing Mechanisms

Figure 4.11 compiles the sequences at the intron splice sites of 5AP and 35A, and compares these with the consensus sequence of Breathnach and Chambon (1981). The latter was based on a study of 90 5' splice junctions and 85 3' splice junctions of various protein-coding genes. Both that study, and the sequences presented in this thesis, show that only the GT at the 5' end of the intron, and the AG at the 3' end, are present at all the sites. The other positions in the consensus sequence represent bases that were present in a greater number of sites than the alternatives. Some of the splice sites of 5AP and 35A have sequences that differ from this consensus. For example, at position 1 of the intron at amino acid 267 of 35A there is a thymine which is found in only 4 out of the 90 sites examined by Breathnach and Chambon. Similarly the adenine at position 17 of the 35A (41/42) intron was found in only 3 of the 85 3' splice sites, and the cytosine at position 8 of the 5AP (41/42) intron was found in only 1 out of 90 5' splice sites. The number of differences of each sequence from the consensus has been shown in Figure 4.11(a), and it is clear that for most of the sites only a few positions differ.

It has been suggested that one or more of the small nuclear RNAs found in the nuclei of eukaryotic cells (Busch *et al.*, 1982) may be involved in the mechanism of splicing. One of these, designated U1,has a sequence which displays complementarity to the consensus sequence for both 5' and 3' splice sites. Consequently, Lerner *et al.*, (1980) and Rogers and Wall (1980) have proposed that the U1 RNA might hydrogen bond with both 5' and 3' sequences in such a way as to hold the two exon ends together for splicing to occur. Mount *et al.* (1983) have recently produced evidence that U1snRNPs will selectively bind 5' splice sites *in*

Fig. 4.11

a

	Exon	/	Intron	/	Exon
5AP(41/42)	CAC CAG	$/^{2}$	GTATC TATC ___	CTTTTTTTAG	$/^{1}$ GGT GT
5AP(150)	CCA CAG	$/^{1}$	GTGAGCATA ___		
35A(Leader)	ACC AAG	$/^{1}$	G TAAGACCT ___	TGTCACACAG	$/^{4}$ AATTA
35A(41/42)	CAC CAG	$/^{2}$	GTTTG TATT ___	CTGTTTTAAG	$/^{2}$ GGT GT
35A(121/122)	CA CAG	$/^{2}$	GTTTGTTTG ___	TTCCTTTTAG	$/^{2}$ ATC AT
35A(267)	CC TGG	$/^{3}$	GTGGGTATT ___		

Exon

GT GT

ATTA

GT GT

TC AT

Figure 4.11(a): Intron Splice Sites

A compilation of the putative intron splice sites contained within the 5AP and 35A genes. The intron in the leader (5' untranslated portion of the transcribed region) of the 35A gene was discovered as a result of the sequencing of primer extension products which is described in Section 5.1.2 (see also, Figure 5.3). The 3' splice sites of the introns at positions 150 of the 5AP gene, and 267 of the 35A gene, are not contained in these clones. The numbers above each splice site give the number of bases surrounding that site which differ from the Breathnach and Chambon consensus sequence (see Figure 4.11b).

Fig. 4.11

b

C/A	A	G/G	T	A/G	A	G	T	___	Py	Py	Py	Py	Py	Py	N	C	A	G/G	G/T
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
37	50	68	90	90	50	61	70	56	63	71	63	73	74	74	—	63	85	85	40	27
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—		—	—	—	—	—
90	90	90	90	90	90	90	90	90	85	85	85	85	85	85		85	85	85	85	85
38				30																32
—				—																—
90				90																85

Figure 4.11(b): The Breathnach and Chambon Splice Site Consensus Sequence

The Breathnach and Chambon consensus sequence (Breathnach and Chambon, 1981); formulated as a result of examining 90 5' splice sites and 85 3' splice sites of various protein-coding genes. The bases are numbered for easier reference. Below the latter are given the fraction of the total number of splice sites where the base shown was found in that position.

vitro. They could not, however, find any binding of U1 snRNPs to 3' sites, and have suggested that this nuclear RNA is only involved in binding the 5' site.

The data currently available on the mechanism of intron splicing has been reviewed elsewhere (Breathnach and Chambon, 1981; Mount and Steitz, 1983), so a detailed survey of this material will not be made here. A brief summary, however, may be of use when considering the significance of these splice site sequences.

The sequence at the splice sites has been shown to be important in determining which site is used, as alteration of a natural splice site has often led to the use of 'cryptic' sites (Felber et al., 1982; Treisman et al., 1982, 1983; Wieringa et al., 1983). That this is not the only factor which determines the choice of splice sites, however, has been indicated by a number of examples where sequences some distance from splice sites affect the efficiency or fidelity of splicing (Spence et al., 1982; Khoury et al., 1979; Langford and Gallwitz, 1983). The creation of new 5' splice sites by mutation has sometimes caused the use of new 3' sites as well (Treisman et al., 1983; Dobkin et al., 1983), and the placing of 5' or 3' splice sites upstream or downstream of their natural counterparts in the rabbit beta-globin gene has been found to result in the outermost sites always being used (Wieringa et al., 1983b). All of these data argue for complex splicing models which do not just involve the splicing machinery first associating with one splice site and then moving one dimensionally in search of a counterpart (as has been suggested previously). Kuhne et al. (1983) have suggested that the secondary and tertiary structure of the RNA be taken into consideration in such models. Such considerations might be responsible, for instance, in the choice of the 3' splice site of the 35A leader

intron (see Figure 4.11), where a sequence more closely resembling the Breathnach and Chambon consensus (CACTGCCTAGGT) can be found between 50 and 100 bases upstream (although it is possible that which bases differ from the consensus sequence is also important in the choice of splice site; not just the number).

The process of splicing is at present poorly understood. Although this work has not been directed at determining the splicing mechanism, it has provided further conformation of the GT/AG rule of intron splicing (see above), and the intron sequences provided in Figures 4.1 and 4.2 may be useful, eventually, in forming a mdoel for this mechanism.

Summary

All of the introns of the 5AP and 35A actin genes obey the GT/AG rule: the intron starts with GT at the 5' end, and finishes with AG at the 3' end. Also, most of the splice sites are very similar to the consensus sequence established by Breathnach and Chambon (1981). However, in at least one case, the 3' splice site of the leader intron in 35A, a sequence more closely resembling the consensus sequence can be found close by. The factors determining the choice of splice sites in eukaryotic genes are at present poorly understood.

5 THE EXPRESSION OF ACTIN GENES IN XENOPUS

5.1 Primer Extension and the Transcriptional Mapping of 35A

5.1.1 Introduction

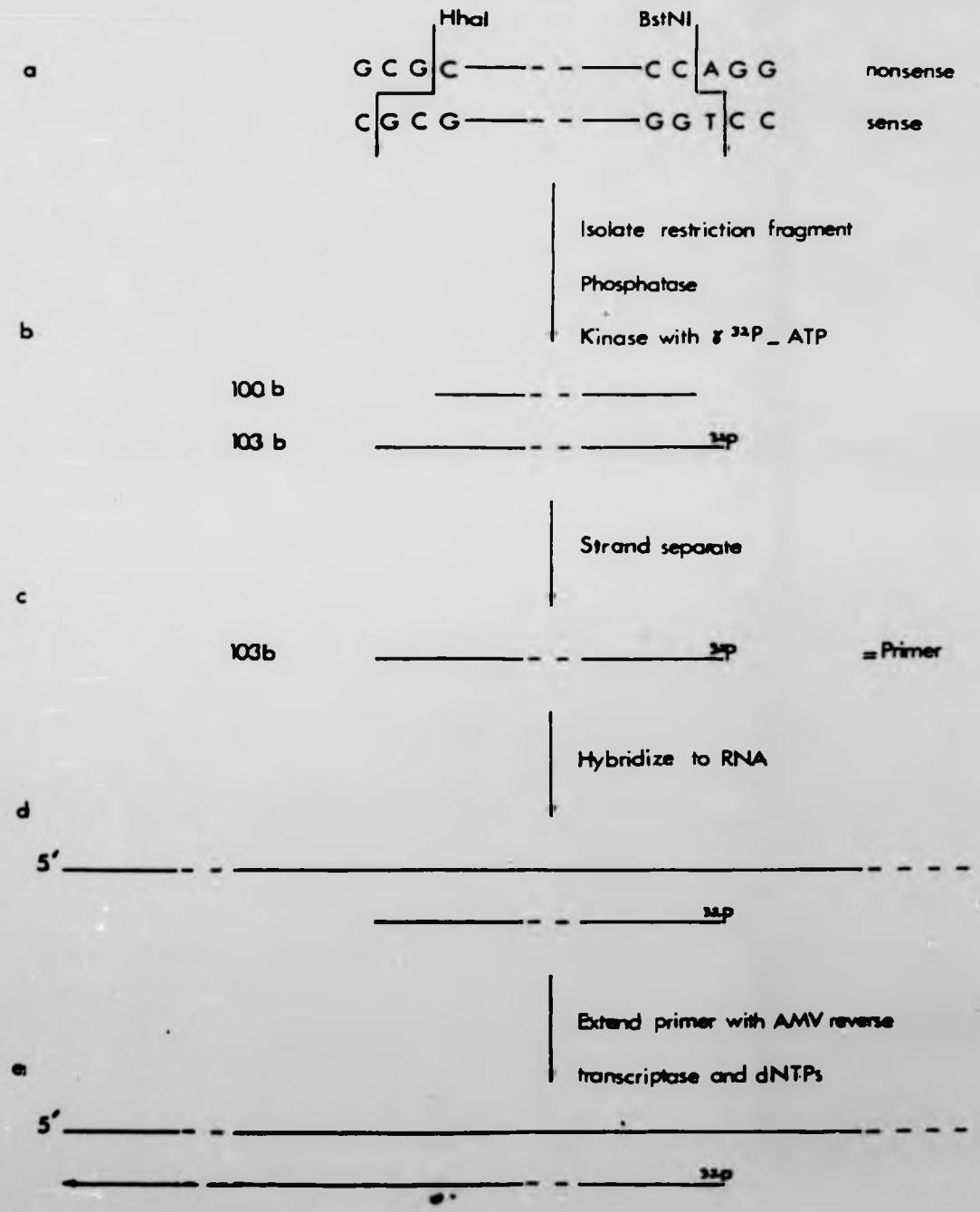
An interesting prospect that arises, once genes have been sequenced, is that of the search for sequence elements that play a role in their expression. This can take the form of sequence comparisons between genes, together with experiments where sequences 5' to a gene are gradually deleted, and the subsequent transcriptional activity of the gene in *in vitro* and *in vivo* systems examined. In this way, several elements have been discovered, which are common to almost all RNA polymerase II-transcribed eukaryotic genes: the 'TATA' and 'CAAT' boxes (Corden *et al.*, 1980; Benoist *et al.*, 1980; Flavell, 1980). A similar approach has been used to search for sequences that are responsible for the expression of a limited set of genes (Davidson *et al.*, 1983). A common feature of all of these sequences is that they are found 5' to the transcriptional initiation site. In order to start looking for similar sequences in these *X. borealis* actin genes, therefore, it was clearly necessary to find the position in the sequence where transcription starts. There are two basic approaches to this: S1 mapping and primer extension. I will briefly describe the principles involved in each, and explain why the latter was chosen for this project.

S1 mapping (Berk and Sharp, 1977) exploits the ability of the nuclease S1 to hydrolyze single stranded nucleic acid, while leaving double stranded nucleic acid intact. Typically, the technique involves the hybridization of a labelled, single-stranded DNA probe to a preparation of mRNA, followed by the degradation of all unhybridized

regions, and assay by gel electrophoresis of the so-called 'protected' regions. An appropriate probe for this purpose would be a restriction fragment that could be labelled at the 5' end of the non-coding strand within the N-terminal exon of the gene, while the other end extends several hundred bases into non-coding regions upstream from the initiation codon. For many genes, such a scheme has enabled the start of transcription to be mapped on a genomic clone. It was known, however, that several vertebrate actin genes possessed introns within the RNA leader sequence (Nudel et al., 1983; Zakut et al., 1982; Fornwald et al., 1982). The simple S1 mapping experiment described above would merely give the position of the 3' end of this intron, and the front portion of the leader sequence would remain to be found; a more difficult task.

An alternative method is that of primer extension. This involves 5' end labelling a segment of sense-strand DNA near the 5' end of the gene, hybridisation to RNA containing that gene product, followed by the extension of the DNA primer on the RNA template with reverse transcriptase (see Figure 5.1). The primer extended DNA can then be sequenced using the Maxam and Gilbert technique. In effect, the sequence of the mRNA can be obtained, and comparison with the appropriate regions of the genomic clone should define the transcriptional start point, as well as any leader introns present. It should also be apparent, that, provided the primer is taken from a protein coding region that is conserved in all the members of the gene family concerned, the number and, in this case, identity of the transcripts from that family, in any given tissue, can be defined. This aspect is another advantage over S1 mapping, where the only information obtained on the expression of genes in tissues is whether, or not, the

Fig. 5.1 Primer Extension



nonsense

sense

nt

= Primer

---- -

everse

---- -

Figure 5.1: Primer Preparation, Hybridization and Extension

The primer used in this work was made from a Hha I - Bst NI fragment (a) which included all but the first 7 amino acids of the first coding exon of the 35A gene (see Figure 4.2). (a) illustrates that the sense strand of this fragment is 3 bases longer than the nonsense strand. To allow for the detection of small quantities of the primer and its extended products, the 5' phosphate of this sense strand was removed with bacterial alkaline phosphatase and replaced with a ^{32}P-labelled phosphate using T4 nucleotide kinase and gamma-labelled ATP (b). This fragment was denatured, and the two strands size-separated by electrophoresis on a urea-(8%) polyacrylamide sequencing-type gel. The labelled 103 base strand was extracted (C) and hybridized to an RNA preparation. After hybridization of the primer to actin mRNA (d), the primer DNA strand was extended in its 3' direction using AMV reverse transcriptase and unlabelled deoxynucltoside triphosphates; the mRNA being the template (e). The products of this reaction were then separated on an 8% sequencing-type gel.

specific gene that has been cloned, is complementary to RNA from those tissues. The primer extension technique, therefore, is a more useful tool, for, in the study of the differential expression of a family of genes, it is obviously essential to know where and when each gene is expressed.

5.1.2 Results and Discussion

The primer for these studies was made from a Hha I/Bst NI fragment (103 base sense strand) which included all but the first 7 amino acids of the first coding exon of the 35A gene (see Figure 4.2). A property of this fragment is that the HhaI site is situated within the ala-leu-val triplet of amino acids that is common to all vertebrate actins, and also it does not include the highly variable N-terminal end of the gene. This makes it a potentially very useful tool as it will probably hybridize to any vertebrate actin mRNA (it hybridized to *X. borealis* skeletal muscle actin RNA - see Section 6).

Figure 5.1 illustrates the preparation of the primer and the principle of its use. In a fragment, such as this, with a 3' overhang (HhaI site) at one end and a 5' overhang (BstNI site) at the other (which can be kinase-labelled), one strand is 3 bases longer than the other. This makes the process of separating the strands very simple, as the double-stranded fragment can be electrophoresed in a sequencing-type, denaturing, acrylamide/ urea gel system which separates the 2 strands on the basis of their size.

Initially, the conditions used for the hybridization of this primer to the RNA were similar to those described in Tsang *et al*. (1982). Figure 5.2(a) shows the primer extension products after this primer had been annealed to *X. borealis* total ovary RNA, overnight in 10 mM Pipes

Fig. 5.2

a

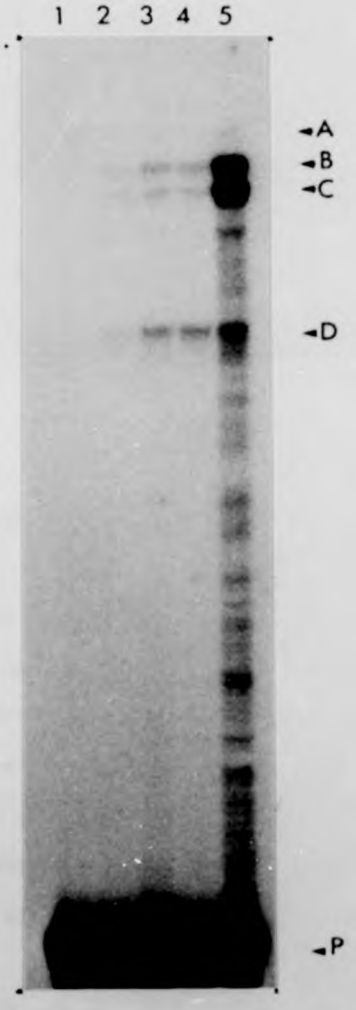


Fig. 5.2

a

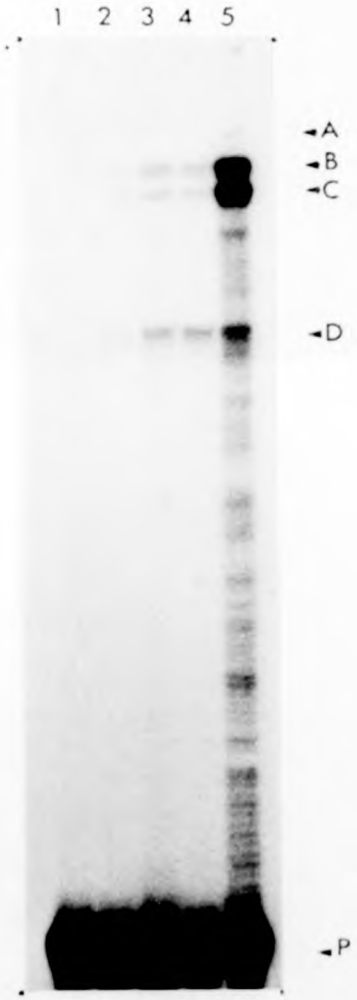


Fig. 5.2

a

Figure 5.2(a): Primer Hybridization to Ovary RNA at Various Temperatures

Single-stranded, 5'-labelled primer (see Figure 5.1) was annealed overnight with 4 µg of X. borealis ovary RNA in 10 mM Pipes pH 6.4, 0.4 M NaCl, 0.2% SDS, in a 10 µ l volume. After the hybridization, the samples were precipitated twice, redissolved in primer extension buffer, and incubated for 60 minutes at 42°C with AMV reverse transcriptase. They were then extracted with phenol/chloroform, ethanol-precipitated twice, and dissolved in formamide sample buffer, before electrophoresis on an 8% sequencing-type gel (see Section 2.7). Untreated primer was run in track 1. Tracks 2 to 5 contain the primer extension products following hybridizations at 50°, 55°, 60° and 70°C respectively. The sets of bands labelled A to D are those that are thought to represent products from discrete mRNA species. The position of unextended primer is marked with a P.

Fig. 5. 2

b

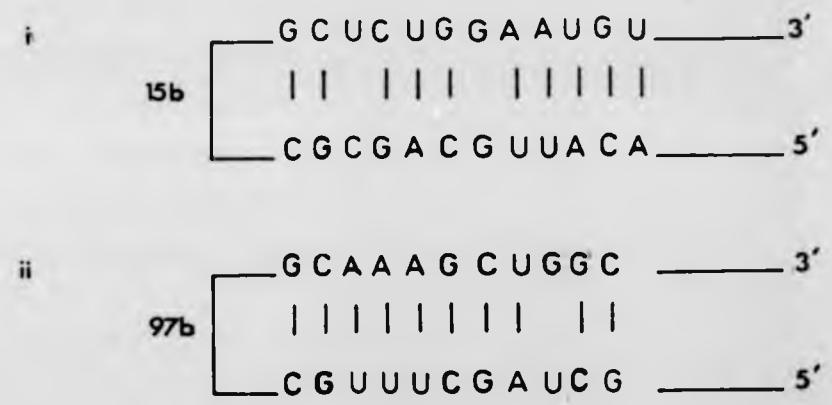

i
G C U C U G G A A U G U 3′
15b
C G C G A C G U U A C A 5′
ii
G C A A A G C U G G C 3′
97b
C G U U U C G A U C G 5′

Figure 5.2(b): Possible Hairpin Loop Structures in the 35A Transcript

Two sections of the X. borealis beta-like mRNA (which is transcribed from the 35A gene) primer-annealing region were found to be complementary to sections further along the mRNA in the 5' direction. The hairpin loops that would be formed from annealing of such sequences are shown. 10 out of 12 bases of sequence (1), from the primer-annealing region, can pair with a sequence 15 bases upstream from it in the protein-coding region. 10 out of 11 bases of sequence (ii), also in the primer-annealing region, can pair with a sequence of the leader region, 97 bases upstream.

Fig. 5.2

c

Fig. 5.2

c

Figure 5.2(c): Primer Hybridization with Varying RNA Concentrations and Hybridization Conditions

X. borealis ovary RNA was hybridized overnight to primer at 70°C in 10 mM Pipes pH 6.4, 0.4 M NaCl, 0.2% SDS (tracks 1 to 3), or at 45°C in the same buffer containing 50% formamide (track 4). The amounts of RNA used in the hybridizations were: 60 µg (tracks 3 and 4) or 30 µg (track 2), each in a 30 µl volume, or 6 µg (track 1), in 10 µl. The primer that had hybridized to RNA was then extended with AMV reverse transcriptase, and the products run on an 8% sequencing-type gel (see Figure 5.2a and Section 2.7). Sets of bands labelled B, C, and D are the same as those described in (a).

pH 6.4, 0.4 M NaCl, 0.2% SDS at 50°, 55°, 60° and 70°. It can be seen that the 70°C hybridization appears to have resulted in the largest yield of primer extended products. I have labelled 4 sets of bands, A to D, which I think may have arisen from discrete RNA species (the only way to conclusively show that each band has arisen from a discrete mRNA species is to sequence each band; as we will see, this has only been done for bands B and C). There appears to have been more hybridization of primer to the messages that give rise to bands A to C at 70°C than at lower temperatures; this was quite surprising, but a similar result has been observed by Professor H. R. Woodland (per. comm.) using a Xenopus histone H4-derived primer. It is possible that this is due to secondary structure at the 5' end of these mRNA species, which may have to be destroyed by high temperature before primer annealing can occur. Figure 5.2(b) shows two examples of sequences in the 35A primer-annealing region which might base-pair with sequences further along the mRNA in the 5' direction.

Figure 5.2(c) demonstrates more clearly that 'bands' B and C are, in fact, each composed of two to three bands. Other workers have found that primer extension on a single RNA species gives rise to two bands one base apart (Luse et al., 1981), and have suggested that this may be an artefact of the primer extension reaction. However, the lower of each of the sets in B and C is clearly more than one base smaller than the upper. Nevertheless, I think it likely that these lower bands are derived from the major species (they show the same behaviour in response to changes in hybridization temperature, for example), and the sets of bands for B and C were each pooled for sequencing.

Because more radioactive label was incorporated into them, the sets of bands B and C were the obvious first choice for sequencing. Figures

5.2(a) and (c) show that while a greater degree of hybridization occurs at 70°C for some RNAs, there is also a large amount of chemical degradation at this temperature. In order to increase the radioactivity incorporated into the sets of bands B and C still further, an experiment was performed to test the effect of using a lower temperature with formamide in the hybridization. Tracks 3 and 4 of Figure 5.2(c) demonstrate that far more extended product was obtained when the hybridization was performed at 50% formamide, 45°C, than at 70°C without formamide; this appears to be due to a decrease in chemical degradation. This modification was used to obtain material from B and C for sequencing.

Figure 5.2(c), tracks 1 to 3, demonstrate the effect of using increasing amounts of RNA in the hybridization reaction. Not surprisingly, there is an initial increase in the products obtained, with increasing RNA concentration. However, a comparison of tracks 2 and 3 reveals that, after saturating the primer with total actin message, band D is lost on increasing the RNA concentration further. This phenomenon is repeatable (e.g. see Figure 5.6b). One possible explanation for this phenomenon might be that this RNA species is only partly complementary to the primer and, on reducing the primer/RNA ratio, this weakly hybridizing species loses out in the competition for primer. If this is true then care must be taken when interpreting experiments that compare the primer extension products obtained with RNA from different tissues; artefactual differences might be obtained, for example, by using different concentrations of RNA from each tissue.

