THE EXPRESSION OF A *XENOPUS BOREALIS* CARDIAC ACTIN GENE IN NORMAL AND TRANSFORMED FROG EMBRYOS

Clive Wilson, B.A. (Hons.) Cambridge

A thesis submitted for the degree of

Doctor of Philosophy

at the University of Warwick

MRC Animal Molecular Genetics Group

Department of Biological Sciences

University of Warwick

Coventry

U.K.

November, 1986.

CONTENTS

TABLE OF TABLES

TABLE OF FIGURES

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Hugh Woodland, not only for his advice, his encouragement and the continuous stream of ideas, but also for his technical assistance in this project. In particular, I am grateful for the many hours he has spent injecting and dissecting embryos. I would also like to give a big thank you to Gareth Cross, who first discovered that frogs had actin genes, for his ideas and assistance throughout the three years, and for the moral support in time of need.

I thank Elizabeth Ballantine, Malcolm McCrae, Alister Wilson and Sabu Sheikh for materials, and Dave Jackson, Bob Old and Aslam Mohammed for useful discussion. I am grateful for materials and information from Tommy Gallacher, Ed Bagenal, Phil Turner, Alistair Chambers and Mark Boardman. I thank Celia Kwasnik for looking after all the froglets and everyone else in the lab. for help over the past three years. I also thank Alistair Chambers for reminding me to work harder, and Phil Gilmartin for still talking to me when things were going badly, and, worse still, when things were going well. I am also grateful to Carolyn Alderson for her extremely rapid and highly accurate preparation of the manuscript in hurried circumstances. I acknowledge the financial support of the MRC.

However, the biggest thanks must go to Mum and Dad for their endless love, encouragement and support. Few of the opportunities I have enjoyed, including the last three years, would have been possible without their help, and I only hope that in the future I will be able to make further use of the chances they have given me.

DECLARATION

All the results presented in this thesis were obtained by the author, apart from those which are specifically indicated in the text. All the injections and almost all the dissections of *Xenopus* embryos were performed by Professor Hugh Woodland. In addition, the results in Section 5.2.1 and Figure 6.1 were obtained by Dr. Gareth Cross during the early part of my studies here, but they provide important background information for the rest of the work.

All sources of information have been specifically acknowledged by means of reference. None of the work contained in this thesis has been used for any previous application for a degree.

Clive Wilson

SUMMARY

The major aim of this project has been to ascertain if the expression of a cloned *Xenopus* gene, which is normally expressed in a tissue-specific fashion during early development, is regulated in the same way as its chromosomal counterpart in micro-injected *Xenopus* embryos.

A number of clones, which contain genes encoding different actin isoforms, have been isolated from *Xenopus borealis* genomic libraries. One of these contains an entire cardiac actin gene, on the basis of the isotype-specific sequence of its encoded product. Indeed, in the adult frog, the chromosomal gene is only expressed in the heart. However, in the embryo, transcripts are also detected in the myotomes, which contain skeletal muscle cells.

Two transcriptional assays have been developed, so that transcripts from the unmodified, cloned *X. borealis* cardiac actin gene can be detected separately from endogenous transcripts in micro-injected *X. laevis* embryos. In such transformed animals, injected linear DNA forms high-molecular-weight extrachromosomal concatenates, which replicate and become relatively evenly distributed throughout all tissues. Properly initiated transcripts from the cloned gene are correctly localised to the myotomes in both neurulae and tadpoles. The temporal regulation of expression also shares strong similarities with that of the endogenous, chromosomal actin gene. In a preliminary investigation of the sequences responsible for this regulation, a fusion construct between the first two actin exons and the last exon of a mouse β-globin gene has been injected. The same wide distribution of DNA, but spatially restricted pattern of expression, as the actin gene is found, whereas transcripts from a histone-globin fusion gene are formed in all tissues.

This is the first report of correct spatial control of expression from an injected, cloned gene in *Xenopus*. I discuss the wider significance of these results for future studies on the early developmental regulation of gene expression.

ABBREVIATIONS

Ac	acetate
ATP, dATP	adenosine triphosphate, deoxyadenosine triphosphate
b	base
bp	base pair
BSA	bovine serum albumin
cDNA	complementary DNA
Ci	curie
CIAP	calf intestinal alkaline phosphatase
cm	centimetre
CTP, dCTP	cytidine triphosphate, deoxycytidine triphosphate
DNA	deoxyribonucleic acid
DTT	dithiothreitol
EDTA	ethylenediaminetetra-acetate
FSH	follicle stimulating hormone
g	gram
GTP, dGTP	guanosine triphosphate, deoxyguanosine triphosphate
hCG	human chorionic gonadotrophin
hr	hour
kb	kilobase
l	litre
m	metre
MBT	mid-blastula transition
min	minute
ml	millilitre
pfu	plaque-forming units
rpm	revolutions per minute

SDS	sodium dodecyl sulphate
SV40	simian virus 40
TEMED	N,N,N',N'-tetramethyl-ethylenediamine
TK	thymidine kinase
Tris	tris(hydroxymethyl)aminomethane
tRNA	transfer RNA
dTTP	deoxythymidine triphosphate
5'/3' UTR	5'/3' untranslated region
v/v	volume per volume
w/v	weight per volume
μg	microgram
μl	microlitre

Clones

actbin	-	see Section 6.2.3
hisbin	-	see Section 6.2.3
λXB-C1	-	see Section 4.1
pXB-C1	-	see Section 4.1
λXB-C2	-	see Section 4.2.2
λXB-C3	-	see Section 4.2.3
pXB-C3.1	-	see Section 4.2.3

CHAPTER 1

INTRODUCTION

SECTION 1.1 INTRODUCTION

A multicellular animal is a clone of cells descended from a single original cell, the fertilised egg. All the daughter cells generally contain the same genetic information, even though their phenotypes differ so widely. This has been best illustrated in *Xenopus laevis* by the nuclear transplantation experiments of Gurdon and his colleagues (see Gurdon, 1974), who have shown that nuclei from a differentiated tissue, like the larval intestine, can be transferred to enucleated eggs, which will then develop into fertile adult frogs. These results clearly suggest that differentiated cells have not lost the genetic information that is required for differentiation into other cell types.

An understanding of the mechanisms by which phenotypic diversity is produced in genotypically identical cells requires answers to come of the most fundamental questions concerning the molecular biology of development. How are tissue-specific gene products expressed in terminal differentiation? What genes commit a cell to a specific fate? Indeed, how are the spatial coordinates of the egg first determined, so that it develops in an asymmetric but organised fashion? The advent of recombinant DNA technology and its application to the study of development in those animals, which are particularly suited to such investigation, has recently started to provide us with some answers to these problems.

The expression of gene products can be controlled at many levels, including translation, RNA or protein stability, and RNA processing or

transport. However, for those genes that have so far been examined, which are mostly expressed in a tissue-specific fashion, it has become increasingly clear that transcriptional regulation plays a key role. Nuclear run-off experiments have shown that transcription rates of some tissue-specific genes are raised in the cell types that express the gene product at high levels (e.g. McKnight and Palmiter, 1979; Jaynes *et al.*, 1986). Furthermore, activation of many genes is generally associated with DNA modifications and structural changes in the chromatin of the gene locus (see Weisbrod, 1982; Section 1.4.1), suggesting that these genes are also at least partly regulated by transcriptional control.

In addition, it has recently become possible to reintroduce cloned genes into cultured cells or whole organisms, and frequently it has been demonstrated that they are expressed in a fashion comparable to their chromosomal counterparts (see Section 1.3). Moreover, using such systems to test constructs in which the upstream sequences in the 5' flanking region of a tissue-specific gene have been fused to the promoter and transcription unit of another non-tissue-specific gene, it has sometimes been possible to show directly that non-transcribed sequences can regulate gene expression (e.g. Walker *et al.*, 1983; Chepelinsky *et al.*, 1985). With the development of these transformation systems, it is now relatively straightforward to begin to delineate sequences which are responsible for transcriptional control. However, different approaches are required to determine the *trans*-acting molecules which bind to the control elements, and, indeed, to isolate any genes which encode regulatory proteins that operate during development.

In this project, I have cloned a complete cardiac actin gene from *Xenopus borealis*, which is normally transcribed only in differentiated muscle cells. It has been reintroduced into fertilised *Xenopus* eggs and I have shown that it is expressed in developing embryos and tadpoles in a comparable spatial and temporal fashion to its chromosomal counterpart. This is the first time that expression from an injected, cloned gene has been demonstrated to be regulated in the correct tissue-specific fashion in *Xenopus*. This organism is particularly suited for studies during early development, so the system described may open avenues for investigation of gene activation events that are not so readily accessible in other vertebrates.

In the opening sections of this chapter, I briefly review the progress which has recently been made in our understanding of the molecular processes in eukaryotic development. Much of this work has been pioneered in lower eukaryotes and relatively simple animals, in particular yeast and *Drosophila*, where developmental mutants provide a means for isolating important regulatory genes. As will be discussed, it is much more difficult to use vertebrates for such studies, and so very few genes in these organisms have yet been positively shown to have a regulatory function. However, vertebrates, like the mouse, have been successfully used to elucidate the sequences required for proper cell-type-specific gene regulation, by introducing cloned genes either into cultured cells or into the whole organism. The results from these studies are summarised, and the general features at both the sequence and structural levels that are associated with expressed genes are discussed, with special reference to their functional significance.

The extensive literature on actin genes and their regulation is reviewed in the later part of the chapter. I also briefly describe the early development of *Xenopus*, and conclude with a short explanation of the aims of the project, commenting on the apparent suitability of the organism for studies on early developmental gene control.

SECTION 1.2 THE STUDY OF MOLECULAR MECHANISMS IN DEVELOPMENT

Although there are rare examples of changes in the genotype of eukaryotic cells during development, like the rearrangement of immunoglobulin genes in the lymphoid cells of vertebrates (see Gough, 1981) and the amplification of the chorion genes in *Drosophila* (Spradling, 1981), it is generally considered that almost all differential gene expression must result solely from changes in the cellular environment of unaltered genes. As it would seem uneconomic for genes, which are coordinately expressed, to each be controlled by different regulators, it has been proposed that small numbers of master regulatory genes might activate batteries of tissue-specific genes in an appropriately determined cell (see Davidson, 1976). Only a few of these regulators, acting in combination, would be needed to specify numerous different phenotypes. Their activity might also be controlled by other regulatory molecules, which affect the ability of the gene regulator to bind to its targets, and therefore alter gene activity in an indirect fashion. At a more fundamental level, it might be expected that other master genes would control regional specification in the developing embryo. Indeed, as the axes of an animal are already fixed in the single-cell fertilised egg, products from some of these genes must be localised even at this early stage.

For many decades, workers have been able to manipulate the cells of developing organisms in order to alter their fates or to reveal commitment to a restricted fate. In recent years, amphibian embryos, especially from *Xenopus*, have proved particularly informative to vertebrate embryologists, because their easy availability, large size

and independent development make investigative techniques like grafting relatively straightforward. Although a number of experiments involving rotation and centrifugation of the single-cell amphibian egg indirectly support the hypothesis that cytoplasmic determinants are present at this stage (reviewed by Gerhart, 1980), other work has clearly demonstrated that the cytoplasmic coordinates defined by these determinants do not alone specify the fate of a cell derived from a specific cytoplasmic region. Cellular interactions are also required as the egg divides, even for such fundamental processes as the formation of mesoderm (see Woodland and Jones, 1985; Section 5.3.1). Therefore, a molecular model of animal development must consider the combined effects of cellular interactions and inherited determinants, which eventually lead to the activation of alternative pathways of differentiation in different cells. Not only DNA-binding proteins, but also molecules which modulate their activity or expression, including those that do not act directly on the nucleus at all (e.g. cell surface molecules), must be investigated if a complete model is to be devised.

How can genes, which take part in events like these, be isolated? Those that are expressed in a tissue-specific fashion in differentiated cells, frequently produce abundant proteins and high message levels. It has therefore been relatively straightforward to isolate them in many vertebrates, including man, the mouse and *Xenopus* by initially screening with tissue-enriched cDNA probes, or using tissue-enriched cDNA libraries, coupled with other selection procedures.

On the other hand, most regulatory molecules are thought to be at very low concentrations in the cell, and so it has not been generally

possible to use the same techniques to find them. Indeed, as different subsets of regulatory gene products probably interact with each other to determine different cell fates, the localisation of any single controlling molecule may not correlate with a specific cellular or positional phenotype (e.g. Wasylyk and Wasylyk, 1986; see Section 1.4.2). Furthermore, if a molecule operates either directly or indirectly by cell interactions, it does not need to be expressed in the cells it affects (e.g. Kilchherr *et al.*, 1986). Therefore, unlike tissue-specific genes, even if regulatory genes were transcribed at high levels, the spatial distribution of their products might not correlate with the regions in which they are thought to function. Only recently have workers attempted to isolate such genes on the basis of localised expression alone (Rebagliati *et al.*, 1985), and the resulting genes have yet to be demonstrated to have any regulatory effects in development.

The isolation of developmentally important genes has chiefly been achieved by the study of organisms which exhibit developmental mutations that can be readily mapped in the genome. Genes involved in cell division (Pringle and Hartwell, 1981) and mating type determination (Nasmyth, 1982) have been characterised in the yeast *Saccharomyces cerevisiae* by this method. Many mutations have been recognised and some of the relevant genes cloned (Greenwald, 1985) in *Caenorhabditis elegans*, a nematode containing only 959 cells, whose fates in normal development can be designated according to a fixed cell lineage (see Lewin, 1984).

Although the relative simplicity of *C. elegans* may well lead to a rapid determination of some of the molecular mechanisms involved in the regulation of its development, even more interest and excitement has

been aroused by the advances made over the last three to four years in the molecular study of the fruit fly, *Drosophila melanogaster*. These investigations have increased our understanding of molecular processes in development and there are some indications that the discoveries can be partly extended to vertebrate development. The results are therefore discussed in the following sub-section.

1.2.1 The Molecular Biology of Development in *Drosophila melanogaster*

Drosophila melanogaster, like other insects, has a segmented structure. Over many years, mutants have been identified which affect this pattern or have other developmental effects. Recently, it has become possible to clone the genes involved in these mutations using the extensive background genetic knowledge already available in *Drosophila*. In particular, extensive studies have been undertaken to isolate the genes that determine pattern in the embryo, both because mutations of these loci are easily recognised, and because the mechanisms behind such early determination events are of special interest to developmental biologists. The resulting discoveries have been reviewed in numerous articles (e.g. Gehring, 1985; Berg, 1986; Akam, 1986; Woodland and Jones, 1986), and will be briefly summarised in this section.

Basically, the genes that direct the establishment of segmentation in *Drosophila* can be divided into three groups, which operate at different levels in the organisation process. First, the maternal-effect genes, which are transcribed prior to fertilisation, set up the axial coordinates of the cell. Some of these must therefore encode axial determinants. For example, the product of the gene *Toll* is thought to

act at a level close to the determination step in the establishment of dorsal-ventral polarity in the egg (Anderson *et al.*, 1985a, 1985b). However, at least for dorsal-ventral pattern formation, many of the maternal-effect gene products probably only act indirectly by activating the specific morphogen (see Woodland and Jones, 1986).

The second group of genes, the segmentation genes (Nüsslein-Volhard and Wieschaus, 1980), are involved in the division of the embryo into segments. They appear to interpret the anterior-posterior coordinates described by the maternal-effect determinants, and are subsequently expressed either in a specific region or in a series of regions, which defines units with the width of a half segment or multiples of a half segment. Once the segmented pattern is established, the third group of genes, which are part of the family of homeotic genes, are expressed in a specific segment or set of segments and, in combination, specify the precise fate of each segment according to its position in the whole embryo. Mutations in homeotic genes lead to the transformation of one body segment into another, or, in other cases, of one tissue-type into another. For example, a dominant mutation of the *Antennapedia* gene (*Antp*), which is normally expressed in the thorax and abdomen, produces abnormal expression of the gene in the head. Consequently, in the mutant animal, a head segment that carries antennae in the wild type organism forms legs like those of the metathoracic segment, which is the only segment in which *Antp* normally induces leg structures.

When the genes involved at all levels of this developmental programme were cloned, it was found that some share a translated region, called the homeo-box (McGinnis *et al.*, 1984a), which is highly conserved at the nucleotide level. This string of 180 bases encodes a polypeptide that

is even more conserved than the RNA sequence, suggesting that the translated product is functionally important. Indeed, it has been shown that this amino acid sequence is related to a domain involved in DNA binding in prokaryotes and also bears strong similarity to a region in the yeast mating type regulatory proteins derived from the *MAT*a1 and *MAT*α2 genes (Laughon and Scott, 1984; Shepherd *et al.*, 1984). The conserved domain forms a structure consisting of two overlapping α-helices linked in a helix-turn-helix arrangement, where the first helix can interact with the DNA backbone, while the second helix lies in the major groove.

Accumulating evidence suggests that the homeo domain, or at least some of the proteins which contain it, will bind to DNA. For example, antibodies to the *Ubx* product (encoded by a homeo-box-containing homeotic gene) react specifically with material in the nuclei of cells (Beachy *et al.*, 1985; White and Wilcox, 1984). Also, fusions of *lac*Z to a cDNA clone, which was derived from the transcripts of the *engrailed* gene (a homeo-box-containing segmentation gene), have been expressed in *E. coli* and the protein product exhibits sequence-specific DNA-binding activity *in vitro*, even when only a small segment containing the homeo domain is present (Desplan *et al.*, 1985). However, the binding selectivity is relatively low, and other regions of the protein involved in the known specificity of *in vivo* interactions have not yet been defined. An alternative approach to *in vitro* studies is to use transformed flies, which are null mutants for the gene of interest. Provided the introduction of the cloned wild-type gene can rescue these mutants (see Section 1.3 for studies on transgenic flies), the clone can be mutated in its coding regions in order to alter the amino acid sequence in putative functional positions. Similar experiments have

recently been performed in yeast with the mating type genes (Porter and Smith, 1986).

Hybridisation studies with homeo-box-specific DNA probes suggest that there are probably more than twenty homeo-boxes in the *Drosophila* genome, twelve of which have now been cloned. At least six are grouped in a region called the *Antennapedia* complex (ANT-C), which contains *Antp*, and a further three are clustered in the bithorax complex (BX-C), which contains *Ubx*. Homeo-boxes are present in both homeotic and segmentation genes. Although the homeotic genes that contain them are generally involved in anterior-posterior segment specification, recently *zerknüllt*, a homeotic gene that is involved in the proper differentiation of dorsally derived embryonic tissues, has been shown to contain a homeo-box (Doyle *et al.*, 1986). Another homeo-box-containing gene called *caudal* (Mlodzik *et al.*, 1985), isolated by cross-hybridisation with homeo-box probes, generates a transient anterior-posterior concentration gradient of transcripts at the pre-blastoderm stage. Its function is unclear but it would seem likely that it is responding directly to the morphogen gradients set up in the unfertilised egg.

Therefore, the characterised homeo-box-containing genes in *Drosophila* represent a group of genes that are all involved in the earliest stages of pattern development. Their products are almost certainly DNA-binding proteins. When homeo-box sequences were also found in many other organisms, including vertebrates (McGinnis *et al.*, 1984b; see Section 1.2.2), it became apparent that an understanding of their function in *Drosophila* might have important implications in the study of development in all animals. Consequently, the investigation of the characterised

Drosophilia genes is continuing to attract considerable attention in many laboratories.

Although the homeo-box-containing loci have been most extensively studied, other regulatory genes that affect development in *Drosophila* have also been characterised. For example, *Krüppel* (a segmentation gene) has been cloned (Preiss *et al.*, 1985), and shown to encode a repeated DNA binding motif with homology to the *Xenopus* transcription factor TFIIIA (Rosenberg *et al.*, 1986). These proteins contain metal-binding domains that are believed to form polypeptide fingers, which can interact with structurally repetitive regions in DNA (Rhodes and Klug, 1986; see Section 1.4.2). A further example of this structure has been found in *Drosophila* (*Serendipity*; Vincent *et al.*, 1985) and another is known in yeast (ADR1; Hartshorne *et al.*, 1986). Moreoever, there appear to be related domains in some mammalian proteins (see Berg, 1986).

The homeotic locus *Notch* is involved in cell determination events as opposed to pattern formation. It produces a number of recessive alleles, which lead to a great increase in the proportion of ectodermal cells which differentiate into neuroblasts (Akam, 1984). Cloning of the major *Notch* transcript (Wharton *et al.*, 1985) has demonstrated that it encodes a protein which shares homologies with a group of mammalian proteins, including epidermal growth factor and the low-density lipoprotein receptor, many of which participate in cell communication processes. These proteins all contain characteristic cysteine-rich peptides of 40-50 amino acids, which are frequently repeated. As yet, it is unclear if the *Notch* product is secreted or is membrane-bound. However, the discovery of a similar peptide motif encoded by a homeotic gene in *C. elegans* (*lin-12*; Greenwald, 1985) would suggest that this may

be a domain that is shared by developmentally regulating genes of other organisms. It also seems likely that it functions via cell-cell interactions.

In *Drosophila*, not all those pattern-determining genes, which have been cloned, have yet been shown to encode products with homology to known regulatory proteins. It would be interesting to test them for properties like DNA-binding activity. However, data from those that do share homologies would suggest that the same basic molecular mechanisms may be involved in the early developmental events in *Drosophila* and vertebrates. The homeo domain is particularly interesting as it is only associated with genes required for pattern formation in *Drosophila*. Furthermore, its highly conserved sequence over 180 nucleotides allows related genes to be rapidly isolated, including those from other organisms. Future studies on *in vitro* binding to DNA and using functional tests with transformed genes should elucidate the other DNA-binding or protein-binding regions present in each homeo domain-containing protein, which account for the unique specificity of binding.

By contrast, the other conserved regulatory motifs do not require such strict amino acid conservation and so related genes cannot be recognised by hybridisation. Indeed, there may also be other helix-turn-helix domains in regulatory proteins that are constructed with different amino acid arrangements compared to the homeo domain and therefore are not isolated by a hybridisation screen with a homeo-box probe.

In situ hybridisation has confirmed that generally the localisation of transcripts from the cloned *Drosophila* regulatory genes correlates with

their function (see Coulter and Wieschaus, 1986). However, transcripts from *paired*, a pair-rule segmentation gene, are found in regions that are not affected by the mutation (Kilchherr *et al.*, 1986). These RNAs may need cell interactions to have an effect, or they may have no clear function in the additional regions of expression. Whatever the explanation, this example does illustrate that localisation of expression and regionalisation of function need not precisely overlap. It is already known that some of the *Drosophila* developmental genes and their products interact with each other (e.g. Struhl and White, 1985; Howard and Ingham, 1986). Indeed, Jürgens (1985) has estimated from genetic data that approximately 40 loci can affect the expression of the bithorax complex.

All these findings suggest that a complex network of control mechanisms is necessary to regulate pattern formation in *Drosophila*. Only functional assays using mutant flies transformed with normal and mutagenised cloned genes are likely to elucidate the interactions involved. Therefore, the availability of mutants is not only important in the isolation of developmental genes in *Drosophila*, but, by providing a null background for expression of a particular gene in some cases, it can allow the investigation of gene function. Without this approach, alternative, and less satisfactory, experiments would need to be performed involving functional tests *in vitro* or in cells that do not normally express the gene (see Section 1.2.2).

1.2.2 The Molecular Biology of Development in Vertebrates

Unlike *Drosophila*, no vertebrate has been screened for developmental

mutants in the exhaustive fashion that would be required to find most of the genes taking part in the establishment of pattern in early embryos. Indeed, in organisms like the mouse, where some developmental mutants do exist, it would be unrealistic to consider such a project, merely because such excessive space and time would be required. Furthermore, even those mutants that do exist cannot be mapped accurately enough on the chromosomes to isolate the relevant gene.

Therefore, until recently, only vertebrate genes expressed in terminal differentiation had been isolated. However, the discovery that mammals, birds and amphibians all contain sequences which hybridise to the homeo-box (McGinnis *et al.*, 1984b) has led to the isolation of numerous genes that may have an important developmental function. Although none of these genes have yet been directly shown to play a role in development, accumulating circumstantial evidence would suggest that they do, and in the mouse, it has been found that two homeo-box genes very roughly map to known developmental mutations (Joyner *et al.*, 1985; Rabin *et al.*, 1985).

The homeo domain is remarkably highly conserved between vertebrates and *Drosophila*. For example, the *Xenopus* gene MM3 shares 59 out of 60 amino acids with *Antennapedia* in the homeo domain (Müller *et al.*, 1984). The homeo-box-containing genes are often clustered (e.g. Harvey *et al.*, 1986; Duboule *et al.*, 1986), a property that is also found in *Drosophila*. Recently, Harvey *et al.* (1986) have injected synthetic RNA made from a *Xenopus* homeo-box-containing cDNA into *Xenopus* oocytes and demonstrated a nuclear location of the protein product.

At a more functional level, many of these genes are expressed in

embryonic development, and in some cases the transcripts are lost before adulthood (e.g. Rubin *et al.*, 1986). Furthermore, a number of homeo-box-containing genes are expressed specifically in the germ cells of the adult as well as in embryos (Rubin *et al.*, 1986; Müller *et al.*, 1984), in a similar way to the *caudal* gene of *Drosophila* (Mlodzik *et al.*, 1985).

Analysis of transcripts from dissected embryos (Jackson *et al.*, 1985) indicates that expression from the mouse homeo-box gene, *Hox2.1*, is preferentially localised to the spinal cord and brain. But *in situ* hybridisation in the mouse has provided the strongest evidence for believing that any of these genes have a developmental function. A probe from the homeo-box gene, *Hox-3*, shows a spatial restriction of transcripts in a specific anterior-posterior region of the mouse central nervous system (Awgulewitsch *et al.*, 1986). The pattern bears a remarkable resemblance to that found for genes like *Antennapedia* in the nervous system of *Drosophila*.

As discussed in the previous section, it has not been possible to use probes encoding other structural motifs to isolate different types of regulatory genes by hybridisation, because of their less stringent requirements for sequence conservation. Alternative methods to find other relevant genes are based on the assumption that they will be expressed in temporally or spatially restricted patterns. A number of studies in *Xenopus* has led to the isolation of a small group of genes with a possible regulatory function.

For example, the gastrula-specific gene, GS17, is expressed only at a very early stage of *Xenopus* embryonic development (Krieg and Melton,

1985) and may therefore be involved in early developmental events. Differential screening of cDNA libraries synthesised from RNAs present in the animal and vegetal poles of the *Xenopus* oocyte has led to the isolation of localised maternal RNAs, which might encode cytoplasmic determinants (Rebagliati *et al.*, 1985). Sargent and Dawid (1983) have used a substracted *Xenopus* cDNA library enriched in early embryo-specific transcripts to find spatially regulated genes. In experiments of this kind, the resulting genes may well encode end-products of a differentiation process, particularly as the RNAs from such genes are likely to be much more abundant than those from regulator genes. Genes expressed in a particular cell type might also be isolated if antibodies, which react with specific tissues, are used to screen a cDNA expression library, or if the proteins with which they react are sequenced and a suitable oligonucleotide probe is synthesised. In this case, if a large panel of different antibodies is available, they may react at different stages on a developmental pathway, and an approach of this kind may therefore select some regulatory genes as well. Such a study is currently in progress on the development of the eye in *Drosophila* (see Beachy, 1985).

However, none of the vertebrate genes isolated has yet been shown to synthesise a product that regulates genes transcribed by RNA polymerase II or that affects development. In *Drosophila*, direct evidence for gene function can be taken from the study of mutants, and rescue experiments using transgenic mutants containing the cloned wild-type genes. Although, as will be discussed in the next section, introduction of cloned DNA into vertebrates like the mouse provides the most effective test for the sequences required to regulate a gene, testing for the function of that gene involves using cells in which the chromosomal gene

is not working. As vertebrate mutants are not readily available, currently the most informative experiments will probably involve expressing the gene in cell types where transcripts are not found in normal animals and observing the effects on development. This could produce mutants in the same way as the dominant *Antennapedia* mutation in *Drosophila*, leading to establishment of altered pattern or different cell phenotypes in the affected region. However, although such assays should provide valuable information on the role of any one particular gene, they could not be expected to completely unravel the effects of a number of different developmental regulators, if they interact in a similarly complex fashion to those which are thought to regulate the bithorax complex (Jürgens, 1985).

As an alternative approach, it has been suggested that anti-sense RNA, introduced into developing organisms, might selectively eliminate or reduce the expression from its complementary RNA and so produce artificial mutants. Izant and Weintraub (1984) found that they could decrease the expression of an injected thymidine kinase gene in mouse L cells by coinjection of a plasmid directing the synthesis of anti-sense RNA. Moreover, Melton (1985) has demonstrated that the translation of injected globin mRNA in *Xenopus* oocytes can be inhibited by introducing anti-sense globin RNA synthesised *in vitro*. However, in *Xenopus*, whose large, independently developing embryos are particularly suited to this type of experiment, no successful results using anti-sense RNA to block the expression of putative developmental genes have yet been reported. By contrast, in *Drosophila*, this technique has been successfully used to produce weak *Krüppel* mutants (Rosenberg *et al.*, 1985).

In view of the apparent complexity of interactions between genes

involved in pattern formation in *Drosophila*, it would seem unlikely that investigations of similar processes during the early development of vertebrates, which do not have the same advantages for these studies, will be able to proceed at a comparable rate. However, although vertebrates are segmental animals, their descriptive and experimental embryology is clearly different from *Drosophila* and it will probably be impossible to extrapolate much more than the basic molecular mechanisms of gene regulation during pattern development from the fruit fly to these higher organisms.

With the recent characterisation of some of the early cell lineages in the development of zebra fish (Kimmel and Warga, 1986), it has been suggested that it may be the most suitable organism to examine both pattern formation and cell determination mechanisms in vertebrates. Its great advantages are that homozygous diploid fish can be produced and identical clones can be made from these fish (Streisinger *et al.*, 1981). These properties, together with the small size, high numbers of progeny and relatively short generation time in this species, may make them ideal candidates to be systematically screened for developmental mutations.

Despite the problems in analysing the normal function of products from cloned genes in vertebrates, introduction of DNA into whole organisms, particularly the mouse, has already provided considerable information on the sequences required to control expression of some genes. This work is briefly summarised in the next sections. Later, I will also discuss some of the recent advances in identifying the *trans*-acting molecules which bind to these sequences. If the RNAs which encode such factors can be cloned as cDNAs, this may provide an alternative and more

practicable means for isolating important regulatory genes involved in vertebrate cell differentiation processes (see Section 1.4.2 and Chapter 8).

SECTION 1.3 THE INTRODUCTION OF CLONED GENES INTO CELLS AND DEVELOPING ORGANISMS

Although sequence homologies between equivalent genes of different species, like the globin, immunoglobulin or actin genes, may indicate those regions necessary for the proper regulation of expression, their conservation may equally well result from other functional requirements. Therefore, tests of mutated genes in whole organisms or cultured cells must be used to delineate the sequences of real interest, unless satisfactory *in vitro* systems can be established. In the former approaches, cloned genes may be marked to distinguish them from chromosomal genes or they may be introduced into heterologous cellular backgrounds.

In yeast, an ideal system for the reintroduction of cloned DNA has been devised (Scherer and Davis, 1979; Orr-Weaver *et al.*, 1981). By linearising the cloned DNA before transformation in a region of homology with its chromosomal counterpart, it is preferentially incorporated into its normal genomic location. Thus, the endogenous sequences are replaced, and so the transgenic yeast differs from the wild-type organism only in the changes made in the introduced gene.

At present it is not possible to replace genes in cells from higher eukaryotes with this level of efficiency, although some recent experiments have suggested that this may soon be plausible, at least in cultured cells. For example, Lin *et al.* (1985) and Thomas *et al.* (1986) have corrected mutant genes, which had been previously integrated into the genome of cell lines. They introduced replacement DNA by

transformation or micro-injection with gene copies defective in another region. Therefore, recombination between the two types of gene had to occur to produce a normal gene. When micro-injection was used, which is a more efficient method for DNA transfer than transformation, one in every thousand cells injected (or one in every hundred in which the DNA was integrated) exhibited productive targeting. Studies by Smithies *et al.* (1985) on the homologous recombination of cloned globin DNA at the normal chromosomal human β-globin locus of human cell lines have suggested that the chromosomal genes may need to be expressed in the transformed cells for the site to be accessible for integration.

Further work is clearly still required to characterise an efficient gene targeting method that is applicable to normal chromosomal loci. Furthermore, even if a suitable techique is devised, it will still only alter one gene copy in the diploid genome, and so it could not be used for direct studies on the function of the gene.

Many transformed cell lines and cell cultures are of limited application to studies of gene regulation. Although some are derived from cells, which can be induced to differentiate *in vitro*, like lymphoma, erythroblast and myogenic cell lines, even these can only be used to study a single step in the developmental process. Moreover, as is discussed with reference to myogenic cells in Chapter 8, the precise state of differentiation of any cell line may be open to question, and cloned genes are not necessarily treated in the same way as their endogenous chromosomal counterparts. Nevertheless, using such transformed cells, some important progress has been made in identifying sequences that regulate the expression of some cell-type-specific genes, like the rat chymotrypsin gene (Walker *et al.*, 1983), the chicken α-

crystallin gene (Chepelinsky *et al.*, 1985) and the muscle-specific actin genes (see Chapter 8). However, those sequences implicated in gene regulation must eventually be tested under more stringent conditions in the whole transgenic or transformed organism.

Both in *Drosophila* and the mouse, systems have been established in which foreign DNA can be integrated into the genome, including that of the germ line. Consequently, transgenic progeny containing the same genomic insertion in all cells can be produced. Transgenic *Drosophila* were first used for studies of gene regulation by Spradling and Rubin (1983), and Goldberg *et al.* (1983). The transposable P element (see Engels, 1983) is generally stable in the *Drosophila* genome, but it will move in the fertilised egg, when a male containing P elements in the germ cells is crossed with a female of an M strain (that does not carry the P transposon). M strains have an "M cytotype", which permits P element transposition, a process that is catalysed by a transposase which is encoded by the normal P transposon (Engels, 1984).

P element vectors have been designed that can incorporate exogenous DNA and will replicate in *E. coli*. If they are injected into *Drosophila* embryos, which are then allowed to develop and mate, transgenic progeny are formed at high frequency. The insertions are at relatively random positions, but integrated genes display two particularly important and useful features. Firstly, only one copy of the gene is usually inserted into the genome. Secondly, even though there is slight variation in the quantitative expression of an incorporated gene (Spradling and Rubin, 1983), chromosomal position does not usually appear to affect the temporal and spatial regulation of the numerous genes tested, and frequently these genes are transcribed at similar rates to the normal

chromosomal genes. These two properties cannot presently be reproduced in any vertebrate, although, now the mechanism by which P element transposition is limited solely to the *Drosophila* germ line has been elucidated (Laski *et al.*, 1986) and the P element has been altered to circumvent this control, it may be possible to use P element transposition for the construction of transgenic organisms in other species.

As reviewed by Palmiter and Brinster (1985), transgenic mice have been produced by both micro-injection of DNA and retroviral infection. The former has been more commonly used, and its major advantages are that there is no constraint on the size of the DNA introduced and typically insertion takes place at the one-cell stage, so the founder mice can be directly analysed. However, usually many copies become integrated at a single site in tandem head-to-tail arrays. Although many reports of correct developmental gene regulation have been published (e.g. Swift *et al.*, 1984; Grosschedl *et al.*, 1984), there is often considerable quantitative variation in expression between individual mice, presumably due to position effects, and the level of transcription per cloned gene copy may be low compared to its endogenous counterpart (e.g. Shani, 1985; Townes *et al.*, 1985). The influence of chromosomal location on gene expression is discussed in Section 1.4. In addition, studies on a number of genes like the β-globin gene (Chada *et al.*, 1985; Townes *et al.*, 1985), and the rat skeletal actin gene (Shani, 1986; see Chapter 8) in transgenic mice have suggested that attached prokaryotic sequences cannot only influence expression quantitatively but they can also alter its localisation. In comparison to the P element system in *Drosophila*, this transgenic system in mice does not therefore provide such a neutral environment in which the introduced gene can operate.

In infected mouse embryos, retroviral vectors generally insert single copies of the cloned sequence, which are flanked by the long terminal repeats of the viral DNA (Harbers *et al.*, 1981). However, infection is usually initiated at later stages in development, so that founder mice are mosaic and only progeny can normally be used in studies of gene regulation. This has meant that reports investigating gene expression by this method are still not commonplace. However, infection of stem cell cultures, particularly haematopoietic progenitor cells, has attracted much more interest as a model system for future human gene therapy (Williams *et al.*, 1986). To this end, defective viruses have been designed, which cannot form infective particles once they have entered their target cell and also lose their enhancer and promoter sequences upon integration (Yu *et al.*, 1986). In addition, retroviruses have been shown to mutate genes by insertion in infected mouse embryos, producing developmental abnormalities (Schnieke *et al.*, 1983).

At present, the micro-injection of DNA into mouse embryos is the major method for producing transgenic animals for the investigation of gene regulation. However, it has now become possible to culture mouse teratocarcinoma cells *in vitro*. These cells can be grafted into normal blastocysts and they can give rise to germ cells in the developing mosaic organism (Stewart and Mintz, 1981). Recently this approach has been used with teratocarcinoma cells that had been infected with retroviral vectors, and the introduced DNA has been shown to be transmitted through the germ-line (Robertson *et al.*, 1986). Clearly, methodology of this kind will allow the pre-selection of appropriately transformed cells for the production of transgenic animals. It may even be possible in the future to target mutated genes in such cultured cells to their normal chromosomal position by the methods described earlier in

this section. This approach would finally give rise to mutant non-mosaic offspring, carrying a defective gene in the proper chromosomal location.

Other transgenic organisms have been constructed by micro-injection techniques, including rabbits, pigs and sheep (Hammer *et al.*, 1985), but they have not been extensively used for gene expression studies. Foreign DNA has also been introduced into *Xenopus* (Rusconi and Schaffner, 1981) and sea urchin (Davidson *et al.*, 1985) eggs. Although genes are thought to become integrated in these systems, to date the published expression studies have concentrated on early development when most of the genes are extrachromosomal. There have been no reports of correct spatial regulation of such injected genes (see Section 7.1 for a more detailed review). As *Xenopus* eggs are large and relatively easy to inject, and they develop rapidly and independently, it would be particularly advantageous to show that expression from exogenous DNA could be controlled properly in the early embryo. The initial activation of developmentally regulated genes might then be studied at stages when other vertebrate embryos, like those of the mouse, are not readily accessible for such investigation.

The established experimental systems described above have already shed considerable light on the mechanisms of gene expression. The following section summarises the progress made in identifying the characteristic properties of active genes, the sequences required to activate them and maintain them in an activated state, and also the molecules which appear to be involved in this activation process. Both general and specific features associated with eukaryotic gene expression are therefore described, concentrating particularly on genes transcribed by RNA

polymerase II. Data on muscle-specific actin genes are reviewed later, in Chapter 8.

SECTION 1.4 THE EXPRESSION OF EUKARYOTIC GENES

1.4.1 The Properties of Active Eukaryotic Genes

The general structure of the prokaryotic gene promoter and the mechanisms by which it is activated were elucidated some years ago (see Miller and Reznikoff, 1978). Basically a binding site for RNA polymerase is present at the start of a gene and the efficiency of binding is governed by specific proteins which interact with regulatory DNA sequences located close by. In one example, the work of Ptashne and his colleagues (Ptashne, 1984) has shown that the stimulation of transcription from the *cI* gene of bacteriophage λ by the lambda repressor involves protein-protein contacts with RNA polymerase. By contrast, the inhibition of expression by this repressor at other promoters is simply a consequence of steric exclusion.

In some respects, it might also be expected that eukaryotic promoters would be activated in a similar way by gene-specific activating proteins. However, as Lin and Riggs (1975) have discussed, on average, eukaryotes have about a thousand times more DNA than prokaryotes and so regulatory factors with roughly the same binding affinities as prokaryotic proteins would be mostly sequestered to non-specific binding sites. As it would seem unlikely that these molecules could have much higher specificities in eukaryotes, it has been suggested that most of the genome must be masked from diffusible factors in any particular cell type. There is good evidence for specific *trans*-acting repressor molecules in some instances (Nasmyth, 1982; Killary and Fournier, 1984), but a more general repression mechanism must exist to explain the

widespread inactivation of the genome. Such control may not operate in *in vitro* transcription systems made from cell-free extracts, and hence this may explain the ability of the latter to transcribe cloned tissue-specific genes from a wide range of cell types (e.g. Luse and Roeder, 1980).

Eukaryotic DNA is packaged by histones into nucleosomes and these in turn form higher order structures by complex folding (see Butler,1984). However, if this packaging is to provide a masking mechanism it must be able to allow different selected regions to be made accessible in different cells during the early stages of gene activation. An additional requirement of eukaryotic gene activation is that it should remain semi-permanent through cell divisions, once differentiation has occurred.

As soon as clones of eukaryotic genes became readily available, it was demonstrated that normal chromosomal genes displayed unique properties when they were activated (see Weisbrod, 1982), and these properties are described in this section. In particular, cell-type-specific genes like the globins have enjoyed considerable attention. One drawback with such studies is that they only recognise changes which correlate with expression and do not shed light on the more important question of the role of these changes in inducing gene activation.

Basically, to understand developmental gene regulation, three problems must be resolved:

(1) how is the target gene made selectively accessible to its binding factors, and how is most of the genome maintained in a repressed state?

(2) how do regulatory molecules act to promote transcription?

(3) how is a gene maintained in an activated state through cell division?

The experiments on the properties of active genes have provided some insight into the mechanisms that may be involved in these processes. Unfolding of chromatin, increased nuclease sensitivity, negative supercoiling and undermethylation are all characteristic properties of many activated gene loci. As will also be discussed in the section, a number of groups have recently attempted to demonstrate a causal role for some of these changes by introducing cloned DNA into cells. However, it is still not possible to propose a general model for developmental gene activation from these results.

Digestion of nuclear chromatin with some nucleases, especially pancreatic DNase I, has shown that active genes are preferentially sensitive to digestion. It is generally found that a wide domain that includes the transcription unit of the gene displays an increased susceptibility, which is not affected by the rate of transcription (Garel *et al.*, 1977), and, in some cases, can merely reflect a potential to be transcribed. In the chick oviduct, for example, the ovalbumin gene remains sensitive to DNase I in the absence of stimulating steriods (Palmiter *et al.*, 1977).

The division of chromatin into domains with apparently different accessibilities to diffusible proteins like DNase I has been correlated with the discovery of loops of DNA, approximately 50-100 kb in length, attached to a nuclear substructure in lampbrush and mitotic chromosomes

(2) how do regulatory molecules act to promote transcription?