Figure 5.3(a) shows the sequence read from an 8% sequencing gel, on which was run the Maxam and Gilbert sequencing products of B and C. The sequence of B exactly matches the coding sequence and first 8 bases of

Fig. 5.3 a

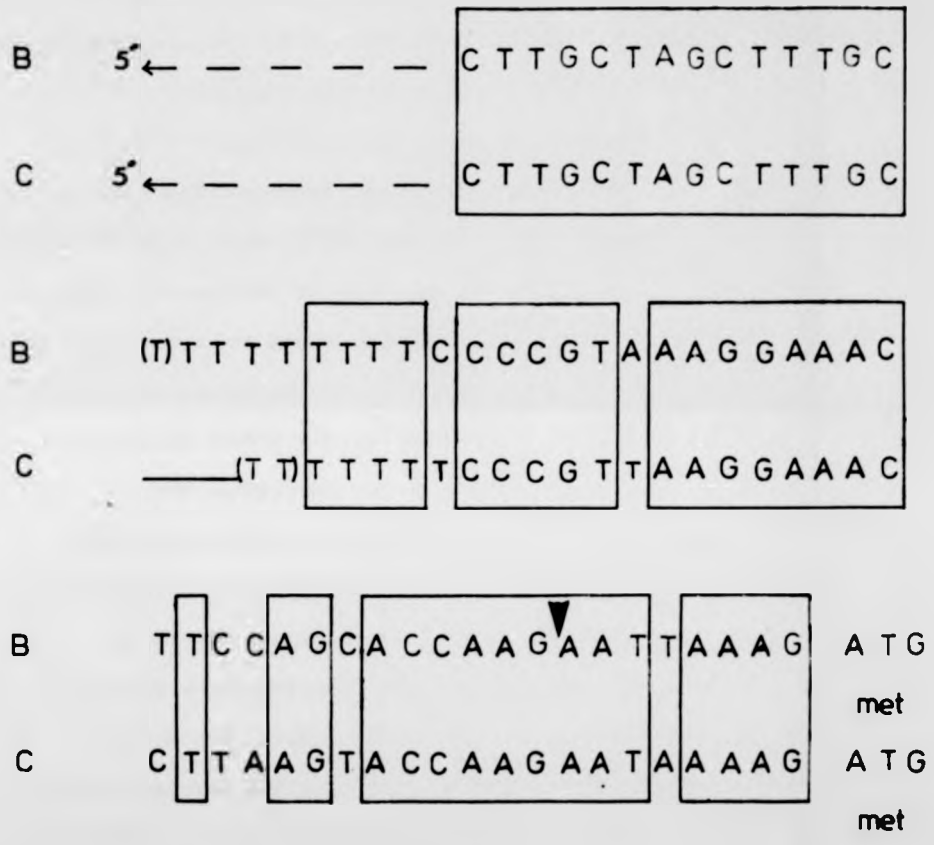

B 5' ← — — — — C T T G C T A G C T T T G C
C 5' ← — — — — C T T G C T A G C T T T G C
B (T) T T T T T T T T C C C C G T A A A G G A A A C
C _____ (T T) T T T T T C C C C G T T A A G G A A A C
B T T C C A G C A C C A A G A A T T A A A G A T G
met
C C T T A A G T A C C A A G A A T A A A A G A T G
met
B G C A G A T G A T G A C A T T G C A G
ala asp asp asp ile ala
C G C A G A A G A A G A G A T T G C A G
ala glu glu glu ile ala

Figure 5.3(a): Sequences of Primer Extension Sets of Bands B and C

The sequences of the primer extended sets of bands B and C are shown. The bracketed Ts indicate that the precise number of Ts was not discernible. There were, however, more Ts in the C sequence than the B sequence, so a gap has been introduced in the former to show the similarity of the sequence 5' to this region. It was not possible to read either of the sequences all the way to the 5' end of the mRNA. In the coding region, the amino acids have been written underneath the corresponding sequence. The regions where the two leaders are homologous have been boxed. The arrow marks the position in the sequence of the B band where an approximately 1200 bp insert was found in the sequence of 35A. Thus, the sequence 5' to this arrow can be seen to read the same as the sequence 5' to a similar arrow in Figure 5.3(b); the 3' sequence reads the same as the sequence 3' to a similar arrow situated 8 bases 5' to the first ATG codon (see Region 5 in Figure 4.2).

Fig. 5.3b

35A: Region 3

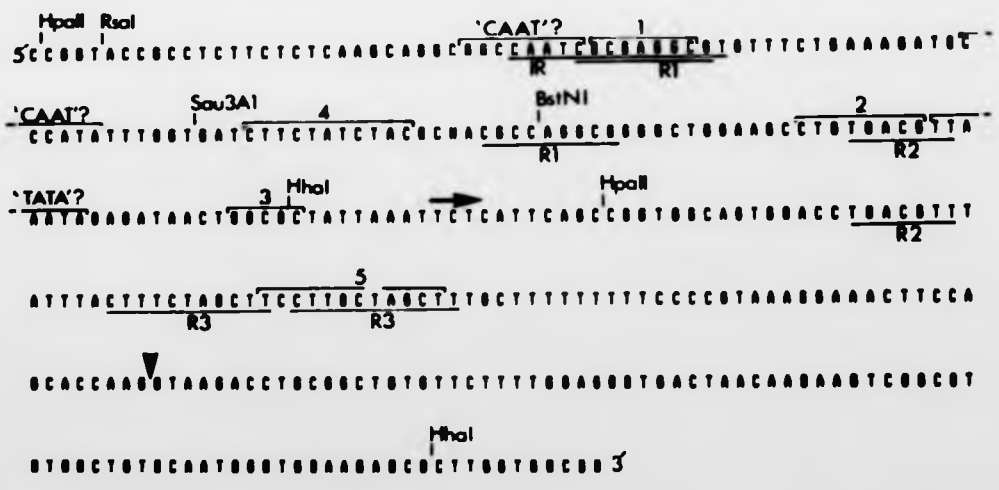


Figure 5.3(b): Sequence of 35A: Region 3

The anti-sense strand sequence of region 3 of 35A (see Figure 3.20) is shown. Each complete line contains 60 bases. The horizontal arrow roughly marks the position of the transcriptional start site. The vertical arrow head marks the 5' splice site of the leader intron. The putative 'TATA' and 'CAAT' boxes are overlined, as are 5 other sequences that are candidates for being elements involved in the control of this gene's expression (see Section 5.1.2). Sequences 1 to 3 are found in similar positions, relative to the transcriptional start, in the rat beta-cytoskeletal actin gene (Nudel et al., 1983), while sequence 4 is present (with one base difference) 300 bases from the start of transcription of the rat gene. Sequence 5 is found in the leader region of the rat beta-gene (11 out of 12 bases) as well as the leader of this gene. The pair of sequences, R2, represents a perfect repeat while the pairs, R1 and R3, are imperfect repeats (8 out of 9, and 10 out of 11 bases, respectively, being repeated). These are discussed in Section 6, as is the sequence IR, the inverted complement of which (with 2 base changes) is present 50 to 100 bases upstream (see Figure 6.3). Most of the restriction enzyme cut sites shown in Figure 3.20 are indicated (except for the Alu I sites, AGCT, and a Hae III site, GGCC, the latter being within the first putative 'CAAT box').

AGATGC

ACGTTA
R2

ACGTTT
R2

CTTCCA

CAGCGT

Figure 5.3(b): Sequence of 35A: Region 3

The anti-sense strand sequence of region 3 of 35A (see Figure 3.20) is shown. Each complete line contains 60 bases. The horizontal arrow roughly marks the position of the transcriptional start site. The vertical arrow head marks the 5' splice site of the leader intron. The putative 'TATA' and 'CAAT' boxes are overlined, as are 5 other sequences that are candidates for being elements involved in the control of this gene's expression (see Section 5.1.2). Sequences 1 to 3 are found in similar positions, relative to the transcriptional start, in the rat beta-cytoskeletal actin gene (Nudel et al., 1983), while sequence 4 is present (with one base difference) 300 bases from the start of transcription of the rat gene. Sequence 5 is found in the leader region of the rat beta-gene (11 out of 12 bases) as well as the leader of this gene. The pair of sequences, R2, represents a perfect repeat while the pairs, R1 and R3, are imperfect repeats (8 out of 9, and 10 out of 11 bases, respectively, being repeated). These are discussed in Section 6, as is the sequence IR, the inverted complement of which (with 2 base changes) is present 50 to 100 bases upstream (see Figure 6.3). Most of the restriction enzyme cut sites shown in Figure 3.20 are indicated (except for the Alu I sites, AGCT, and a Hae III site, GGCC, the latter being within the first putative 'CAAT box').

the 5' upstream sequence of the 35A gene. Approximately 1200 bases 5' to this, the sequence of 35A again matches that of the B band (for the 51 bases of the latter that can be read). From this, I conclude that the 35A actin gene is expressed in the oocyte.

As Figure 4.11 shows, the approximately 1200 bp intron possesses the terminal 5' GT and 3' AG associated with almost all introns (in fact, the precise boundaries of the intron were fixed by applying this rule; Breathnach and Chambon, 1981). An intron located within the mRNA leader, at a similar position, has been found in all the other vertebrate actins that have been studied so far (Nudel et al., 1983; Zakut et al., 1982; Fornwald et al., 1982).

Although the reading of the primer extended sequence became difficult after the region shown in Figure 5.3(a), the rest of the sequence of B could be roughly identified with that of 35A so as to give an approximate position for the start of transcription. As Figure 5.3(b) (the sequence of region 3 in Fig. 3.20) shows, this lies close to the sequence TCATTCA, which exactly fits the consensus sequence found for histone gene cap sites (PyCATTCPu; Hentschel and Birnstiel, 1981). This is surprising as this consensus sequence is thought to be specific to histone genes. Most other genes do not conform to any specific consensus sequence at all, except that transcription usually starts at an A residue (in the anti-sense strand) which tends to be surrounded by pyrimidines (PyNNNPyAPyPyPyPyPy; see Breathnach and Chambon, 1981). Although the exact transcriptional start point is hard to determine, it appears to be several bases upstream from this histone-like sequence, and therefore the latter resemblance may be coincidental.

Approximately 25 to 30 bases upstream from the start of transcription lies the beginning of the sequence 5'-TTAAATA-3'. This is

similar, but not identical, to the consensus sequence (TATA$^{A}_{T}A^{A}_{T}$A) found in this position (the initial T, 26 to 34 bp from the transcriptional start point, Breathnach and Chambon, 1981) in nearly all eukaryotic polymerase II genes that have been sequenced so far. The so-called 'TATA box' is believed to be the dominant element in determining the transcriptional start site of these genes (Grosveld et al., 1982; Grosschedl and Birnstiel, 1980a; Benoist and Chambon, 1981; Fromm and Berg, 1982; Dierks et al., 1983). Its deletion often results in in vivo initiation at multiple start points, although Dierks et al. (1983), using the rabbit beta-globin gene, have found that, in its absence, transcripts start at the 'cap site', the normal initiation point. They have suggested that the latter region also exerts a weak effect on deciding the initiation point. However, on insertion of linker sequences inbetween the 'TATA box' and the 'cap site' of the rabbit beta-globin gene, Dierks et al. found the start point to be always 26 to 28 nucleotides downstream from the 'TATA box'; indicating the overriding influence of the latter.

In addition to determining the site of the start of transcription, the 'TATA box' is also required for the efficient transcription (that is to say, its removal lowers the level of transcription) of the beta-globin gene (Dierks et al., 1983; Grosveld et al., 1982), and probably other genes. These include the sea urchin histone H2A (Grosschedl and Birnstiel, 1980a), HSV TK (McKnight and Kingsbury, 1982), conalbumin (Grosschedl et al., 1981) and adenovirus EIA and major late promoters (Osbourne et al., 1982; Moreau et al., 1981). Deletion of the 'TATA box', in these in vivo studies, does not lead to complete transcriptional inactivation, however, as does the Pribnow box promoter in a bacterial gene (Miller, 1978), and so it is presumably not the sole

promoter-like element in a eukaryotic gene.

Another conserved sequence found in many eukaryotic protein-coding genes is the so-called 'CAAT box'. This region is usually found 70 to 90 base pairs upstream from the start of transcription, and has the canonical sequence 5' - GG$^{C}_{T}$CAATCT - 3' (Benoist et al., 1980; Efstratiadis et al., 1980). Deletion of this sequence has been found to decrease transcription in vivo (i.e. on transformation of the altered gene into cell lines) of alpha- and beta-globin genes (Dierks et al., 1981, 1983; Grosveld et al., 1982a, 1982b; Melton et al., 1981). The sequence GGCCAATC occurs in the 35A gene at a position where the first G is approximately 115 to 120 bases from the start of transcription. This oligonucleotide most closely resembles the canonical sequence, although another oligonucleotide, GCCCATA, lies 32 bases closer to the transcriptional start, at the expected position. Either of these sequences may function as a promoter element, although not necessarily so. The deletion or mutation of the CCAAT-like sequences of the sea urchin histone H2A gene (Grosschedl and Birnstiel, 1980a) actually leads to a 2-fold increase in the rate of transcription, while McKnight and Kingsbury (1982) showed that most of the nucleotides within a CCAAT-like sequence in the HSV TK promoter region could be substituted without loss of promoter function. Possibly, these CAAT-like sequences are not functional homologues of the globin CAAT box. It may be unwise, therefore, to equate either of these 35A sequences with a promoter element solely on the basis of partial homology to the consensus sequence.

In addition to these elements, various other upstream regions appear to be important for maximum transcriptional activity. Thus, the beta-globin genes of mouse, human, goat and rabbit contain a sequence

approximately 100 bases upstream from the cap site with the consensus sequence 5' - CCTCACCCTG - 3' (Dierks et al., 1983). Such sequences tend to be gene-type specific, however, as studies on the HSV TK gene (McKnight and Kingsbury, 1982), the sea urchin H2A gene (Grosschedl and Birnstiel, 1980b, 1982), and the SV40 early region (Benoist and Chambon, 1981; Fromm and berg, 1982) have shown. It is perhaps coincidental, therefore, that the 35A sequence contains the sequence 5' - CCAGACCCTG - 3', 150 to 200 bp from the capsite (see Fig. 6.3).

The most obvious method of trying to locate sequences important in the transcription of actin genes, is to compare the region 5' to the cap site of 35A with similar regions of other actin genes. A comparison of 35A with the rat beta-actin gene (Nudel et al., 1983) reveals several interesting short sequences shared by both. One of these 5' - GGCGC - 3' is situated approximately midway between the 'TATA box' and the cap site in both genes. Another, 5' - GCGAGGC - 3', also lies in similar positions in both genes; the last C lies approximately 103 bases from the cap site of the 35A gene, and 110 bases from that of the rat gene. If the GCCCATA sequence were to represent the true 'CAAT box' in the 35A gene, these sequences would also be located the same distance (approx. 20 bases to the 5' side) from the 'CAAT box' in both genes. The sequence CTGTGPuCG lies close to the 5' side of both 'TATA boxes'; immediately in front, in the case of the 35A gene, and 3 bases in front, in the case of the rat gene. These sequences appear significant because of the similarity of their positions in these two genes (both encoding cytoskeletal actins). They are also absent from the rat skeletal muscle actin gene of Zakut et al. (1982), and so possibly may be involved in the control of expression of cytoskeletal actin genes.

Sequences involved in the control of expression of other gene-types

have been described in Section 1.3. Recently, a candidate for such a control element has been found in skeletal muscle actin genes. Ordahl and Cooper (1983) have shown that a 20 base-pair region is highly conserved in the 5' upstream regions of the chick and rat genes. The surrounding non-coding sequence shows relatively low homology between the genes, and the localization of the conserved sequence (surrounding and including the 'CAAT box') is the same in both genes. This sequence is also absent from the rat and chicken beta-cytoskeletal actin genes, which, as we have seen, have a different pattern of expression to the muscle actin genes.

Other homologies exist between the frog and rat genes, but these do not occur in similar positios. For example, the sequence 5' - CTTCTCTCTAC - 3' occurring approximately 300 bases upstream from the cap site of the rat gene, is similar to the sequence 5' - CTTCTATCTAC - 3', approx. 71 bases 5' to the frog cap site. Obviously, sequence similarities are likely to be found if a large enough region of DNA is searched. However, the recent discovery of promoters, called enhancers, which appear to function regardless of orientation to, and distance from, a gene (within limits), makes it unwise to disregard all such sequences, merely because of their different positions relative to the cap site (Khoury and Gruss, 1983). The immunoglobulin enhancer is especially interesting as it appears to function in a tissue-specific manner (Gillies et al., 1983).

Determining the function, if any, of these sequences will depend on experiments where they are deleted or mutated before being recloned and introduced back into cells to study their pattern of expression. Of course, before this can be done, it must be shown that unaltered actin gene clones can be used to transform cells and obtain normal

transcriptional regulation. With muscle-specific actins, normal regulation consists of transcriptional activation only at myogenesis (see Section 1.3). With cytoskeletal actin genes (such as the 35A gene), however, we do not yet definitely know of any transcriptional regulation in vertebrates, and further experiments are required in this area, before such experiments can be performed.

As explained in Section 1.3, the expression of cytoskeletal actin genes in several circumstances may be controlled at the level of mRNA translation. It is possible that the regions of homology between the rat beta-actin gene and the 35A gene leader regions, TCCTTGC and AGCT (joined in the rat gene, but with a T separating them in the frog gene; see Fig. 5.3b), are involved in such a form of control. A similar sequence is not found in the leaders of the rat and chicken alpha-actin genes (Zakut et al., 1982; Fornwald et al., 1982). It is also possible, however, that this sequence is merely involved in the interaction of ribosomes with the mRNA, as 6 out of the first 7 bases of this region of homology are complementary to the conserved, purine-rich, 3' region of eukarotic 18S (including Xenopus) ribosomal RNAs (Hagenbuchle et al., 1978; Maden et al., 1982). This complementarity is illustrated in Fig. 5.4(a). The precise role that would be played by such an interaction of this sequence with a ribosome is not clear, however. Most eukaryotic messages do not possess leader sequences with any significant complementarity to the 18S RNA (De Wachter, 1979) although there are a few exceptions, such as an ovalbumin mRNA (Kuebbing and Liarakos, 1978) and an adenovirus major late mRNA (Ziff and Evans, 1978), with regions of 6 to 9 bases complementary to the conserved 3' end of the 18S rRNA (Kozak, 1983). These, however, are not thought to serve a function analagous to the 'Shine and Dalgarno' sequence of prokaryotic messages

(a 7 base sequence situated 10 bases upstream from the initiator AUG which is complementary to the prokaryotic 16S ribosomal RNA, and is thought to be involved in ribosome binding; Shine and Dalgarno, 1974; Steitz and Jakes, 1975; Kozak, 1983) as they do not occur near the AUG codon. The region of complementarity to the 18S RNA found here is 54 bases upstream from the first AUG of the 35A gene and 37 bases upstream from the initiator AUG of the rat beta-gene. A considerable body of circumstantial evidence (reviewed by Kozak, 1978, 1980, 1981, 1982, 1983) appears to support a model whereby eukaryotic 40S ribosomal subunits bind initially at the 5' end of the mRNA molecule before migrating towards the interior and stopping at the first AUG (although there are several instances where initiation does not occur at the 5'-proximal AUG triplet-reviewed by Kozak, 1983; in the case of the 35A gene there are no AUG codons preceeding the initiating AUG). Kozak (1982) has suggested that, if the regions complementary to the 18S RNA have any functional significance, their interaction with rRNA might enhance the efficiency of translation, perhaps by stabilizing a 'pre-initiation complex'. There is, as yet, no evidence for such an interaction, however, and the possibility that the region of homology found here (between the rat beta- and 35A gene leaders) is involved in translational control cannot be ruled out.

Summary

The actin gene found on 35A is expressed in the X. borealis oocyte. There is an approximately 1200 bp long intron in the 5' untranslated region of the primary transcript. The start of transcription has been roughly mapped, and candidates for the 'TATA' and 'CAAT boxes' have been found. Several short sequences have been found which are also found in

Figure 5.4

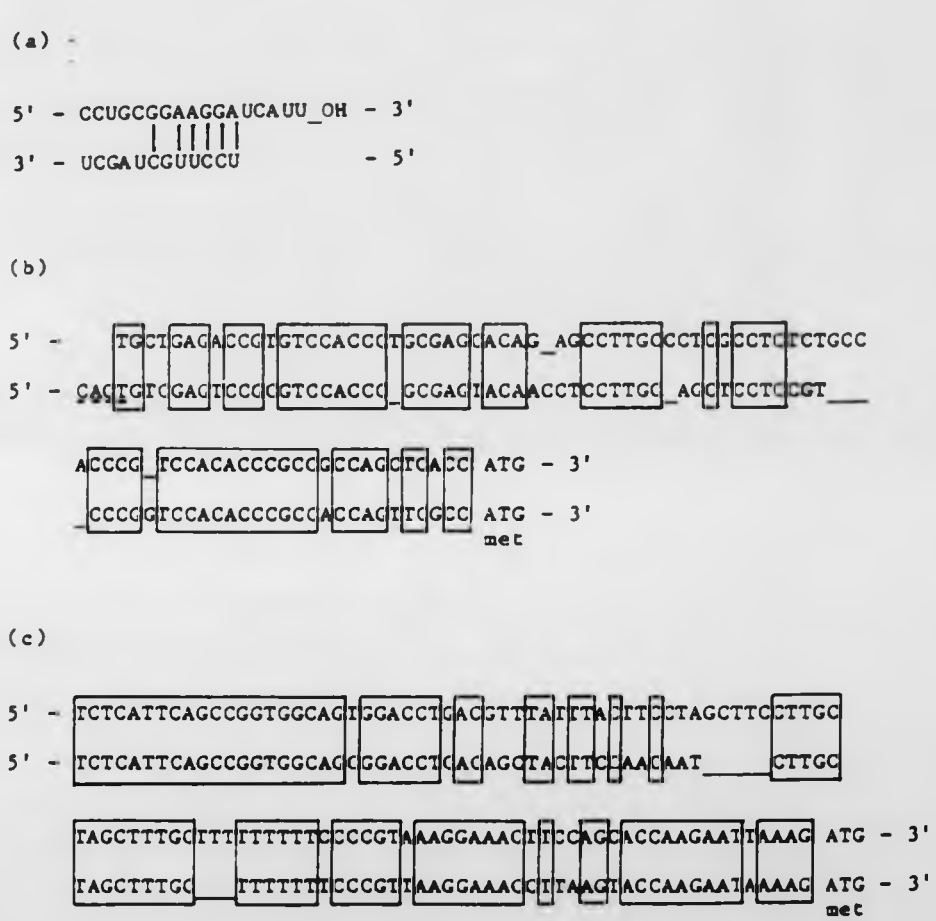

(a)
5' - CCUGCGGAAGGAUCAUU_OH - 3'
3' - UCGAUCGUUCCU - 5'
(b)
5' - TGCTGAGACCGTGTCCACCCTGCGAGCACAG_AGCCTTGCCCTCGCCTCTCTGCC
5' - CACTGTCGAGTCCCCGTCCACCC_GCGAGTACA_ACCTCCTTGC_AGCTCCTCCGT___
ACCCC_TCCACACCCGCCGCCAGCTCACC ATG - 3'
_CCCCGTCCACACCCGCCACCAGTTCGCC ATG - 3'
met
(c)
5' - TCTCATTCAGCCGGTGGCAGTGGACCTGACGTTTATTTACTTCCTAGCTTCCTTGC
5' - TCTCATTCAGCCGGTGGCAGCGGACCTCACAGCTACTTCC_AACAAT____CTTGC
TAGCTTTGCTTTTTTTTTCCCGTAAAGGAAACTTCCAGCACCAAGAATTAAAG ATG - 3'
TAGCTTTGC___TTTTTTTCCCGTTAAGGAAACCTTAAGTACCAAGAATAAAAG ATG - 3'
met

Figure 5.4: Sequence Comparison

In (a) the 3' region of 18S rRNA that has been found to be conserved in several eukaryotes (Hagenbuchle et al., 1978) is shown to be partly complementary to the region of the 35A gene's leader (bottom sequence) that is also found in the rat beta-gene's leader (see Figure 5.3b). In (b) the region 5' to the actin-like sequence of the human cytoskeletal actin pseudo - gene, pseudo-1 (Moos and Gallwitz, 1983), was compared with the leader region of the rat beta-actin gene (lower sequence). The cap site of the rat gene is underlined with a broken line. In (c) the leader region of the 35A gene (upper sequence) was compared with a putative X. laevis gamma-cytoskeletal actin cDNA clone (T. Mohun, per. comm.). Gaps have been introduced into the sequences to achieve maximum homology, and homologous regions are boxed.

TGCC

T___

GC

GC

ATG - 3'

ATG - 3'
met

a similar position in the rat beta-cytoskeletal actin gene, but not the rat skeletal muscle actin gene.

5.2 Band C: A Gamma-Actin Gene Product?

As Figure 5.3(a) shows, the sequence of primer extension band C differs from that of B in coding for 3 glutamate residues at positions 2 to 4, instead of 3 aspartate residues. In this respect is is identical to the mammalian gamma-cytoskeletal actin, although, of course, there may be differences elsewhere in the protein.