(3) how is a gene maintained in an activated state through cell division?

The experiments on the properties of active genes have provided some insight into the mechanisms that may be involved in these processes. Unfolding of chromatin, increased nuclease sensitivity, negative supercoiling and undermethylation are all characteristic properties of many activated gene loci. As will also be discussed in the section, a number of groups have recently attempted to demonstrate a causal role for some of these changes by introducing cloned DNA into cells. However, it is still not possible to propose a general model for developmental gene activation from these results.

Digestion of nuclear chromatin with some nucleases, especially pancreatic DNase I, has shown that active genes are preferentially sensitive to digestion. It is generally found that a wide domain that includes the transcription unit of the gene displays an increased susceptibility, which is not affected by the rate of transcription (Garel *et al.*, 1977), and, in some cases, can merely reflect a potential to be transcribed. In the chick oviduct, for example, the ovalbumin gene remains sensitive to DNase I in the absence of stimulating steriods (Palmiter *et al.*, 1977).

The division of chromatin into domains with apparently different accessibilities to diffusible proteins like DNase I has been correlated with the discovery of loops of DNA, approximately 50-100 kb in length, attached to a nuclear substructure in lampbrush and mitotic chromosomes

(Paulson and Laemmli, 1977). This putative scaffold for general chromatin organisation (Jackson, 1986) has come to be known by numerous names, like the nuclear matrix, cage and scaffold, as well as the nucleoskeleton, reflecting the fact that many different techniques have been used to prepare it. Most of these have steps involving treatment in non-physiological hypertonic or hypotonic conditions, which has led to scepticism about the validity of this approach. However, Jackson and Cook (1985a) have developed a method using isotonic conditions where cells are encapsulated in agarose beads, thus preventing the clumping of nuclei that is normally observed at these salt concentrations. Experiments in this system (Jackson and Cook, 1985b) have tentatively demonstrated that RNA polymerase II is preferentially bound to the nuclear matrix. Other workers (Ciejek *et al.*, 1983; Jost and Seldran, 1984) have shown by different methods that the entire transcribed region and some flanking sequences of genes become associated with the nuclear matrix when the genes are activated. Such data suggest that the matrix structure may take part in transcription processes, as well as providing a framework on which the genome is packaged (see Section 1.4.2 for further details).

The mechanism by which chromatin is retained in a repressed, as opposed to an activated, state is still unclear (reviewed by Brown, 1984, and Weintraub, 1985). Most, if not all, transcribed genes are packaged into nucleosomes (protein complexes consisting of the histones H2A, H2B, H3 and H4), so a restriction in accessibility must be explained at a higher structural level. Some studies have suggested that certain high mobility group nuclear proteins (particularly HMGs 14 and 17) may associate specifically with active genes to induce DNase I sensitivity (e.g. Weisbrod and Weintraub, 1979). However, this phenomenon has only

been demonstrated in a few selected cases. At a more general level, Weintraub (1984) has found that the histone H1 is bound to inactive regions of chromatin in such a way that it holds together adjacent nucleosomes, while in active chromatin this does not occur.

The most extensive evidence for the participation of H1 in repression has accumulated from studies of the 5S RNA genes in *Xenopus*. They are activated by a transcription complex containing at least three components including TFIIIA, a DNA-binding protein with a metal-binding finger structure (Miller *et al.*, 1985; see Section 1.4.2). This factor binds an internal control region in 5S RNA genes, and the complex potentiates correct initiation by RNA polymerase III (see Brown, 1984).

There are two types of 5S gene. The 800 somatic-type genes are expressed in the oocyte and throughout development in somatic cells. By contrast the 40,000 oocyte-specific genes are only transcribed in oocytes and briefly during early embryonic development. Not only can an active 5S transcription complex be formed *in vitro* using cloned DNA and nuclear extracts (Segall *et al.*, 1980), but Bogenhagen and his colleagues (1982) have found that somatic 5S RNA genes from isolated somatic cell chromatin can also be transcribed *in vitro* by addition of RNA polymerase III alone. However, oocyte 5S RNA genes from the same cells were stably repressed by chromatin, which required perturbation (i.e. a high salt wash) for added transcription factors to form an active complex. Further investigations (Schlissel and Brown, 1984) have implicated histone H1 as the repressor, because its binding correlates with gene inactivation, and the purified protein can be added back to washed chromatin to reestablish repression.

There is only a 4-5 fold transcriptional bias *in vitro* for somatic 5S genes over oocyte 5S genes. However, Brown and Schlissel (1985) have demonstrated a 25- to 200-fold higher affinity of TFIIIA for the former in injected embryos, although this is still insufficient to account for the absolute preference for somatic genes *in vivo* in later embryos and somatic cells.

An attractive feature of the *in vitro* system, which might be important *in vivo*, is that the repressed and activated states are stable and therefore do not appear to be in direct equilibrium. Of course, at some time during development, perhaps when DNA replication takes place, these two conformations must be allowed to interchange. Indeed Gilbert (1986) has shown that oocyte 5S RNA genes replicate after their somatic counterparts in the somatic cell cycle. He therefore suggested that the former may become inactivated because transcription factors are all sequestered by the somatic genes, so that no new active transcription complexes can form on the oocyte genes. This may explain the complete repression of these genes during early embryogenesis, and the fact that this level of repression has not yet been reproduced with cloned 5S DNA. Such a hypothesis for repression is supported by the demonstration that oocyte-type 5S RNA genes, which have been translocated in a somatic cell line to a site where they are replicated early in the cell cycle, are expressed (Guinta *et al.*, 1986).

The structure of repressed oocyte 5S genes can also be destabilised by transplantation of somatic cell nuclei into *Xenopus* oocytes (Korn and Gurdon, 1981), although this effect varies between different oocyte batches. Experiments by Blau *et al.* (1983) and Chiu and Blau (1984), in which the fusion of differentiated muscle cells with non-muscle cells to

form stable heterokaryons led to expression of muscle-specific genes from the latter cell type, further suggest that repression need not be irreversible even if replication does not take place.

Other characteristic features have also been associated with active chromatin. As well as the DNase I-sensitive domains spanning the whole gene locus, specific nucleotides or small regions of DNA become hypersensitive to this enzyme (see Elgin [1984] for review). These are particularly prevalent within about one kilobase upstream of the cap site of many genes (e.g. Shermoen and Beckendorf, 1982; McGhee *et al.*, 1981) and presumably coincide with areas where normal chromatin structure is disrupted. Indeed, in many examples, the hypersensitive sites are in DNA regions, which have been shown by deletion analysis to affect gene expression. In the simian virus SV40 (Jakobovits *et al.*, 1980; Jongstra *et al.*, 1984) and the chick β-globin gene (McGhee *et al.*, 1981), a hypersensitive region correlates with a nucleosome-free gap. Emerson and Felsenfeld (1984) have isolated a protein from chick erythrocyte nuclei that binds to the β-globin promoter in this area and induces the formation of the appropriate hypersensitive sites in chromatin (see Section 1.4.2).

In some cases, sequences sensitive to single-strand-specific endonucleases, like S1 nuclease and *Bal*31 nuclease, are observed in supercoiled cloned DNA as well as active chromatin close to the positions of the DNase I hypersensitive sites (Mace *et al.*, 1983; Larsen and Weintraub, 1982; Weintraub, 1983; Nickol and Felsenfeld, 1983; Schon *et al.*, 1983). However, they do not appear to represent single-stranded stretches, at least *in vitro*. In the case of the chick β-globin gene,

the sensitive regions found *in vitro* with pure plasmid do not fully correlate with those detected in chromatin from erythrocytes (Weintraub, 1983; Schon *et al.*, 1983). It therefore remains unclear if the unusual susceptibility of DNA to nucleases at the 5' flanking region of active genes reflects a different DNA structure (and whether this is in any way inherent to the DNA itself), or if the binding of specific proteins in chromatin can produce these differential sensitivities without significantly altering the DNA conformation.

Although one report appeared to address this question by demonstrating that regions which react with antibodies to Z-DNA coincide with active chromatin, this observation has now been shown to be the result of an artifact of sample preparation (Stollar, 1986). However, it has recently been demonstrated that the binding of two molecules of the *Drosophila* heat shock transcription factor (HSTF) to the 5' flanking region of the *hsp*70 heat shock gene leads to bending of heat shock promoter DNA (Shuey and Parker, 1986). In this case, specific changes in nuclease sensitivity produced by binding would therefore appear to be at least partly accounted for by concomitantly induced changes in DNA structure.

At the DNA level there is some evidence that negative supercoiling, which is known to stabilise S1 nuclease-sensitive regions *in vitro* and might favour many of the conformations that would be sensitive to DNase I, can also activate genes *in vivo*. On a gross scale, most of the negative superhelical tension in eukaryotic chromosomal DNA, unlike that in prokaryotic chromosomes, is taken up by coiling around the nucleosome (see Sinden *et al.*, 1980), but the psoralen binding technique used to demonstrate this cannot detect localised stresses. Harland *et al.*

(1983) have injected the herpes simplex virus thymidine kinase (TK) gene into *Xenopus* oocytes as either circular or linear plasmid molecules. They found that the circular form was transcribed at least 500 times more efficiently. Furthermore, if the circular molecules were allowed to be assembled into transcription complexes in the oocyte and then they were linearised by injection of a restriction enzyme, subsequent expression was inhibited. However, such dramatic differences between circular and linear molecules have not always been observed in oocytes (Gurdon and Melton, 1981; Probst *et al.*, 1979), and I have found significant transcription from the linear form of an injected histone-globin fusion gene (see Section 7.2.8) in these cells.

Other indirect evidence supporting a role for supercoiling has come from experiments with injected circular 5S RNA genes in oocytes (Ryoji and Worcel, 1985), where transcriptional activity correlates with the introduction of torsional stress in minichromosomes. A similar phenomenon has also been observed with SV40 minichromosomes (Luchnik *et al.*, 1985). For the 5S genes, injection of novobiocin (an inhibitor of topoisomerase II; see Section 1.4.2) into oocytes leads to inhibition of transcription from introduced genes, suggesting that DNA gyration is required for expression (Ryoji and Worcel, 1984). *In vitro*, Kmiec and Worcel (1985) have demonstrated that TFIIIA will induce negative superhelical tension on binding to 5S genes.

The other major feature of DNA in active chromatin is its methylation state (see Bird, 1984, 1986). Between 60% and 90% of all CpG dinucleotides in vertebrate DNA are methylated at the 5 position on the cytosine ring and this constitutes most of the base methylation found in

the cell. One attractive aspect of using this mechanism to delineate transcribed loci is that methylation is inherited on DNA replication, because the hemimethylated DNA produced is rapidly fully methylated in all cells.

Numerous studies have found that some CpG sites are differentially demethylated in active chromatin, particularly in the 5'-flanking regions of genes like chick β-globin (McGhee and Ginder, 1979) and ovalbumin (Mandel and Chambon, 1979). In these studies and in many others, restriction enzymes that are inhibited by methylation at their recognition site have been used to test for demethylation, and therefore only a limited proportion of all the CpGs are examined. This may explain the failure to observe differences in some activated genes (Macleod and Bird, 1983; Gerber-Huber *et al.*, 1983). However, it has now become possible to examine all the methylatable sites by genomic sequencing (Church and Gilbert, 1984), which may clarify the situation.

More recent investigation has suggested that many of the unmethylated CpGs occur at high density in particular regions of DNA. Stretches of several hundred bases in length have been found ('HTF islands'; Bird *et al.*, 1985) that are methylation-free even in sperm, where general methylation is at its highest. These islands are frequently found at the 5' ends of housekeeping, non-tissue-specific and viral genes (Bird, 1986), but only at the ends of a few tissue-specific genes. In one example of the former group, the hamster APRT (adenine phosphoribosyl transferase) gene, specific artificial methylation in the HTF island significantly reduces transcription of transfected DNA (Keshet *et al.*, 1985), a result that has also been demonstrated for some viral genomes (Kruczek and Doerfler, 1983; Harbers *et al.*, 1981). From an

evolutionary point of view, as eukaryotic DNA repair mechanisms can replace 5-methylcytosine with a thymine residue, dense clusters of CpGs should remain much more stable in an unmethylated form. If the regions were to become methylated they would mutate with time and the sequence of the 5' region of the gene would change. Therefore, it has been suggested that the original function of HTF islands might have been as an evolutionary silencing mechanism for gene loci.

There is still only limited evidence for a few specific genes that undermethylation at HTF islands can also directly lead to expression. In an alternative scenario, which would link undermethylation to expression, non-nucleosomal specific transcription-activating factors might bind to the islands and prevent access of methylating enzymes. The role of methylation in controlling expression of tissue-specific genes, which often contain a more limited array of methylation-free CpGs, is even more uncertain. Many of the early experiments which suggested that there might be a direct causal link involved treatment of cells with 5-azacytidine (see Jones, 1985). This chemical inhibits DNA methylation and so its induction of differentiation in some cells was interpreted as an immediate consequence of undermethylation. However, it undoubtedly has other effects on cells (see Kolata, 1984), which might be responsible for the changes observed.

Experiments of this kind have, therefore, been superceded by studies of the relationship between the methylation and expression of single well-characterised genes. Busslinger *et al.* (1983), for example, found that methylation at the 5' end of a human γ-globin gene inhibited transcription in transformed mouse L cells. Keshet *et al.* (1986) have also introduced fully methylated and unmethylated cloned tissue-specific

genes (skeletal α-actin and β-globin) into L cells. They have found that methylation affects the properties of the resulting integrated genes, so that they no longer exhibit the nuclease sensitivities, which are characteristic of the activated state. At first sight, these experiments would appear to support a role for methylation in tissue-specific expression, but L cells represent a fibroblast line, which may well not contain any of the globin-specific factors and actin-specific factors found in erythroid cells and myogenic cells respectively. The genes may therefore be under a type of housekeeping control in these examples.

Indeed, recent transformation experiments with a methylated fusion gene containing the 5' end of the rat skeletal actin gene have demonstrated that L cells and myoblasts treat such a methylated construct differently (Yisraeli *et al.*, 1986). In the former cell type, as with the experiments already discussed, expression is inhibited by methylation, but in the latter cells, specific sites are demethylated, which correspond to those modified on gene activation *in vivo*, and the fusion is transcribed at a normal level. In this experiment, it is unclear if the loss of methylated cytosines preceded expression.

In summary, a number of general structural changes have been directly correlated with the activation of a gene locus. Many of these may be forerunners of events that involve more gene-specific *trans*-acting factors. There is accumulating evidence that some of the features, like the alterations in histone H1 binding, supercoiling and methylation can, at least in some cases, directly affect gene expression by releasing the locus from a general repression mechanism. Changes in methylation state

may be particularly important as they can in principle be inherited on cell division, an essential requirement in eukaryotic developmental gene regulation.

However, taking methylation as an example, if the initial activation of a gene resulted from demethylation (i.e. this occurred before any gene-specific factors were bound), presumably highly specific demethylases would be needed to ensure proper gene regulation. A model of this kind would seem highly unlikely. In an alternative and more plausible mechanism, perhaps a more general effect like the unfolding of a chromatin domain might be the first step in gene activation. Such a process might take place during DNA replication, and might be induced by the binding of a tissue-specific factor or possibly multiple factors to form a relatively stable complex. This could allow other structural changes to take place which might further stabilise the active state. For example, demethylation might occur, because particular methylation sites were protected by factors. As other regulators might eventually need to interact for transcription to occur, this model would also explain the fact that DNase I sensitivity, a property that may correlate with changes in the chromatin domain, sometimes only reflects a potential to be transcribed. Certainly an explanation of this nature accounts for the considerable difficulty in demonstrating a primary causal event which leads to gene expression, because it suggests that a number of changes can all have at least a stabilising effect. The confirmation that chromatin domains play an important and fundamental role will require a more detailed understanding of their structure. Recent results indicate that this may not be long in coming (see Section 1.4.2).

1.4.2 The Sequences Required for the Transcription of Eukaryotic Genes by RNA Polymerase II and the Molecules which Interact with Them

In the light of the previous discussion, eukaryotic gene activation could be broadly divided into two cooperative forms of regulatory mechanism, the release from general repression and the enhancement of transcriptional initiation by binding of activating molecules to specific sequences in the gene. The cloning of complete gene loci has led to considerable progress in our understanding of these processes, and some of the relevant results are considered here.

There is accumulating evidence (see Section 1.4.1) that the nuclear matrix is not an artifact of preparation procedures, and that it may act as a framework on which transcription processes take place. However, nothing is known about the interactions between transcription complexes and the matrix.

In addition, it has been suggested that permanent attachments to the nuclear matrix may serve as the anchorage sites at the ends of chromatin domains. They have been found at the 5′ and 3′ ends of the few complete gene loci that have been closely examined (Gasser and Laemmli, 1986b), but one has also been demonstrated within a transcription unit (Cockerill and Gerrard, 1986). Gasser and Laemmli (1986b) have noted that, when the *Drosophila* segmentation gene, *fushi tarazu*, was reintroduced into flies by P element transformation (Hiromi *et al.*, 1985), the removal of the 3′ attachment site could lead to position effects. This result suggests that the domain defined by the anchorage

sites may have an important functional role. However, one problem with the present methodology available for detecting these regions is that the sites can bind to the nuclear matrix during preparation procedures, so it cannot be guaranteed that they are truly permanent in all cells.

The sequences of some of the putative attachment sites, known as MARs (matrix association regions) or SARs (scaffold association regions), which are close to identified genes, are now being elucidated in both mammals and *Drosophila* (Cockerill and Gerrard,1986; Gasser and Laemmli, 1986a, 1986b). Each covers a stretch of roughly 200 base pairs consisting of A/T-rich DNA. Interestingly, they all usually contain more than one 15 base sequence with a very high degree of homology to the consensus recognition site for *Drosophila* topoisomerase II (Sander and Hsieh, 1985). This enzyme can introduce negative supercoils into DNA and has recently been demonstrated to be a major structural component in interphase nuclei (Berrios *et al.*, 1985) and mitotic chromosomes (Earnshaw *et al.*, 1985), where it may be localised at the bases of DNA loops (Earnshaw and Heck, 1985). It could easily be envisaged that topoisomerase II might be responsible for the introduction of superhelical tension into individual domains that would favour gene activation. Such a role is supported by experiments in which transcription from injected 5S genes in oocytes is blocked by novobiocin (Ryoji and Worcel, 1984) which specifically inhibits the enzyme. Other consensus sequences have also been observed in MARs, but no function has yet been assigned to them. More remarkably, developmentally regulated transcriptional enhancer sequences appear to be localised adjacent to MARs (Cockerill and Garrard, 1986; Gasser and Laemmli, 1986b), and this applies even when the enhancer is inside the transcription unit (Cockerill and Gerrard, 1986). This may be

particularly relevant to the mechanism of action of enhancers (see below), which can stimulate gene expression at considerable distances from the transcription initiation site.

The sequences implicated in recruitment of RNA polymerase II to a suitable promoter, after a gene locus has been derepressed, have recently been reviewed (Dynan and Tjian, 1985; Voss *et al.*, 1986; Ptashne, 1986). They can be divided into regions that are common to all or many genes, and those which are not and may therefore bind to specific regulatory molecules (see Figure 1.1). The latter include some elements that must reside near the start of transcription to exert an effect and others, the enhancers, which can operate at much greater distances.

The general elements were first recognised by sequence comparisons of many different genes transcribed by RNA polymerase II (Breathnach and Chambon, 1981), and include the 'TATA' box (with consensus sequence 5'-TATAAA-3') positioned about 25-30 base pairs upstream from the cap site of most genes. Further upstream, other common motifs are also found in some genes, like the 'CAAT' box, the 'TGGCA' sequence (Dynan, 1985) and 5'-GGGCGG-3' (or the 'GC' box). Like specific regulatory elements, the function of these sequences can be deduced by mutation or deletion in the cloned gene, which is then tested either in transformed cells or *in vitro*. The removal of the 'TATA' box often does not markedly affect the rate of transcription but it does lead to inaccurate initiation (Grosschedl and Birnstiel, 1980; Benoist and Chambon, 1981), suggesting that the role of this element is to position RNA polymerase at the correct start site. The 'CAAT' box appears to stimulate the level of

transcription (but not its accuracy) in many cases (Grosveld *et al.*, 1982a, 1982b), but in a few examples (McKnight and Kingsbury, 1982) it has less obvious effects. The sequence 5'-GGGCGG-3' is discussed in more detail in the following paragraphs.

Upstream sequences involved in the specific activation of many cell-type-specific genes have also been characterised by studies in transformed cells and *in vitro* (e.g. Bergman *et al.*, 1984; see Chapter 8). Generally, relatively short regulatory sequences have been identified, some of which can only operate in close proximity to the transcription initiation site. These sequences are frequently shared in homologous genes from different species (see Chapter 8) and, in a few cases, they are also found in genes of the same species that are regulated in a similar way (Friedman and Stark, 1985). The few examples of such elements, which are discussed in the following pages, are selected because there has been some progress in characterising the factors that interact with them. Putative actin-specific elements are considered in detail in Chapter 8.

In a small number of cases, *in vitro* transcription systems have been developed, in which the transcriptional apparatus behaves in a similar fashion to *in vivo*, directing correctly initiated transcription and sometimes requiring specific factors from particular cell types to stimulate expression of selected genes. It has therefore become possible to at least partially purify the active components of such cell extracts (e.g. Parker and Topol, 1984a; Jones *et al.*, 1985). The regions to which these factors bind can be elucidated *in vitro* by DNase I footprinting or other similar methods. Furthermore footprinting

techniques, like genomic sequencing coupled with dimethyl sulphate (DMS) methylation (Church *et al.*, 1985), DNase I digestion (Zinn and Maniatis, 1986; Emerson *et al.*, 1985), and exonuclease III digestion (Wu, 1984a), can be used on native chromatin to confirm that the contacts are shared *in vivo*.

Parker and Topol (1984a) have described a factor in *Drosophila*, which interacts with the 'TATA' box, and Jones *et al.* (1985) have found a protein (CTF) in HeLa cells that binds to the region containing the 'CAAT' box in the herpes simplex I virus thymidine kinase (TK) promoter. Graves *et al.* (1986) have identified a similar 'CAAT' box-binding protein (CTB) in rat liver nuclei. In the TK gene, the 'CAAT' box has been shown to be unnecessary for transcription in *Xenopus* oocytes (McKnight and Kingsbury, 1982), but this of course may not apply to expression in other cell types. Alternatively, sequences adjacent to the 'CAAT' box may be more important in the binding of these two factors.

More detailed work has been undertaken on the factor, known as Sp1, that binds at a short DNA stretch including the sequence 5'-GGGCGG-3' (Dynan and Tjian, 1983; see McKnight and Tjian, 1986). There are six tandem binding sites with different affinities for Sp1 in the 21 bp repeat upstream region of the SV40 early promoter. Deletion studies suggest that they operate independently (Dynan and Tjian, 1985), but that the entire region is vital for high level transcription (Benoist and Chambon, 1981). Further experiments have revealed that this repeated 5'-GGGCGG-3' motif is also present at multiple upstream sites in the herpes virus TK gene and immediate early gene 3, as well as a promoter element in the monkey genome and in the mouse dihydrofolate reductase

gene (Jones and Tjian, 1985; Jones *et al.*, 1985; Dynan *et al.*, 1985; Dynan *et al.*, 1986). In all these cases it has been demonstrated that Sp1 can interact with these sequences *in vitro*. However, unlike the multiple repeats in SV40, the binding sites do not overlap with each other. Numerous other examples of 5'-GGGCGG-3'-containing cellular promoters have been described (see Dynan and Tjian, 1985) indicating that this sequence is a general upstream regulatory element, but it has not yet been directly demonstrated that any of these other promoters can bind to Sp1. However, at least for the TK gene, which is the most extensively characterised, it is clear that the other factor (called CTF) positioned at the 'CAAT' box between the two Sp1 molecules is also required for optimal expression (Jones *et al.*, 1985), and it will be interesting to discover if this interacts with any of the other related promoters.

Experiments have now been performed. which shed some light on the way in which Sp1 operates to enhance the rate of transcription in transformed cells (Takahashi *et al.*, 1986). They are based on the principle that if two binding factors interact with each other directly, increasing their spacing by odd multiples of half turns of DNA should inhibit this process much more than increasing it by even multiples, because in the former case the orientation of the two factors is affected. Indeed, Takahashi and his colleagues have shown that insertions between the 'TATA' box and the 21 bp repeat regions in SV40 that maintain the helical pitch have little effect on the efficiency of transcription, whereas those which alter the pitch decrease the efficiency by up to 10 fold. It therefore appears likely that protein-protein contacts are required for Sp1 to be effective, although other explanations for this phenomenon are possible. Furthermore, as additional intervening whole

turns are allowed, it must be assumed that the inserted DNA can be looped out when the protein interactions are established. Such a model may well be applicable to other general and tissue-specific factors that operate in close proximity to promoter sequences. However, similar experiments have not yet been performed with other characterised systems.

A recent report (Briggs *et al.*, 1986) has described the purification of Sp1 by sequence-specific DNA affinity chromatography. Like many transcription factors it is at particularly low cellular concentrations, and so very large quantities of starting material were required just to obtain a few micrograms of the final product. However, now purification has been achieved, microsequencing of the protein should be possible, and it should not be long before the Sp1 gene is isolated.

A few factors that bind to highly specific upstream control elements have also been identified. Two groups (Parker and Topol, 1984b; Wu, 1984b) have isolated a factor that attaches to a specific conserved sequence in the heat shock genes of *Drosophila*. Parker and Topol used an *in vitro* transcription system to assay for the factor, whereas Wu merely tested for specific DNA binding, which mirrored the binding observed *in vivo*. In another different approach, Emerson and Felsenfeld (1984) have partially purified a factor which recreates the DNase I hypersensitive region in the chick adult β-globin gene. To do this, the mixture of gene and factor were incubated in a *Xenopus* oocyte extract containing histone assembly factors (Laskey *et al.*, 1977) with added chicken histones. Further work (Emerson *et al.*, 1985) has demonstrated that the footprint created by *in vitro* binding of the factor closely resembles the pattern from the isolated chromatin of erythroid cells.

Assays, like those described above, which reproduce some property observed *in vivo*, are clearly important methods for confirming that a putative factor is active in normal cells. However, the complex mixture of proteins in a crude nuclear extract frequently makes it difficult to recognise a specific component by these tests. An obvious alternative is to select factors solely on the basis of their binding to a specific DNA fragment, and only after some purification to test for similarities to *in vivo* interactions. Until recently, the nitrocellulose filter assay (Jones and Berg, 1966) was the only suitable technique for these studies, but it suffered from a number of problems. For example, it was never clear if DNA would form normal complexes when it was bound to a filter, and different types of complex could not be distinguished.

The field has now been revolutionised by the use of gel retardation electrophoresis (see Garner and Revzin, 1986), which is not affected by these problems. Basically, the relevant region of DNA (present in a large excess of competitor DNA) is mixed with a nuclear extract, and then the sample is fractionated on a low ionic strength polyacrylamide gel. After Southern blotting and hybridisation of the filter to a specific probe, the stretch of DNA under examination can be visualised. If complexes have formed, these retard the migration of the DNA.

In eukaryotes, a fine example of the application of this technique is illustrated by the study of mammalian immunoglobulin genes. An octamer DNA sequence, which is present in immunoglobulin gene promoter sequences and in the heavy chain transcriptional enhancer, has been shown to be involved in the lymphoid-specific transcription of immunoglobulin genes by transformation studies (see Bergman *et al.*, 1984). Gel retardation

assays have identified sequence-specific binding factors, one of which is found in non-lymphoid as well as lymphoid cells (IgNF-A; Singh *et al.*, 1986), while the other is only produced by lymphoid cells (IgNF-B; Staudt *et al.*, 1986; see also Landolfi *et al.*, 1986). As yet, however, neither of these factors have been shown to stimulate immunoglobulin gene expression directly.

The existence of a putative binding factor outside lymphoid cells is not totally surprising. The octamer motif is not limited to immunoglobulin genes, but is also found, for example, in *Xenopus* U1 and U2 small nuclear RNA genes, where it is required for accurate transcription (Mattaj *et al.*, 1985). Although it is reassuring that a lymphoid-specific factor is synthesised as well in this case, it has already been explained in a previous section that any particular regulatory factor need not be limited to the cell types in which a particular gene that it activates is expressed (see Section 1.2) As a further example of this, Plumb *et al.* (1985) have found that the globin-specific factor described by Emerson and Felsenfeld (1984) is weakly produced by thymus cells as well as erythrocytes.

In addition to the vertebrate factors isolated on the basis of their binding activity or their stimulatory effect on gene expression, some regulators have been directly cloned by studies of mutations in other eukaryotes that have a more complete genetic map (see Section 1.2.1). A number of these share characteristics that are associated with DNA binding. The homeo domain of many developmental regulators in *Drosophila* and of proteins in other organisms, for example, almost certainly has a DNA-binding function. However, other regions in each protein presumably provide specificity, either by selective DNA

interaction or by allowing specialised contacts with other regulatory factors. The metal-binding fingers of the *Krüppel* gene product are also characteristic of a protein that operates by direct DNA interaction.

However, the best characterised eukaryotic regulatory protein is TFIIIA. Unlike all other vertebrate factors so far recognised, its synthesis is at such high levels in one cell type (it constitutes up to 10% of the total protein in small oocytes) that it was relatively easy to obtain a partial amino acid sequence, synthesise an oligonucleotide probe and subsequently isolate a cDNA made from the message encoding TFIIIA (Ginsberg *et al.*, 1984).

Although TFIIIA interacts with internal control regions in the 5S genes, which are transcribed by RNA polymerase III, mounting evidence suggests that the basic design of its binding domain (see Section 1.2.1) may be a common motif for other regulators, including those that are transcribed by RNA polymerase II. Aside from the presence of binding fingers in other eukaryotic proteins (e.g. Rosenberg *et al.*, 1986), Rhodes and Klug (1986) have recently identified a 5 bp repetitive element in the 5S RNA internal control region at both the DNA sequence and structural levels, which appears to correlate with a binding site for one finger of the TFIIIA molecule. The sequence periodicity is not reflected by obvious homologies, but by a strong tendency for between one and three guanine residues to occur every 5½ nucleotides (half a turn of double-helical DNA). Not only has such a repetitive element been recognised in other genes transcribed by RNA polymerase III, but a similar periodicity is also found in Sp1 binding sites, suggesting that its binding domain may also be based on the finger motif.

A further group of transcriptional control elements are the enhancers. First discovered in SV40, they have now been recognised in other viral and many cellular genes, including, for example, the immunoglobulin and globin genes (see Voss *et al.*, 1986). They can be active in all cells, or, in other examples, they appear to be entirely tissue-specific. Enhancers are characterised by their ability to function at distances of well over 1 kb either upstream or downstream from the cap site. However, it is now clear that the other specific property of bidirectionality, which they alone were thought to have, is not unique to enhancers, but is also found for other regulatory sequences (e.g. Everett *et al.*, 1983). Furthermore, other upstream elements can be separated, albeit to a lesser extent, from the TATA box without affecting efficiency, as discussed previously (Takahashi *et al.*, 1986).

The mechanism by which enhancers can act at such distances is still not determined (see Ptashne, 1986). *In vivo*, many enhancers are relatively close to the promoters they affect, but this is not always the case (e.g. Choi and Engel, 1986). Recent evidence on the SV40 enhancer (Takahashi *et al.*, 1986), which is directly adjacent to the 21 bp repeats of the SV40 promoter, suggests that direct protein-protein interactions may be involved, as adjustment of the spacing between these two regions by odd numbers of half turns decreases transcriptional efficiency.

Furthermore, elegant experiments by Ptashne and his coworkers (Brent and Ptashne, 1985; Keegan *et al.*, 1986) have demonstrated that the yeast GAL4 protein, which binds to an enhancer-type element (UAS_G), can be separated into two distinct functional domains. One is responsible for interaction with the UAS_G and may be comprised of finger binding domains

(Hartshorne *et al.*, 1986). The other can be fused to alternative DNA-binding domains and will still activate yeast promoters if the appropriate binding sequence is placed upstream of the promoter. Therefore at least in this case, interaction with DNA itself would not appear to account for enhancer activity, and it must be proposed that protein-protein contacts can take place even over long distances, presumably by looping of the DNA in between the two regulatory elements.

Such evidence argues against an alternative model, where the enhancer directly affects DNA supercoiling. Of course indirect effects via interaction with other factors could alter supercoiling even in the case of GAL4. But further work by Plon and Wang (1986) in which the SV40 enhancer was topologically separated from an attached gene promoter has shown that it can still function normally when alterations in the supercoiling of the enhancer region are unable to change the superhelical tension in the gene. As distal promoters can be separated from enhancers by other promoters (Atchison and Perry, 1986), it would also seem unlikely that the enhancer serves as a bidirectional entry point for RNA polymerase II or other transcriptional factors, or indeed operates by any mechanism that involves a propagation of a signal up to the nearest promoter.

Although limited activation of transcription can apparently be reproduced by enhancers *in vitro* (Sassone-Corsi *et al.*, 1985), its magnitude compares poorly with transformed cells, and the enhancer will not operate at distances from the promoter that are acceptable *in vivo*. The latter property therefore requires conditions that cannot yet be reproduced outside the cell. In this respect, the linkage of the enhancer to nuclear matrix association regions (Gasser and Laemmli,

1986; see earlier in this section) may be an important factor in bringing the enhancer and the promoter into close proximity in a looping model of activation, which could not currently be reproduced *in vitro*.

The SV40 enhancer is strongly activated in most cell types, but other viral and cellular enhancers are much more restricted. Evidence for the existence of limiting *trans*-acting cellular factors first came from competition experiments *in vivo* and *in vitro* (Sassone-Corsi *et al.*, 1985; Scholer and Gruss, 1985) with the SV40 enhancer. Few site-specific, enhancer-binding, activating factors have yet been directly identified. However, the glucocorticoid receptor is known to interact with the enhancers of murine mammary tumour virus (Payvar *et al.*, 1983; Scheidereit *et al.*, 1983) and the human metallothionein II_A gene (Karin *et al.*, 1984), and recently a factor has been defined that binds to the immunoglobulin heavy-chain enhancer (Weinberger *et al.*, 1986).

However, some generally important features of enhancers have become apparent. As they show a wide range of tissue-specificities in their activities, it is not surprising that no overall consensus sequence has been demonstrated. Nevertheless, Voss *et al.* (1986) have suggested that small elements may be conserved in various enhancers, even though they are flanked by non-conserved regions.

DNase I footprinting and DMS methylation protection experiments with nuclear extracts from a number of cell types (e.g. Davidson *et al.*, 1986) have shown that different *trans*-acting factors may bind to the SV40 enhancer in different cells. Furthermore, activation of many enhancers may require the binding of a series of different factors along

their length. For example, the heavy chain immunoglobulin enhancer contains four similar binding regions (Ephrussi *et al.*, 1985; Church *et al.*, 1985), but these are recognised by different factors (Weinberger *et al.*, 1986). Indeed, mutational analysis has demonstrated that part of this enhancer can activate transcription in all cell types, not just lymphoid cells (Wasylyk and Wasylyk, 1986), and one of the binding factors can interact with viral enhancers. Similarly, deletion analysis of the human β-interferon enhancer has revealed a constitutive enhancer element with homologies to viral enhancers, and a downstream region, which reduces transcriptional activation in cells that are not stimulated by viral infection or double-stranded RNA (Goodbourn *et al.*, 1986).

Therefore, binding sites for both general and specific *trans*-acting factors exist in enhancers in a comparable way to the situation at the 5' flanking regions of many genes. The best characterised examples of tissue-specific or inducible enhancers support a model in which constitutive enhancer elements are controlled by repressors interacting at adjacent sites.

There may also be only a limited period of time during which the enhancer is needed to operate. Immunoglobulin heavy chain genes that have lost their enhancers *in vivo* by deletion in lymphoid cells still express at high levels (Klein *et al.*, 1984; Wabl and Burrows, 1984). But, when the resulting gene is cloned and reintroduced into cell lines, it is not transcribed (Klein *et al.*, 1985), suggesting that the role of the enhancer is for gene activation and not maintenance of expression. In other transfection experiments by Mason *et al.* (1985), it has been shown that the immunoglobulin heavy chain enhancer can be replaced by

the SV40 enhancer and the promoter will still confer tissue-specificity. Taken together, these reports would tend to suggest that an enhancer activity is necessary, at least during immunoglobulin heavy chain gene activation, but surprisingly this activity does not need to be tissue-specific.

Figure 1.1 presents a diagram which summarises the arrangement of a gene transcribed by RNA polymerase II and the sequences involved in its regulation. The most noticeable feature of such a gene is that it can be influenced by a large number of different control sequences and each of these may require a group of *trans*-acting factors to set them into operation. This property may be an important factor in the inheritance of transcriptional activity during DNA replication, as the numerous gene regulators can be divided between the two resulting duplexes, and they may therefore maintain both daughter duplexes in an active configuration.

A system of multiple activation steps can also account for the stringent regulation of genes in specific cell types. However, in order to keep the number of factors within reasonable limits, some activators must almost certainly interact with batteries of genes. At present, however, only limited homologies have been recognised between different genes which follow a similar developmental pattern of expression. Many genes that are regulated by α-interferon share a common sequence in their promoter regions (Friedman and Stark, 1985), and putative homologous sequences are shared by some muscle-specific genes (see Chapter 8). Generally it is possible that the shared elements may not be as obvious as these examples, particularly if they are encroached upon by other

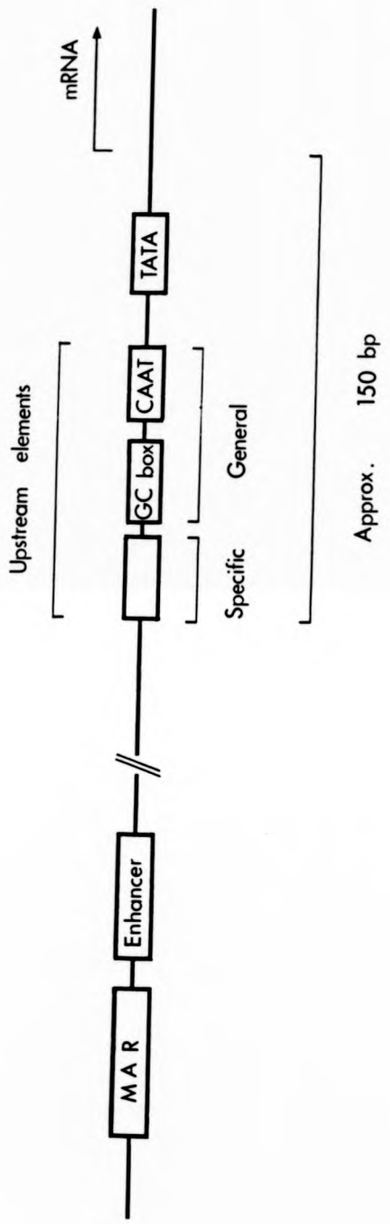


Figure 1.1 Schematic Diagram of the General Organisation of a Gene Transcribed by RNA Polymerase II

This diagram illustrates the different types of control sequences that can affect the expression from a gene locus which is transcribed by RNA polymerase II. In addition to the 'TATA' box, which is usually positioned about 25-30 bases upstream of the start of transcription (Breathnach and Chambon, 1981), there are commonly a number of other upstream elements, typically within the first 150 bp of the initiation site. Some of the latter are shared by many genes (e.g. the 'CAAT' box [Breathnach and Chambon, 1981] and the 'GC' box [Dynan and Tjian, 1983], otherwise known as 5'-GGGCGG-3'). However, others are much more specific and may serve to regulate the gene in a tissue-specific fashion. For example, a specific sequence in *Drosophila* heat shock genes has been demonstrated to bind a heat shock transcription factor which can activate transcription *in vitro* (Parker and Topol, 1984b).

The enhancer sequence (see Voss *et al.*, 1986), unlike the other regulatory elements, can be positioned at considerable distances (i.e. a few kilobases) from the start of transcription, either upstream or downstream, without affecting its activity in the appropriate cell type. Recent evidence (e.g. Gasser and Laemmli, 1986b) has suggested that enhancers may frequently be adjacent to a nuclear matrix association region (MAR). This may have important consequences in accounting for the way in which the enhancer element functions (see text).

control elements.

Clearly, in comparison to other lower eukaryotic organisms, vertebrates are generally much more suited to studies, where the regulated gene is examined, as opposed to studies where the function of a regulator is under investigation (see Section 1.2). When tissue-specific or inducible factors are recognised in vertebrates by transcription studies, they usually appear to be maintained at very low concentrations. It has therefore not yet been possible to clone them, either by the use of antibodies or microsequencing of the protein. As has been demonstrated with TFIIIA and the homeo-box genes, cloning is a necessary step if the mechanism of action of a regulator is to be elucidated. One of the most important consequences of the work in this thesis is that, in confirming that genes transformed into *Xenopus* can be regulated normally, it suggests a novel method by which the regulatory factors may be cloned (see Chapter 8).

SECTION 1.5 THE STRUCTURE, EXPRESSION AND FUNCTION OF ACTIN PROTEINS AND THEIR GENES

1.5.1 The Functions of Actin in Eukaryotic Cells

Actin is a structural protein found ubiquitously in eukaryotic cells. It exists in a small number of very closely related isoforms, which are expressed from different genes (see Section 1.5.3) and are localised to different cell types. All these isoforms can be maintained as monomers (G-actin) with a molecular weight of approximately 42,000 daltons, or they can polymerise to form filaments (F-actin).

In sarcomeric muscle (skeletal and heart muscle), F-actin is relatively stable and is arranged in a regular array of thin filaments, which interdigitate with a similarly regular set of thick myosin filaments. Although the exact details of the mechanism of muscular contraction remain controversial (e.g. see Clarke, 1984), Huxley (1969) demonstrated many years ago that shortening was achieved by the sliding of thin filaments over thick filaments. Myosin is a complex of light- and heavy-chain proteins, and is attached to actin filaments by a globular domain that is hinged to the rest of the complex. It is believed that a cyclic process of binding, rotation about the hinge and dissociation drives the sliding action of contraction. Such a mechanism requires the hydrolysis of ATP, and is inhibited by decreasing intracellular calcium ion concentration. In molluscs, calcium directly affects the binding affinity of myosin for actin in muscle, but in vertebrate striated muscle the regulation is mediated by the control proteins troponin and tropomyosin, which can block the myosin binding sites on actin

molecules.