Like the beta-type actin, encoded by the 35A sequence, this gamma-like actin has an alanine codon immediately after the first methionine. This has been found also to be the case in a X. laevis cDNA clone encoding a putative gamma-cytoskeletal actin (T. Mohun, per. comm.). This codon is not present in the rat beta-actin gene (Nudel et al., 1983), or the cDNA clones encoding beta- and gamma-cytoskeletal actins of humans (Gunning et al., 1983). All vertebrate muscle actin genes so far examined have a cysteine codon in this position. The implications of this have been discussed in section 4.1.

Figure 5.3(a) also shows the high homology found between the beta- and gamma-like RNA species (bands B and C) in those parts of their 5' untranslated regions ('leaders') that have been sequenced. There are several possible explanations for this observation. The genes giving rise to these RNAs may have arisen from a duplication event in the recent past or, alternatively, this event may have occurred some time ago and a recent gene conversion event has equalized the leader regions of the two genes. Another possibility is that this region of the leader is functionally important and so has been conserved for this reason.

If the latter explanation is true, then comparisons of these sequences with genes from more distantly related species might also expect to reveal some homologies in this region. Figure 5.3(b) shows the only region of homology between leader regions of the 35A gene (beta-like) and the rat beta-gene (Nudel et al., 1983). The homology in this region is surprising, and is perhaps a result of some functional role for this sequence (see Section 5.1.2). However, the majority of the sequence conserved between the X. borealis beta- and gamma-like gene leaders is not found in the rat beta-gene. If this sequence is conserved for some functional role, then this role must have changed since the divergence of the mammalian and amphibian ancestral lines. It is interesting, however, that a comparison between this rat beta-gene sequence and the sequence of the human cytoskeletal actin pseudogene, pseudo-1 (Moos and Gallwitz, 1983), reveals a very high degree of homology in the leader region (Figure 5.4b). Furthermore, the pseudogene shows a high degree of replacement site divergence from the other vertebrate cytoskeletal actins, which would be unusual for an expressed gene, and has, almost certainly, occurred since the creation of the pseudogene. It seems likely, therefore, that a large amount of the differences that are observed between the rat and human leader regions have also occurred since this event. The homology between these two species, which can be assumed to have diverged at the time of mammalian radiation, 85 million years ago (Romero-Herrara et al., 1973), is therefore even more striking than at first appearance. Hanauer et al. (1983) have recently found a high degree of homology between the leader sequence of the rat skeletal muscle actin gene (Zakut et al., 1982) and a human skeletal muscle actin cDNA clone (Hanauer et al., 1983). This is in contrast to the high divergence of most of the 3'

untranslated region. It seems that the leader region of all actin genes may evolve very slowly (i.e. more slowly than if the changes were neutral).

Is this the explanation for the similarities between the X. borealis beta- and gamma-like gene leaders? It may be contributory. The leader region of the X. borealis gamma-like gene that has been sequenced (Figure 5.3b) is identical to that of the X. laevis gamma-like sequence of T. Mohun (per. comm.). The C peptide of the preproinsulin genes (Perler et al., 1980), on average, accumulates silent site changes with a U.E.P. of 0.7 (U.E.P. = unit evolutionary period, the time in millions of years for the fixation of 1% changes between the two lines). If there were no functional constraints on the gamma-like actin leader sequence then I might have expected roughly 11% divergence between the sequences in the 8 million years since X. laevis and X. borealis diverged (Bisbee et al., 1977). This calculation, however, assumes that the rate of silent site fixation is the same for all genes in all species, and the high degree of variation found for different genes by Perler et al. (1980) makes this assumption doubtful. Nevertheless, the similarity between the two leaders does suggest conservation.

Is there any evidence to suggest that the similarity between the leaders of the X. borealis beta- and gamma-like genes is not the result of functional conservation? Figure 5.4(c) shows that there is a high degree of homology between the X. borealis beta-like leader and the leader of the X. laevis gamma-like gene leader cloned by Dr. T. Mohun (per comm.). However, the silent site divergence between the two coding regions of these genes is surprisingly small; only 29%, when corrected for multiple substitutions (Perler et al., 1980). Although, as explained in Section 3.2, the silent site substitutions do not make an

untranslated region. It seems that the leader region of all actin genes may evolve very slowly (i.e. more slowly than if the changes were neutral).

Is this the explanation for the similarities between the X. borealis beta- and gamma-like gene leaders? It may be contributory. The leader region of the X. borealis gamma-like gene that has been sequenced (Figure 5.3b) is identical to that of the X. laevis gamma-like sequence of T. Mohun (per. comm.). The C peptide of the preproinsulin genes (Perler et al., 1980), on average, accumulates silent site changes with a U.E.P. of 0.7 (U.E.P. = unit evolutionary period, the time in millions of years for the fixation of 1% changes between the two lines). If there were no functional constraints on the gamma-like actin leader sequence then I might have expected roughly 11% divergence between the sequences in the 8 million years since X. laevis and X. borealis diverged (Bisbee et al., 1977). This calculation, however, assumes that the rate of silent site fixation is the same for all genes in all species, and the high degree of variation found for different genes by Perler et al. (1980) makes this assumption doubtful. Nevertheless, the similarity between the two leaders does suggest conservation.

Is there any evidence to suggest that the similarity between the leaders of the X. borealis beta- and gamma-like genes is not the result of functional conservation? Figure 5.4(c) shows that there is a high degree of homology between the X. borealis beta-like leader and the leader of the X. laevis gamma-like gene leader cloned by Dr. T. Mohun (per comm.). However, the silent site divergence between the two coding regions of these genes is surprisingly small; only 29%, when corrected for multiple substitutions (Perler et al., 1980). Although, as explained in Section 3.2, the silent site substitutions do not make an

accurate evolutionary clock, and so the exact time since their divergence cannot be calculated, it would seem that these genes may not have arisen from the same event that gave rise to the mammalian beta- and gamma-cytoskeletal genes. Alternatively, a gene conversion event may have recently equalized the whole or part of these genes. Whatever the cause, the point to be made is that it is not just the leader regions, in this case, that show unexpected homology, and so it may be erroneous to attribute a special function to the leader sequence.

Clearly, it is difficult to gauge the contributions of factors such as gene conversion, recent duplication, and conservation for functional reasons, to the homology found between Xenopus cytoskeletal actin mRNA leader regions. The data on actin genes from other species, however, does suggest that the sequence of actin leaders from distantly related species can be highly conserved; presumably (as gene conversion and unequal crossover events cannot equalize genes from different species) for functional reasons.

5.3 Are The Same Cytoskeletal Actin Genes Expressed In All Tissues?

The two mammalian cytoskeletal actin isoforms have been found in all non-muscle tissues so far examined, although the ratio of these two proteins varies from tissue to tissue (Garrels and Gibson, 1976; Vandekerckhove and Weber, 1978b). Bravo et al., (1981) have shown them to be both expressed in a single cell. The 3 X. laevis cytoskeletal actins have also been shown to exist in a number of non-muscle tissues: liver, brain, oocyte and kidney (Vandekerckhove et al., 1981). The ratio of these amphibian actins also varies from tissue to tissue. Therefore, it seems a general rule in vertebrates that cytoskeletal

actins are not tissue-specific, that all are co-expressed in many tissues, and that the degree of the expression varies.

It is pertinent to ask whether the same cytoskeletal actin genes are being expressed in each of these tissues. It is conceivable, for example, that many such genes might exist, each being associated with a different group of tissue-specific genes which are all activated in the appropriate tissue. Another possibility is that the same genes are transcribed differently in different tissues; for example, using different promoters. This is the situation for the Drosophila alcohol dehydrogenase gene (Benyajati et al., 1983), and the mouse alpha-amylase 1 gene (Young et al., 1981).

Figure 5.5(a) shows the primer extension products obtained after hybridization of the primer with ovary poly(A)+ and poly(A)-, kidney, and XTC cell RNA (a fibroblast-derived cell line) from X. laevis. The same pattern of bands can be seen in tracks 2 to 4 (XTC cell and ovary RNA) and the upper two bands in these tracks, at least, are just visible in track 1 (kidney RNA). It seems that at least one gene is expressed in all three tissues (it is tempting to say two, but it is impossible, without sequencing material from each band, to be sure that the lower band is not a result of degradation of the RNA that gave rise to the upper band). This conclusion, of course, assumes that it is unlikely for two different actin genes to produce RNA with the same length of leader region. I consider this a reasonable assumption, except for those cytoskeletal genes that have arisen from a very recent duplication event. Thus, tracks 5 and 6 of Figure 5.5 show that the primer extension products of X. borealis ovary RNA (run on the same gel) can be clearly distinguished from those of X. laevis; the two species probably having diverged only 8 million years ago (Bisbee et al., 1977).

Fig. 5.5

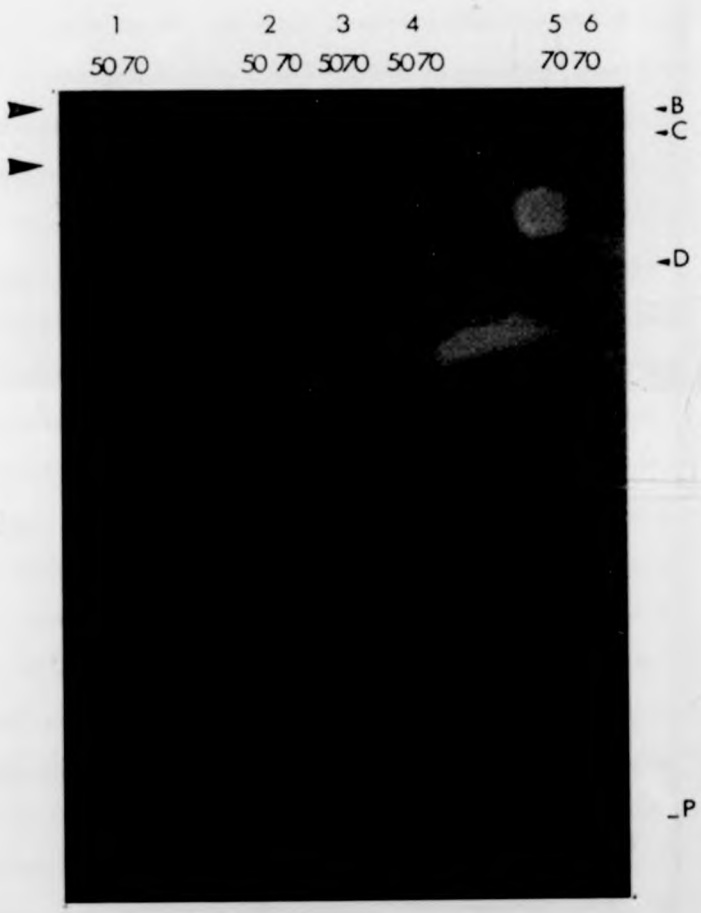


Fig. 5.5

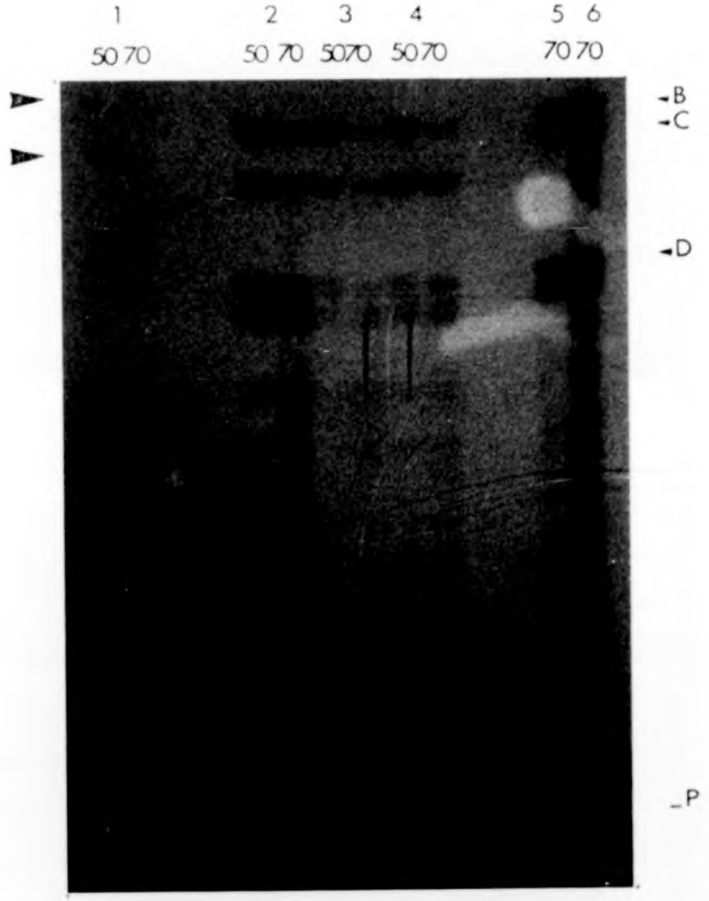


Fig. 5.5

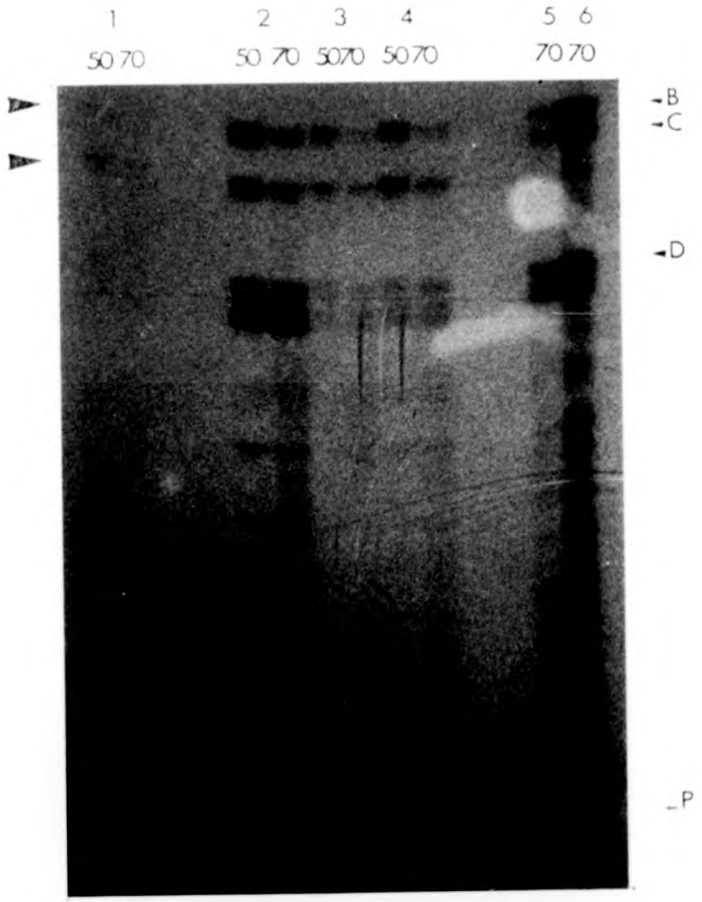


Fig. 5.5

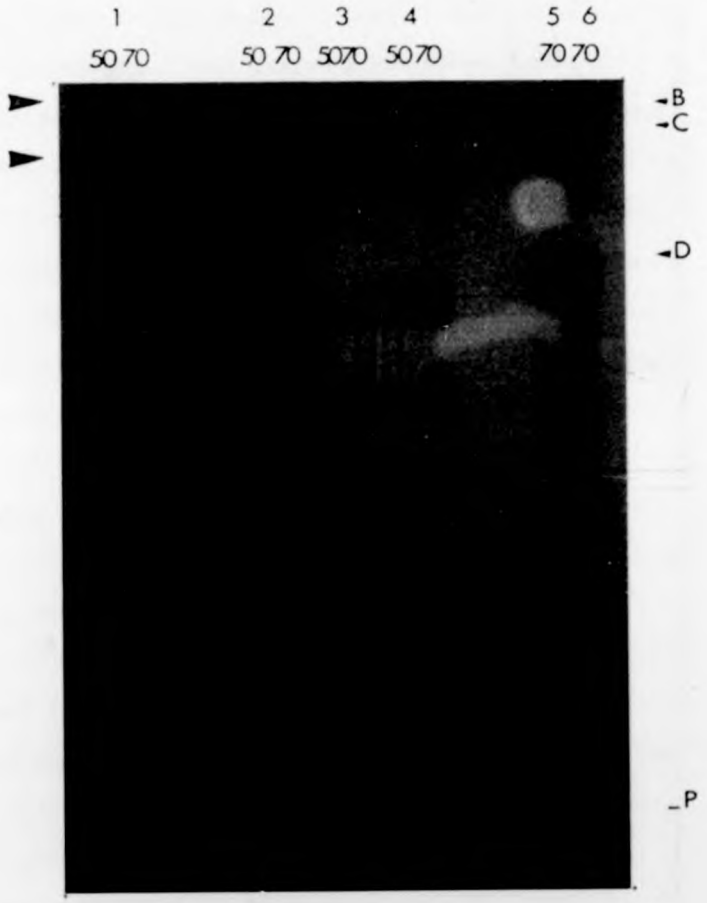


Figure 5.5: Primer Extension with RNA From Various Non-Muscle Tissues

Labelled primer was hybridized to 10µg X. laevis kidney total RNA (1), XTC cell (fibroblast derived) total RNA (2), ovary poly(A) + RNA (3), and ovary poly(A) - RNA (4), as well as 10 µg and 50 µg, respectively, of X. borealis total ovary RNA (5 and 6). The latter two hybridizations were performed at 70°C (without formamide) while the hybridizations to X. laevis RNA were performed at 50°C and 70°C, for each RNA. The primer extension products were run on an 8% sequencing-type gel. X. borealis ovary bands (B to D) are labelled in the same way as in Figure 5.2. The large arrows mark the positions in the X. laevis kidney tracks of the 2 faint sets of bands which are present in the other X. laevis tracks. The position of unextended primer is labelled P.

Nevertheless, these results do not conclusively prove that the same cytoskeletal actin gene is transcribed in all these tissues, although this seems to be most likely, and they do provide no evidence to the contrary. They also provide evidence against the possibility that different promoters are used for the same gene in different tissues, assuming that different length leaders would result from this.

5.4 Actin Expression in Skeletal Muscle

Figure 5.6(a) shows the primer extension products resulting from the hybridization of the cytoskeletal actin-derived primer to X. borealis skeletal muscle RNA. A range of temperatures from 50° to 70°C was used for the hybridization. Three groups of bands are indicated which could represent distinct RNA species. Set (ii) was more prevalent when the hybridization was performed at the lowest temperature. This was in contrast to the other 2 groups, which were more prevalent at 70°C. Figure 5.6(b) shows that group (i) runs at the same position as the set of bands called C in Figure 5.2(a), and that group (iii) runs at a different position to set D (the hybridizations in this experiment were performed at 70°C, so group (ii) was not seen). No set of bands corresponding to set B could be detected in experiments using skeletal muscle RNA (see Figure 5.6b), although in the original autoradiograph it was possible to detect a faint band running at the same position as the lower, fainter, band of the B set. As in non-muscle tissues, the precise number and clear identification of the genes that are expressed in skeletal muscle must await the sequencing (preferably individually) of all of these bands. The experiments described here appear to show that, in addition to the gamma-like cytoskeletal gene (whose product

Fig.5.6

a

Fig.5.6

a

Figure 5.6(a): Primer Hybridization to Muscle RNA at Varying Temperatures

5 µg of total X. borealis skeletal muscle RNA was hybridized overnight with 5'-labelled primer (see Figure 5.1) at 50°C (track 1), 55°C (track 2), 60°C (track 3), 65°C (track 4), and 70°C (track 5) in the absence of formamide. The hybridizations were done in 10 µl volumes. The primer that had hybridized to RNA was extended with AMV reverse transcriptase and the products separated on an 8% sequencing-type gel (see Section 2.7). Three sets of bands (labelled i to iii) are thought to represent distinct RNA species. The unextended primer is labelled P.

Fig. 5.6

b

Fig. 5.6

b

Figure 5.6(b): Comparison of Primer Extension Products of Muscle and Ovary RNA

6 µg (track 1) and 30 µg (track 2) of *X. borealis* ovary poly(A) + RNA and 5 µg of *X. borealis* skeletal muscle total RNA (track 3) were hybridized overnight with 5'-labelled primer (see Figure 5.1) at 70°C, in the absence of formamide. The primer which hybridized to RNA was extended with AMV reverse transcriptase and the products separated on an 8% sequencing-type gel (see Section 2.7). The ovary RNA bands labelled B, C and D, and the skeletal muscle bands labelled i and iii, are the same as those in Figure 5.2 and Figure 5.6(a), respectively. The unextended primer is labelled P.

appears as the C set of bands; see Section 5.2), at least 2 other actin genes are expressed in skeletal muscle.

It appears that the gamma-like actin mRNA species is far more abundant in skeletal muscle than the beta-like mRNA species (corresponding to the B set of bands), whereas the primer extension products of these two species, when using oocyte RNA, appear to be approximately equal in abundance (although, because the primer was made from a beta-like gene, and may not hybridize as well to a gamma-like message, this does not necessarily mean that the 2 messages are equal in abundance in the oocyte). Garrels and Gibson (1976) were unable to detect any cytoskeletal actins in adult muscle tissues, by Coomassie blue staining. This is an insensitive method, however, and Pardo et al. (1983) and Lubit and Schwarz (1980) have used antibodies to detect two different cytoskeletal actins in skeletal muscle cells (which are not components of the muscular contractile mechanism). They were, however, unable to prove, that these were the same proteins as those found in non-muscle tissues. My results would appear to support the view that whole skeletal muscle contains cytoskeletal actin and show that at least one species of muscle mRNA is the same as that found in the oocyte. It is possible, however, that the cytoskeletal actin mRNA species detected in this experiment is not found in muscle cells, but originates from non-muscle cells (e.g. fibroblasts and endothelial cells) present in muscle tissue. Also, it is possible that the presence of two non-cytoskeletal actin mRNA species, instead of the expected one, would be due to contamination with smooth muscle actin message from arteries etc. To remove such doubts, these experiments should be repeated using RNA from a purified myoblast cell line where differentiation has been induced (as in Caravatti et al., 1982).

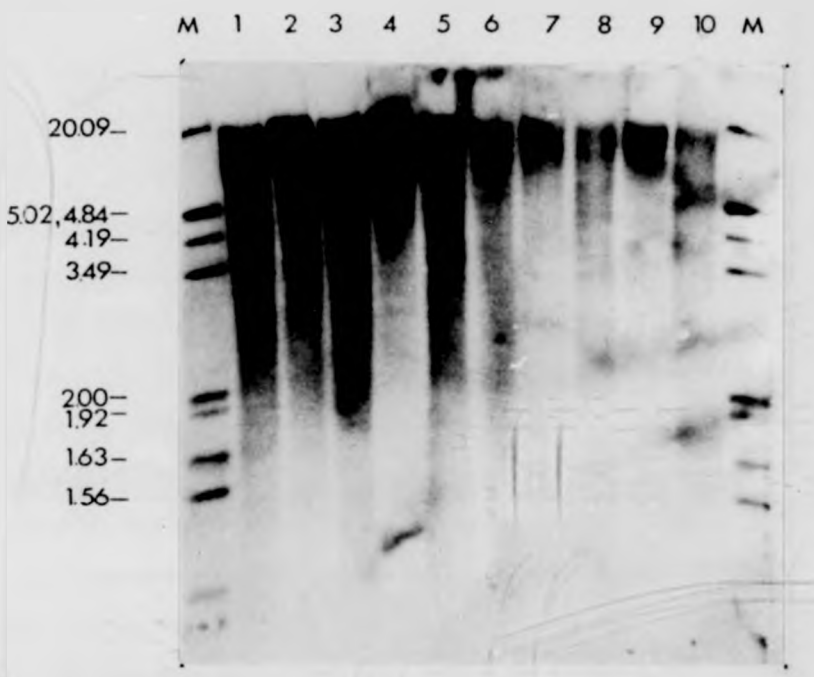
6 REPETITIVE ELEMENTS IN THE p35A INSERT

Figure 6.1(a) shows the result of hybridizing nick-translated insert from p35A to a Southern blot of X. borealis and X. laevis genomic DNA digested with various 6-cutter restriction enzymes. Figure 6.1(b) shows the result of a similar experiment, using nick-translated p5AP DNA as the probe. Using hybridization conditions of 68°C, 3 x SSC, and washing for 1 hour at 65°C in 0.5 x SSC, the hybridization with 35A resulted in a smear in each track, while hybridization with 5AP gave only discrete bands. This smear indicates that the p35A insert contains at least one sequence that is repeated many times elsewhere in the X. borealis and X. laevis genomes. The smear was not due to contaminating nuclease activity, as the same DNA digests were run on the gel which, once blotted, was used for the hybridization with p5AP. Of course, with such experiments, the genomic sequences that hybridize need not be identical to the repetitive elements in 35A. It is also possible that the 5AP insert contains repetitive elements that have diverged from its related sequences in the rest of the genome to a greater extent than those in 35A. Post hybridization washing of the nitrocellulose at lower stringency conditions (i.e. lower temperature and higher salt concentrations) might possibly reveal such sequences in 5AP.