As with the mollusc muscle, attachment of myosin to actin and subsequent movement in smooth muscle and non-muscle cells is effected by the regulation of thick filaments by calcium. However, the exact processes involved are much less clear. One established part of the mechanism (see Adelstein, 1982) involves the activation of myosin light-chain kinase by interaction with a complex between calcium and a calcium-binding protein called calmodulin. This leads to the phosphorylation of myosin, which stimulates the actin-activated myosin ATPase. More recent evidence (see Bretscher, 1986) has suggested that another protein called caldesmon may interact with actin (and not myosin) to block actomyosin formation. This activity can be inhibited either by direct binding of free caldesmon to calcium-calmodulin or by reversible phosphorylation of caldesmon via a calcium-calmodulin-regulated specific kinase.

Apart from the obvious role of actin in muscle contraction, the protein is also a vital component of the cytoskeleton in non-muscle cells. It is involved in different aspects of cell motility, including cell movement, cytokinesis, cytoplasmic transport, secretion and phagocytosis (e.g. Goldman *et al.*, 1976; Stossel, 1984; Wang, 1984). It is organised in the cell in a number of filamentous networks, and other structures like stress fibres (e.g. Lazarides and Weber, 1974), which are usually associated with the cell membrane. Polymerised actin filaments are in dynamic equilibrium with the cytoplasmic monomeric pool (Pollard and Craig, 1982) and a series of actin-binding proteins (Craig and Pollard, 1983) have been identified, which can affect the equilibrium position of this process. Tubulin polymerisation is also controlled in a similar way, and Kirschner and Mitchison (1986) have recently suggested that

simple molecular interactions with proteins that bind tubulin can explain many of the processes involved in the complex microtubule reorganisation that occurs during mitosis. Furthermore, Vale *et al.* (1985) have identified specific proteins which direct movement in opposite directions along microtubules in the squid axon. It may therefore be possible in the future to account for apparently complex actin-related cytoskeletal events by reductionist model systems of a similar kind (e.g. Bray, 1985).

Non-structural roles for actin have also been proposed. In particular, it has been shown that RNA polymerase II requires cytoskeletal β- and γ-actins as cofactors for correct initiation of transcription *in vitro* (Egly *et al.*, 1984), and it has also been suggested that actin may regulate poly(A) polymerase (Schröder *et al.*, 1982). Furthermore, Scheer *et al.* (1984) found that injection of actin-specific antibodies or actin-binding proteins into oocytes of Pleurodeles led to an inhibition of transcription by RNA polymerase II, a retraction of lampbrush loops and general chromosome condensation.

Possibly another relevant observation is that mutant actin genes in the indirect flight muscle of *Drosophila* not only cause flightlessness, but also activate heat shock genes constitutively (Hiromi and Hotta, 1985). However, although it has been confirmed by P element transformation that it is the mutant protein which is responsible for this reaction (Hiromi *et al.*, 1986), it has still not been demonstrated that actin directly affects transcription at the heat shock locus. Indeed the activation may result from a ubiquitin-related response to the synthesis of an unstable protein in the cell, and therefore may not be associated with the action of nuclear actin at all.

1.5.2 The Actin Isotypes and the Localisation of their Expression

The amino acid sequences of actins in many eukaryotes have now been determined both directly and by DNA sequence analysis. The level of sequence conservation is extremely high. Even the actins of yeast and mammals differ in only about 10% of their amino acid residues (Gallwitz and Sures, 1980). Actin is known to interact with numerous different molecules in the cell, and the exceptional conservation may reflect the preservation of many different binding domains in the protein. In this section, I will concentrate particularly on vertebrate actins, briefly summarising the established categorisation of mammalian isoforms and the localised expression of these proteins, and discussing recently discovered exceptions to the general model.

The existence of multiple actin isotypes in mammals was first recognised from variability of protein migration on isoelectric focussing gels (Garrels and Gibson, 1976; Whalen *et al.*, 1976). By this criterion, at least three different molecules (α, β and γ) were identified. However, by sequencing different isoforms from a number of tissues, Vandekerckhove and Weber (1978a, 1978b, 1978c, 1979) have shown that mammals synthesise at least six actins (i.e. more than one actin protein can migrate to the same position on an isoelectric focussing gel; see Table 1.1). Classically there is an α-actin sequence specific to adult skeletal muscle and another one that is specific to heart muscle. Smooth muscle contains two isoforms, an α-type that predominates in the aorta for example, and a γ-type that is at higher relative levels in the stomach. Finally two cytoskeletal actins (a β-type and a γ-type) have also been recognised. The latter have been shown to be coexpressed in a

Table 1.1 The Classical Mammalian Actin Isoforms

	Type of Muscle	Isoform	N-Terminal Amino Acids of Mature Proteins	Tissues of Origin
MUSCLE FORMS	Sarcomeric	α-skeletal	H_2N-A-G-A-G	skeletal muscle
		α-cardiac	H_2N-A-A-G-G	cardiac muscle
	Smooth	α-smooth	H_2N-G-G-G-A	aorta
		γ-smooth	H_2N-G-G-G	stomach
NON-MUSCLE FORMS		β-cytoskeletal	H_2N-A-A-A	most tissues
		γ-cytoskeletal	H_2N-G-G-G	most tissues

Table 1.1 The Classical Mammalian Actin Isoforms

The six classical actin isoforms in mammals have been defined by three different criteria; firstly the migration of the protein on isoelectric focussing gels (α, β or γ: Garrels and Gibson, 1976; Whalen *et al.*, 1976), secondly the N-terminal amino acid sequence of the mature protein (see Vandekerckhove and Weber, 1978a) and thirdly the localisation of expression. All these characteristics are listed here. The N-terminal amino acid sequence (A = aspartic acid; G = glutamic acid) is essentially responsible for changes in the isoelectric point of the actin molecule and hence the migration of the protein during isoelectric focussing (e.g. all the α-types have an extra acidic amino acid, and so they migrate further in the acidic direction on the gel).

For the muscle isoforms, the tissues of origin are adult tissues in which the particular isoform is the most abundant. However, as discussed in the text, there is not an absolute distinction between the tissues of expression for the two sarcomeric genes or the two smooth muscle genes, as assessed by both RNA (e.g. Minty *et al.*, 1982; Gunning *et al.*, 1983) and protein (Buckingham *et al.*, 1986) distributions. Furthermore, evidence from non-mammalian vertebrates suggests that actins with other N-terminal sequences can adequately replace or complement the standard six isoforms (see Section 1.5.2).

single cell type (Bravo *et al.*, 1981).

Like actin, different isoforms of other muscle-specific proteins, including myosin and tropomyosin, are found in skeletal, cardiac and smooth muscle cells, as well as non-muscle cells (see Buckingham and Minty, 1983). However, the same actin has been isolated from fast and slow skeletal muscle, whereas myosin, tropomyosin and troponin have separate fast and slow isoforms. The six actin isoforms are totally conserved in all warm-blooded vertebrates examined so far. Furthermore, the conservation of amino acid sequences between different isoforms is also extremely high. For instance, only 4 out of 375 amino acids differ between cardiac and skeletal actins, and, in the most widely divergent example, only 24 and 25 amino acids differ between the cytoskeletal, and the cardiac and skeletal actins respectively (Vandekerckhove and Weber, 1979). A particularly divergent region of the actin molecule is the N-terminus. The first three or four amino acids of the mature protein are characteristic of each isoform. They are always acidic residues (either glutamate or aspartate), but their exact sequence varies between different isoforms (Table 1.1). These changes essentially account for the variable mobility of the proteins on isoelectric focussing gels. Vandekerckhove and Weber (1981a) have consequently developed a simple assay involving protein-chemical analysis of the N-terminal tryptic peptide to identify other types of actin in different organisms.

More recent studies have demonstrated that both the strict tissue-specificity of expression of the sarcomeric actins and the existence of only six isoforms of actin described in the preceding paragraphs represent an oversimplification of the situation in vertebrates. The

cardiac and skeletal actin genes are co-expressed in the skeletal and heart muscle of the fetal mouse (Minty *et al.*, 1982). Indeed, in mouse fetal skeletal muscle, cardiac actin transcripts are a major component (30-40% of the total muscle actin message in 17-20 day old fetuses). Moreover, in chicken skeletal muscle, they account for more than 90% of actin mRNA during early development (Paterson and Eldridge, 1984). As a general rule, it would appear that the mRNAs for both sarcomeric isoforms become increasingly localised to either skeletal or heart muscle during development, and this is also observed in *Xenopus* (see Chapter 5). However, even in some vertebrates at the adult stage, the two genes may still be coexpressed in both tissue types, for example in the heart and skeletal muscle of man (Gunning *et al.*, 1983) and in the rat heart (Mayer *et al.*, 1984).

Therefore, like the two smooth muscle isoforms and the two cytoskeletal isoforms (see Vandekerckhove and Weber, 1981a), the sarcomeric actins are coexpressed at variable ratios within different cell types, assuming both messages are translated at comparable rates in any one tissue. In this regard, Vandekerckhove *et al.* (1986) have recently confirmed that the changes in the ratio of the two transcripts in skeletal muscle during the development of vertebrates like the chick and the mouse also appear to be reflected by similar changes in the ratio of the two proteins. In addition to the coexpression of similar isoforms, cytoskeletal actins are also found in muscle (Hall *et al.*, 1981), although they are not associated with the sarcomeres, but with the membrane structures in the cells.

As well as the standard six isoforms of actin, Bergsma *et al.* (1985)

have recently recognised a transcript encoding another actin subtype in chicken. Its translated amino acid sequence is generally consistent with that of a cytoskeletal actin and the localisation of its RNA expression supports this view, although no translation product has yet been detected. The major divergence lies at the N-terminus, where a different arrangement (type 5; Vandekerckhove and Weber, 1981b; see Table 1.2) of glutamates and aspartates is found in the first three acidic amino acids, and an alanine is encoded between this region and the N-terminal methionine. It is assumed that this amino acid is lost by processing during formation of the mature protein.

The possible diversity in the N-terminal structure of vertebrate actins becomes more apparent as more distantly related organisms are examined. In particular, the study of cytoskeletal actins in Amphibia led Vandekerckhove and Weber (1981b) to propose that virtually any order of acidic amino acid residues at the N-terminus of these proteins might be functional. Eight different arrangements are possible (types 1-8; see Table 1.2); they identified five of these in Amphibia, three of which are those present in the chicken, and we have recently found another type in *X. borealis* (type 7; see Appendix I). Indeed some Amphibia, like *X. laevis*, do not appear to need both the β- and γ-type actins (types 1 and 8), but have alternative cytoskeletal actin genes instead. Like the third chicken cytoskeletal isoform, some of these genes encode an N-terminal alanine (Cross *et al.*, 1986; see Table 1.2). At least in the case of *X. laevis*, the protein data of Vandekerckhove and Weber (1981b) demonstrate that this is subsequently removed.

At present it is extremely difficult to devise an evolutionary plan that would explain the diversity of vertebrate cytoskeletal actins (see Cross

Table 1.2 Variation in the N-Terminus of Vertebrate Cytoskeletal Actins

Type	N-terminal amino acids of mature protein	mammals	chick	*Xenopus laevis*	*Xenopus borealis*	*Rana pipiens*	*Triturus cristatus*
1(β)	H_2N-A-A-A	+	+		+ (Ala)		+
2	H_2N-A-A-G						
3	H_2N-A-G-A						+
4	H_2N-G-A-A			+		+	
5	H_2N-A-G-G		+ (Ala)	+ (Ala)		+	
6	H_2N-G-A-G						
7	H_2N-G-G-A				+		
8(γ)	H_2N-G-G-G	+	+	+ (Ala)	+ (Ala)		+

Table 1.2 Variation in the N-Terminus of Vertebrate Cytoskeletal Actins

Vandekerckhove and Weber (1981b) have suggested eight possible cytoskeletal actin isoforms, which differ only in the sequence of the first three acidic residues at the N-terminus of the mature protein (types 1 to 8). These are listed, together with their proposed N-terminal structure (A = aspartic acid; G = glutamic acid). Results from Vandekerckhove and Weber (1981b), Bergsma *et al.* (1985) and Cross *et al.* (1986) [Appendix I] suggest that at least six of these isoforms are found *in vivo*, and the organisms in which they have been demonstrated are marked. As can be seen, the cytoskeletal isoforms of amphibia are extremely variable.

Those genes which encode an alanine after the N-terminal methionine, and before the acidic residues, are also indicated (Ala: Cross *et al.*, 1986; Bergsma *et al.*, 1985). The genes of *Rana pipiens* and *Triturus cristatus* have not yet been examined. In the case of the *X. laevis* type 5 and type 8 actins, it is known that the alanine is removed during maturation of the protein.

et al., 1986 for discussion - Appendix I). However, although warm-blooded vertebrates may all share two common and major isoforms, it is clear that at least in amphibians the diagnostic N-terminal sequences of these molecules are not absolutely necessary for proper function.

Other studies by Vandekerckhove and Weber (1984) on the muscle actins of lower vertebrates, chordates and invertebrates have suggested that a muscle-type actin like that found in warm-blooded vertebrates arose prior to the divergence of jawless fish from the rest of the craniates. However, it was not present in invertebrates, which synthesise muscle actins that are more closely related to the vertebrate cytoskeletal actins. Both a sarcomeric and a smooth muscle actin protein are observed in a shark (Condrichthyes), and separate skeletal and cardiac isotypes are found in reptiles. They have identical N-termini to those of higher vertebrates. By contrast, some Amphibia, specifically *X. laevis* (Vandekerckhove and Weber, 1981b; Stutz and Spohr, 1986) and *X. borealis* (Chapters 4 and 5; M. Boardman, per. comm.) also have a skeletal and cardiac isotype, but the skeletal actin N-terminus differs from that of its vertebrate counterpart. It would therefore appear possible that the evolution of two distinct sarcomeric actin isoforms in Amphibia occurred separately from other higher vertebrates, like reptiles and mammals, presumably after the two lines diverged. This would account for the unusual skeletal actin in Amphibia.

In conclusion, the simple model, which was generally accepted only a few years ago, of six vertebrate actins with the two sarcomeric actins being expressed specifically either in skeletal or in cardiac muscle cells does not correlate with more recent observations. The way in which the

isoforms have diverged and evolved is still unclear, but the newly discovered anomalies pose two major questions. Firstly, how are actin messages regulated so that they are coexpressed at some stages of development but differentially expressed at others? Present information bearing on this question is summarised later in the chapter and in Chapter 8. This phenomenon clearly makes the actins a particularly interesting group of related proteins in which to study differential expression. Secondly, if a short length of coding region, which is characteristic of a specific isoform in most vertebrates, can be replaced in other vertebrates, or additional isoforms can also be expressed, how important is the amino acid sequence it encodes to the function of the protein?

The striking isotype-specific conservation of warm-blooded vertebrate actins not surprisingly led many workers to conclude that the subtle differences between the isoforms had to be related to vital aspects of their individual functions. By the same reasoning, the identical regions were probably involved in equally important interactions, in which all the isotypes participated in a similar fashion. However, the alternative explanation for any differences is that they occur in parts of the actin molecule where protein function is not affected, and the conservation of the isotype-specific amino acid sequences in a single isotype is related to factors other than the function of the protein. Although such a hypothesis is partly supported by the discovery in Amphibia of unusual actin molecules, which vary even between closely related species, it could always be argued that these are the result of important adaptational changes, which are not required by other vertebrates.

However, additional indirect evidence bears some further light on this important question. For example, invertebrates, including *Drosophila* (Fyrberg *et al.*, 1981), express muscle-specific actins, which have an amino acid sequence more closely related to vertebrate cytoskeletal actins than muscle actins, but they can still produce functional myofibrillae. Furthermore, the yeast genome only encodes one actin gene (Ng and Abelson, 1980; Gallwitz and Seidel, 1980), and does not require the two cytoskeletal isoforms that are found in vertebrates. Indeed, a recent study using antibodies specific to the γ-cytoskeletal actin suggests that it is distributed in similar cellular locations to β-cytoskeletal actin (Otey *et al.*, 1986). This result implies that there is no subcellular sorting of non-muscle actins. As the ratio of the vertebrate cytoskeletal isoforms is highly variable between different cell types (Garrels and Gibson, 1976; Vandekerckhove and Weber, 1981a) and under changing physiological conditions (e.g. Leavitt *et al.*, 1980), it is possible that the presence of two genes in these organisms might really be required for the regulation of overall cytoskeletal actin gene expression in the large variety of cell environments.

If the differences between actin molecules do not affect the function of the protein, they should all be able to integrate into the same structures. Pardo *et al.* (1983) found that γ-actin expressed in mouse diaphragm muscle cells was localised to mitochondria and not present in sarcomeres. They concluded that subcellular sorting of muscle and non-muscle actins did take place, and this might be construed as a demonstration that these different isotypes cannot be mixed in the same structures. However, such selectivity need not be determined at the protein level. Recently Lawrence and Singer (1986) have shown that the intracellular distribution of cytoskeletal actin, which favours

localisation at cell extremities, is reflected in the distribution of its mRNA. A similar mechanism regulated by any conserved region in the message might lead to sorting of different isotypes.

Experiments leading to a resolution of the whole problem will involve the expression of transformed engineered actin genes in various cell types. When the cloned human cardiac actin gene was reintroduced into mouse L cells, it expressed the normal protein and the latter participated in a similar fashion to the endogenous cytoskeletal actins in the formation of the cytoskeleton, as judged by Triton X-100 insolubility (Gunning *et al.*, 1984b). This suggests that there is no clear subcellular sorting of the muscle-specific isoform away from cytoskeletal structures when it is expressed in non-muscle cells. However, although no adverse effects on the cell were observed, such an approach does not directly investigate the capability of the protein to completely replace the endogenous actin at a functional level, because the endogenous isoform is still present.

In one example at least, this problem may well be overcome in the near future by using a null allele of the indirect flight muscle (IFM) actin gene in *Drosophila* (see Hiromi *et al.*, 1986). The resulting mutant flies can be rescued by P element transformation with functional copies of the IFM actin gene, which encodes a muscle-specific isotype more closely related to vertebrate cytoskeletal actins than vertebrate muscle actins. It will be informative to discover if vertebrate muscle actin coding sequences linked to the appropriate IFM regulatory sequences will function in the same way. These experiments again illustrate the advantages of using an organism, in which mutations that inactivate a specific gene locus are frequently available, in order to investigate

the function of the gene product.

1.5.3 The Structure and Organisation of the Actin Genes, and their Regulated Expression

The preceding section has considered the multiple isotypes of actin which are synthesised in vertebrates. There are two non-exclusive mechanisms by which these different proteins might be encoded in the genome. First, they may be expressed from the same transcription unit, and the resulting RNA is then differentially processed to produce message for the appropriate isoform. The expression of a number of developmentally regulated muscle proteins is at least partly controlled in this way, including that of the myosin heavy chain (Rozek and Davidson, 1983), myosin light chains (e.g. Nabeshima *et al.*, 1984; Robert *et al.*, 1984), tropomyosin (Basi *et al.*, 1984) and troponin T (Medford *et al.*, 1984). However, there is no evidence for such a mechanism in the actins. The alternative system is based on a series of genes, each encoding different isoforms. Such multigene families are found for many of the other muscle proteins as well as actins, and also encode other important tissue-specific proteins like the globins, immunoglobulins and vitellogenins.

The high amino acid sequence conservation of all actins has meant that DNA probes from a particular actin gene can usually be used to screen for genes synthesising any form of actin. Indeed, even actin cDNAs from organisms as distantly related as *Dictyostelium* can act as suitable probes for the vertebrate genome (see Cross, 1984). Consequently, rough estimates of actin gene number can be made simply by counting the

hybridising bands on genomic Southern blots (see Buckingham and Minty [1983] for review). In the mouse (Minty *et al.*, 1983) and man (Humphries *et al.*, 1981; Engel *et al.*, 1981), there are at least twenty different gene copies. However, the fact that the hamster can survive with far fewer genes (Dodemont *et al.*, 1982) suggests that they may not all be necessary. Indeed, at least in man, many of the copies appear to be dispersed, processed β-actin pseudogenes (Ng *et al.*, 1985).

Several vertebrate actin genes have now been cloned and sequenced, and a summary of the sequence data currently available is presented in Table 1.3. As this thesis does not directly address the question of gene evolution and structure in *Xenopus*, I will only briefly summarise the general conclusions which have been drawn from this evidence. Further discussion of conserved regions in the promoter of muscle-specific actin genes may be found in Chapter 8.