The finding of repetitive sequences near the 35A actin gene came as no surprise, as it has been known for some time that eukaryotic genomes contain a substantial proportion of repetitive sequences (see Lewin, 1980b, and Section 1.3 for details of X. laevis), and that some of these are interspersed with single copy DNA (Lewin, 1980c; Jelinek and Schmid, 1982). Thus, 50% of the X. laevis genome is composed of repetitive sequences, of average length 300 bases, and of non-repetitive sequences

Fig.6.1

a


M 1 2 3 4 5 6 7 8 9 10 M
2009
502,484
419
349
200
192
163
156

b

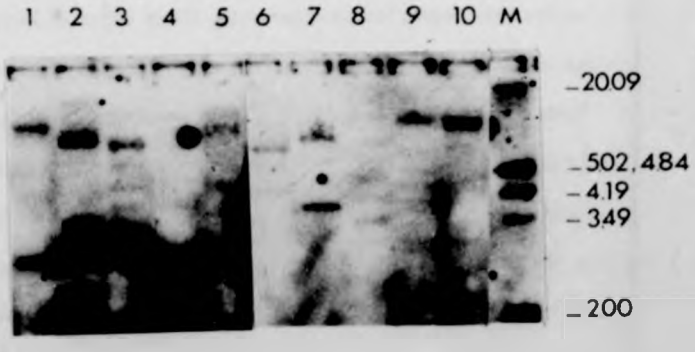
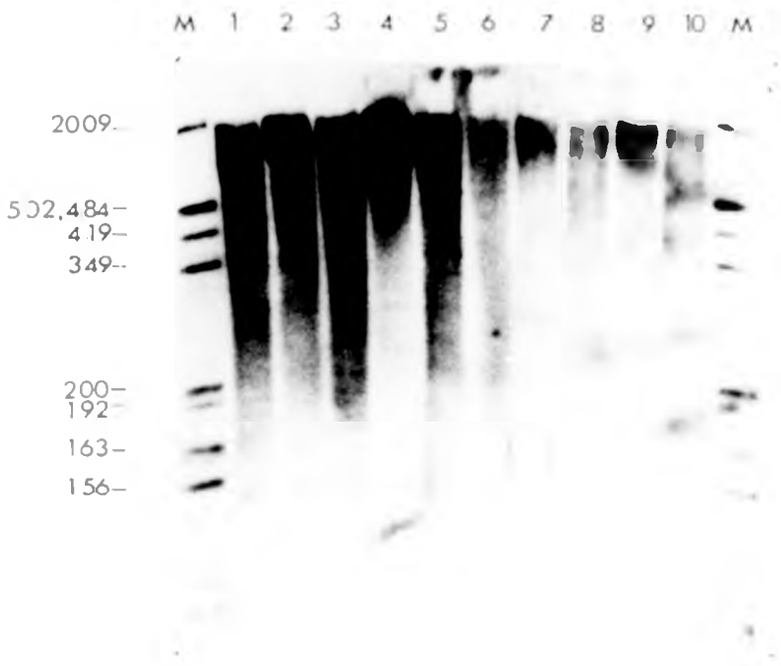

1 2 3 4 5 6 7 8 9 10 M
2009
502,484
419
349
200

Fig 6 1

a


M 1 2 3 4 5 6 7 8 9 10 M
2009
502,484
419
349
200
192
163
156

b

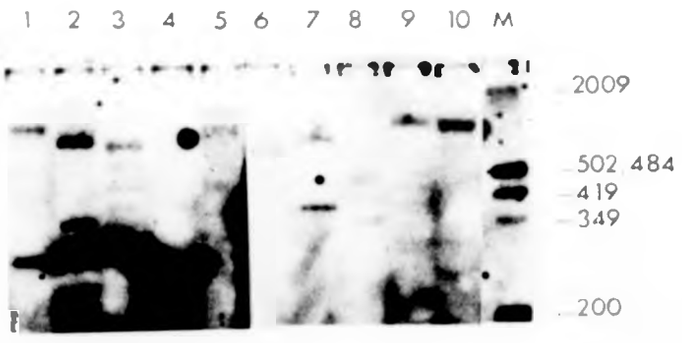

1 2 3 4 5 6 7 8 9 10 M
2009
502 484
419
349
200

Figure 6.1: Genomic Blots Probed with 35A and 5AP

Approximately 4 µg of X. borealis or X. laevis DNA (each from one individual) was digested with Eco RI (1 + 6), Sst I (2 +7), Hind III (3 + 8), Kpn I (4 + 9), and Pst I (5 + 10). Each digest was then divided in half and each half run on a separate 1.5% agarose gel. In each gel, tracks 1 to 5 contained X. borealis DNA, and 6 to 10, X. laevis DNA. The DNA from each gel was blotted onto nitrocellulose.

The blot shown in (a) was probed with nick-translated, ^{32}P-labelled p35A insert. The blot shown in (b) was probed with nick-translated ^{32}P-labelled p 5AP. The hybridizations were carried out overnight at 68°C in 3 x SSC plus Denharts additives. The filters were washed in 0.5 x SSC, at 65°C, for 1 hour. A mixture of a ^{32}P-end-labelled Hpa II digest of pAT153 and an Eco RI/Hind III digest of wild type lambda DNA run in the tracks labelled, M, as size markers. The sizes (in kilobases) of these are shown.

Fig. 6.2

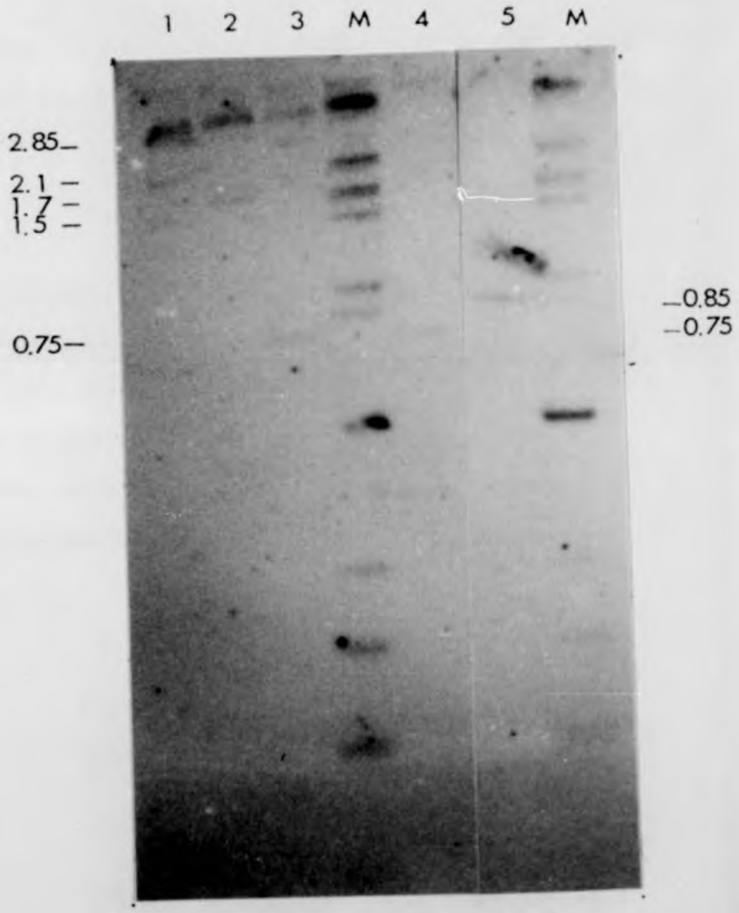


Fig. 6.2

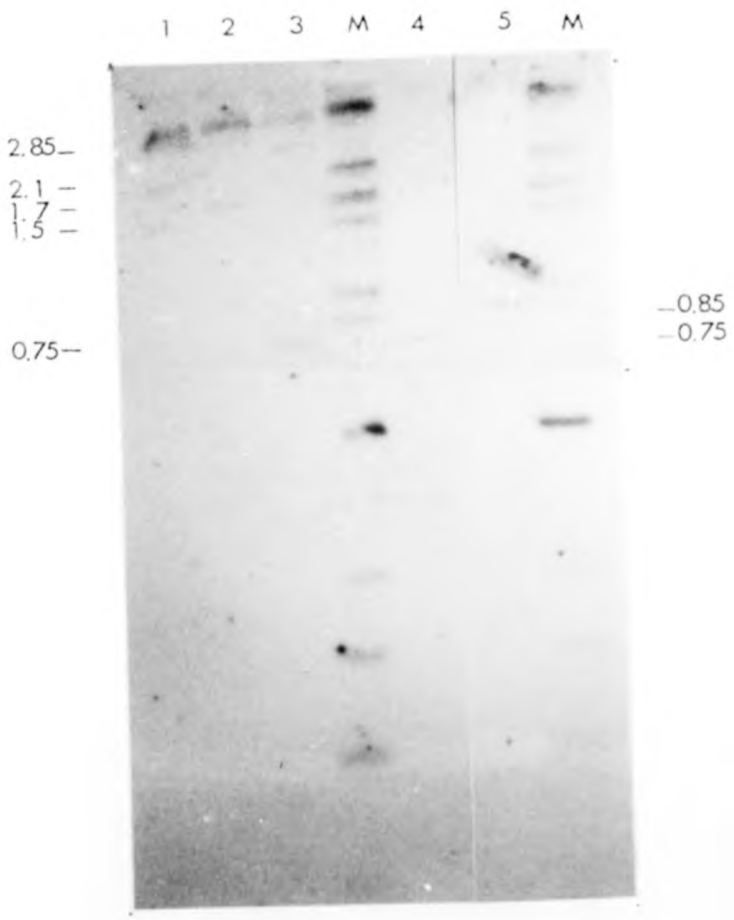


Fig. 6.2

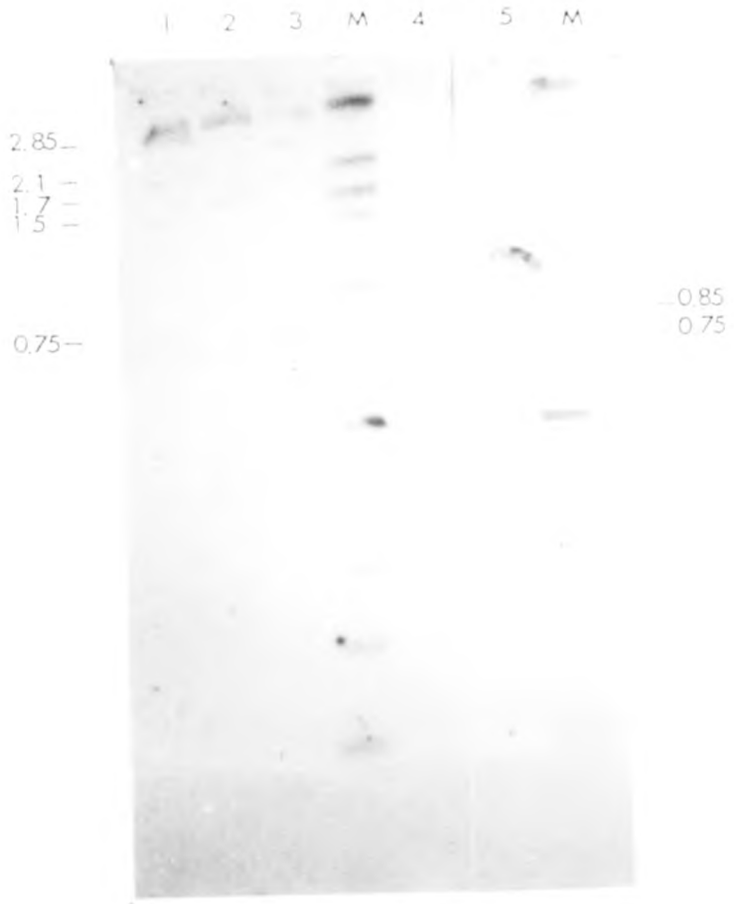


Figure 6.2: Localization of Repetitive DNA on the 35A Map

Some p35A was digested with Eco RI (until the digestion was complete). Approximately 2 to 3 µg of this mixture was digested with (1) Sst I, (2) Xba I, (3) Kpn I, (4) Hpa II, or (5) Hha I and the products electrophoresed on a 3% agarose gel. The DNA was blotted onto nitrocellulose and probed with nick-translated, ^{32}P-labelled X. borealis genomic DNA. The hybridization was at 60°C in 3 x SSC, 0.2% Denhardts, and the filters were washed in the same conditions. A mixture of a ^{32}P-end labelled Hinf I digest of pAT153 and an EcoRI/Hind III digest of wild-type lambda DNA was run in the tracks labelled M, as size markers (see Appendix 1). The sizes of the hybridizing bands are shown.

Figure 6.2: Localization of Repetitive DNA on the 35A Map

Some p35A was digested with Eco RI (until the digestion was complete). Approximately 2 to 3 µg of this mixture was digested with (1) Sst I, (2) Xba I, (3) Kpn I, (4) Hpa II, or (5) Hha I and the products electrophoresed on a 3% agarose gel. The DNA was blotted onto nitrocellulose and probed with nick-translated, ^{32}P-labelled X. borealis genomic DNA. The hybridization was at 60°C in 3 x SSC, 0.2% Denhardts, and the filters were washed in the same conditions. A mixture of a ^{32}P-end labelled Hinf I digest of pAT153 and an EcoRI/Hind III digest of wild-type lambda DNA was run in the tracks labelled M, as size markers (see Appendix 1). The sizes of the hybridizing bands are shown.

of average length 800 bases (Davidson et al., 1973). The possible role of such repetitive sequences in gene regulation has been discussed, briefly, in Section 1.3. It was of interest to find out where the repetitive elements on 35A were situated, relative to the gene.

Figure 6.2 shows an autoradiograph resulting from an experiment designed to localize the repetitive DNA in 35A. This involved the electrophoresis of various restriction digest products of 35A (the whole plasmid, p35A, was first restricted with EcoRI, the digestion checked, and then the resulting mixture used for further digestions) on an agarose gel, followed by Southern blotting. The nitrocellulose filter was hybridized overnight with nick-translated X. borealis genomic DNA at low stringency (60°C, 3 x SSC) and then washed under the same conditions. The rationale of this experiment was that only the repetitive elements in 35A would be abundant enough in the radioactive probe to produce detectable hybridization. Examination of Figure 6.2 reveals that with the three 6-cutter digests, each of which produces only 2 fragments, both fragments have hybridized with the genomic DNA. However, the degree of hybridization differs over the insert and the greater intensity of the 0.75 kb Kpn I band correlates with that of the 2.1 kb Sst I band, the 1.7 kb Xba I band, the 0.75 kb Hpa II band, and a 0.85 kb Hha I band. It would appear, then, that there may be more than one repetitive element in 35A. One element lies somewhere in the 0.75 kb Kpn I band, 5' to the transcriptional start of the actin gene, while it is possible that at least one other element is located somewhere in the 1.5 kb Sst I fragment, presumably within an intron, and is either less abundantly represented in X. borealis genome than the first, or less homologous to the rest of its family of repetitive sequences.

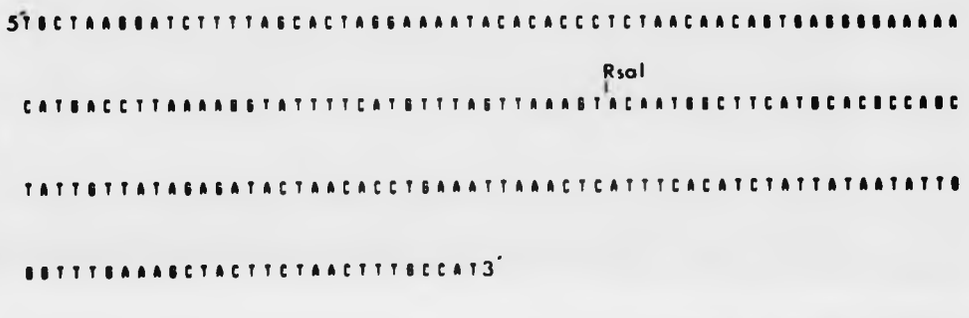
However, while the stronger hybridizing region 5' to the gene (a

result which is repeatable) can best be explained as being due to a repetitive element localized there, the absence of the necessary control from the above experiment (e.g. a region of 5AP containing actin coding sequence) makes the existence of the second proposed region rather more uncertain. Thus, it is possible that the region of the clone within the coding sequence hybridized to the actin coding sequences in the probe and not to any repetitive elements contained in the introns.

Does the DNA sequence tell us anything about the location and character of these repetitive sequences? Figure 6.3 shows the sequence of areas 1 and 2 from Figure 3.20 which are situated in the 0.75 kb Kpn I fragment containing the repetitive DNA which hybridizes to genomic DNA to the greatest extent. It is obviously not known that these two stretches of sequences actually include the repetitive element. Nevertheless, a computer comparison was made between this region, as well as other non-coding regions of 35A, and the sequences of several cloned Xenopus repetitive sequences. One of these was cloned from X. laevis and sequences related to part of this element have been found to be expressed during X. laevis development (Spohr et al., 1981, 1982). The other clone sequence has been found to be present in several Xenopus species, including X. borealis (Lam and Carroll, 1983). None of these sequences could be found in any great length in the sequence of 35A, although it is still possible that such sequences might exist in the 900 bases of 35A that have not been sequenced. Some smaller regions of homology between 35A and these sequences can be found. The largest of these is the sequence 5' - GAGNAGANGCNCTTAAGANTTNA - 3', which is shared by the sequence of Spohr et al. (1981) and the leader intron of the 35A gene (region 5, Figure 3.20, Figure 4.2). There is very little homology in the area outside this sequence. Numerous other short stretches of

Fig. 6.3

35A: Region 1



Region 2

Figure 6.3: **Sequence of 35A: Regions 1 and 2**

The anti-sense sequence of regions 1 and 2 of 35A (see Figure 3.20) are shown. Each complete line contains 60 bases. The inverted complement (with 2 base changes) of the underlined sequence, IR, is found 50 to 100 bases downstream (see Figure 5.3b and Section 6). The overlined sequence is the same, in 8 out of 10 positions, as the consensus sequence found 100 bases upstream from the cap site of several mammalian globin genes (see Section 5.1.2).

homology (10 to 20 bases) can be found, but the significance of such sequences is not clear. These homologies may be accidental, or an indication that these regions of the X. borealis clone may have originally arisen by the insertion of elements resembling these cloned repetitive sequences. Such short homologies may be indicative of the ways in which such non-coding sequences have arisen, yet may not necessarily be the sequences responsible for the hybridization of 35A to many genomic DNA restriction fragments.

A computer analysis of the 35A sequence reveals many small sequences (5 to 10 base-pairs) that are repeated within the cloned DNA. These are too numerous to be presented here and their significance is doubtful. Some repeats occur within a short enough length of sequence to appear to be not the result of chance. Thus, the sequence 5' - CTTNCTAGCTT - 3', which occurs in the leader region (Figure 5.3b), is repeated one base downstream, and the sequence 5' - ACTGNCTAGGT - 3', in the leader intron (Figure 4.2), is repeated 26 bp downstream (contained within this space, incidentally, is the perfect palindrome ACCATGCATGGT). Several elements containing sequences which I have suggested might be conserved in the rat beta-actin, are also repeated nearbye in the genome (see Figure 5.3b). These include the sequence immediately 5' to, and including part of, the putative 'TATA box': TGACGTT, which is repeated within the leader region, and the sequence 5' - CGCGAGGCG - 3', which is present as an imperfect repeat, CGCCAGGCG, approximately 50 bases 3' to this. The significance of such repeats is not clear. It is possible that such sequences are analogous to the short direct repeats, 8 to 20 bp long, that flank the Alu I family of short repetitive elements (Jelinek and Schmid, 1982), as well as several 'processed' pseudogenes (e.g. Moos and Gallwitz, 1983; Lee et al.,

1983). These are thought to be the results of integration of these DNA segments into a staggered host chromosomal break, followed by the repair of the target site (Van Arsdell et al., 1981). However, it seems unlikely that part of the 'TATA box' could consist of sequence resulting from the recent insertion of such an element (presumably the insertion would have to be recent for the direct repeat to remain unaltered).

Also of interest, is the sequence 5' - CAATCGCGAGGCGT - 3', which contains one of the putative 'CAAT boxes' (see Figure 5.3b), and the sequence 5' - GCGAGGC - 3' which was discussed in Section 5.1.2. This sequence is an imperfect inverted repeat of the sequence 5' - ACGCCTCCCTATTG - 3', which is situated 50 to 100 bases 5' to this (see Figure 6.3). Because of their existence in, or near, a region of DNA known to be repeated elsewhere in the genome, it is possible that these sequences are analogous to those inverted repeats bordering the (direct) Long Terminal Repeats (LTRs) of retroviruses, the Tyl sequences of yeast, and the long-period dispersed repeats of Drosophila (Jelinek and Schmid, 1982). The latter two types of sequence are thought to be transposable because they are not always present at the same chromosomal locations, and have been found in circular extrachromosomal form (Flavell and Ish-Horowicz, 1981). The double stranded form of retroviruses can integrate into the host cell chromosome, and may be stably transmitted to progeny cells (Hughes et al., 1979). These elements do differ from the sequences found in 35A, however, in that their LTRs are 300 to 600 bases long, and their inverted repeats have always been found to be bordered by the dinucleotides TG, at the 5' end, and CA, at the 3' end.

Summary

Part of the insert of p35A is repeated elsewhere in the X. borealis (and X. laevis) genome. The 0.75 kb Kpn I fragment has been shown to contain such repetitive DNA, but it is possible that other regions of 35A may do so as well.

7. ACTIN GENE NUMBER

7.1 How Many Actin Genes Are There In X. borealis?

Figure 7.1(a) shows the autoradiograph from an experiment where the products of various 6-cutter digests of genomic DNA from a single X. borealis frog have been separated on an agarose gel, blotted, and then probed under low stringency hybridization conditions with a fragment containing only actin-coding sequence from 5AP (from amino acids 52 to 137). The probe was made by digesting the 0.54 kb Hinf I fragment with BstN I, and extracting the 258 bp band from a 5% acrylamide gel (see Figure 3.18). Figure 7.1(b) shows the result of washing the blot shown in (a) at a higher temperature, before re-exposing.

What do these results tell us? Comparison of Figures 7.1(a) and (b) reveal that several strong bands present in the former, disappear in the latter, while others remain. Thus, in (b), two strong bands remain in tracks 1 and 2 (EcoRI and Sst I digests, respectively) while 3 remain in track 3 (Hind III digest). These bands, I believe, reveal that 2, or perhaps 3, genes exist in the X. borealis genome that have a high degree of homology with the probe. Of course, one of these represents the putative muscle-specific gene of which 5AP is part, and the probe was made from; one of the strong bands in the EcoRI track (track 1) is approximately the same size as the 2.6 kb EcoRI insert cloned in λ 5AP. The discrepancy in number of bands, in (b), between track 3 and the others, is possibly the result of the probe hybridizing to two halves of one gene in track 3, or alternatively, to two genes not being separated in tracks 1 and 2. If there is only one gene that is very similar to that cloned in 5AP then this is possibly the result of a recent duplication event. It is thought that the ancestor of X. borealis (and

Fig. 7.1

a

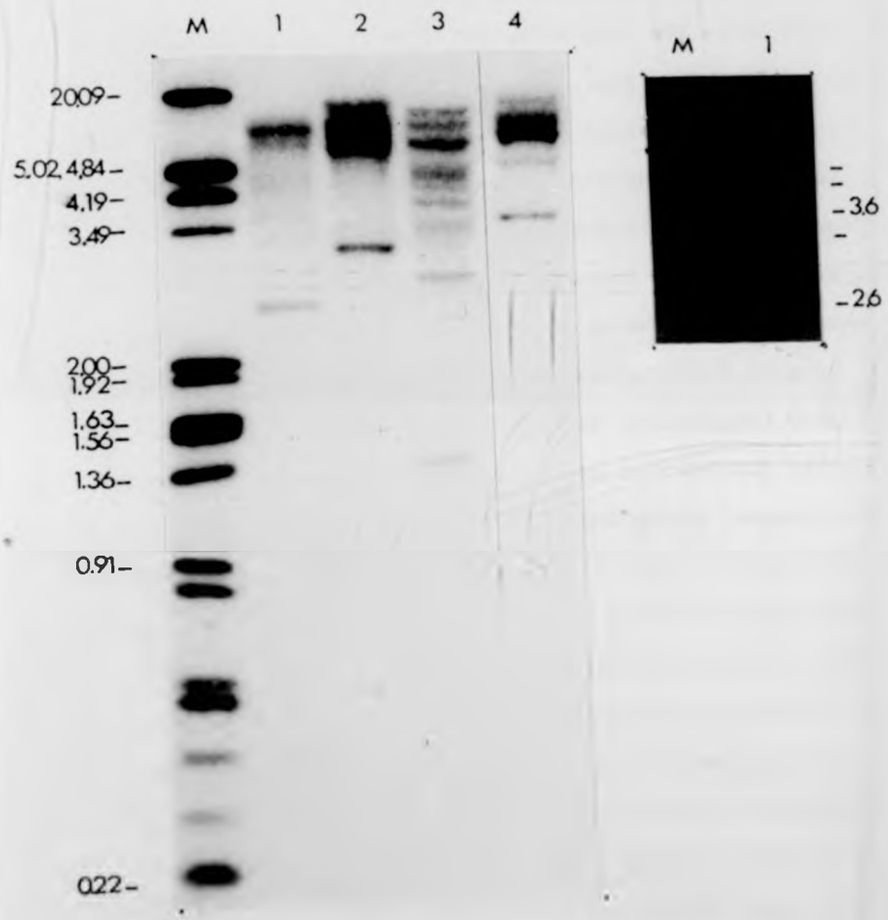


Fig. 7.1

a

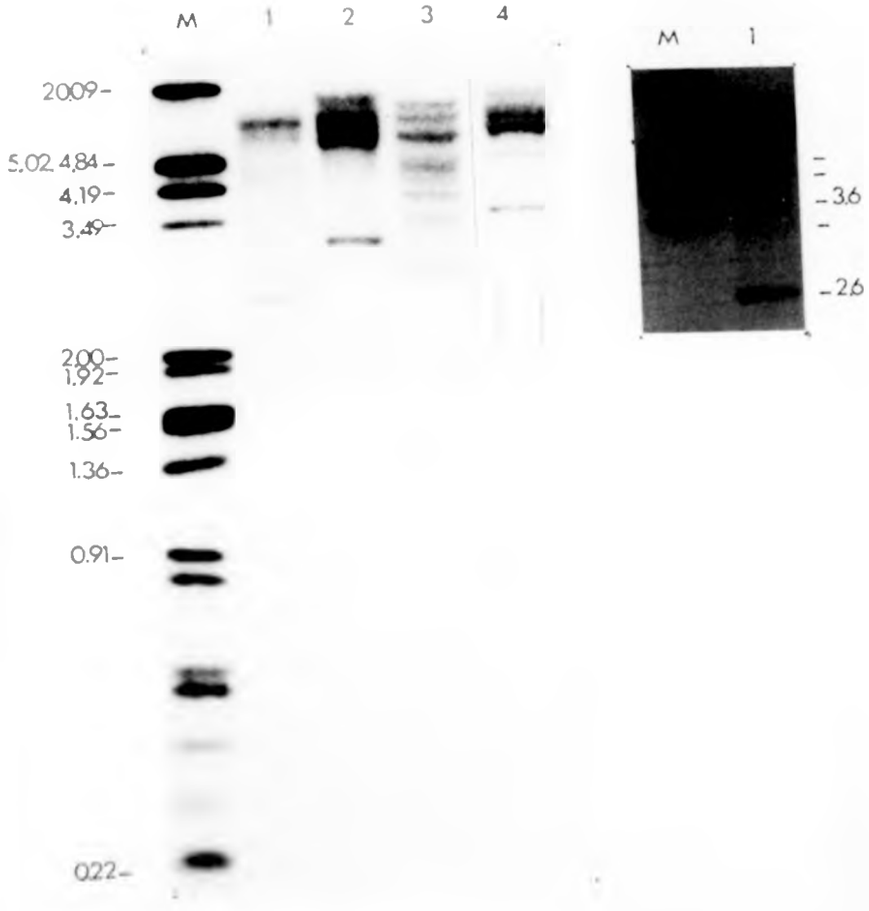


Fig. 7.1

b

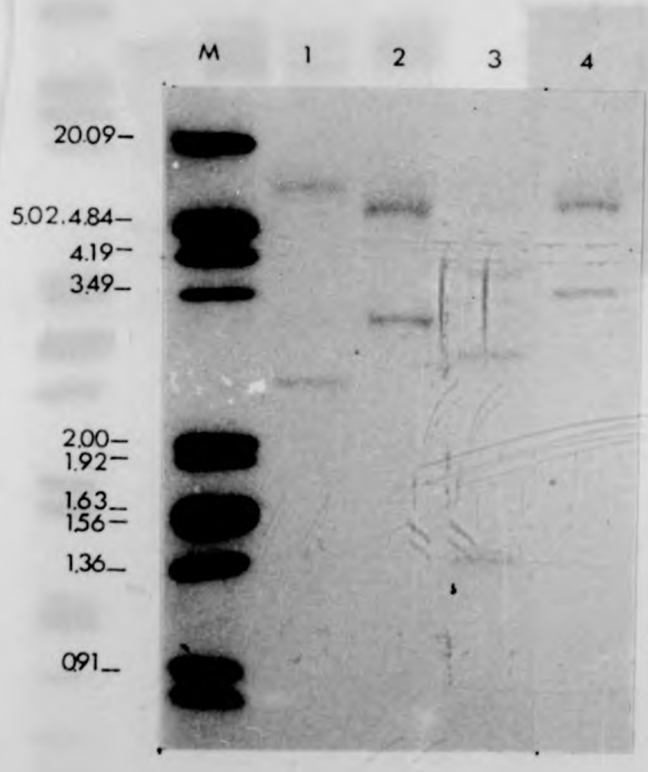


Fig. 7.1

b

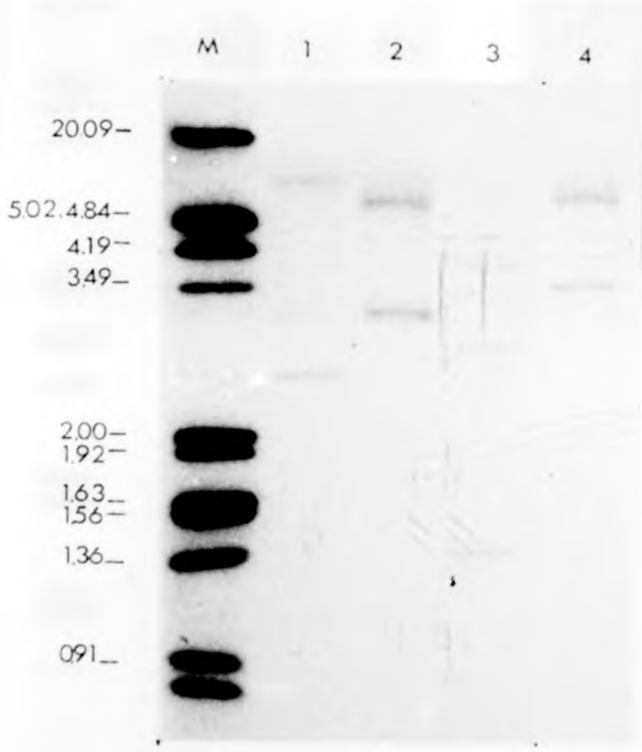


Figure 7.1: Genomic Blot Probed with Actin-Coding Region

5.6 µg of X. borealis genomic DNA (from one individual) was digested with EcoRI (track 1), Sst I (track 2), or Hind III (track 3). For track 4, 4 µg of X. laevis genomic DNA (from one individual) was digested with Sst I. The digestion products were run on a 1.5% agarose gel, and then blotted onto nitrocellulose. The blot was then probed with a nick-translated, ^{32}P-labelled, Bst NI-Hinf I fragment containing actin-coding sequence (amino acids 52 to 137) from the p5AP insert (see Figure 3.18). This hybridization was overnight in 50% formamide, 3 x SSC, 10% dextran sulphate (see Methods) at 42°C. The final wash of the filter was in 0.1 x SSC at (a) 0°C, and (b) 60°C. The markers (M) were a mixture of a radiolabelled Hinf I digest of pAT153 and of an Eco RI/Hind III digest of wild type lambda (see Appendix 1). In (a), a longer exposure of track 1 is included to visualize the fainter Eco RI bands. The band sizes are in kilobases.

X. laevis; two strong bands also remain in the X. laevis Sst I track after stringent washing; see track 4) has undergone a tetraploidization event since its divergence from a common ancestor with Xenopus tropicalis (which contains approx. half the genome size of the others), about 35 million years ago (Tymowska and Fischberg, 1973; Theibaud and Fischberg, 1977). This is a possible candidate for the gene duplication event that may have given rise to these results. An alternative, however, is that these genes are similar because of a recent gene conversion event (or events). At present, such possibilities are difficult to distinguish.

It is probably worthwhile pointing out that the detection of a 2.6 kb EcoRI band strongly homologous to the 5AP-derived probe, and a weaker EcoRI band, in (a), which is approximately the same size as the 3.6 kb insert from 35A, is useful confirmation that the DNA cloned in 5AP and 35A did originate from X. borealis. This is not a trivial point, as several X. laevis genomic libraries, using the same vector, were present in this laboratory at the time of the isolation of these clones, and contamination with pre-existing clones has caused much trouble in the past.

How many actin genes are there, in total, in X. borealis? There are approximately 9 to 12 bands in track 3 (Hind III digest) of Figure 7.1(a), while only 5 in track 2 (Sst I digest). The greater intensity of the latter bands, however, implies that some of these may represent more than one gene (although two, of course, are strong because of a greater homology between them and the probe), although, equally, some of the bands in the Hind III track may well arise from the hybridization of the probe to 2 portions of one gene. It is obviously possible to produce only a very rough estimate of actin gene number in X. borealis from a result like this; say between 5 and 15.

How could a more accurate estimation be produced? One of the problems with the method used here is the possibility that a restriction enzyme might cut one or more of the actins in the region to which the probe hybridizes. For this reason, it is obviously desirable to use as small a fragment of sequence as possible. However, the probe actually used will hybridize to DNA either side of the intron at amino acid 121 in the case of the cytoskeletal actin genes; this increases the chance that such genes will be cut by a restriction enzyme to show up as two bands. A useful test will be to repeat the experiment using a different coding-region probe; if the same bands were produced, for a given restriction enzyme, then it would be unlikely that this problem would apply.

There are several other problems with the interpretation of this experiment. As mentioned already, the number of genes represented by each band is unknown. Also, there is the possibility that some bands might be the result of only partial digestion of the DNA by the restriction enzyme. Finally, in this case, the counting of bands would be made easier by gaining further separation of the digested DNA in the agarose gel.

An alternative method of estimating gene number is in widespread use. This involves the hybridization of a radioactive probe to a known amount of genomic DNA, followed by a comparison of the amount of radioactivity bound to the genomic DNA with that bound to varying amounts of the DNA used to make the probe. In this way, the number of gene equivalents contained in a known amount of genomic DNA, and hence in one haploid genome's worth, can be calculated (for example, see Turner and Woodland, 1983). Although some of the problems described above would not be encountered in this method, others would still

remain; in particular, the weaker hybridization of the probe to less-homologous sequences would tend to give an underestimate.

To summarize: the experiments described here place the haploid number of actin genes in X. borealis as roughly between 5 and 15. I have also indentified at least one other gene which is similar in sequence to the putative muscle-specific gene cloned in 5AP.

7.2 Actin Gene Number In Other Species

How many actin genes do other organisms possess? With the exception of yeast, which has only a single gene (Ng and Abelson, 1980; Gallwitz and Seidel, 1980), other organisms have multigene families of varying sizes. The lower eukaryotes tend to have fewer genes than the higher eukaryotes, with the exception of Dictyostelium, which has 17 (Kindle and Firtel, 1978; McKeown et al., 1978). Thus Acanthamoeba has 3 (or more) genes (Nellen and Gallwitz, 1982), Oxytricha has 2 to 3 (Kaine and Spear, 1980) and Physarum, 4 or more (Schedl and Dove, 1982). The nematode Caenorhabditis elegans has 4 actin genes (Files et al., 1983), while Drosophila has 6 (Fyrberg et al., 1980). The deuterostomes, with the exception of hamster, which has only approximately 5 genes (Dodemont et al., 1982), have a greater number of genes still. Thus the chicken has been estimated to have 10 genes by Cot analysis (Schwartz and Rothblum, 1980), or 4 to 7 genes by Southern blot analysis (Cleveland et al., 1980; Zimmer and Schwartz, 1982). The sea urchin Strongylocentrotus purpuratus has 11 to 20 genes (Durica et al., 1980; Scheller et al., 1981; Davidson et al., 1982), while mice, rats and human have an even larger number: greater than 20 (Engel et al., 1981; Humphries et al., 1981; Soriano et al., 1982; Dodemont et

al., 1982).

Why do some mammals have so many actin genes? The low number of actin genes in hamster makes it seem unlikely that all of these genes of rats, mice and humans encode proteins with distinct functions. In humans, 8 different actin gene sequences homologous to both beta- and gamma-actin mRNAs have been cloned so far (Engel et al., 1982), while in chickens there appear to be 1 beta- and 3 gamma-like actins (Cleveland et al., 1980). The reason for this multiplicity in cytoskeletal actin genes is not clear. It is possible that there may be more cytoskeletal isoforms than at first thought (brain and nucleus types? - Section 1.2), as was found in amphibians (Vandekerckhove et al., 1981), although this explanation seems unlikely to explain the large number of genes in humans (see below). In fact, very few of these genes have been shown to be expressed, so it is possible that many of the rat, mouse and human genes are pseudogenes. The latter have been found in many other gene families (Jacq et al., 1977; Proudfoot, 1980; Little, 1982) and consist of genomic DNA sequences which show considerable homology to functional genes, but which no longer give rise to functional RNA transcripts. Recently, Moos and Gallwitz (1983) have identified several beta- and gamma-like actin pseudogenes in the human genome, and postulated that a large number of the human actin gene copies represent pseudogenes related to cytoskeletal actins. These cloned pseudogenes were of the type that lack introns, and are flanked by direct repeats, with a poly(dA) tract preceeding the 3' repeat. Such pseudogenes have been termed 'processed genes' (Hollis et al., 1982) and are thought to derive from the reintegration, into the germline DNA, of a cDNA transcript from an mRNA (which, of course, has been processed and is, therefore,

intronless). The direct repeats may have arisen from a transposon-like insertion mechanism of the cDNA copies (see Calos and Miller, 1980). Such a phenomenon is not unique to actin genes; it has also found in the case of the human immunoglobulin (Hollis et al., 1982), tubulin (Wilde et al., 1982A, 1982b), metalothionein (Karin and Richards, 1982), dihydrofolate reductase (Chen et al., 1982), and the rat tubulin (Lemischka and Sharp, 1982) gene families. Cytoskeletal actin messages have been found to be abundant in many cell types, including the Xenopus oocyte (Ballantine et al., 1979), so it is not difficult to see how a cDNA transcript may find its way into the chromosomes of the germline. The fact that some processed genes, such as immunoglobulins, would not be expected to be expressed in any germ line cell, indicates that more than one mechanism may be involved. Alternatively, the obvious explanation may be incorrect in the case of genes encoding abundant mRNAs in germline cells. An alternative explanation is that the pseudogenes derive from infection with a retrovirus. Such a mechanism would enable the transmission of abundant mRNA sequences from any cell-type capable of viral infection.

A second class of pseudogenes has been encountered in other gene families (Little, 1982). These usually contain introns, and are thought to be derived by the duplication of a functional gene followed by subsequent mutation and inactivation of one of the resultant pair of genes (presumably because there is no selective pressure to retain both in a functional state). Two of the Dictyostelium actin genes have been found not to be expressed, and not to possess 'TATA' boxes (see Section 5.1.2), and are thought to be of this type (functional Dictyostelium actin genes have no introns and so this criterion is of no use here) because of their close linkage to other actin genes (McKeown and Firtel,

1981a; McKeown and Firtel, 1981b; McKeown and Firtel, 1982).

It can be seen that, while these studies of actin gene number are interesting from the point of view of the evolution of the eukaryotic genome in general, they are not greatly important to the developmental biologist. While the number obtained is useful as a maximum value, the presence of a large number of pseudogenes can be greatly misleading. What is important is the number of sequences actually expressed. This may be studied by S1 analysis using unique regions of particular cloned genes, and by primer extension, using probes from regions of conserved sequence.

8 PROSPECTS

This project has involved the isolation and sequencing of the 5' portions of 2 Xenopus borealis actin genes; one that encodes an amino acid sequence similar to a mammalian cytoskeletal actin, and the other, a sequence similar to a mammalian muscle-specific actin. Using primer extension, I have shown that the former gene is expressed in the oocyte, and have mapped its transcriptional start point in the genomic clone. I have also shown that at least one cytoskeletal actin gene is expressed in several different X. laevis non-muscle tissues. In this section I will briefly consider several experiments that could be performed to continue this work.

One of the original aims of this project was to examine the transcriptional control of a muscle-specific gene. Although a large amount of the coding and 5'-flanking regions of a putative muscle-specific gene have been sequenced, this gene has not yet been transcriptionally mapped, or the site of its expression determined. The primer constructed from the cytoskeletal gene hybridizes to muscle-specific RNA (see Section 5.4) and it should be possible to use this to obtain the sequence of the 5' translated (the first 25 bases) and untranslated regions of actin mRNAs from various muscle tissues of Xenopus. This should identify the tissue where the putative muscle-specific gene is expressed (as well as giving useful information on the expression of other muscle-specific actin genes; for example, it would be useful to know if the distribution of isoforms between the tissues is the same as that found in mammals). As with the cytoskeletal actin gene, comparison of this primer extended sequence with that of the genomic clone could reveal the transcriptional start point on the

latter. The sequence in the promoter region of this gene would then be compared with that of similarly expressed genes of other species. Thus, any conserved sequences, possibly with a regulatory function, could be identified.

It may be necessary, for future experiments with the muscle-specific actin gene, to isolate clones containing DNA fragments adjacent (in the genome) to the one isolated here. Thus, for example, the discovery of leader introns in several vertebrate muscle-specific actin genes, as well as the X. borealis cytoskeletal gene described here, increases the likelihood that the muscle actin clone I have isolated does not include the promoter (i.e. this gene might contain a leader intron greater than 900 bases, the amount of DNA between the coding region and the 5' end of the insert). If this was found to be true, then the 0.45 kb Hind III fragment of this insert (containing intron sequence only) could be used as a gene-specific probe for screening a genomic library. This is because, when used as a probe in hybridizations to Southern blots of X. borealis genomic DNA, only one Pst I, Hind III, or Eco RI fragment is found to hybridize.

Once it has been transcriptionally mapped, it would be interesting to inject the cloned muscle-specific gene into Xenopus laevis fertilized eggs. If this gene is not expressed in the egg, but is expressed in muscle tissue, later in development (it should be possible to distinguish its transcripts from the endogenous ones by primer extension or S1 analysis), then the role of various sequences in this regulation could be tested (by mutating or deleting these sequences, recloning the altered genes, and observing the expression of these genes during development). In addition, it may be possible to induce the expression of the muscle-specific gene in non-muscle tissues, such as the oocyte,

by the addition of various components from muscle cells, and so, the proteins involved in the gene-regulation might eventually be isolated. Blau et al. (1983) have shown that cytoplasmic factors exist in differentiated mouse muscle cells which will activate the expression of several muscle-specific genes in human amniocytes when the latter are fused to them (the nuclei remaining separate).

It would be interesting to extend the investigation (using primer extension) of cytoskeletal actin expression to other tissues; for example, the brain. The possibility of a brain synaptosome-specific actin isoform has already been mentioned in Section 1.2 (Marotta et al., 1982), and an mRNA for another cytoskeletal protein, beta-tubulin, has recently been shown to be present only in brain tissue (Bond and Farmer, 1983).

One of the major conclusions of this thesis stems from the similarity of the intron positions, and encoded amino acid sequences, of the cytoskeletal and putative muscle-specific actin genes to those of their mammalian counterparts. This suggests that these 2 types of gene began to evolve separately before the divergence of the Amphibia from the rest of the vertebrates. The cloning and sequencing of the 3' portions of these genes would obviously reveal how great the similarity between amphibian and mammalian actin genes is.

Another interesting finding has been that X. borealis contains an expressed cytoskeletal actin gene which is identical, in its first 3 acidic amino acid residues, to the mammalian beta-isoform. It has been reported (Vandekerckhove et al., 1981) that this isoform cannot be found in the closely related species X. laevis. I have suggested that, if true, this indicates that such differences between isoforms, in the first 3 acidic residues, may have no functional significance. However,

the conservation of the mammalian beta- and gamma-cytoskeletal isoforms, as well as the apparently separate localization of these isoforms in mammalian skeletal muscle (Pardo *et al.*, 1983), makes this conclusion rather surprising. It would be useful, therefore, to try and confirm the findings of Vandekerckhove *et al.* (1981) (which were based on the sequencing of N-terminal peptides) by sequencing the primer extension products obtained after hybridizing the *X. borealis* primer used in this study to *X. laevis* oocyte RNA.

APPENDIX 1: DNA Marker Fragment Sizes

Several gels shown in this thesis include tracks containing DNA size markers. For the sake of clarity the sizes of these fragments have not always been included in the Figures concerned. For easy reference, the fragment sizes (in kilobases) of various marker mixtures have been summarised here:

Hind III/Eco RI digest of 'wild-type' lambda (lambda cI857S7)

20.09, 5.02, 4.84, 4.19, 3.49, 2.00, 1.92, 1.56, 1.36, 0.91, 0.81, 0.567, 0.153.

Hind III digest of 'wild-type' lambda

23.72, 9.46, 6.67, 4.26, 2.25, 1.96, 0.59, 0.10.

Hinf I digest of pAT 153 (Twigg and Sherrat, 1980)

1.631, 0.517, 0.396, 0.298, 0.221, 0.220, 0.154, 0.145, 0.75.

Hpa II digest of pAT153

0.622, 0.492, 0.404, 0.242, 0.238, 0.217, 0.201, 0.190, 0.160, 0.147, etc.

APPENDIX 2: A Method for Calculating Silent Site Sequence Divergence, taken from Perler et al. (1980)

The nucleotide substitutions are assumed to be Poisson-distributed. First, the number of potential silent sites are counted. The possible changes at each position of a codon are considered and scores are assigned as shown in Fig. A1. The number of sites that can afford one, two, or three changes are counted separately, and each of the three categories are averaged between the two sequences under comparison.

Then the two sequences are compared codon by codon. Each silent substitution is scored with one point and categorized according to the type of site in which it has occurred. For example, if the codons compared are Pro CCC - Pro CCT, there is one change of category 3 of silent substitutions; the codons Leu CTC - Phe TTT have one change of replacement substitutions (first codon position), a half change of category 1 of silent substitutions (third codon position); the change in the third position of codons Ile ATC - Thr ACG is half silent (category 3), and half replacement.

Finally, the sum of silent substitutions in each category is divided by the sum of the corresponding potential sites. Each percentage change (z) is then corrected for multiple events by one of the following formulae:

Category 1: $3 \times [-1/2 \ln(1-2z)]$

Category 2: $3/2 \times [-2/3 \ln(1-3/2z)]$

Category 3: $[-3/4 \ln(1-4/3z)]$

The brackets are correction formulas relating the observed frequency to the Poisson-distributed frequency of mutational hits for

mutations restricted to one, two, or three substitutions, respectively. The factors outside the brackets correct these restricted values up to a total mutation rate, assuming that transitions and transversions are equally probable.

For each pair of sequences the overall divergence is the average of two weighted averages of the corrected percentages, using separately as weighing factors the number of sites or the numbers of changes in each category.

Figure A1: Scores of Silent Sites of Codons

	A	B	C
Met, Trp	0	0	0
Phe, Tyr, His, Gln, Glu, Asn Asp, Cys, Lys, Ser AGY	0	0	1
Ile	0	0	2
Val, Pro, Thr, Ala, Gly, Ser, UCX Leu CUY, Arg CGY	0	0	3
Leu CUR, Arg CGR	1	0	3
Leu UUR, Arg AGR	1	0	1

The number of potential silent changes is shown for each codon (ABC). Y is a pyrimidine, R is a purine and X is any base.

REFERENCES

Abreu, S. L. and Brinster, R. L. (1978) Actin and tubulin synthesis in murine blastocyst outgrowth. Expl. Cell. Res. 115, 89-94.

Adams, J. W., Kaufman, R. E., Kretschmer, P. J., Harrison, M. and Nienhuis, A. W. (1980) A family of long reiterated DNA sequences, one copy of which is next to the human beta-globin gene. Nucl. Acid. Res. 8, 6113-6128.

Akam, M. (1983) Decoding the Drosophila complexes. T.I.B.S. 8 No. 5, 173-177.

Akusjarvi, G. and Persson, H. (1981) Controls of RNA splicing and termination in the major late adenovirus transcription unit. Nature 292, 420-426.

Alton, T. H. and Lodish, H. F. (1977) Translational control of protein synthesis during the early stages of differentiation of the slime mould Dictyostelium discoideum. Cell 12, 301-310.