As different vertebrate actin isoforms are so highly homologous, the coding sequences of their parent genes also share considerable similarities. Even many of the silent sites in different messages from the same species have been shown to be maintained at anomalously high frequency (Schüler *et al.*, 1983; Erba *et al.*, 1986; Alonso *et al.*, 1986). Although this may be related to a direct functional requirement like the conservation of secondary structure in the message, it would seem more likely that recombination events between the different genes (possibly leading to other functional consequences) are largely responsible for these observations. However, comparison of coding sequences for the same isoform in different vertebrates (e.g. Cross, 1984; Alonso *et al.*, 1986) suggests that conservation mechanisms have been employed during the evolution of divergent species to retain

Table 1.3 Reports of Sequence Data from Vertebrate Actin Genes

ISOFORM	HUMAN	MOUSE	RAT
α-skeletal	Hanauer *et al.* (1983) -C Gunning *et al.* (1984a) -C	Hu *et al.* (1986)	Zakut *et al.* (1982)
α-cardiac	Hamada *et al.* (1982) Minty and Kedes (1986)	Minty *et al.* (1982) -C	
α-smooth	Ueyama *et al.* (1984)		
γ-smooth			
type 1(β) cytoskeletal	Ng *et al.* (1985) Nakajima-Iijima *et al.* (1985)		Nudel *et al.* (1983)
type 8(γ) cytoskeletal	Erba *et al.* (1986) -C Tokunaga *et al.* (1985) -P		
other cytoskeletal			

	CHICKEN	*X. LAEVIS*	*X. BOREALIS*
α-skeletal	Fornwald *et al.* (1982)	Stutz and Spohr (1986) Mohun *et al.* (1984)	
α-cardiac	Chang *et al.* (1985) Eldridge *et al.* (1985)	Stutz and Spohr (1986) Mohun *et al.* (1984)	
α-smooth			
γ-smooth			
type 1(β) cytoskeletal	Kost *et al.* (1983)	not detected	Cross *et al.* (1986)
type 8(γ) cytoskeletal		Cross *et al.* (1986) -C	Cross *et al.* (1986) -C
other cytoskeletal		Cross *et al.* (1986) -C	Cross *et al.* (1986) -C

Table 1.3 Reports of Sequence Data from Vertebrate Actin Genes

This table lists the relevant publications containing sequence data from vertebrate actin genes. When no gene sequences are available, cDNA or partial cDNA sequences have been noted (marked -C). Furthermore, the γ-cytoskeletal actin gene sequence is from a processed pseudogene (marked -P). General conclusions on the conservation of sequences in the homologous genes from different species are discussed briefly in Sections 1.5.3 and 8.2.

specific nucleotides at selected silent sites. It is unclear if the gene sequences encoding the N-terminal amino acids (i.e. alanine and cysteine), which are subsequently removed to produce the mature protein, are also important in their own right (see Appendix I).

Examination of the remainder of the transcription unit reveals further homologies in the 5' and 3' untranslated regions (5' and 3' UTRs). In all vertebrates examined, the 3' UTR sequence, which is considerably longer than the 5' UTR, is characteristic of the isotype that the message encodes (Gunning *et al.*, 1984a; Yaffe *et al.*, 1985; Cross *et al.*, 1986 - Appendix I). Few significant homologies are shared between the 3' UTRs of messages for different actin isoforms in the same species. The conserved segments between species may span the entire 3' UTR (e.g. the mammalian cytoskeletal β-actin genes), or encompass a more limited section of the sequence (e.g. the 3' part in the cardiac and skeletal actin genes). Even with the cytoskeletal β-actin gene, homologies between the human and the more distantly related chicken are more heavily localised at the proximal end of the 3' UTR (Yaffe *et al.*, 1985), and this pattern also appears to apply for the cytoskeletal γ-actin genes (Erba *et al.*, 1986).

Similar studies of the 5' UTR do reveal some conserved sequences (e.g. Cross *et al.*, 1986), but its length is essentially too short and variable to make the same diagnostic comparisons. The positions of introns are also entirely conserved between species in genes encoding a single vertebrate isoform (see Buckingham and Minty, 1983), but the pattern is altered between different isoforms. The diversity of intron position is even greater in lower eukaryotes, and it remains unclear if present-day introns have evolved from an elimination or an additive

mechanism, or both.

Generally no conservation of intron sequences has been observed in the vertebrate actin genes, although Ng *et al.* (1985) and Nakajima-Iijima *et al.* (1985) have discovered sequence homologies in the first and third introns of the human and rat cytoskeletal β-actin genes. Not surprisingly though, much more extensive similarities have been reported in the promoter regions of many vertebrate actin genes. Again these are largely isotype-specific (e.g. Nakajima-Iijima *et al.*, 1985; see Section 8.2), but there are some sequences, which are shared between the skeletal and cardiac actin genes (e.g. Minty and Kedes, 1986), and between the sarcomeric actins and other muscle-specific genes (see Section 8.2). The latter may therefore be involved in coordinate gene regulation. Generally actin genes are unclustered (see Buckingham and Minty, 1983) and current evidence suggests that clustering of muscle-specific genes is not a general mechanism for coordinating expression.

The significance of all these sequence conservations can only be verified at a functional level. A number of recent studies have investigated the regulation of skeletal and cardiac actin gene expression, particularly in transformed, cultured myoblasts (e.g. Bergsma *et al.*, 1986; Minty and Kedes, 1986; see Section 8.2), and the conclusions are summarised in Chapter 8. These reports have basically focussed on the 5' flanking region of both genes in a search for elements which developmentally regulate transcription. However, this is only one level at which control may take place, and conserved regions in the transcription unit may affect properties like RNA stability, processing or transport (see Section 1.1).

The evidence to date from transformation experiments (see Section 8.2) and the detection of increased DNase I sensitivity in the chromosomal skeletal actin gene on differentiation of cultured myoblasts into myotubes (Carmon *et al.*, 1982) strongly implicates transcription as a major regulatory mechanism. This, of course, does not exclude additional controls at the post-transcriptional level. However, the suggestion that actin genes might be preferentially amplified during muscle differentiation (Zimmer and Schwartz, 1982) does not appear to be reproducibly supported by more recent experiments (Hayward and Schwartz, 1986).

Whatever the mechanisms are which lead to muscle development and the activation of actin genes, the studies of Blau *et al.* (1983) and Chiu and Blau (1984), discussed briefly in Section 1.4.1, indicate that other differentiated cell types can be reprogrammed to form muscle-specific products on fusion to myotubes. Therefore, at least in these cases, some of the factors, which initiate such changes, must be diffusible and are present even in terminally differentiated muscle cells. Now that upstream elements are being recognised in muscle genes that are required for tissue-specific expression, the next steps must be to identify the factors which bind them, to attempt to clone these factors in order to examine the way their expression is controlled, and to elucidate sequences in other genes that are affected by the factors. Examination of post-transcriptional controls will require the design of constructs which test the regulatory influence of transcribed sequences in the actin gene and show that their effects are not directed at the transcriptional level.

If factors involved in the regulation of actin gene expression in

vertebrates are to be cloned, assay systems are required to test for these molecules. Aside from *in vitro* transcription systems, the *Xenopus* oocyte can offer a neutral environment for genes, which might be affected by added extracts from other cell types (see Section 8.3). Consequently, the demonstration of correct cardiac actin gene regulation in injected *Xenopus* embryos and oocytes would have two important consequences. For, as well as allowing the elucidation of the sequences directing developmental regulation of embryonic expression, it might also provide a system in which the relevant *trans*-acting factors could be identified.

The sarcomeric actin genes of *Xenopus* are of particular interest, as they are transcribed at a very early stage of development relative to other vertebrates. The following section gives a brief outline of embryogenesis in *Xenopus*, as a background to the expression work that is described in this thesis.

SECTION 1.6 EARLY DEVELOPMENT IN *XENOPUS*

The early stages of *Xenopus* embryogenesis share many general features common to all vertebrates This organism has enjoyed particular attention because the large embryos are readily available throughout the year and they develop rapidly and independently. The following section briefly summarises the embryological changes most relevant to an understanding of the work described in this thesis. A much more detailed discussion is presented in the account by Nieuwkoop and Faber (1956) from which the drawings of developing embryos are taken (Figure 1.2) to illustrate the various stages referred to below.

The unfertilised *Xenopus* egg is a large, asymmetric cell. At its vegetal pole, many yolk platelets are aggregated, while at the animal pole, these platelets are much less dense and the region is darkly pigmented by granules at the cell surface. Although this organisation does lend some polarity to the egg, only on fertilisation are the precise axes of the future embryo determined. The successful sperm always enters in the animal hemisphere and reorganisation of the cytoskeleton before the first cell division leads to the formation of a lightly pigmented band (the grey crescent) on the opposite side of the egg, which is approximately centred on the site where cells will later invaginate to produce the internal structures of the embryo.

The first few hours of development (until stage 8) involve twelve very rapid, synchronous cell divisions (the cleavage stage), each of which, excluding the first division, takes only about 35 minutes on average (Newport and Kirschner, 1982a; this is somewhat slower than described by

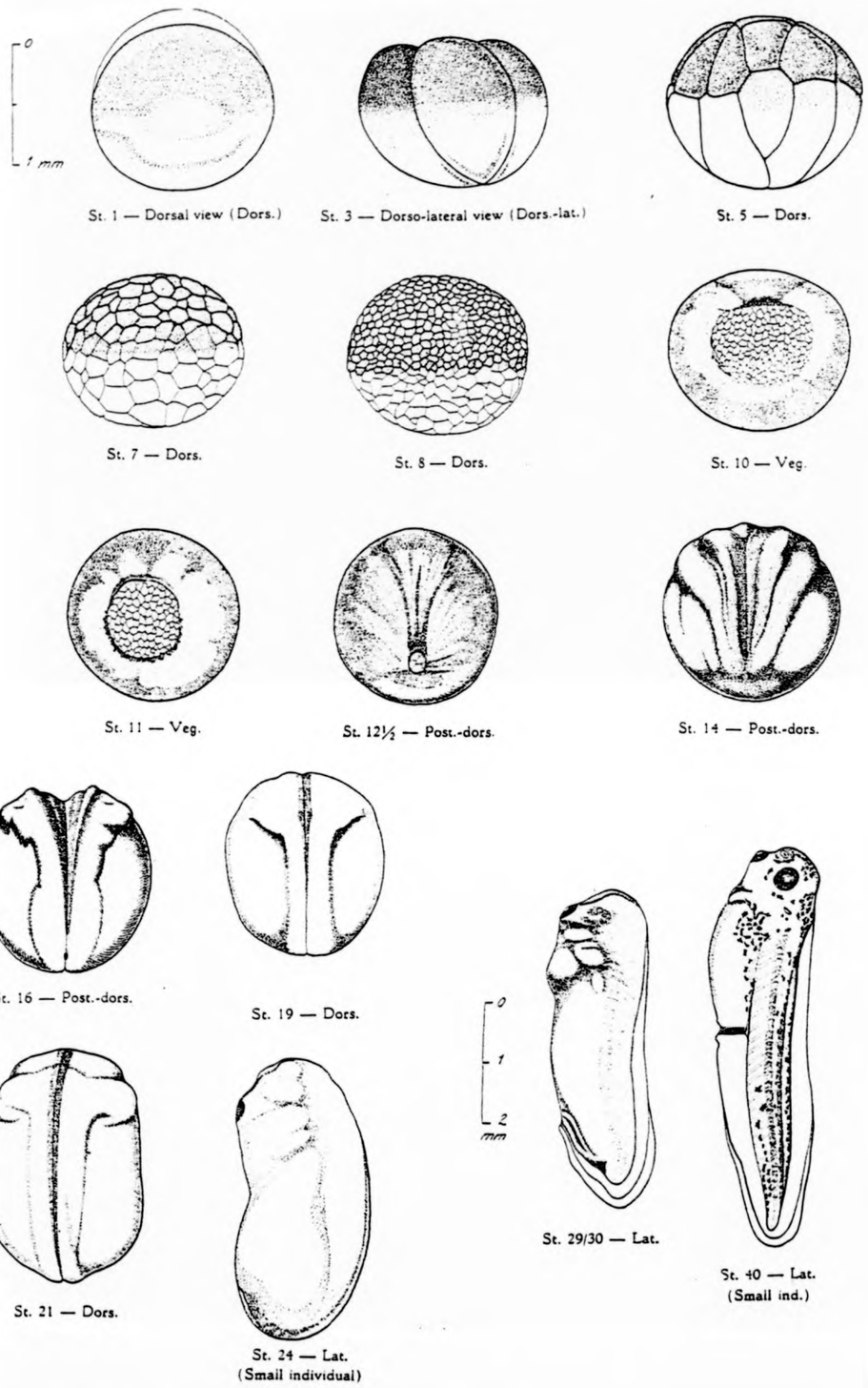


Figure 1.2 Early Stages of *Xenopus laevis* Development

The drawings illustrate numerous stages in the early development of *X. laevis*. They are listed below, together with the approximate time at which each stage is reached after fertilisation (normalised to 23°C; see Nieuwkoop and Faber, 1956).

Stage 1 = single-cell (0-1 hour); stage 3 = 4-cell (2 hr); stage 5 = 16-cell (2 hr 45 min); stage 7 = early blastula (4 hr); stage 8 = mid-blastula (5 hr); stage 10 = early gastrula (9 hr); stage 11 = mid-gastrula (11 hr 45 min); stage 12½ = late gastrula (14 hr 15 min); stage 14 = early neurula (16 hr 15 min); stage 16 = mid-neurula (18 hr 15 min); stage 19 = mid-/late neurula (20 hr 45 min); stage 21 = late neurula (22 hr 30 min); stage 24 = early tailbud tadpole (26 hr 15 min); stage 29/30 = tailbud tadpole (35 hr); stage 40 = swimming tadpole (66 hr).

The last two diagrams (stage 29/30 and stage 40) are drawn to a different scale than the others. The relevance of the various stages is discussed in the text. In particular, it should be noted that cardiac actin gene transcripts are initially detected at stage 12½, approximately 4 hours before the first α-actin protein is observed and more than 8 hours before myofibrillae are formed (see Section 5.2.4). The drawings were all taken from Nieuwkoop and Faber (1956).

Nieuwkoop and Faber, 1956). No new transcription is observed during this period. However, at stage 8, the mid-blastula transition (MBT), transcription of many genes is activated or reactivated, and cell division becomes asynchronous. This transformation has been suggested to be associated with the attainment of a critical level of DNA in the whole embryo (Newport and Kirschner, 1982b), which perhaps depletes a gene regulator by stoichiometric binding as the genome is replicated.

Animal pole cells from a blastula embryo will develop only into ectodermal structures (e.g. epidermis) when cultured in isolation, although it seems that even this differentiation process may require permissive cell interactions (Jones and Woodland, 1986). However, interaction of these same cells with other parts of the whole embryo is thought to lead to the establishment of mesodermal tissue (see Slack, 1983; Sections 1.2 and 5.3.1), so that many cells in the equatorial region of the blastula are committed to form mesoderm by the early gastrula (stage 10; Slack and Forman, 1980). As discussed in Section 5.3.1, recent studies suggest that this process involves the induction of ectodermal cells by endoderm, although such a mechanism may not account for all the mesoderm formation *in vivo*.

During gastrulation, cells invaginate into the embryo at a position in the vegetal region opposite the sperm entry point, known as the blastopore (in fact, these movements first begin in the blastula, but they are less obvious at this early stage). Endodermal and mesodermal cells gradually move inside the embryo to locations where they will subsequently form the internal organs. It is at the end of this process (stage 12½) that some mesodermal cells, which will later form the myotomes, begin to express the cardiac actin gene (see Section 5.2.4).

Over the following hours, neurulation takes place, leading to the eventual creation of the neural tube and primitive nervous system. The first somite is recognised at stage 17, just after α-actin protein is initially detected (stage 16; Sturgess *et al.*, 1980). However, it is some hours later (the late neurula; stage 21) before myofibrillae are observed in the myotomes (see Nieuwkoop and Faber, 1956). Therefore the activation of the cardiac actin gene during early *Xenopus* development occurs about eight hours before morphological differentiation of muscle tissue takes place. This contrasts with other organisms like *Drosophila* (Fyrberg *et al.*, 1983; Sanchez *et al.*, 1983) and sea urchins (Shott *et al.*, 1984), as well as mammalian myogenic cell lines (e.g. see Minty *et al.*, 1982; Hayward and Schwartz, 1986), where the appearance of actin gene transcripts and sarcomeres are much more closely coordinated. This unusual asynchrony in *Xenopus* makes the study of its sarcomeric actin genes of particular developmental interest for two reasons. First, actin gene expression acts as an early marker for muscle commitment and second, it may be activated by gene regulators which operate at the very earliest stages of muscle differentiation.

Figure 1.2 also shows drawings of two tadpole stages, the tailbud tadpole (stage 30) and the swimming tadpole (stage 40). The primitive heart is already visible in the swimming tadpole.

SECTION 1.7 THE AIMS OF THE PROJECT

The introduction of cloned genes into whole organisms is an important approach for the elucidation of sequences which developmentally regulate expression (see Section 1.3). Although a number of cell lines can be used to find such regions relatively rapidly, these studies cannot test for the same stringency of tissue-specificity as can be tested for in all the cell types of the entire animal. Furthermore, cultured cells, which can undergo differentiation, may not develop in an identical fashion to their counterparts *in vivo* (e.g. Hayward and Schwartz, 1986; see Section 8.2), and so these cells cannot be employed to examine all developmental changes in gene expression during the differentiation process. It is therefore imperative at some stage to verify results from cultured cells in transformed or transgenic animals.

Although the transgenic systems in *Drosophila* and the mouse are well established, the large size of *Xenopus* embryos and their rapid, independent development during embryogenesis appear to make them a particularly suitable vertebrate in which to investigate early gene expression. However, even though fertilised *Xenopus* eggs are much easier to micro-inject than mouse eggs, the apparent potential for such studies has remained unfulfilled. The principle aim of this project was to test if transformed *Xenopus* embryos would show appropriate regulation of expression from injected genes during the earliest stages of development.

The sarcomeric actin genes in *Xenopus* are activated many hours before muscle cells are first observed on the basis of morphological criteria

SECTION 1.7 THE AIMS OF THE PROJECT

The introduction of cloned genes into whole organisms is an important approach for the elucidation of sequences which developmentally regulate expression (see Section 1.3). Although a number of cell lines can be used to find such regions relatively rapidly, these studies cannot test for the same stringency of tissue-specificity as can be tested for in all the cell types of the entire animal. Furthermore, cultured cells, which can undergo differentiation, may not develop in an identical fashion to their counterparts *in vivo* (e.g. Hayward and Schwartz, 1986; see Section 8.2), and so these cells cannot be employed to examine all developmental changes in gene expression during the differentiation process. It is therefore imperative at some stage to verify results from cultured cells in transformed or transgenic animals.

Although the transgenic systems in *Drosophila* and the mouse are well established, the large size of *Xenopus* embryos and their rapid, independent development during embryogenesis appear to make them a particularly suitable vertebrate in which to investigate early gene expression. However, even though fertilised *Xenopus* eggs are much easier to micro-inject than mouse eggs, the apparent potential for such studies has remained unfulfilled. The principle aim of this project was to test if transformed *Xenopus* embryos would show appropriate regulation of expression from injected genes during the earliest stages of development.

The sarcomeric actin genes in *Xenopus* are activated many hours before muscle cells are first observed on the basis of morphological criteria

(see Section 1.6), and only shortly after these cells become committed to myogenic differentiation. They are therefore particularly good examples of the type of gene that might be suitable for micro-injection studies. The early aims of my project were to isolate a complete cardiac actin gene from a genomic library and characterise its normal expression during early development.

On completion of these experiments, I then intended to study the expression of the gene on injection into embryos and oocytes, and, if it was regulated properly, to further analyse the sequences responsible.

The following chapters discuss the results from all these lines of investigation. They report for the first time that transformed *Xenopus* can be used to study tissue-specific gene expression. I also describe preliminary experiments which locate the important regulatory regions of the cardiac actin gene to sequences upstream of the second intron. The potential of this system to further characterise these control regions is discussed. In addition, a novel method to clone the genes which encode the *trans*-acting factors that activate the cardiac actin gene is considered.

CHAPTER 2

MATERIALS AND METHODS

SECTION 2.1 MATERIALS

2.1.1 General Materials

Restriction enzymes were purchased from Amersham International Ltd., Anglian Biotechnology Ltd., the Bethesda Research Laboratory (U.K.), Boehringer-Mannheim (London), New England Biolabs and Northumbria Biochemicals Ltd.. *E. coli* DNA polymerase I and *E. coli* DNA polymerase I Klenow fragment were obtained from Boehringer-Mannheim or Anglian Biotechnology Ltd. Reverse transcriptase was bought from Life Sciences Inc.; T4 polynucleotide kinase was purchased from Amersham International Ltd. and calf intestinal alkaline phosphatase (CIAP) from Boehringer-Mannheim. Pancreatic DNase I, RNase A and S1 nuclease were all obtained from the Sigma Chemical Company. T4 DNA ligase was purchased from Anglian Biotechnology Ltd., the Bethesda Researach Laboratory or Amersham International Ltd., and T4 DNA polymerase was acquired from the Bethesda Research Laboratory.

All radioisotopes were obtained from Amersham International Ltd. at the following specific activities: [α-^{32}P] NTP at approximately 3000 Ci/mmole dissolved at 10 mCi/ml as the triethylammonium salt, and [γ-^{32}P] ATP at 6500 Ci/mmole dissolved at 10 mCi/ml as the triethylammonium salt.

X-ray film was obtained from Fuji Photo Film Co. Ltd., and nitrocellulose filters from Schleicher and Schull (W. Germany). Type II agarose for electrophoresis was purchased from the Sigma Chemical Company, and acrylamide from BDH Chemicals Ltd. (AnalaR grade).

Extracts for bacteriological media were acquired from Difco Laboratories and Oxoid Ltd., while all other chemicals were purchased from BDH Chemicals, the Sigma Chemical Company and Fisons Scientific Apparatus.

Xenopus laevis were obtained from the South African Snake Farm, Fishhoek, South Africa, but *Xenopus borealis* (found in the wild in Kenya) were bred in the laboratory. The hormones FSH (Folligon) and hCG (Chorulon) were purchased from Intervet Laboratories Ltd.. Some of the control tadpole RNA from *X. laevis* and *X. borealis* was kindly supplied by Ms. Elizabeth Ballantine.