Alving, R. E. and Laki, K. (1966) N-terminal sequence of actin. Biochemistry 5, 2597-2601.

Anderson, D. M., Scheller, R. H., Posakony, J. W., McAllister, L. B., Trobert, S. G., Beall, C., Britten, R. J. and Davidson, E. H. (1981) Repetitive sequences of the sea urchin genome. Distribution of members of specific repetitive families. J. Mol. Biol. 145, 5-28.

Arnheim, N., Krystal, M., Schmickel, R., Wilson, G., Ryder, O. and Zimmer, E. (1980) Molecular evidence for genetic exchanges among ribosomal genes on nonhomologous chromosomes in man and apes. PNAS 77, 7323-7327.

Artavanis-Tsakonas, S., Schedl, P., Mirault, M. E., Moran, L. and Lis, J. (1979) Genes for the 70,000 dalton heat shock protein in two cloned D. melanogaster DNA segments. Cell 17, 9-18.

Artymiuk, P. J., Blake, C. C. F. and Sippel, A. E. (1981) Genes pieced together - exons delineate homologous structures of diverged lysozymes. Nature 290, 287-288.

Aviv, H., Volloch, Z., Bastos, R. and Levy, S. (1976) Biosynthesis and stability of globin mRNA in cultured erythroleukaemic Friend cells. Cell 8, 495-503.

Bains, W. (1982) The structure of cloned histone genes of Xenopus borealis. Ph.D. Thesis, University of Warwick.

Ballantine, J. E. M., Woodland, H. R. and Sturgess, E. A. (1979) Changes in protein synthesis during the development of Xenopus laevis. J. Embryol. exp. Morph. 51, 137-153.

Barrie, P. A., Jeffreys, A. J. and Scott, A. F. (1981) Evolution of the beta-globin gene cluster in man and primates. J. Mol. Biol. 149, 319-336.

Bastos, R. N., Volloch, Z. and Aviv, H. (1977) Messenger RNA population analysis during erythroid differentiation: a kinetical approach. J. Mol. Biol. 110, 191-203.

Bautch, V. L., Storti, R. V., Mischke, D. and Pardue, M. L. (1982) Organization and expression of Drosophila tropomyosin genes. J. Mol. Biol. 162, 231-250.

Bell, G. I., Pictet, R. and Rutter, W. J. (1980a) Analysis of the regions flanking the human insulin gene and sequence of an Alu family member. Nucl. Acid. Res. 8, 4091-4109.

Bell, G. I., Pictet, R. L., Rutter, W., Cardell, B., Tischer, E. and Goodman, H. M. (1980b) The nucleotide sequence of the human insulin gene. Nature 284, 26-32.

Bender, W., Davidson, N., Kindle, K. L., Taylor, W. C., Silverman, M. and Firtel, R. A. (1978) The structure of M6, a recombinant plasmid containing Dictyostelium DNA homologous to actin messenger RNA. Cell 15, 779-788.

Bender, W., Akam, M., Karch, F., Beachy, P. A., Peifer, M., Spierer, P., Lewis, E. B. and Hogness, D. S. (1983) Molecular genetics of the bithorax complex in Drosophila melanogaster. Science 221, 23-29.

Bennetzen, J. L. and Hall, B. D. (1982) Codon selection in yeast. J. Biol. Chem. 257, 3026-3031.

Benoist, C., O'Hare, K., Breathnach, R. and Chambon, P. (1980) The ovalbumin gene sequence of putative control regions. Nucl. Acid. Res. 8, 127-142.

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

Benton, W. D. and Davis, R. W. (1977) Screening lambda-gt recombinant clones by hybridization to single plaques in situ. Science 196, 180-182.

Benyajati, C., Spoerel, N., Haymerle, H., Ashburner, M. (1983) The mRNA for alcohol dehydrogenase in Drosophila melanogaster differs in its 5' end in different developmental stages. Cell 33, 125-133.

Berk, A. J. and Sharp, P. A. (1977) Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease digested hybrids. Cell 12. 721.

Bernstein, S. I., Glenn, A. W., Emerson, C. P. and Donady, J. J. (1981) Isolation, identification, and chromosomal mapping of a Drosophila myosin heavy chain. In Abstracts of the Cold Spring Harbor Meeting on Molecular and Cellular Control of Muscle Development Sept. 1981, p32. Pub. C. S. H. Lab, N.Y.

Bird, A. P., Taggart, M. H. and Smith, B. A. (1979) Methylated and unmethylated DNA compartments in the sea urchin genome. Cell 17, 889-901.

Bird, A., Taggart, M. H. and Macleod, D. (1981) Loss of rDNA methylation accompanies the onset of ribosomal gene activity in early development of Xenopus laevis. Cell 26, 381-390.

Bisbee, C. A., Baker, M. A., Wilson, A. C., Hadji-Azimi, I. and Fischberg, M. (1977) Albumin phylogeny of the clawed toad X. laevis. Science 195, 785-787.

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

Blau, H. M., Chiu, C.-P. and Webster, C. (1983) Cytoplasmic activation of human nuclear genes in stable heterocaryons. Cell 32, 1171-1180.

Bolivar, F. (1978) Derivatives of plasmid pBR322 carrying unique EcoRI sites for the selection of EcoRI-generated recombinant DNA molecules. Gene 4, 121-136.

Bond, J. F. and Farmer, S. R. (1983) Regulation of tubulin and actin mRNA production in rat brain: expression of a new beta-tubulin mRNA with development. Mol. and Cell. Biol. 3, 1333-1342.

Bravo, R., Fey, S. J., Small, J. V., Larsen, P. M. and Celis, J. E. (1981) Co-existence of three major isoactins in a single sarcoma 180 cell. Cell 25, 195-202.

Breathnach, R., Benoist, C., O'Hare, K., Gannon, F., Chambon, P. (1978) Ovalbumin gene: Evidence for a leader sequence in mRNA and DNA sequences at the exon-intron boundaries. P.N.A.S. 75, 4853-4857.

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

Bremer, J. W., Busch, H. and Yeoman, L. C. (1981) Evidence for a species of nuclear actin distinct from cytoplasmic and muscle actins. Biochemistry 20, 2013-2017.

Britten, R. J. and Davidson, E. H. (1969) Gene regulation for higher cells: a theory. Science 165, 349-357.

Brown, D. D. and Dawid, I. B. (1968) Specific gene amplification in oocytes. Science 160, 272-280.

Buckley, I. K. and Porter, K. R. (1967) Cytoplasmic fibrils in living cultured cells. Protoplasma 64, 349-380.

Burke, J. F. and Ish-Horowicz, D. (1982) Expression of Drosophila heat shock genes is regulated in rat-1 cells. Nucl. Acid. Res. 10 3821-3830.

Busch, H., Reddy, R., Rothblum, L., Choi, Y. C. (1982) SnRNAs, snRNPs, and RNA processing. Ann. Rev. Biochem. 51, 617-654.

Cacares, A., Payne, M. R., Binder, L. I. and Steward, O. (1983) Immunocytochemical localization of actin and microtubule-associated protein MAP2 in dendritic spines. P.N.A.S. 80, 1738-1742.

Calabretta, B., Robberson, D. C., Barrera-Saldana, H. A., Lambrout, T. P. and Saunders, G. F. (1982) Genome instability in a region of human DNA enriched in Alu repeat sequences. Nature 296, 219-225.

Calos, M. P. and Miller, J. H. (1980) Transposable elements. Cell 20, 579-595.

Caravatti, M., Minty, B., Montarras, R. D., Weydert, A., Cohen, P., Daubas, P. and Buckingham, M. (1982) Regulation of muscle gene expression: the accumulation of messenger RNAs coding for muscle-specific proteins during myogenesis in a mouse cell line. J. Mol. Biol. 160, 59-76.

Carmon, Y., Czosnek, H., Nudel, U., Shani, M. and Yaffe, D. (1982) DNAase I sensitivitiy of genes expressed during myogenesis. Nucl. Acid. Res. 10, 3085-3098.

Chao, M. J., Mellon, P., Charney, P., Maniatis, T. and Axel, R. (1983) The regulated exprssion of beta-globin genes introduced into mouse erythroleukaemia cells. Cell 32, 483-493.

Chen, M., Shimada, t., Moulton, A. D., Harrison, M. and Nienhuis, A. W. (1982) Intronless human dihydrofolate reductase genes are derived from processed RNA molecules. P.N.A.S. 79, 7435-7439.

Chisholm, R. L. (1982) Methylation and developmental regulation of gene expression. TIBS 7, No. 12, 421.

Clarke, L. and Carbon, J. (1976) A clone bank containing synthetic ColE1 hybrid plasmids representative of the entire E. coli genome. Cell 9, 91.

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 alpha- and beta-tubulin and beta- and gamma-actin genes using specific cloned cDNA clones. Cell 20, 95-105.

Cleveland, D. W., Hughes, S. H., Stubblefield, E., Kirschner, M. W. and Varmus, H. E. (1981) Multiple alpha and beta tubulin genes represent unlinked and dispersed gene families. J. Biol. Chem. 256, 3130-3134.

Clewell, D. B. and Helinski, D. R. (1970) Properties of supercoiled DNA-protein relaxation complex and strand specificity of the relaxation event. Biochemistry 9, 4428-4440.

Cochet, M., Gannon, F., Maroteaux, R. H. C., Perrin, F. and Chambon, P. (1979) Organization and sequence studies of the 17-piece chicken conalbumin gene. Nature 282, 567-574.

Collins, J. H. and Elzinga, M. (1975) The primary structure of actin from rabbit skeletal muscle: completion and analysis of the amino acid sequence. J. Biol. Chem. 250, 5915-5920.

Cooke, N. E., Coit, D., Shine, J., Baxter, J. D. and Martial, J. A. (1981) Human prolactin: cDNA structural analysis and evolutionary comparisons. J. Biol. Chem. 256, 4007-4016.

Cooper, A. D. and Crain, W. R. (1982) Complete nucleotide sequence of a sea-urchin actin gene. Nucl. Acid. Res. 10, 4081-4092.

Corces, V., Pillieer, A., Axel, R. and Meselson, M. (1981) Integration, transcription and control of a Drosophila heat-shock gene in mouse cells. PNAS 78, 7038-7042.

Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C. and Chambon, C. (1980) Promoter sequences of eukaryotic protein-coding genes. Science 209, 1406-1414.

Craig, S. W. and Pollard, T. D. (1982) Actin-binding proteins. T.I.B.S. 7, No. 3, 88-92.

Craik, C. S., Buchman, S. R. and Beychock (1980) Characterisation of globin domains: heme binding to the central exon product. PNAS 77, 1384-1388.

Czosnek, H., Nudel, U., Shani, M., Barker, P. E., Prautcheva, D. D., Ruddle, F. H. and Yaffe, D. (1982) The genes coding for the muscle contractile proteins, myosin heavy chains, myosin light chain 2, and skeletal muscle actin are located on three different mouse chromosomes. Embo. J. 1, 1299-1305.

Daubas, P., Caput, D., Buckingham, M. and Gros, F. (1981) A comparison between the synthesis of contractile proteins and the accumulation of their translatable mRNAs during calf myoblast differentiation. Dev. Biol. 84, 133-143.

Davidson, E. H., Hough, B. R., Amenson, C. S. and Britten, R. J. (1973) General interspersion of repetitive with nonrepetitive sequence elements in the DNA of Xenopus. J. Mol. Biol. 77, 1-24.

Davidson, E. H., Hough, B. R., Klein, W. H. and Britten, R. J. (1975) Structural genes adjacent to interspersed repetitive DNA sequences. Cell 4, 217-238.

Davidson, E. H. and Britten, R. J. (1979) Regulation of gene expression: possible role of repetitive sequences. Science 204, 1052-1059.

Davidson, E. H., Thomas, T. L., Scheller, R. H. and Britten, R. J. (1982) The sea urchin actin genes, and a speculation on the evolutionary significance of small gene families. Pp 177-191 In: Genome Evolution, ed. by Dover, G. A. and Flavell, R. B. Academic Press.

Davidson, E. H. and Posakony, J. W. (1982) Repetitive sequence transcripts in development. Nature 297, 633.

Davidson, E. H., Jacobs, H. T. and Britten, R. J. (1983) Very short repeats and coordinate induction of genes. Nature 301, 468-470.

Davison, B. L., Egly, J. M., Mulvihill, E. R. and Chambon, P. (1983) Formation of stable preinitiation complexes between eukaryotic class beta-transcription factors and promoter sequences. Nature 301, 680-686.

Denhart, D. T. (1966) A membrane-filter technique for the detection of complementary DNA. B.B.R.C. 23, 641-646.

Dermann, E., Krauter, K., Walling, L., Weinberger, C., Ray, M. and Darnell, J. E. (1981) Transcriptional control in the production of liver-specific mRNAs. Cell 23, 731-739.

De Wachter, R. (1979) Do eukaryotic mRNA 5' noncoding sequences base-pair with the 18S ribosomal RNA 3' terminus? Nucl. Acid. Res. 7, 2045-2054.

Di Berardino, M. A. and Hoffner, N. J. (1983) Gene reactivation in erythrocytes: nuclear transplantation in oocytes and eggs of Rana. Science 219, 862-864.

Dickerson, R. E. (1971) The structure of cytochrome C and the rates of molecular evolution. J. Mol. Evol. 1, 26-45.

Dierks, P., van Ooyen, A., Montei, N. and Weissman, C. (1981) DNA sequences preceeding the rabbit beta-globin gene are required for formation in mouse L cells of beta-globin RNA with the correct 5'-terminus. PNAS 78, 1411-1415.

Dierks, P., van Ooyen, A., Cochran, M. D., Dobkin, C., Reiser, J. and Weissman, C. (1983) Three regions upstream from the cap site are required for efficient and accurate transcription of the rabbit beta-globin gene in mouse 3T6 cells. Cell 32, 695-706.

Dobkin, C., Pergolizzi, R. G., Bahre, P. and Bank, A. (1983) Abnormal splice in a mutant human beta-globin gene not at the site of a mutation. PNAS 50, 1184-1188.

Dodemont, H. J., Soriano, P., Quax, W. J., Ramaekers, F., Lenstra, J. A., Groenen, M. A. M., Bernardi, G. and Bloemendal, H. (1982) The genes coding for the cytoskeletal proteins actin and vimentin in warm blooded vertebrates. The Embo J. 1, 167-171.

Donahue, T. F., Davis, R. S., Lucchini, G. and Fink, G. R. (1983) A short nucleotide sequence required for regulation of HIS4 by the general control system of yeast. Cell 32, 89-98.

Doolittle, W. F. (1978) Genes in pieces: were they ever together? Nature 272, 581-582.

Doolittle, W. F. (1982) Selfish DNA after fourteen months. In: Genome Evolution (ed. Dover, G. A. and Flavell, R. B.) Academic Press.

Doolittle, W. F. and Sapienza, C. (1980) Selfish genes, the phenotype paradigm and genome evolution. Nature 284, 601-603.

Dover, G. and Coen, E. (1981) Springcleaning ribosomal DNA: a model for multigene evolution? Nature 290, 731-732.

Durica, D. S., Schloss, J. A. and Crain, W. R. (1980) Organization of actin gene sequences in the sea urchin: molecular cloning of an intron-containing DNA sequence coding for a cytoplasmic actin. PNAS 77, 5683-5687.

Efstratiadis, A., Posakony, J. W., Maniatis, T., Lawn, R. M., O'Connell, C., Spritz, R. A., De Riel, J. K., Forget, B. G., Weissman, S. M., Slightom, J. L., Blechl, A. E., Smithies, O., Baralle, F. E., Shoulders, C. C. and Proudfoot, N. J. (1980) The structure and evolution of the human beta-globin gene family. Cell 21, 653-668.

Ehrlich, M. and Wang, R. Y.-H. (1981) Role of DNA methylation in expression? Science 212, 1350-1357.

Eickbush, T. H. and Kafatos, F. C. (1982) A walk in the chorion locus of Bombyx mori. Cell 29, 633-643.

Elder, J. T., Pan, J., Duncan, C. H. and Weissman, S. M. (1981) Transcriptional analysis of interspersed repetitive polymerase II transcriptional units in human DNA. Nucl. Acid. Res. 9, 1171-1189.

Elgin, S. C. R. (1981) DNAase I hypersensitive sites of chromatin. Cell 27, 413-415.

Elzinga, M. and Lu, R. C. (1976) In: Contractile systems in non-muscle tissues (Perry, S. V., Maryreth, A. and Adelstein, R. S. eds.) pp 29-37. North-Holland, Amsterdam.

Engel, J. N., Gunning, P. W. and Kedes, L. M. (1981) Isolation and characterisation of human actin genes. PNAS 78, 4674-4678.

Engel, J., Gunning, P. and Kedes, L. (1982) Human cytoplasmic actin proteins are encoded by a multigene family. Mol. and Cell. Biol. 2, 674-684.

Enquist, N., Tiemeier, D., Leder, P., Weisberg, R. and Sternberg, N. (1974) Safer derivatives of bacteriophage lambda-gt lambda-c for use in cloning recombinant DNA molecules. Nature 259, 596-598.

Farmer, S. R., Nan, K. M., Ben-Ze'ev, A. and Penman, S. (1983) Regulation of actin mRNA levels and translation responds to changes in cell configuration. Mol. and Cell. Biol. 3, 182-189.

Felber, B. K., Orkin, S. H. and Hamer, D. H. (1982) Abnormal RNA slicing causes one form of alpha-thalassaemia. Cell 29, 895-902.

Files, J. G., Carr, J. and Hirsch, D. (1983) The actin gene family of Caenorhabditis elegans. J. Mol. Biol. 164, 355-375.

Findly, R. C. and Pederson, T. (1981) Regulated transcription of the genes for actin and heat-shock proteins in cultured Drosophila cells. J. Cell. Biol. 88, 323-328.

Finnegan, D. J., Will, B. H., Bayew, A. A., Bowcock, A. M. and Brown, L. (1982) Transposable DNA sequences in eukaryotes. In: Genome Evolution (ed. by Dover, G. A. and Flavell, R. B.). Pp 29-40, Academic Press.

Firtel, R. A. (1981) Multigene families encoding actin and tubulin. Cell 24, 6-7.

Flavell, R. A. (1980) The transcription of eukaryotic genes. Nature 285, 356-357.

Flavell, A. J. and Ish-Harowicz, D. (1981) Extra chromosomal circular copies of the eukaryotic transposable element Copia in cultured Drosophila cells. Nature 292, 591-594.

Fornwald, J. A., Kuncio, G., Peng, I. and Ordahl, C. P. (1982) The complete nucleotide sequence of the chick alpha-actin gene and its evolutionary relationship to the actin gene family. Nucl. Acid. Res. 10, 3861-3876.

Fromm, M. and Berg, P. (1982) Deletion mapping of DNA required for SV40 early region promoter function in vivo. J. Mol. Appl. Genet. 1, 457-481.

Fyrberg, E. A. and Donady, J. J. (1979) Actin heterogeneity in primary embryonic culture cells from Drosophila melanogaster. Dev. Biol. 68, 487-502.

Fyrberg, E. A., Kindle, K. L. and Davidson, N. (1980) The actin genes of Drosophila: a dispersed multigene family. Cell 19, 365-378.

Fyrberg, E. A., Bond, B. J., Hershey, N. D., Mixter, K. S. and Davidson, N. (1981) The actin genes of Drosophila: protein coding regions are highly conserved but intron positions are not. Cell 24, 107-116.

Fyrberg, E. A., Mahaffey, J. W., Bond, B. J. and Davidson, N. (1983) Transcripts of the six Drosophila actin genes accumulate in a stage- and tissue-specific manner. Cell 33, 115-123.

Gaetjens, E. and Barany, M. (1966) N-acetylaspartic acid in G-actin. Biochim. Biophys. Acta 117, 176-183.

Gallwitz, D. and Seidel, R. (1980) Molecular cloning of the actin gene from the yeast Saccharomyces cerevisiae. Nucl. Acid. Res. 8, 1043-1059.

Gallwitz, D. and Sures, I. (1980) Structure of a split yeast gene: complete nucleotide sequence of the actin gene in Saccharomyces cerevisiae. PNAS 77, 2546-2550.

Gallwitz, D., Perrin, F. and Seidel, R. (1981) The actin gene in the yeast Saccharomyces cerevisiae - 5' and 3' end-mapping, flanking and putative regulatory sequences. Nucl. acid. Res. 9, 6339-6350.

Garel, A. and Axel, R. (1976) Selective digestion of transcriptionally active ovalbumin genes from oviduct nuclei. PNAS 73, 3966-3970.

Garel, J. P., Mandel, P., Chavancy, G and Daillie, J. (1970) Functional adaptation of tRNAs to fibroin biosynthesis in the silk gland of Bombyx mori. FEBS Letts. 7, 307-329.

Garel, A., Zolan, A. and Axel, R. (1977) Genes transcribed at diverse rates have a similar conformation in chromatin. PNAS 74, 4867-4871.

Garrels, J. I. and Gibson, W. (1976) Identification and characterization of multiple forms of actin. Cell 9, 793-805.

Gerber-Huber, S., May, F. E. B., Westley, B. R., Felber, B. K., Hosbach, H. A., Andres, A. C. and Ryffel, G. U. (1983) In contrast to other Xenopus genes the oestrogen-inducible vitellogenin genes are expressed when totally methylated. Cell 33, 43-51.

Gilbert, W. (1978) Why genes in pieces? Nature 271, 501.

Gillies, S. D., Morrison, S. L., Oi, V. T. and Tonegawa, S. (1983) A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. Cell 33, 717-728.

Goldman, B. D. and Knipe, D. M. (1972) Functions of cytoplasmic fibers in non muscle mobility. CSH Symp. Quart. Biol. 37, 523-534.

Gouy, M. and Gautier, C. (1982) Codon usage in bacteria: correlation with gene expressivity. Nucl. Acid. Res. 10, 7055-7074.

Grantham, R., Gautier, C. and Goug, M. (1980) Codon frequencies in 119 individual genes confirm consistant choices of degenerate bases according to genome type. Nucl. Acid Res. 8, 1893-1912.

Grantham, R., Gautier, C. and Goug, M. (1981) Codon catalog usage is a genome strategy modulated for gene expressivity. Nucl. Acid. Res. 9, r43-r74.

Grez, M., Land, H., Giesecke, K., Schutz, G., Jung, A. and Sippel, A. E. (1981) Multiple mRNAs are generated from the chicken lysozyme gene. Cell 25, 743.

Grosschedl, R. and Birnstiel, M. L. (1980a) Identification of regulatory sequences in the prelude sequences of an H2A histone gene by the study of specific deletion mutants in vivo. PNAS 77, 1432-1436.

Grosschedl, R. and Birnstiel, M. L. (1980b) Spacer DNA sequences upstream of the TATAAATA sequence are essential for promotion of H2A histone gene transcription in vivo. PNAS 77, 7102-7106.

Grosschedl, R., Wasylyk, B., Chambon, P. and Birnstiel, M. L. (1981) Point mutation in the TATA box curtails expression of sea urchin H2A histone genes in vivo. Nature 294, 178-180.

Grosschedl, R. and Birnstiel, M. L. (1982) Delimitation of far upstream sequences required for maximal in vivo transcription of the H2A histone gene. PNAS 79, 297-301.

Grosveld, G. C., de Boer, E., Shewmaker, C. K. and Flavell, R. A. (1982a) DNA sequences necessary for transcription of the rabbit beta-globin gene in vivo. Nature 295, 120-126.

Grosveld, G. C., Rosenthal, A. and Flavell, R. A. (1982b) Sequence requirements for the transcription of the rabbit beta-globin gene in vivo: the -80 region. Nucl. Acid. Res. 10, 4951-4971.

Groudine, M. and Weintraub, H. (1981) Activation of globin genes during chicken development. Cell 24, 393-401.

Groudine, M. and Weintraub, H. (1982) Propagation of globin DNAase I-hypersensitive sites in the absence of factors required for induction: a possible mechanism for determination. Cell 30, 131-139.

Groudine, M., Holtzer, H., Scherrer, K. and Therroath, A. (1974) Lineage dependent transcription of globin genes. Cell 3, 243-248.

Groudine, M., Eisenmann, R. and Weintraub, H. (1981) Chromatin structure of endogenous retroviral genes and activation by an inhibitor of DNA methylation. Nature 292, 311-317.

Gruenbaum, Y., Naveh-Many, T., Cedar, H. and Razin, A. (1981) Sequence specificity of methylation in higher plant DNA. Nature 292, 860-862.

Gunning, P., Ponte, P., Okayama, H., Engel, J., Blau, H. and Kedes, L. (1983) Isolation and characterization of full-length cDNA clones for human alpha-, beta-, and gamma-actin mRNAs: skeletal but not cytoplasmic actins have an amino-terminal cysteine that is subsequently removed. Mol. and Cell. Biol. 3, 787-795.

Gurdon, J. B. (1974) The control of gene expression during development. Clarendon Press, Oxford.

Gurdon, J. B., Woodland, H. R. and Lingrel, J. B. (1974) The translation of mammalian globin mRNA injected into fertized eggs of Xenopus laevis. I: Message stability in development. Dev. Biol. 39, 125-133.

Guyette, W. A., Matuski, R. J. and Rosen, J. M. (1979) Prolactin-mediated transcriptional and post-transcriptional control of casein gene expression. Cell 17, 1013-1023.

Hagenbuchle, O., Santer, M., Steitz, J. A. and Mans, R. J. (1978) Conservation of the primary structure at the 3' end of the 18S rRNA from eukaryotic cells. Cell 13, 551-563.

Hall, Z. W., Lubit, B. W. and Schwartz, J. H. (1981) Cytoplasmic actin in postsynaptic structures at the neuromuscular junction. J. Cell Biol. 90, 789-792.

Hamada, H., Perrino, M. G., Kakunaga, T. (1982) Molecular structure and evolutionary origin of human cardiac muscle actin gene. PNAS 79, 5901-5905.

Hanauer, A., Levin, M., Heilig, R., Daegelen, D., Kalm, A. and Mandel, J. L. (1983) Isolation and characterization of cDNA clones for human skeletal muscle alpha-actin. Nucl. Acid. Res. 11, 3503-3516.

Harpold, M. M., Evans, R. M., Salditt,-Georgiff, M. and Darnell, J. E. (1979) Cell 17, 1025-1035.

Harpold, M., Wilson, M. and Darnell, J. E. (1981) Mol. and Cell. Biol. 1, 188-198.

Hastings, K. E. M. and Emerson, C. P. (1982) cDNA clone analysis of six co-regulated mRNAs encoding skeletal muscle contractile proteins. PNAS 79, 1553-1557.

Heath, J. P. and Dunn, G. A. (1978) Cell to substratum contacts of chicken fibroblasts and their relation to the microfilament system. A correlated interference-reflection and high-voltage electron microscopic study. J. Cell. Sci. 29, 197-212.

Heilig, R., Perrin, F., Gannon, F., Mandel, J. L. and Chambon, P. (1980) The ovalbumin gene family: structure of the X gene and evolution of duplicated split genes. Cell 20, 625-637.

Hentschel, C. C. and Birnstiel, M. L. (1981) The organization and expression of histone gene families. Cell 25, 301-313.

Hentschel, C., Probst, E. and Birnstiel, M. L. (1980) Transcriptional fidelity of histone genes injected into Xenopus oocyte nuclei. Nature 288, 100-102.

Hofer, E. and Darnell, J. E. (1981) The primary transcription unit of the mouse beta-major globin gene. Cell 23, 585-593.

Hoffman-Lieberman, B and Sachs, L. (1978) Regulation of actin and other proteins in the differentiation of myeloid leukaemic cells. Cell 14, 825-834.

Hollis, G. F., Hieter, P. A., McBride, O. W., Swan, D. and Leder, P. (1982) Processed genes: a dispersed human immunoglobulin gene bearing evidence of RNA-type processing. Nature 296, 321-315.

Horovitch, S. J., Storti, R. V., Rich, A. and Pardue, M. L. (1979) Multiple actins in Drosophila melanogaster. J. Cell Biol. 82, 86-92.

Hughes, S. M., Voght, P. K., Stubblefield, E., Robinson, H., Bishop, J. M. and Varmus, H. E. (1979) Organization of endogenous and exogenous viral and linked non-viral sequences. CSHSQB 44, 1077-1089.

Humphries, S. E., Whittall, R., Minty, A., Buckingham, M. E. and Williamson, R. (1981) There are approximately 20 actin genes in the human gene. Nucl. Acid. Res. 9, 4895-4908.

Huxley, H. E. (1969) The mechanism of muscular contraction. Science 164, 1356-1366.

Hyldig-Nielsen, J. J., Jensen, E. O., Paludon, K., Wiborg, O., Garrett, R., Jorgensen, P. and Marcker, K. A. (1982) The primary structure of two leghaemoglobin genes from soybean. Nucl. Acid. Res. 10, 689-701,

Igo-Kemenes, T., Horz, W. and Zachau, H. G. (1982) Chromatin. Ann. Rev. Biochem. 51, 89-121.

Ikemura, T. (1981a) Correlation between the abundance of E. coli transfer RNAs and the occurrence of the respective codons in its protein genes. J. Mol. Biol. 146, 1-21.

Ikemura, T. (1981b) Correlation between the abundance of E. coli transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the E. coli translation system. J. Mol. Biol. 151, 389-409.

Ikemura, T. (1982) Correlation between the abundance of yeast transfer RNAs and the occurrence of the respective codons in protein genes. J. Mol. Biol. 158, 573-597.

Jacq, C., Miller, J. R. and Brownlee, G. G. (1977) A pseudogene structure in 5S DNA of Xenopus laevis. Cell 12, 109-120.

Jacob, F. and Monod, J. (1961) Genetic regulatory mechanisms in the synthesis of proteins. J. Mol. Biol. 3, 318-356.

Jantzen, H. (1981) Control of actin synthesis during the development of Acanthamoeba castellanii. Dev. Biol. 82, 113-126.

Jeffreys, A. J. (1981) Recent studies of gene evolution using recombinant DNA. In: Genetic Engineering (ed. by Williamson, R.), vol. 2, Academic Press, London, New York.

Jeffreys, A. J. (1982) Evolution of globin genes. In: Genome Evolution (ed. by Dover, G. A. and Flavell, R. B.) p. 157, Academic Press.

Jeffreys, A. J. and Harris, S. (1982) Processes of gene duplication. Nature 296, 9-10.

Jelinek, W. R. and Schmid, C. w. (1982) Repetitive sequences in eukaryotic DNA and their expression. Ann. Rev. Biochem. 51, 813-844.

Jelinek, W. R., Toomey, t. P., Leinwand, L., Duncan, C. H., Biro, P. A., Choudary, P. W., Weissman, S. M., Rubin, C. M., Hozeck, C. M., Deininger, P. L. and Schmid, C. W. (1980) Ubiquitous interspersed repeated sequences in mammalian genomes. PNAS 77, 1398-1402.

Jones, C. W. and Kafatos, F. C. (1980) Coordinately expressed members of the two chorion multigene families are clustered, alternatively and divergently oriented. Nature 284, 635-658.

Jukes, T. H. (1980) Silent nucleotide substitutions and the molecular evolutionary clock. Science 210, 973-978.

Kafatos, F. C., Efstratiadis, A., Forget, B. G. and Weissman, S. M. (1977) Molecular evolution of human and rabbit beta-globin mRNAs. PNAS 74, 5618-5622.

Kaine, B. P. and Spear, B. B. (1980) Putative actin genes in the macronucleus of Oxytricha fallax. PNAS 77, 5336-5340.

Kaine, B. P. and Spear, B. B. (1982) Nucleotide sequence of a macronuclear gene for actin in Oxytricha fallax. Nature 295, 430-432.

Kakiuchi, S. and Sobue, K. (1983) Control of the cytoskeleton by calmodulin and calmodulin-binding proteins. TIBS 8, No. 2 59-62.

Karin, M. and Richards, R. I. (1982) Human metallothionein genes - primary structure of the metallothionein-II gene and a related processed gene. Nature 299, 797-802.

Kedes, L. H. (1979) Histone genes and histone messengers. Ann. Rev. Biochem. 48, 837-870.

Keller, L. R. and Emerson, C. P. (1980) Synthesis of adult myosin light chains by embryonic muscle cultures. PNAS 77, 1020-1024.

Kimura, M. (1981) Estimation of evolutionary distances between homologous nucleotide sequences. PNAS 78, 454-458.

Kindle, K. L. and Firtel, R. A. (1978) Identification and analysis of Dictyostelium actin genes, a family of moderately repeated genes. Cell 15, 763-778.

Khoury, G., Gruss, P., Dhar, R. and Lai, C. (1979) Processing and expression of early SV40 mRNA. Cell 18, 85-92.

Khoury, G. and Gruss, P. (1983) Enhancer elements. Cell 33, 313-314.

Kimmel, A. R. and Firtel, R. A. (1979) A family of short, interspersed repeat sequences at the 5' end of a set of Dictyostelium single-copy mRNAs. Cell 16, 787-796.

Kondoh, H., Yasuda, K. and Okada, T. S. (1983) Tissue-specific expression of a cloned chick delta-crystallin gene in mouse cells. Nature 301, 440-442.

Konkel, D. A., Maizel, J. V. and Leder, P. (1979) The evolution and sequence comparison of two recently diverged mouse chromosomal beta-globin genes. Cell 18, 865-873.

Korn, E. D. (1978) Biochemistry of actomyosin-dependent cell mobility. PNAS 75, 588-599.

Kozak, M. (1978) How do eukaryotic ribosomes select initiation regions in messenger RNA. Cell 15, 1109-1123.

Kozak, M. (1980) Evaluation of the 'scanning model' for initiation of protein synthesis in eukaryotes. Cell 22, 7-8.

Kozak, M. (1981) Mechanism of mRNA recognition by eukaryotic ribosomes during initiation of protein synthesis. Curr. Topics Microbiol. Immunol. 93, 81-123.

Kozak, M. (1982) Recognition of initiation sites in eukaryotic messenger RNAs. In: Protein Biosynthesis in Eukaryotes (ed. Perez-Bercoff), pp 167-189, Plenum Press.

Kozak, M. (1983) Comparison of initiation of protein synthesis in prokaryotes, eukaryotes, and organelles. Microbiol. Rev. 47, 1-45.

Kuebbing, D. and Liarakos, C. (1978) Nucleotide sequence at the 5' end of ovalbumin mRNA from chicken. Nucl. Acid. Res. 5, 2253-2266.

Kuhne, T., Wieringa, B., Reiser, J. and Weissman, C. (1983) Evidence against a scanning model of RNA splicing. EMBO J. 2, 727.

Laemlli, U. K. (1970) Cleavage of structural proteins during assembly of the head of bacteriophage T4. Nature 227, 680-685.

Lam, B. S. and Carroll, D. (1983) Tandemly repeated DNA sequences from Xenopus laevis. II. Dispersed clusters of a 388 base-pair repeating unit. J. Mol. Biol. 165, 587-597.

Longford, C. J. and Gallwitz, D. (1983) Evidence for an intron-contained sequence required for the splicing of yeast RNA polymerase II transcripts. Cell 33, 519-527.

Laskey, R. A. and Mills, A. D. (1975) Quantitative film detection of tritium and carbon-14 in polyacrylamide gels by fluorography. Eur. J. Biochem. 56, 335-341.

Lazarides, E. (1976) Actin, alpha-actinin and tropomyosin interaction in the structural organization of actin filaments in non muscle cells. J. Cell. Biol. 68, 202-219.

Lazarides, E. and Weber, K. (1974) Actin antibody: the specific visualization of actin filaments in non-muscle cells. PNAS 71, 2268-2272.

Lee, M. G.-S., Lewis, S. A., Wilde, C. D. and Cowan, W. J. (1983) Evolutionary history of a multigene family: an expressed human beta-tubulin gene and three processed pseudogenes. Cell 33, 477-487.

Leigh Brown, A. J. and Ish-Horowicz (1981) Evolution of the 87A and 87C heat-shock loci in Drosophila. Nature 290, 677-682.

Lemischka, I. R., Farmer, S., Racaniello, V. R. and Sharp, P. (1981) Nucleotide sequence and evolution of a mammalian alpha-tubulin mRNA. J. Mol. Biol. 151, 101-120.

Lemischka, I. and Sharp, P. A. (1982) The sequences of an expressed rat alpha-tubulin gene and pseudogene with an inserted repetitive element. Nature 300, 330-335.

Lerner, M. R., Boyle, J. A., Mount, S. M., Wolin, S. L. and Steitz, J. A. Are snRNPs involved in splicing? Nature 283, 220-224.

Lewin, B. (1974) Gene Expression vol 1: Bacterial genomes. Wiley-Interscience, New York.

Lewin, B. (1980b) Gene Expression2- 2nd edition. p 965, Wiley-Interscience, New York.

Lewin, B. (1980c) Gene Expression2- 2nd edition. p 529, Wiley-Interscience, New York.

Lewin, B. (1980d) Heterogeneous nuclear RNA. In: Gene Expression2- 2nd edition. pp 746-755, Wiley-Interscience, New York.

Lewis, R. A., Wakinoto, B. T., Denell, R. E., and Kaufman, T. C. (1980) Genetic analysis of the antennapedia gene complex (ant-c) and adjacent chromosomal regions of Drosophila melanogaster. II. Polytene chromosome segments 84A-84B1,2. Genetics 95, 383-397.

Little, P. F. R. (1982) Globin pseudogenes. Cell 28, 683-684.

Lodish, H. F. and Small, B. (1976) Different lifetimes of reticulocyte messenger RNA. Cell 7, 59-65.

Lodish, H. F. (1974) Model for the regulation of mRNA translation applied to haemoglobin synthesis. Nature 251, 385-388.

Lomedico, P., Rosenthal, N., Efstratiadis, A., Gilbert, W., Kolodner, R. and Tizard, R. (1979) The structure and evolution of the two nonallelic rat preproinsulin genes. Cell 18, 545-558.

Long, E. O. and Dawid, I. B. (1980) Repeated genes in eukaryotes. Ann. Rev. Biochem. 49, 727-764.

Lowenhaupt, K. and Lingrel, J. B. (1978) A change in the stability of globin mRNA during the induction of murine erythroleukemia cells (Friend cells). Cell 14, 337-344.

Lubit, B. W. and Schwartz, J. H. (1980) An anti-actin that distinguishes between cytoplasmic and skeletal muscle actins. J. Cell. Biol. 86, 891-897.

Luse, D. S., Haynes, J. R., Van Leeuwen, D., Schon, E. A., Cleary, M. L., Shapiro, S. G., Lingrel, J. B. and Roeder, R. G. (1981) Transcription of the beta-like globin genes and pseudogenes of the goat in a cell-free system. Nucl. Acid. Res. 9, 4339-4354.

Maden, B. E. H., Hall, L, M. and Salim, M. (1982) Ribosome formation in the eukaryotic nucleolus: recent advances from sequence analysis. In: The

Nucleolus (ed. Jordan, E. G. and Cullis, C. A.) pp 87-112, Cambridge University Press.

Mandel, M. and Hiya, A. (1970) Ca-dependent bacteriophage DNA infection. J. Mol. Biol. 53, 159-162.

Mandel, J. L. and Chambon, P. (1979) DNA methylation: organ-specific variations in the methylation pattern within and around ovalbumin and other chicken genes. Nucl. Acid. Res. 7, 2081-2103.

Maniatis, T., Jeffrey, A. and Van de Sande, H. (1975) Chain length determination of small double and single-stranded DNA molecules by polyacrylamide gel electrophoresis. Eur. J. Biochem. 14, 3787-3794.

Mantei, N. and Weissman, C. (1982) Controlled transcription of a human alpha-interferon gene introduced into mouse L-cells. Nature 297, 128-132.

Margiarotti, G., Lefebre, P. and Lodish, H. F. (1982) Differences in the stability of developmentally regulated mRNAs in aggregated and disaggregated Dictyostelium discoideum cells. Dev. Biol. 89, 82-91.

Marotta, C. A., Strocchi, P. and Gilbert, J. M. Microheterogeneity of brain cytoplasmic and synaptoplasmic actins. J. Neurochem. 30, 1441-1451.

Martin, G. R., Smith, S. and Epstein, C. J. (1978) Protein synthetic patterns in teratocarcinoma stem cells and mouse embryos at early stages of development. Dev. Biol. 66, 8-16.

Marx, J. L. (1981) Genes that control development. Science 213, 1485-1488.

Maxam, A. M. and Gilbert, W. (1978) Sequencing end-labelled DNA with base-specific chemical cleavages. Meth. in Enz. 65, 499-559.

Mayo, K. E., Warren, R. and Palmiter, R. D. (1982) The mouse metallothionein-I gene is regulated by cadmium following transfection into human or mouse cells. Cell 29, 99-108.

McClelland, M. and Ivarie, R. (1982) Asymmetrical distribution of CpG in an 'average' mammalian gene. Nucl. Acid Res. 10, 7865-7877.

McGhee, J. V., Wood, W. I., Dolan, M., Engel, J. D. and Felsenfeld, G. (1981) A 200 base pair region at the 5' end of the chicken adult beta-globin gene is accessible to nuclease digestion. Cell 27, 45-55.

McGhee, J. D. and Ginder, G. D. (1979) Specific DNA methylation sites in the vicinity of the chicken beta-globin genes. Nature 280, 419-420.

McKeown, M., Taylor, W. C., Kindle, K. L., Firtel, R. A., Bender, W. and Davidson, N. (1978) Multiple, heterogeneous actin genes in Dictyostelium. Cell 15, 789-800.

McKeown, M. and Firtel, R. A. (1981a) Differential expression and 5' end-mapping of actin genesin Dictyostelium. Cell 24, 799-807.

McKeown, M. and Firtel, R. A. (1981b) Evidence for sub-families of actin genes in Dictyostelium as determined by comparison of 3' end sequences. J. Mol. Biol. 151, 593-606.

McKeown, M. and Firtel, R. A. (1982) Actin multigenic family of dictyostelium. In: The Organization of the Cytoplasm. CSHSQB 46, 495-506.

McKnight, S. L. and Kingsbury, R. (1982) Transcriptional control signals of a eukaryotic protein-coding gene. Science 217, 316-324.

Melton, P., Parker, V., Gluzman, Y. and Maniatis, T. (1981) Identification of DNA sequences required for transcription of the human alpha 1-globin gene in a new SV40 host-vector system. Cell 27, 279-288.

Messing, J. (1981) In: Proceedings of the Third Cleveland Symposium on Macromolecules (ed. Walton, A. G.) p 143. Elsevier, Amsterdam.

Messing, J. B., Crea, R. and Seeburg, P. H. (1981) A system for shotgun DNA sequencing. Nucl. Acid Res. 9, 309-321.

Messing, J. and Vieira, J. (1982) A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. Gene 19, 269-272.

Miller, J. H. (1978) In: The Operon (eds. Miller, J. H. and Resznikoff, W. S.) Cold Spring Harbor Laboratory, Cold Spring Harbor, New York) p 66.

Minty, A. J., Alonso, S., Caravatti, M. and Buckingham, M. E. (1982) A fetal skeletal muscle actin mRNA in the mouse and its identity with cardiac mRNA. Cell 30, 185-192.

Minty, A. J., Alonso, S., Guenet, J. L. and Buckingham, M. E. (1983) Number and organization of actin-related sequences in the mouse genome. J. Mol. Biol. 167, 77-101.

Mirault, M. E., Goldschmidt-Clermont, M., Artavanis-Tsakonas, S. and Sched, P. (1979) Organization of the multiple genes for the 70K dalton heat-shock protein in Drosophila melanogaster. PNAS 76, 5254-5258.

Miyata, T., Yasunaga, T. and Nishida, T. (1980) Nucleotide sequence divergence and functional constraint in mRNA evolution. PNAS 77, 7328-7332.

Montarras, D. and Fiszman, M. Y. (1982) A new muscle phenotype is expressed by subcultured quail myoblasts from future fast and slow muscles. Submitted to J. Biol. Chem.

Moore, G. W., Goodman, M., Callahan, C., Holmquist, R. and Moise, H. (1976) Stochastic versus augmented maximum parsimony method for estimating superimposed mutations in the divergent evolution of protein sequences. J. Mol. Biol. 105, 15-37.

Moos, M. and Gallwitz, D. (1983) Structure of two human beta-actin-related processed genes, one of which is located next to a simple repetitive sequence. The EMBO J. 2, 757-761.

Moreau, P. Hen, R., Wasylyk, R., Everett, R., Gaub, M. P. and Chambon, P. (1981) The SV40 72 base pair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. Nucl. Acid. Res. 9, 6047-6068.

Mount, S. and Steitz, J. (1983) RNA processing: lessons from mutant globins. Nature 303, 380-381.

Mount, S. M., Pettersson, I., Hinterberger, M., Karmas, A. and Steitz, J. A. (1983) The U1 small nuclear RNA-protein complex selectively binds a 5' splice site in vitro. Cell 33, 509-518.

Murray, N. E. and Murray, K. (1974) Manipulation of restriction targets in phage lambda to form receptor chromosomes for DNA fragments. Nature 251, 476-481.

Nagata, S., Mantei, N., Weissman, C. (1980) The structure of one of the 8 or more distinct chromosomal genes for human interferon alpha. Nature 287, 401-408.

Nellen, W. and Gallwitz, D. (1982) Actin genes and actin mRNA in Acanthamoeba castellanii. Nucleotide sequence of the split actin gene I. J. Mol. Biol. 159, 1-18.

Ng, R. and Abelson, J. (1980) Isolation and sequence of the gene for actin in Saccharomyces cerevisiae. PNAS 77, 3912-3916.

North, G. (1983) Cloning the genes that specify fruit flies. Nature 303, 134-136.

Nudel, U., Zakut, R., Shani, M., Neuman, S., Levy, Z. and Yaffe, D. (1983) The nucleotide sequence of the rat cytoplasmic beta-actin gene. Nucl. Acid. Res. 11, 1759-1771.

O'Farrell, P. H., Kutter, E. and Nakanishi, M. (1980) A restriction map of the bacteriophage T4 genome. Molec. Gen. Genet. 179, 421-435.

Ohno, S. and Taniguchi, T. (1982) Inducer-responsive expression of the cloned human interferon beta-1 gene introduced into cultured mouse cells. Nucl. Acid Res. 10, 967-977.

Ordahl, C. P. and Cooper, T. A. (1983) Strong homology in promoter and 3'-untranslated regions of chick and rat alpha-actin genes. Nature 303, 348-349.

Orgel, L. and Crick, F. H. C. (1980) Selfish DNA: the ultimate parasite. Nature 204, 604-607.

Osbourne, T.-F., Gaynor, R. B. and Berk, A. J. (1982) The TATA homology and the mRNA 5' untranslated sequence are not required for expression of essential adenovirus E1A functions. Cell 29, 139-148.

Overbeek, P. A., Merlino, G. T., Peters, N. K., Cohn, V. H., Moore, G. P. and Kleinsmith, L. J. (1981) Characterization of five members of the actin gene family in the sea urchin. BBA 656, 195-205.

Page, G. S., Smith, S. and Goodman, H. M. (1981) DNA sequence of the rat growth hormone gene: location of the 5' terminus of the growth

hormone mRNA and identification of an internal transposon-like element. Nucl. Acid Res. 9, 2087-2104.

Palmiter, R. D. (1975) Quantitation of parameters that determine the rate of ovalbumin synthesis. Cell 4, 189-197.

Palmiter, R. D., Chen, H. Y. and Brinster, R. L. (1982) Differential regulation of metalothionein-thymidine kinase fusion genes in transgenic mice and their offspring. Cell 29, 701-710.

Panet, A. and Cedar, H. (1977) Selective degradation of integrated murine leukemia proviral DNA by deoxyribonucleases. Cell 11, 933-940.

Pardo, J. V., Pittenger, M. F. and Craig, S. W. (1983) Subcellular sorting of isoactins: selective association of gamma-actin with skeletal muscle mitochrondria. Cell 32, 1093-1103.

Paterson, B. M., Roberts, B. E. and Yaffe, D. (1974) Determination of actin messenger RNA in cultures of differentiating embryonic chick skeletal muscle. PNAS 71, 4467-4471.

Patient, R. K., Elkington, J. A., Kay, R. M. and Williams, J. G. (1980) Internal organization of the major adult alpha- and beta-globin genes of X. laevis. Cell 21, 565-573.

Pearson, W. R., Mukai, T. and Morrow, J. F. (1981) Repeated DNA sequences ner the 5'-end of the silk fibroin gene. J. Biol. Chem. 256, 4033-4041.

Pelham, H. R. B. (1982) A regulatory upstream promoter element in Drosophila Hsp 70 heat-shock gene. Cell 30, 517-528.

Pelham, H. R. B. and Jackson, R. J. (1976) An efficient mRNA-dependent translation system from reticulocyte lysates. Eur. J. Biochem. 67, 247-256.

Perler, F., Efstratiadis, A., Lomedico, P., Gilbert, W., Kolodner, R. and Dodgson, J. (1980) The evolution of genes: the chicken preproinsulin gene. Cell 20, 555-566.

Pitha, P. M., Ciufo, D. M., Kellum, M., Raj, N. B. K., Reyes, G. R. and Hayward, G. S. (1982) Induction of human beta-interferon synthesis with poly (rI-rC) in mouse cells transfected with cloned cDNA plasmids. PNAS 29, 4337-4341.