2.1.2 Commonly Used Solutions

These solutions are frequently referred to in the text of this chapter, and their compositions, and, where appropriate, the procedures for their preparation are described below:

1.	Chloroform (for DNA and RNA extraction)	= 24:1 (v/v) chloroform: isoamyl alcohol
2.	10 x core buffer	= 0.5 M Tris-HCl, pH 8.0, 0.5 M NaCl, 0.1 M $MgCl_2$
3.	LB (L-broth)	= 10 g/l tryptone, 5 g/l yeast extract 0.2 M NaCl
4.	Phenol (neutralised, for DNA and RNA extraction)	= neutralised with 1 M Tris-HCl (pH 8.0), followed by 0.1 M Tris-HCl (pH 8.0), and containing 0.1% (w/w) 8-hydroxyquinoline

5.	SM	- 0.1 M NaCl, 10 mM $MgSO_4$, 0.05 M Tris-HCl, pH 7.5 and 2% gelatin
6.	3 M sodium acetate (NaAc), pH 6.3	- 3 M NaAc brought to pH 6.3 with acetic acid
7.	20 x SSC	- 3 M NaCl, 0.3 M trisodium citrate
8.	TE	- 10 mM Tris-HCl, pH 8.0, 1 mM EDTA

2.1.3 Commonly Used Bacterial Strains and Vectors

Bacterial strains

K803 - *hsdR*$^-$,*hsdM*$_k^-$,*gal*$^-$,*met*$^-$,*supE* (Wood, 1966)

Q358 - *hsdR*$^-$,*hsdM*$_k^+$,*supE*,$\phi 80^r$ (Karn *et al.*, 1980)

Q359 - *hsdR*$^-$,*hsdM*$_k^+$,*supE*,$\phi 80^r$,P2 (Karn *et al.*, 1980)

Vectors

Lambda vectors

EMBL 3 - Frischauf *et al.*, 1983.

λ47.1 - Loenen and Brammar, 1980.

Plasmid vectors

pEMBL 8/9 - Dente *et al.*, 1983.

pEMBL 18/19 - constructed from pEMBL 8/9 with the polylinker of M13 mp 18/19

SECTION 2.2 GENERAL MOLECULAR TECHNIQUES

2.2.1 Small-Scale Preparation of Plasmid DNA

A suitable colony was picked from a plate, and used to inoculate a 10 ml aliquot of LB (10 g/l tryptone, 5 g/l yeast extract and 0.2 M NaCl), containing the appropriate selective antibiotic. The culture was incubated overnight, and centrifuged for 10 minutes at 2000 g in a bench-top centrifuge. The pellet was resuspended in 2 ml of TE (10 mM Tris-HCl, pH 8.0, 1 mM EDTA). Then 1 ml of the suspension was centrifuged for 2 min in a 1.6 ml Eppendorf tube. The pellet was resuspended in 0.8 ml of STET (8% w/v sucrose, 5% v/v Triton X-100, 50 mM EDTA, 50 mM Tris-HCl, pH 8.0) by sucking up and down in a pipette, and 20 μl of 20 mg/ml lysozyme, freshly made in water, was added. The mixture was incubated for 5 min at 96^{o}C, allowed to cool, and then 1 μl of 2 mg/ml DNase-free RNase A was added before vortexing.

The solution was centrifuged for 15 min, and the supernatant was decanted into another 1.6 ml Eppendorf tube. After one extraction with an equal volume of phenol/chloroform, 160 μl of 3 M NaAc (pH 6.3) and 480 μl of isopropanol were added. The mixture was incubated at room temperature for 5 min, and centrifuged for 5 min. Following a wash step with 80% ethanol, the pellet was dried and redissolved in 50 μl of TE. Typically, yields of 2-5 μg of DNA were obtained.

2.2.2 Large-Scale Preparations of Plasmid DNA

An adaptation of the 'cleared lysate' method of Clewall and Helinski (1970) was performed. A 10 ml inoculum, containing the appropriate transformed colony in LB with added antibiotic, was incubated overnight. Then 0.5 litres of LB, also containing the antibiotic, was inoculated with this suspension and left to grow for at least 5 to 6 hours, until it had reached stationary phase. Solid chloramphenicol was added to a concentration of 100 μg/ml and the mixture was shaken at 37^{o}C overnight. The bacteria were recovered by centrifugation at 1,700 g (2,600 rpm in the MSE Mistral 6 L with the 4 x 500 ml swing-out rotor) for 45 min. The pellet was resuspended on ice in 12 ml of a solution containing 25% w/v sucrose and 0.05 M Tris-HCl, pH 8.0. It was then transferred to an Oakridge tube. After addition of 2.5 ml of lysozyme solution (10 mg/ml lysozyme freshly added to 10 mM Tris-HCl, pH 8.0), the mixture was incubated for 5 min on ice, with intermittent swirling. Then 2.5 ml of 0.5 M EDTA (pH 8.0) was added, the solution was left for a further 5 minutes, and 20 ml of Triton lysis solution (0.1% v/v Triton X-100, 0.06 M EDTA, 0.05 M Tris-HCl, pH 8.0) was introduced. After vigorous mixing, the lysed bacteria were pelleted by centrifugation at 38.000 g (18,000 rpm in the MSE HS18 with the 8 x 50 fixed angle rotor) for 1 hour at 4^{o}C. The supernatant was decanted into two polycarbonate Oakridge tubes, and 2 ml of 3 M NaAc (pH 6.3) was added to each. The solutions were extracted once with phenol/chloroform and once with chloroform, dialysed for 2 hours at room temperature in 5 litres of TE and then made up to 0.3 M NaAc. The nucleic acids were precipitated with ethanol in 250 ml plastic centrifuge pots.

The precipitate was pelleted by centrifugation at 15,000 g (10,000 rpm

in the MSE HS18 with the 6 x 250 fixed angle rotor). The pot was drained, and the pellet briefly allowed to air dry before resuspension in 10 ml of 0.1 x SSC (1 x SSC = 0.15 M NaCl, 0.015 M sodium citrate). The solution was made up to 0.3 M NaAc, 50 μl of 20 mg/ml RNase A was added and the mixture was incubated for 1 hour at 37^{o}C. It was then extracted once with phenol/chloroform and once with chloroform, before ethanol precipitation. The washed and dried pellet was redissolved in 5 ml of 0.1 x SSC. The DNA was purified on a caesium chloride gradient by standard methods (see Maniatis *et al.*, 1982). The lower band containing supercoiled plasmid was drawn into a syringe attached to steel tubing that was inserted into the gradient from above. Although the method takes about four days to complete, the yields are extremely high (between 2 and 6 mg of DNA for a 500 ml initial culture).

2.2.3 Restriction Digests

Basically, these followed manufacturers' instructions, although, wherever possible (i.e. *Acc*I, *Bam*HI, *Eco*RI, *Hinc*II, *Hind*III, *Hinf*I, *Pst*I, *Pvu*II, *Sac*I, *Sal*I, *Sau*3A, *Sph*I, *Sst*I), core buffer (50 mM Tris-HCl, pH 8.0, 50 mM NaCl, 10 mM $MgCl_2$) was usually used as the incubation medium. With plasmid DNA, particularly that from small-scale preparations, the solution was made up to 3 mM spermidine, but this was not added for λ DNA and genomic DNA, because spermidine preferentially precipitates high-molecular-weight nucleic acids.

2.2.4 Agarose Gel Electrophoresis

Usually, approximately 150-170 ml agarose gels were poured on 20 cm x 15 cm glass plates. The agarose was dissolved in TAE, made from a 20 x TAE stock (0.8 M Tris, 0.4 M NaAc, 0.04 M EDTA, brought to pH 8.3 with glacial acetic acid). On cooling to about 45-50^{o}C, ethidium bromide was added from a 10 mg/ml stock solution to a concentration of 1 μg/ml, and the gel was then poured.

Gels were electrophoresed in horizontal tanks, submerged in TAE (containing 1 μg/ml ethidium bromide). For plasmid and λ DNA, a current of 200 mA over a few hours was suitable, but genomic DNA was electrophoresed overnight at 40-60 mA. All samples were dissolved in loading buffer, added from a 6 x loading buffer stock (0.25% bromophenol blue, 40% w/v sucrose). For λ DNA samples, the cohesive ends were denatured at 65^{o}C for 10 min before loading. The appropriate % agarose was selected according to the size of fragments to be separated (see Maniatis *et al.*, 1982). The bands were visualised under ultraviolet illumination.

2.2.5 Isolation of DNA Fragments from Agarose Gels

Although low-melting-point agarose was occasionally used for fragments of less than 1 kb, usually the appropriate band was removed by the paper/dialysis membrane technique described by Maniatis *et al.* (1982). This was modified by omitting SDS from the elution buffer.

2.2.6 Southern Blots

This method followed the basic protocol of Southern (1975), adapted by Maniatis *et al.* (1982). For all genomic Southern blots, and any other cases where high-molecular-weight DNA was to be transferred, the gels were treated with 0.25 M HCl prior to alkali denaturation (Wahl *et al.*, 1979) in order to partially hydrolyse the DNA by acid depurination. The transfer buffer was 20 x SSC. Before placing the nitrocellulose on it, the neutralised gel was wetted with 2 x SSC to remove excess salt. Usually the transfer of DNA was allowed to proceed overnight, although, for large amounts of specific fragments from plasmid DNA, as little as 4 hours was required. When the filter was removed from the gel, it was soaked in 2 x SSC for 5 minutes before drying and baking at 80°C for 2 hours.

2.2.7 Labelling of DNA by Nick Translation

This was a modification of the method of Rigby *et al.* (1977). A total reaction volume of 20 μl was set up (taking into account the subsequent addition of 7 units *E. coli* DNA polymerase I), containing nick-translation buffer (NTB) from a 5 x NTB stock (0.25 M Tris-HCl, pH 7.2, 0.5 mM DTT, 0.05 M $MgSO_4$ and 250 μg/ml BSA), 1 μM dATP, dCTP and dTTP, approximately 0.02 units DNase I, 1-2 μl (3-6 pmoles) of high specific activity [α-^{32}P] GTP, and no more than 50 ng of DNA. If [α-^{32}P] CTP was used, unlabelled dGTP replaced unlabelled dCTP in the reaction. The mixture was incubated for 15 min at 37°C before 7 units *E. coli* DNA polymerase I was added. After vortexing, the solution was incubated for 3 hours at 16°C, and then extracted once with phenol/chloroform. The

labelled DNA was separated from unincorporated [α-^{32}P] dNTP by chromatography on a column of Sephadex G-50.

2.2.8 Hybridisation of Nitrocellulose Filters to Labelled Probes

Filters were prehybridised for at least 3 hours at 65°C in a solution of 3 x SSC, 0.1 x Denhart's additives (0.02% Ficoll, 0.02% polyvinyl-pyrollidone, 0.02% BSA) and 0.1 mg/ml poly(A), added at approximately 0.05 ml/cm^2 of nitrocellulose. For genomic Southern blots, the solution was replaced by a fresh aliquot of the same mixture after this time. The nick-translated probe was denatured at 100°C for 10 min, quenched on ice for at least 5 min, and then, for example, no more than 20 ng of DNA (usually approximately 10^6-10^7 cpm) was added to a 20 cm x 15 cm nitrocellulose filter. Hybridisation at 65°C for 16 hours was followed by washing in appropriate conditions (typically 2 x SSC at 60°C for 2 x 10 min). The damp filters were sealed in plastic bags and examined by autoradiography at -70°C.

2.2.9 Subcloning into Plasmid Vectors

Preparation of vector

Usually 10 μg of vector DNA was restricted with the appropriate enzymes in a 20 μl volume. If only one enzyme was used, the vectors cohesive ends were phosphatased. On completion of digestion, the reaction was made up to 0.1% w/v SDS in a total volume of 50 μl of 0.05 M Tris-HCl, pH 8.0, containing approximately 2.5 units of calf intestinal alkaline phosphatase (CIAP). This mixture was incubated at 37°C for 30 min,

before being made up to 1% w/v SDS and 0.3 M NaAc, pH 6.3. The solution was extracted once with phenol, once with chloroform and once with ether, and finally precipitated with ethanol. The vector DNA was redissolved at an appropriate concentration.

If two enzymes were used for vector preparation, the DNA fragments were separated on an agarose gel, and the vector DNA was recovered, as described in section 2.2.5.

Ligations

These were performed in a 10 μl volume of 1 mM rATP and C buffer (10 x C buffer = 0.5 M Tris-HCl, pH 7.5, 0.1 M $MgCl_2$ and 10 mM DTT). Even if the vector DNA was phosphatased, usually an equimolar ratio of target to vector molecule was used, at a final vector concentration of 5 μg/ml. An excess of 0.5 units of DNA ligase was added, and the reaction was incubated at 14^{o}C for at least 4 hours or overnight.

Transformation

Standard transformation procedures were employed using calcium chloride, as described by Maniatis *et al.* (1982).

2.2.10 Polyacrylamide-Urea Denaturing Gels

Gels prepared for the analysis of samples from S1 nuclease and primer extension reactions, as well as those for isolation of DNA fragments, were made using glass plates separated by 0.4 mm plasticard spacers. Generally 45 ml of solution were required for each gel. For an X% acrylamide gel, the mixture contained X% acrylamide, 1/20 X% bis-

acrylamide, 8 M urea and TBE (10 x TBE = 0.89 M Tris, 0.89 M boric acid, and 0.02 M EDTA, at pH 8.3 by addition of solid boric acid or Tris). Polymerisation occurred on addition of 0.3 ml of 6.6% ammonium persulphate and 0.03 ml of TEMED. The acrylamide concentration was selected for maximum separation of DNA fragments (see Maniatis *et al.*, 1982), and ranged from 6-20%. The gels were electrophoresed vertically in TBE, usually at a constant current of 26 mA, although for high acrylamide concentrations, this was reduced to prevent overheating. On completion of electrophoresis, the gel was covered with cling film, and the bands were detected by autoradiography (generally with an intensifying screen) at -70°C, without fixing or drying the gel beforehand.

2.2.11 Isolation of DNA Fragments from Polyacrylamide Gels

This method was adapted from Maxam and Gilbert (1978). A gel slice containing the DNA fragment was excised, using the developed autoradiograph as a stencil. It was transferred to a 1.6 ml Eppendorf tube and a suitable volume of PAGEB (0.5 M NH_4Ac, 0.01 M MgAc, 0.1% SDS, 0.1 mM EDTA) was added (for example, 200 μl for a slice of up to 1 cm in width). The tube was incubated overnight at 37°C, and the solution was then transferred to another large Eppendorf tube. Approximately 10 μg of E. coli tRNA was added as carrier, and the DNA was precipitated with ethanol. The redissolved fragment could subsequently be used, for example, as a primer in primer extension reactions.

2.2.12 Sequencing

Both the chemical base modification procedure according to Maxam and Gilbert (1978), and the enzyme method of Sanger *et al.* (1980) were used to sequence DNA from actin clones. Using the former method, end-labelled primer extension products were also sequenced by Dr. Gareth Cross.

2.2.13 Labelling of DNA Fragments with T4 Polynucleotide Kinase

The appropriate fragment or mixture of fragments was first phosphatased in 50 μl CIAP buffer (10 x CIAP buffer = 0.5 M Tris, pH 9.0, 10 mM $MgCl_2$, 1 mM $ZnCl_2$, 10 mM spermidine), by addition of 1 unit of CIAP and subsequent incubation at 37°C for 30 min. A further unit of CIAP was added and the mixture left at 37°C for another 15 min. For blunt ends or 3' overhangs, the second incubation step and the last 15 min of the first incubation were performed at 56°C.

Then 10 μl of 10 x TE and 5 μl of 10% SDS was added, and the volume was made up to 100 μl with water. The solution was left at 68°C for 15 min to destroy the enzyme, and, after addition of 3 M NaAc (pH 6.3) to 0.3 M final concentration, it was extracted twice with phenol/chloroform and once with chloroform. The DNA was then precipitated with ethanol.

The pellet was redissolved in a total volume of 20 μl of kinase buffer (10 x kinase buffer = 0.7 M Tris-HCl, pH 8.0, 0.1 M $MgCl_2$, 50 mM DTT), supplemented with approximately 10 units of T4 polynucleotide kinase and 40 μCi of high specific activity [γ-^{32}P] ATP. The mixture was incubated

for 40 min at 37°C and then the volume was raised to 100 µl, containing 0.3 M NaAc (pH 6.3). Phenol was added to destroy the enzyme, followed by 10 µg of *E. coli* tRNA. The solution was extracted with the phenol and then with chloroform. The labelled fragment was separated from unincorporated [γ-^{32}P] ATP by chromatography on a Sephadex G-50 column.

2.2.14 S1 Nuclease Analysis

The construction and labelling of specific probes is described in Section 2.5.1 of this chapter. A mixture of labelled probe and RNA (normally two embryo equivalents) was precipitated with ethanol. The pellet was washed with 80% ethanol, and resuspended in 15 µl of hybridisation buffer (40 mM Pipes, pH 6.4, 1 mM EDTA, 0.4 M NaCl, 80% deionised formamide). This was denatured at 80°C for 10 minutes and hybridised at 60°C for 3 hours or overnight in a sealed capillary tube. The solution was then added to 150 µl of S1 buffer (0.28 M NaCl, 0.05 M NaAc, pH 4.6, 4.5 mM $ZnSO_4$. 20 µg/ml sonicated, denatured salmon sperm DNA), usually containing 100 units of S1 nuclease (Sigma). Normally this was incubated at 30°C for 30 minutes, after which the reaction was stopped by the addition of 3 µl 0.4 M EDTA, pH 8.0. The sample was extracted once with phenol and ethanol precipitated. The pellet was then washed and resuspended in 3 µl of formamide loading buffer (98% deionised formamide, 4% w/v Ficoll, 0.01% bromophenol blue, 0.01% xylene cyanol FF), denatured at 80°C for 10 minutes, quenched on ice and loaded on an 8% polyacrylamide-urea gel. For the differential actin S1 assay higher concentrations of S1 nuclease were used (100-300 units) and digestion was performed at 50°C.

Gel slices containing protected probe were excised from gels, added to triton/toluene scintillant and counted.

For all probes, optimal hybridisation and S1 digestion conditions were initially determined. It was also shown that there was sufficient probe for only a small fraction to be hybridised to RNA in each reaction. Signal intensities were therefore proportional to the level of message in the assay, as the probe was in excess.

2.2.15 Primer Extension Analysis

The construction and labelling of specific primers is described in Section 2.5.2 of this chapter. Embryo RNA was mixed with the primer in a total volume of 8 μl. Then 2 μl of 5 x hybridisation buffer (2 M NaCl, 50 mM Pipes, pH 6.4) was added, and the solution sealed in a capillary tube. The sample was heated to 80^{o}C for 10 minutes, then hybridised at the appropriate temperature for 3 hours. It was added to 80 μl of freshly made primer extension mix (0.5 mM dATP, dCTP, dGTP and dTTP, 10 mM DTT, 100 mM Tris-HCl, pH 8.0, 12 mM $MgCl_2$, 25 μg/ml actinomycin D) containing 10 units of reverse transcriptase, and incubated for 1 hour at 42^{o}C. The RNA was hydrolysed by boiling for 5 minutes after adding 10 μl of 1 M NaOH. The solution was then mixed with 10 μl 1 M HCl to neutralise and 12 μl 3 M NaAc (pH 6.3), phenol extracted once and precipitated with ethanol. After washing in 80% ethanol, the pellet was resuspended in 3 μl of formamide loading buffer. The sample was denatured at 90^{o}C for 5 minutes, quenched on ice and loaded on a 15% or 8% polyacrylamide-urea gel, as appropriate.

For all primers, optimal hybridisation conditions were initially determined, and the primer was added in excess, so the assay was quantitative.

2.2.16 Synthesis of Labelled Markers

Approximately 0.25 μg of restricted λ DNA was incubated for 5 min at 37^{o}C in TA buffer (10 x TA buffer = 0.33 M Tris-Ac, pH 7.9, 0.65 M KAc, 0.1 M MgAc, 5 mM DTT and 1 mg/ml BSA), containing 1 unit of T4 DNA polymerase. Then 10 μCi of [α-^{32}P] GTP, and 1 μl of a nucleotide mix of 2 mM dATP, dCTP and dTTP was added. The mixture was incubated at 37^{o}C for a further 30 min. It was extracted once with phenol/chloroform, and the labelled DNA was separated from unincorporated [α-^{32}P] dGTP by chromatography on a Sephadex G-50 column.

SECTION 2.3 BACTERIOPHAGE λ LIBRARY CONSTRUCTION AND SCREENING, AND OTHER RELATED TECHNIQUES

2.3.1 Plating of Bacteriophage λ

Plating bacteria

A single bacterial colony, or a loop from a glycerol stock of a bacterial strain was inoculated in 50 ml of LB-Mg (10 g/l tryptone, 5 g/l yeast extract, 0.2 M NaCl and 10 mM $MgSO_4$) and grown overnight. The suspension was centrifuged for 10 minutes at 2000 g in a bench-top centrifuge, and the pellet resuspended in 20 ml of a 10 mM $MgSO_4$ solution. This stock could be stored for a number of days at 4°C, although highest plating efficiencies were achieved on the day of preparation.

Plates

A solution of LB-Mg containing 15 g/l bacto-agar was autoclaved, allowed to cool to approximately 50°C, and poured into Petri dishes (30 ml for 8 cm diameter round plates or 60 ml for 10 cm x 10 cm square plates). Before use, the Petri dishes were dried in a low temperature oven under ultraviolet light.