Pollack, Y., Stein, R., Razin, A. and Cedar, H. (1980) Methylation of foreign DNA sequences in eukaryotic cells. PNAS 77, 6463-6467.

Pollard, T. D. and Craig, S. W. (1982) Mechanism of actin polymerization. TIBS 7, No. 2, 55-58.

Pollard, T. D. and Weihing, R. R. (1974) Actin and myosin in cell movement. CRC Crit. Rev. Biochem. 2, 1-65.

Prentki, P., Karch, F., Iida, S., Meyer, J. (1981) The plasmid cloning vector pBR325 contains a 482 base-pair long inverted duplication. Gene 14, 289-299.

Proudfoot, N. J., Shander, M. H. M., Manley, J. L., Gefter, M. L. and Maniatis, T. (1980) Structure and *in vitro* transcription of human globin genes. Science 209, 1329-1336.

Putney, S. D., Herlihy, W. C. and Schimmel, P. (1983) A new troponin T and cDNA clones for 13 different muscle proteins, found by shot-gun sequencing. Nature 302, 718-723.

Proudfoot, N. (1980) Pseudogenes. Nature 286, 840-841.

Razin, A. and Riggs, A. D. (1980) Role of DNA methylation in expression? Science 210, 604-610.

Ricciardi, R. P., Miller, J. S. and Roberts, B. E. (1979) Purification and mapping of specific mRNAs by hybridization-selection and cell-free translation. PNAS 76, 4927-4931.

Richardson, C., Copello, J., Cochran, M. P., Armentrout, R. W. and Brown, R. O. (1980) Partial sequence analysis of *Xenopus* alpha- and beta-globin mRNA as determined from recombinant DNA plasmids. Dev. Biol. 78, 161-172.

Riddle, V. G. H., Dubrow, R. and Pardec, A. B. (1979) Changes in the synthesis of actin and other cell proteins after stimulation of serum-arrested cells. PNAS 76, 1298-1302.

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. Mol. Biol. 113, 237-251.

Robert, B., Weydert, A., Caravatti, M., Minty, A., Cohen, A., Daubas, P., Gros, F. and Buckingham, M. (1982) A cDNA recombinant plasmid complementary to mRNAs for both light chains 1 and 3 of mouse skeletal muscle myosin PNAS 79, 2437-2441.

Rogers, J. and Wall, R. (1980) A mechanism for RNA splicing. PNAS 77, 1877-1879.

Romero-Herrera, A. E., Lehmann, H., Josey, K. A. and Friday, A. E. (1973) Molecular evolution of myoglobin and the fossil record: a phylogenetic synthesis. Nature 246, 389-395.

Rosenthal, E., Hunt, T. and Ruderman, J. V. (1980) Selective translation of mRNA controls the pattern of protein syntehsis during early development of the surf clam Spisula solidissima. Cell 20, 487-496.

Royal, A., Garapin, A., Caru, B., Pernin, F., Mandel, J. L. LeMeur, M., Bregegere, F., Gannon, F., Le Pennec, J. P., Chambon, P. and Kourilsky, P. (1979) The ovalbumin gene region: common features in the organisation of three genes expressed in the chick oviduct under hormonal control. Nature 279, 125-132.

Ryffel, G. U., Meullener, D .B., Wyler, T., Wahli, W. and Weber, R. (1981) Transcription of single-copy vitellogenin gene of Xenopus involves expression of middle-repetitive DNA. Nature 291, 429-431.

Saborio, J. L., Seguia, M., Flores, M., Garcia, R., and Palmer, E. (1979) Differential expression of gizzard actin genes during chicken embryogenesis. J. Biol. Chem. 254, 11,119-11,125.

Salditt-Georgieff, M., Harpold, M., Wilson, M. and Darnell, J. E. (1981) Molec. Cell. Biol. 1, 179-187.

Sanchez, F., Tobin, S. L., Rdest, U., Zulauf, E. and McCarthy, B. J. (1983) Two Drosophila actin genes in detail: gene structure, protein structure and transcription during development. J. Mol. Biol. 163, 533-551.

Sanger, F., Nicklen, S. and Coulson, A. R. (1977) DNA sequencing with chain terminating inhibitors. PNAS 74, 5463-5467.

Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H. and Roe, B. A. (1980) Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. J. Mol. Biol. 143, 161-178.

Sargent, T. D., Jagodzinski, L. L., Yang, M. and Bonner, J. (1981) Fine structure and evolution of the rat serum albumin gene. Mol. Cell. Biol. 1, 871-883.

Schedl, T. and Dove, W. F. (1982) Mendelian analysis of the organization of actin sequences in Physarum polycephalum. J. Mol. Biol. 160, 41-57.

Scheller, R. H., Constantini, F. D., Kozlowski, M. R., Britten, R. J. and Davidson, E. H. (1978) Specific representation of cloned repetitive

DNA sequences in sea urchin RNAs. Cell 15, 189-203.

Scheller, R. H., McAllister, L. B., Crain, W. R., Durica, D. S., Posakony, J. W., Thomas, T. L., Britten, R. J. and Davidson, E. H. (1981) Organization and expression of multiple actin genes in the sea urchin. Mol. Cell. Biol. 1, 609-628.

Schibler, U., Pittet, A.-C., Young, R. A., Hagenbuchle, O., Tosi, M., Gellman, S. and Wellauer, P. K. (1982) The mouse alpha-amylase multigene family. Sequence organization of members expressed in the pancreas, salivary gland and liver. J. Mol. Biol. 155, 247-267.

Schimke, R. T. and Doyle, D. (1970) Control of enzyme levels in animal tissues. Ann. Rev. Biochem. 39, 929-976.

Schuler, M. A. and Keller, E. B. (1981) The chromosomal arrangement of two linked actin genes in the sea urchin S. purpuratus. Nucl. Acid. Res. 9, 591-604.

Schuler, M. A., McOsker, P. and Keller, E. B. (1983) DNA sequence of two linked actin genes of the sea urchin S. purpuratus. Mol. and Cell. Biol. 3, 448-456.

Schwartz, R. and Rothblum, K. (1980) Regulation of muscle differentiation: isolation and purification of chick actin messenger ribonucleic acid and quantitation with cDNA probes. Biochemistry 19, 2506-2514.

Schwartz, R. and Rothblum, K. (1981) Gene switching in myogenesis: differential expression of the chicken actin multigene family. Biochemistry 20, 4122-4129.

Shah, D. M., Hightower, R. C. and Meagher, R. B. (1982) Complete nucleotide sequence of a soybean actin gene. PNAS 79, 1022-1026.

Shani, M., Zevin-Sonkin, D., Saxel, O., Carmon, Y., Katcoff, D., Nudel, U. and Yaffe, D. (1981) 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.

Sharp, P. A., Sugden, B. and Sambrook, J. (1973) Detection of two restriction endonuclease activities in Haemophilus parainfluenzae using analytical agarose-ethidium bromide electrophoresis. Bioch. J. 12, 3055-3063.

Sharp, P. A. (1983) Conversion of RNA to DNA in mammals: Alu-like elements and pseudogenes. Nature 301, 471-472.

Shatkin, A. J. (1976) Capping of eucaryotic mRNAs. Cell 9, 645-653.

Shen, C. K. J. and Maniatis, T. (1980a) The organization of repetitive sequences in a cluster of rabbit beta-like globin genes. Cell 19, 379-391.

Shen, C. K. J. and Maniatis, T. (1980b) Tissue-specific DNA methylation in a cluster of rabbit beta-like globin genes. PNAS 77, 6634-6638.

Schwartz, R. and Rothblum, K. (1981) Gene switching in myogenesis: differential expression of the chicken actin multigene family. Biochemistry 20, 4122-4129.

Shah, D. M., Hightower, R. C. and Meagher, R. B. (1982) Complete nucleotide sequence of a soybean actin gene. PNAS 79, 1022-1026.

Shani, M., Zevin-Sonkin, D., Saxel, O., Carmon, Y., Katcoff, D., Nudel, U. and Yaffe, D. (1981) 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.

Sharp, P. A., Sugden, B. and Sambrook, J. (1973) Detection of two restriction endonuclease activities in Haemophilus parainfluenzae using analytical agarose-ethidium bromide electrophoresis. Bioch. J. 12, 3055-3063.

Sharp, P. A. (1983) Conversion of RNA to DNA in mammals: Alu-like elements and pseudogenes. Nature 301, 471-472.

Shatkin, A. J. (1976) Capping of eucaryotic mRNAs. Cell 9, 645-653.

Shen, C. K. J. and Maniatis, T. (1980a) The organization of repetitive sequences in a cluster of rabbit beta-like globin genes. Cell 19, 379-391.

Shen, C. K. J. and Maniatis, T. (1980b) Tissue-specific DNA methylation in a cluster of rabbit beta-like globin genes. PNAS 77, 6634-6638.

Shiba, T. and Saigo, K. (1983) Retrovirus-like particles containing RNA homologous to the transposable element Copia in Drosophila melanogaster. Nature 302, 119-124.

Shine, J. and Dalgarno, L. (1974) The 3'-terminal sequence of E. coli 16S ribosomal RNA: complementary to nonsense triplets and ribosome binding sites. PNAS 71, 1342-1346.

Slightom, J. L., Blechl, A. E. and Smithies, O. (1980) Human fetal $^{G}\gamma$ - and $^{A}\gamma$ -globin genes: complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. Cell 21, 627-638.

Small, J. V., Isenberg, G. and Celis, J. E. (1978) Polarity of actin at the leading edge of cultured cells. Nature 272, 638-639.

Smith, D. W. E. and McNamara, A. L. (1971) Specialization of rabbit reticulocyte transfer RNA content for haemoglobin synthesis. Science 171, 577-579.

Smith, H. O. and Birnstiel, M. L. (1976) A simple method for DNA restriction site mapping. Nucl. Acid. Res. 3, 2387-2398.

Smith, S. S. and Thomas, C. A. Jr. (1981) The two dimensional restriction analysis of Drosophila DNAs: males and females. Gene 13, 395-408.

Soeiro, R., Vaughn, M. H., Warner, J. R. and Darnell, J. E. (1968) The turnover of nuclear DNA-like RNA in HeLa cells. J. Cell. Biol. 39, 112-

118.

Soriano, P., Szabo, P. and Bernardi, G. (1982) The scattered distribution of actin genes in the mouse and human genomes. EMBO J. 1, 579-583.

Southern, E. M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98, 503-517.

Southern, E. M. (1979) Measurement of DNA length by gel electrophoresis. Analyt. Bioch. 100, 319-323.

Spence, S. E., Pergolizzi, R. G., Donovan-Peluse, M., Kosche, K. A., Dobkin, C. S. and Bank, A. (1982) Five nucleotide changes in the large intervening sequence of a beta-globin gene in a beta-plus thalassemia patient. Nucl. Acid. Res. 10, 1283-1294.

Spiegelman, B. M. and Farmer, S. R. (1982) Decreases in tubulin and actin gene expression prior to morphological differentiation of 3T3 adipocytes. Cell 29, 53-60.

Spofford, J. B. 91976) Position-effect variegation in Drosophila. In: The Genetics and Biology of Drosophila (eds. Ashburner, M. and Novitski, E.), Vol Ic, pp 955-1009. Academic Press, London, New York and San Francisco.

Spohr, G., Reith, W. and Sures, I. (1981) Organization and sequence analysis of a cluster of repetitive DNA from Xenopus laevis. J. Mol. Biol. 151, 573-592.

Spohr, G., Reith, W. and Sures, I. (1982) Structural analysis of a cDNA clone from Xenopus laevis containing a repetitive sequence element. Dev. Biol. 94, 71-78.

Spradling, A. L. and Mahowald, A. F. (1980) Amplification of genes for chorion proteins during oogenesis in Drosophila melanogaster. PNAS 77, 1096-1100.

Spradling, A. C. (1982) Chorion gene amplification during the development of Drosophila folicle cells. In: Gene Amplification (ed. Schimke, R. T.), CSH, 1982.

Stalder, J., Larsen, A., Engel, J. D., Dolan, M., Groudine, M. and Weintraub, J. (1980) Tissue-specific DNA cleavages in the globin chromatin domain introduced by DNAase I. Cell 20, 451-460.

Stein, J. P., Catterall, J. F., Kristo, R., Means, A., R. and O'Malley, B. W. (1980). Ovomucoid intervening sequences specify functional domains and generate protein polymorphism. Cell 21, 681-687.

Steitz, J. A. and Jakes, K. (1975) How ribosomes select initiator regions in mRNA: base-pair formation between the 3' terminus of 16S rRNA and mRNA during initiation of protein synthesis in E. coli. PNAS 72, 4734-4738.

Stockdale, F. E., Saden, H. and Raman, N. (1981) Slow muscle myoblasts differentiating in vitro synthesize both slow and fast myosin light chains. Dev. Biol. 82, 168-171.

Storb, U., Wilson, R., Selsing, e. and Walfield, A. (1981) Rearranged and germline immunoglobulin K genes. Different states of constant K genes in immuno-competent and non-immune cells. Biochemistry 20, 990-996.

Storti, R. V., Horovitch, S. J., Scott, M. P., Rich, A. and Pardue, M. L. 91978) Myogenesis in primary cell cultures from Drosophila melanogaster: protein synthesis and actin heterogeneity during development. Cell 13, 589-598.

Sturgess, E. A., Ballantine, J. E. M., Woodland, H. R., Mohun, P. R., Lane, C. D., Dimitriadis, G. J. (1980) Actin synthesis during the early development of Xenopus laevis. J. Embryol. exp. Morph. vol. 58, 303-320.

Suck, D., Kabsch, W. and Mannlerz, H. G. (1981) Three-dimensional structure of the complex of skeletal muscle actin and bovine pancreatic DNAase I at 6A reduction. PNAS 78, 4319-4323.

Sutoh, K. (1982) Identification of myosin-binding sites on the actin sequence. Biochemistry 21, 3654-3661.

Sutter, D. and Doeffler, W. (1980) Methylation of integrated adenovirus in DNA sequences in transformed cells is inversely correlated with viral gene expression. PNAS 77, 253-256.

Theibaud, C. H. and Fischberg, M. (1977) DNA content in the genus Xenopus. Chromosoma 59, 253-257.

Thireos, G., Griffin-Shea, R. and Kafatos, F. C. (1982) Expression and amplification of chorion genes in Drosophila melanogaster. In: Gene Amplification (ed. Schimke, R. T.), CSH.

Tiemeier, D., Enquist, L. and Leder, P. (1976) Improved derivative of a phage lambda for cloning recombinant DNA. Nature 263, 526-527.

Tonegawa, S., Maxam, A. M., Tizard, A., Bernard, O. and Gilbert, W. (1978) Sequence of a mouse germline gene for a variable region of an immunoglobulin light chain. PNAS 75, 1485-1489.

Torok, I., Mason, P. J., Karch, F., Kiss, I. and Udvardy, A. (1982) Extensive regions of homology associated with heat-induced genes of loci 87A7 and 87C1 in Drosophila melanogaster. In: Heat Shock From Bacteria to Man (ed. Schlesinger, M. J., Ashburner, M. and Tissieres, A.), pp 19-26, CSH.

Treisman, R., Proudfoot, N. J., Shander, M. and Maniatis, T. (1982) A single base change at a splice site in a beta^{o}-thalassemic gene causes abnormal RNA splicing. Cell 29, 903-911.

Treisman, R., Orkin, S. H. and Maniatis, T. (1983) Specific transcription and RNA splicing defects in five cloned beta-thalassaemia genes. Nature 302, 591.

Tsang, A. S., Mahbubani, H. and Williams, J. G. (1982) Cell-type-specific actin mRNA populations in Dictyostelium discoideum. Cell 31, 375-382.

Turner, P. C. and Woodland, H. R. (1982) H3 and H4 cDNA sequences from Xenopus: a sequence comparison of H4 genes. Nucl. Acid. Res. 10, 3769-3780.

Turner, P. C. and Woodland, H. R. (1983) Histone gene number and organization in Xenopus: Xenopus borealis has a homogeneous major cluster. Nucl. Acid. Res. 11 971-986.

Twigg, A. J. and Sherrat, D. (1980) Trans-complementable copy-number mutants of plasmid ColEl. Nature 283, 216-218.

Tymowska, J. and Fischberg, M. (1973) Chromosome complements of the genus Xenopus. Chromosoma 44, 335-342.

Van Arsdell, S. W., Denison, R. A., Bernstein, L. B., Weiner, A. M., Manser, T. and Gesteland, R. F. (1981) Direct repeats flank three small nuclear RNA pseudogenes in the human genome. Cell 26, 11-17.

Van der Ploeg, L. H. T. and Flavell, R. A. (1980) DNA methylation in the human gamma, sigma, and beta globin locus in erythroid and non-erythroid tissues. Cell 19, 947-958.

Van der Ploeg, L. H. T., Konings, A., Oort, M., Roos, D., Bernini, L. and Flavell, R. A. 91980) Gamma-beta-thalassaemia studies showing that deletion of the gamma- and delta-genes influences beta-globin gene expression in man. Nature 283, 637-642.

Vandekerckhove, J. and Weber, K. (1978a) At least six different actins are expressed in a higher mammal: an analysis based on the amino acid sequence of the amino-terminal tryptic peptide. J. Mol. Biol. 126, 783-802.

Vandekerckhove, J. and Weber, K. (1978b) Actin amino acid sequences. Comparison of actins from calf thymus, bovine brain, and SV40-transformed mouse 3T3 cells with rabbit skeletal muscle actin. Eur. J. Biochem. 90, 451-462.

Vandekerckhove, J. and Weber, K. (1978c) Mammalian cytoplasmic actins are the products of at least two genes and differ in primary structure in at least 25 identified positions from skeletal muscle actins. PNAS 75, 1106-1110.

Vandekerckhove, J. and Weber, K. (1978d) The amino acid sequence of actin from chicken skeletal muscle actin and chick gizzard smooth muscle actin. FEBS Letts. 102, 219-222.

Vandekerckhove, J. and Weber, K. (1978e) The amino acid sequence of Physarum actin. Nature 276, 720-721.

Vandekerckhove, J. and Weber, K. (1979) The complete amino acid sequences of actins from bovine aorta, bovine heart, bovine fast skeletal muscle, and rabbit slow skeletal muscle. Differentiation 14, 123-133.

Vandekerckhove, J. and Weber, K. (1980a) Vegetative Dictyostelium cells containing 17 actin genes express a single major actin. Nature 284,

475-477.

Vandekerckhove, J., Leavitt, J., Kakunaga, T. and Weber, K. (1980b) Coexpression of a mutant beta-actin and the two normal beta- and gamma-actins in a stably transformed human cell line. Cell 22, 893-899.

Vandekerckhove, J. and Weber, K. (1981) Actin typing on total cellular extracts: a highly sensitive protein-chemical procedure able to distinguish different actins. Eur. J. Biochem. 113, 595-603.

Vandekerckhove, J., Franke, W. W. and Weber, K. (1981) Diversity of expression of non-muscle actin in Amphibia. J. Mol. Biol. 152, 413-426.

Van Ooyen, A., Van den Berg, J., Mantei, N. and Weissman, C. (1979) Comparison of total sequence of a cloned rabbit beta-globin gene and its flanking regions with a homologous mouse sequence. Science 206, 337-344.

Volloch, V. and Housman, V. (1981) In: Organization and Expression of Globin Genes (eds. Stamatoyannopoulos, G. and Nienhaus, A. W.) pp 251-268, Liss, New York.

Wahli, W., David, I. B., Wyler, T., Weber, R. and Ryffel, G. U. (1980) Comparative analysis of the structural organization of two closely related vitellogenin genes. Cell 20, 107-117.

Wasylyk, B. and Chambon, P. (1981) A T to A base substitution and small deletions in the conalbumin TATA box drastically decrease specific *in vitro* transcription. Nucl. Acid. Res. 9, 1813-1824.

Weeds, A. (1982) Actin-binding proteins-regulators of cell architecture and motility. Nature 296, 811-816.

Weintraub, H. and Groudine, M. (1976) Chromosomal subunits in active genes have an altered conformation: globin genes are digested by deoxyribonucleaseI in red blood cell nuclei, but not in fibroblast nuclei. Science 193, 848-856.

Weintraub, H. (1983) A dominant role for DNA secondary structure in forming hypersensitive structures in chromatin. Cell 32, 1191-1203.

Whalen, R. G., Butler-Brown, G. S. and Gros, F. (1976) Protein synthesis and actin heterogeneity in calf muscle cells in culture. PNAS 73, 2018-2022.

Whalen, R. G., Butler-Browne, G. S. and Gros, F. (1978) Identification of a novel form of myosin light chain present in embryonic skeletal muscle tissue and cultured muscle cells. J. Mol. Biol. 126, 415-431.

Whalen, R. G., Sell, S. M., Butler-Browne, G. S., Schwartz, K., Bouveret, P. and Pinset-Harstrom, I. (1981) Three myosin heavy-chain isozymes appear sequentially in rat muscle development. Nature 292, 805-809.

Weisbrod, S. (1982) Active chromatin. Nature 297, 289-295.

Wieringa, B., Meyer, F., Reisser, J. and Weissman, c. (1983) Unusual splice sites revealed by mutogenic inactivation of the rabbit beta-

globin gene. Nature 301, 38-43.

Wild, C. D., Crowther, C. E., Cripe, T. P., Lee, M. G. and Cowan, N. J. (1982a) Evidence that a human beta-tubulin pseudogene is derived from its corresponding mRNA. Nature 297, 83-85.

Wild, C. D., Crowther, C. E. and Cowan, N. J. 91982b) Diverse mechanisms in the generation of human beta-tubulin pseudogenes. Science 217, 549-552.

Wilson, A. C., Carlson, S. S. and White, T. J. (1977) Biochemical evolution. Ann. Rev. Biochem. 46, 573-639.

Wiskocil, R., Bensky, F., Dower, W., Goldberger, R. F., Gordon, J. I. and Deeley, R. G. (1980) Coordinate regulation of two oestrogen - dependent genes in avian liver. PNAS 77, 4474-4478.

Wold, B. J., Lein, W. M., Hough-Evans, B. R., Britten, R. J. and Davidson, E. H. (1978) Sea urchin embryo mRNA sequences expressed in nuclear RNA of adult tissues. Cell 14, 941-950.

Woodland, H. R., Gurdon, J. B. and Lingrel, J. B. (1974) The translation of mammalian globin mRNA injected into fertilized eggs of Xenopus laevis. II. The distribution of globin synthesis in different tissues. Dev. Biol. 39, 134-140.

Woodland, H. R. (1980) Histone synthesis during the development of Xenopus. FEBS Letts. 121, 1-7.

Wu, C., Bingham, P. M., Litvak, K. J., Holmgren, B. and Elgin, S. C. R. (1979) The chromatin structure of specific genes. I. Disruption of chromatin structure during gene activity. Cell 16, 807-814.

Yamawaki-Katoaka, Y., Miyata, T. and Honjo, T. (1981) The complete nucleotide sequence of mouse immunoglobulin gamma-2a gene and evolution of heavy chain genes: further evidence for intervening sequence-mediated gene transfer. Nucl. Acid. Res. 9, 1365-1381.

Young, N. S., Beng, E. J., Kantor, J. A., Kretschmer, P. and Nieuhuis, A. W. (1978) Hemoglobin switching in sheep: only the gamma gene is in active conformation in fetal liver, but all the beta and gamma genes are in the active conformation in bone marrow. PNAS 75, 5884.

Young, R. A., Hagenbuchle, O. and Schibler, U. (1981) A single mouse alpha-amylase gene specifies two different tissue-specific mRNAs. Cell 23, 451-458.

Young, R. R., Scott, R. W., Hamer, D. H. and Tilghman, S. M. (1982) Construction and expression in vivo of an internally deleted mouse alpha- fetoprotein gene: presence of a transcribed Alu-like repeat within the first intervening sequence. Nucl. Acid. Res. 10, 3099-3116.

Zakut, R., Shani, M., Givol, D., Neuman, S., Yaffe, D. and Nudel, U. (1982) The nucleotide sequence of the rat skeletal muscle actin gene. Nature 298, 857-859.

Ziff, E. B. and Evans, R. M. (1978) Coincidence of the promoter and capped 5' terminus of RNA from the adenovirus-2 major late transcription

unit. Cell 15, 1463-1475.

Zimmer, E. A., Martin, S. L., Beverley, S. M., Kan, Y. W. and Wilson, A. C. (1980) Rapid duplication and loss of genes coding for the alpha chains of hemoglobin. PNAS 77, 2158-2162.

Zimmer, W. and Schwartz, R. J. (1982) Amplification of chicken actin genes during myogenesis. In: Gene Amplification (ed. R. T. Schimke) pp 137-146. Pub. Cold Spring Harbor Laboratory.

Zuker, C. and Lodish, H. F. (1981) Repetitive DNA sequences cotranscribed with developmentally regulated Dictyostelium discoideum mRNAs. PNAS 78, 5386-5390.