Plating bacteriophage λ

The method of Maniatis *et al.* (1982) was used, mixing 100 μl aliquots of phage diluted in SM (0.1 M NaCl, 10 mM $MgSO_4$, 0.05 M Tris-HCl, pH 7.5 and 2% gelatin) with 100 μl of plating bacteria. If the plates were to be screened by the method of Benton and Davis (1977), the samples were added to 0.7% agarose, which was then poured on to the bottom agar

medium (4 ml for 8 cm diameter round Petri dishes or 8 ml for 10 cm x 10 cm square plates).

2.3.2 Large-Scale Preparation of Bacteriophage λ DNA

Preparation of bacteria

A single bacterial colony or a loop from a glycerol stock was inoculated in 10 ml of LB-Mg and grown overnight. Then 0.5 ml of this suspension was added to 50 ml LB-Mg, and the bacteria were allowed to divide until approximately mid-log phase, when $OD_{600} \simeq 0.5$ (usually 2½ to 3 hours). The cells were pelleted by centrifugation in Oakridge tubes at 6000 g (6,000 rpm in the MSE HS18 with the 6 x 250 angle rotor), and resuspended in 12.5 ml of 10 mM $MgSO_4$.

Infection with bacteriophage

To start the phage infection, 0.1 ml aliquots of the bacterial suspension (approximately 2×10^8 cells) were added to 0.1 ml samples of a number of different phage dilutions in SM (containing between 10^6 and 10^{10} phage per ml), followed by a further 0.5 ml of SM. The mixture was gently shaken and then incubated at 37°C for 25 minutes with further intermittent shaking. Each infected aliquot was added to 500 ml of prewarmed superbroth (35 g/l bactotryptone, 20 g/l yeast extract, 0.1 M NaCl, 0.01 M $MgSO_4$, pH 7.0 with NaOH), and left shaking overnight at 37°C.

If the ratio of phage to bacterial cells is optimal, the lysed cells produced may form stringy clumps overnight. However, particularly with superbroth, these clumps are not always observed. Therefore, generally

20 ml of chloroform were added to all the cultures, which were then shaken for a further 30 minutes at 37°C. The culture containing the highest level of lysed cells was selected, and the excess chloroform was removed with a Pasteur pipette.

Purification of bacteriophage DNA

This essentially followed the method of Maniatis *et al.* (1982), except that the solution was only treated with RNase A at 0.5 μg/ml, and not with DNase I. The bacteriophage in both caesium chloride gradients was collected by inserting a probe from above and withdrawing the phage band into a syringe.

The solution recovered from the second gradient was dialysed overnight in SM. The volume was measured, and 0.5 M EDTA was added to a concentration of 20 mM, followed by proteinase K from a 10 mg/ml stock to a concentration of 50 μg/ml. The mixture was adjusted to 0.5% SDS by addition of 20% SDS, and incubated at 37°C for 1 hour. It was then extracted once with phenol/chloroform, once with chloroform, and once with ether, before precipitation with ethanol. The purified DNA pellet was redissolved overnight in TE. Provided the correct ratio of bacteriophage to bacteria has been selected, this method typically produces at least 1 mg of DNA from a 500 ml culture.

2.3.3 Small-Scale Preparation of Bacteriophage λ DNA

Small quantities (about 50 μg) of bacteriophage DNA, suitable for analysis by restriction digestion and agarose gel electrophoresis, could be rapidly made by a simple plate lysate method. This DNA is generally

unsuitable for subcloning, as the preparation contains a high proportion of *E. coli* chromosomal DNA.

The purified phage stock was titred over a wide range by plating as described in Section 2.3.1. In this case though, both top and bottom agar layers were made with 0.7% agarose, not agar. A dilution was selected, which was sufficient to just produce confluent plaques over the whole of the plate after overnight incubation at 37°C (about 2×10^4 plaques per 8 cm diameter round plate, but this is somewhat dependent on plaque size). It is essential not to use a titre more than 5-10 times higher than this, or the cells are all lysed before they form a complete lawn.

Two plates, infected at the correct dilution, were each covered with 5 ml of SM, and gently swirled for 10 minutes. The liquid was then transferred to an Oakridge tube using a Pasteur pipette, and the top agarose scraped off with a sterile spatula into the same tube. This suspension was centrifuged at 7,600 g (8,000 rpm in the MSE HS18 with the 8 x 50 angle rotor) for 5 min at 4°C. The supernatant was gently poured into another Oakridge tube and kept on ice, while a further 8 ml of SM was added to re-extract bacteriophage from the agarose pellet. The agarose was repelleted by centrifugation at 7,600 g for 5 min, and the resulting supernatant was added to the original supernatant. To remove the small pieces of agarose remaining in this fraction, the solution was centrifuged at 27,000 g (15,000 rpm in the MSE HS18 with the 8 x 50 angle rotor) for 10 mins at 4°C. The supernatant was decanted and centrifuged again at 78,000 g (25,000 rpm in the MSE 65 with the 8 x 50 angle rotor) to pellet the phage particles. This process can damage the bacteriophage tails making them uninfective, but

it does not affect the packaged DNA.

The pellet was resuspended in 0.5 ml of SM, transferred to a 1.6 ml Eppendorf tube and centrifuged in a microfuge for 20 secs to remove debris. The supernatant was extracted twice with phenol, once with phenol/chloroform and once with chloroform, and then precipitated with ethanol. After centrifugation and an 80% ethanol wash, the resulting pellet was usually resuspended in about 0.4 ml of TE.

2.3.4 Preparation of EMBL 3 DNA for Library Construction

The method was after Frischauf *et al.* (1983). About 40 μg of vector DNA was digested with an excess of *Eco*RI and *Bam*HI in a 0.5 ml reaction volume of core buffer. The solution was made up to 0.3 M NaAc using a 3 M NaAc (pH 6.3) stock, and extracted once with phenol, once with chloroform and once with ether. Finally large DNA was precipitated selectively by addition of 0.6 vol. isopropanol and incubation for 15 minutes on ice. The washed pellet was redissolved in 0.1 x TE at a concentration of 0.5 mg/ml. Generally, the precipitation step was inefficient, and only about 50% of the original DNA was recovered. This was taken into account in subsequent ligations.

2.3.5 Preparation of Genomic DNA for Library Construction

The procedures employed are discussed in detail in Section 3.2.1. *Sau*3A was used to partially digest *X. borealis* genomic DNA (see Section 2.4.5). If the restricted fragments were to be separated on a sucrose

density gradient, 350 μg of DNA was digested in 2.7 ml total volume, after characterising the appropriate conditions with a pilot time-course. The reaction was stopped by addition of 0.5 M EDTA (pH 8.0) to a concentration of 15 mM and the enzyme heat-inactivated by incubation at 70^{o}C for 10 minutes. The solution was immediately loaded on to a 10-40% or 20-40% sucrose density gradient, as described by Maniatis *et al.* (1982). The gradient was fractionated by lowering a probe from above down to the bottom of the tube, and slowly pumping the solution (at approximately 1.5 ml/min) into a series of 1.6 ml Eppendorf tubes. Usually about 0.7 ml fractions were collected, and the OD_{260} was measured. A 20 μl aliquot of some fractions was added to 20 μl H_2O, and 8 μl of 6 x bromophenol blue agarose gel loading buffer (see Section 2.2.4). Although the salt concentration in this mixture was high (about 0.4 M NaCl), the DNA could be electrophoresed, without dialysis, on a 3 x TAE agarose gel, run overnight at 80 mA (see Figure 3.5a).

The most suitable fractions for cloning were dialysed in 4 litres of TE at 4^{o}C for 16 hours with one buffer change after 6 hours. The samples were extracted once with phenol and twice with chloroform, before precipitation with ethanol. The pellets were allowed to redissolve in 0.1 ml of 0.1 x TE overnight, and the DNA size was confirmed by electrophoresis on a 1 x TAE agarose gel.

All genomic DNA preparations were phosphatased by the standard technique (see Section 2.2.13), using 13 μg of DNA and a total of 10 units of CIAP. The enzyme was heat-inactivated at 68^{o}C in the presence of 0.5% SDS, and the solution was extracted and precipitated, as described in Section 2.2.13. The pellet was usually redissolved in 0.1 x TE at a DNA concentration of 0.5 mg/ml.

density gradient, 350 μg of DNA was digested in 2.7 ml total volume, after characterising the appropriate conditions with a pilot time-course. The reaction was stopped by addition of 0.5 M EDTA (pH 8.0) to a concentration of 15 mM and the enzyme heat-inactivated by incubation at 70^{o}C for 10 minutes. The solution was immediately loaded on to a 10-40% or 20-40% sucrose density gradient, as described by Maniatis *et al.* (1982). The gradient was fractionated by lowering a probe from above down to the bottom of the tube, and slowly pumping the solution (at approximately 1.5 ml/min) into a series of 1.6 ml Eppendorf tubes. Usually about 0.7 ml fractions were collected, and the OD_{260} was measured. A 20 μl aliquot of some fractions was added to 20 μl H_2O, and 8 μl of 6 x bromophenol blue agarose gel loading buffer (see Section 2.2.4). Although the salt concentration in this mixture was high (about 0.4 M NaCl), the DNA could be electrophoresed, without dialysis, on a 3 x TAE agarose gel, run overnight at 80 mA (see Figure 3.5a).

The most suitable fractions for cloning were dialysed in 4 litres of TE at 4^{o}C for 16 hours with one buffer change after 6 hours. The samples were extracted once with phenol and twice with chloroform, before precipitation with ethanol. The pellets were allowed to redissolve in 0.1 ml of 0.1 x TE overnight, and the DNA size was confirmed by electrophoresis on a 1 x TAE agarose gel.

All genomic DNA preparations were phosphatased by the standard technique (see Section 2.2.13), using 13 μg of DNA and a total of 10 units of CIAP. The enzyme was heat-inactivated at 68^{o}C in the presence of 0.5% SDS, and the solution was extracted and precipitated, as described in Section 2.2.13. The pellet was usually redissolved in 0.1 x TE at a DNA concentration of 0.5 mg/ml.

2.3.6 Preparation of *in vitro* Packaging Extracts

This followed the procedure 'Protocol II' of Maniatis *et al.* (1982), using the *E. coli* strains BHB2690 (prehead donor) and BHB2688 (packaging protein donor). In both preparations, once the induced cells had been pelleted, all further manipulations were performed on ice in the cold room (a very thick overcoat is essential). The sonicated extract (BHB 2690) was stored in 15 μl aliquots, and the freeze/thaw lysate (BHB2688) was stored in 10 μl aliquots.

2.3.7 Packaging *in vitro*

'Protocol II' of Maniatis *et al.* (1982) was employed. Typically it was found that a mixture of one sonicated extract and four freeze/thaw lysates, left for 2-3 hours, produced the highest packaging efficiencies of up to 5×10^7 pfu/μg.

2.3.8 Ligation of Vector and Genomic DNA

Vector DNA was always ligated at a final concentration of 50 μg/ml. A number of concentrations of genomic DNA were used (e.g. see Table 3.1), but the optimal molar ratio of left λ arm: insert: right λ arm was found to be 2:1:2, as expected (i.e. for 20 kb inserts, the optimal concentration of genomic DNA was about 12 μg/ml). The ligations were essentially performed as described in Section 2.2.9 at 14^{o}C, but 3 units of T4 DNA ligase were used for every 0.5 μg of vector DNA. For packaging, 5 μl aliquots of the ligation mixtures were directly

transferred to the combined packaging extracts (see Section 2.3.7).

2.3.9 Plating of Genomic Libraries

After packaging, as described in Section 2.3.7, the library was diluted with 0.5 ml SM to give a final volume of approximately 0.6 ml. A drop of chloroform was added, and the solution was vortexed and then centrifuged for 30 s in an Eppendorf centrifuge. The supernatant was titred, using the bacterial strains K803, Q358 and Q359 (see Section 3.2.2). Routinely the titre of recombinants was approximately 4×10^5 μg^{-1} of vector DNA (10^5 pfu in each packaged mix). The complete library was plated and screened without further amplification. As plaque size was highly variable, to avoid competition between different recombinants, the library was plated at a titre below that required to produce confluent plaques (typically $1\text{-}2 \times 10^4$ pfu per 10 cm x 10 cm square Petri dish).

2.3.10 Screening of Genomic Libraries

This procedure followed the method of Maniatis *et al.* (1982), except 10 cm x 10 cm square Petri dishes were used, and the filters were washed in 2 x SSC after neutralisation. The location marks were made by stabbing a needle through the nitrocellulose into the agar. The dried, baked filters were then hybridised to a suitable probe by the technique discussed in Section 2.2.8. Radioactive ink was used to label the location marks on the filter for subsequent autoradiography.

2.3.11 Isolation and Rescreening of Putative Positive Clones

Location marks on the autoradiograph were aligned with those of the appropriate Petri dish on a light box. For the first screen, a circular plug of agar, approximately 5 mm in diameter and centred on the positive hybridisation signal, was removed using a cut blue tip attached to a Gilson 1 ml automatic pipette. This plug was placed in a 1.6 ml microfuge tube containing 1 ml of SM and a drop of chloroform. It was allowed to stand at room temperature for 1 hour, and then the liquid was recovered and stored at $4^{o}C$ as a phage stock. On average, this solution contained phage from 25-50 plaques, so at least one further purification was required. Different dilutions of the stocks were plated for a second screen (typically 100 μl of a 2×10^{-4} dilution in SM produced about 10^{3} plaques), and for any further screens required.

When a single plaque could finally be selected, it was cut out of the agar with a Pasteur pipette, and phage were extracted from the small plug by the method described in the last paragraph. An average titre for the final stock solution was approximately 3×10^{6} pfu/ml.

2.3.12 Preservation of Amplified Stocks of Genomic Libraries

After screening and picking of positive clones, each square Petri dish was overlayed with 10 ml of SM, and left overnight at $4^{o}C$ on a horizontal surface. The solution was pipetted into a conical flask, and chloroform was added to 5% (v/v) final concentration. The top agarose was also scraped off with a sterile spatula and placed in a separate flask. Approximately 150 ml of SM was added to these scrapings,

together with 15 ml of chloroform. Both the supernatant and the agarose stocks were stored at 4°C.

SECTION 2.4 EMBRYOLOGICAL METHODS; PREPARATION OF NUCLEIC ACIDS FROM FROGS AND EMBRYOS

2.4.1 Artificial Fertilisation

For micro-injection, embryos were obtained by artificial fertilisation. The testes were removed from an adult male and kept at 4°C in high-salt MBS (110 mM NaCl,1 mM KCl, 2.4 mM $NaHCO_3$, 0.33 mM $Ca(NO_3)_2$, 0.41 mM $CaCl_2$, 0.8 mM $MgSO_4$, 15 mM Tris-HCl, pH 7.5). Females were stimulated by injection of 100 i.u. of FSH (Folligon), followed by an injection of 650 i.u. of hCG (Chorulon) on the afternoon of the next day. Ice was added to the water bucket and the frogs were left overnight. They usually started to lay eggs during the following morning. The females were transferred to MBS (88 mM NaCl, 1 mM KCl, 2.4 mM $NaHCO_3$, 0.33 mM $Ca(NO_3)_2$, 0.41 mM $CaCl_2$, 0.8 mM $MgSO_4$, 15 mM Tris-HCl, pH 7.5).

For fertilisation, most of the MBS was drained from Petri dishes containing roughly 50 unfertilised eggs. A teased testis was briefly agitated over the eggs. After 30 s, approximately 4 volumes of distilled water were added, and after a further 5 minutes, the Petri dish was filled with water. Fertilised eggs rotated within about 15 min, so that the dark animal pole was uppermost.

2.4.2 Micro-injection of Fertilised Eggs

The cytoplasm of embryos was injected either before the first cell division or at the two-cell stage. The embryos were first dejellied in

2% cysteine-HCl, pH 8.0 and washed in MBS. In some experiments the eggs were injected under this medium, but in others MBS was supplemented with 5% Ficoll to prevent protrusion of the embryo through the hole made by the microinjection needle. Typically 20 nl of solutions containing DNA at 30-50 μg/ml were injected into each fertilised egg.

The embryos were transferred to 0.1 x MBS at the blastula stage to permit gastrulation. They were staged according to Nieuwkoop and Faber (1956), and developmental times in the results are normalised to 23^{o}C (for a brief summary of *Xenopus* development see Section 1.6).

2.4.3 Dissection of Embryos

Dissections were undertaken at the blastula, mid-neurula and tadpole stages (see Chapter 7). The vitelline membranes were removed under MBS containing 5% Ficoll, and the embryos dissected with forceps, a narrow-gauge needle and a mounted nose hair. For dissection of neurulae, collagenase was added to 1 mg/ml. Incubations were terminated by homogenisation in RNA extraction medium (Section 2.4.4).

2.4.4 Preparation of Nucleic Acids from Embryos

Four to ten whole embryos or pooled fragments from a similar number of dissected embryos were homogenised in 0.27 ml of a solution containing 10 mM Tris-HCl, pH 7.4, 1.5 mM $MgCl_2$, 10 mM NaCl and 1% SDS with 0.03 ml of a 10 mg/ml stock of proteinase K added (Probst *et al.*, 1979). The homogenate was incubated for 40 minutes at room temperature, then

extracted with phenol/chloroform, and chloroform. It was finally precipitated with ethanol. The pellet was washed in 80% ethanol and resuspended in 2.5 μl of TE per embryo or embryo fragment. Particularly at later stages of development, DNA was sometimes removed from the sample by incubation with pancreatic DNase I.

2.4.5 Preparation of DNA from Adult Frog Erythrocytes

A female frog was anaesthetised with MS222, and blood was collected by heart puncture. The cells were washed twice in heparinised NMT (0.1 M NaCl, 0.01 M Tris-HCl, pH 8.0, 3 mM $MgCl_2$ and 13 μg/ml heparin) and twice in non-heparinised NMT, pelleting the cells each time at 2,300 g (3,000 rpm in the Mistral 6 L with the 8 x 50 swing-out rotor) for 10 min at 4°C. The cells were finally resuspended in 5 ml of NMT, and an equal volume of NMT containing 1% NP40 was added. The mixture was left for 10 min on ice and then centrifuged at 4,100 g (4,000 rpm in the Mistral 6 L with the 8 x 50 swing-out rotor) for 5 min at 4°C The pelleted nuclei and cell debris were resuspended in 5 ml of TNE (0.01 M NaCl, 0.001 M EDTA, 0.01 M Tris-HCl, pH 8.0) by vigorous vortexing. RNase A was added to 40 μg/ml and the solution was incubated for 1 hour at 37°C. Proteinase K was then added to 40 units/ml and SDS to a concentration of 0.5%, before a further overnight incubation at 37°C in a shaking incubator (at approximately 90 rpm). The viscous solution produced was extracted twice with phenol and twice with chloroform. The tubes were not vortexed, but they were inverted gently for up to 30 min to improve the efficiency of each extraction.

The sample was precipitated with 2 volumes of ethanol in a conical

flask, and the DNA was spooled from the mixture into a separate tube. Any visible supernatant was removed, but the DNA was not allowed to dry. It was dissolved in 15 ml of TE by gentle shaking at 4°C for several days.

2.4.6 Preparation of RNA from Adult Tissues

Although the method described in Section 2.4.4 has been used, a guanidinium isothiocyanate protocol was particularly effective with muscle, which was difficult to homogenise directly. The tissue was frozen in liquid nitrogen, and, while still submerged under liquid nitrogen, it was crushed into a fine powder with a pestle and mortar. The powder was transferred to a homogeniser containing 21 ml of homogenisation mix (4 M guanidinium isothiocyanate, 20 mM NaAc, pH 5.0, 1 M β-mercaptoethanol). Six separate 5 ml tubes, containing 1.2 ml of 5.7 M CsCl dissolved in 0.1 M EDTA, pH 8.0, were filled by addition of 3.5 ml of the homogenised mixture. The tubes were centrifuged at 40,000 rpm for 24 hours in a Beckman SW50.1 rotor at 20°C. The pellet was redissolved in 0.4 ml of 0.1 x TE, and the solution was extracted with 0.4 ml of a 4:1 (v/v) mixture of chloroform:butanol. The aqueous layer containing purified RNA was precipitated with 2.5 volumes of ethanol.

SECTION 2.5 TRANSCRIPTIONAL ANALYSIS - PROBES AND PRIMERS

The restriction maps of the actin, actbin and hisbin clones are presented in Figures 2.1 and 2.2. Their construction and organisation are discussed in Chapters 4 and 6. This reference section summarises the structure and synthesis of probes and primers used in experiments described in Chapters 5, 6 and 7 of this thesis.

2.5.1 Probes for S1 Nuclease Analysis

The actin probe (see Section 5.2.3) was made by isolating the *Hin*fI fragment of pXB-C1 containing the second exon of the cardiac actin gene (Figure 2.1). This fragment was restricted with *Bst*N1, treated with alkaline phosphatase, and then 5' end-labelled with T4 polynucleotide kinase and [γ-^{32}P] ATP. The mixture of resulting products was used for S1 analyses, and included a labelled strand of approximately 600 bases, which hybridised to both *X. borealis* and *X. laevis* cardiac actin RNA. A region of 147 bases was protected from S1 nuclease digestion by this RNA, under normal conditions (see Section 5.2.3).

The mouse globin probe (see Section 6.2.4) was made from a *Sau*3A fragment which spans the 3' processing and polyadenylation site of the globin gene (Figure 2.2). It was 3'-end-labelled using the Klenow fragment of *E. coli* DNA polymerase I and [α-^{32}P] GTP. Correctly processed mouse globin transcripts protect 243 bases of the 392 base probe.

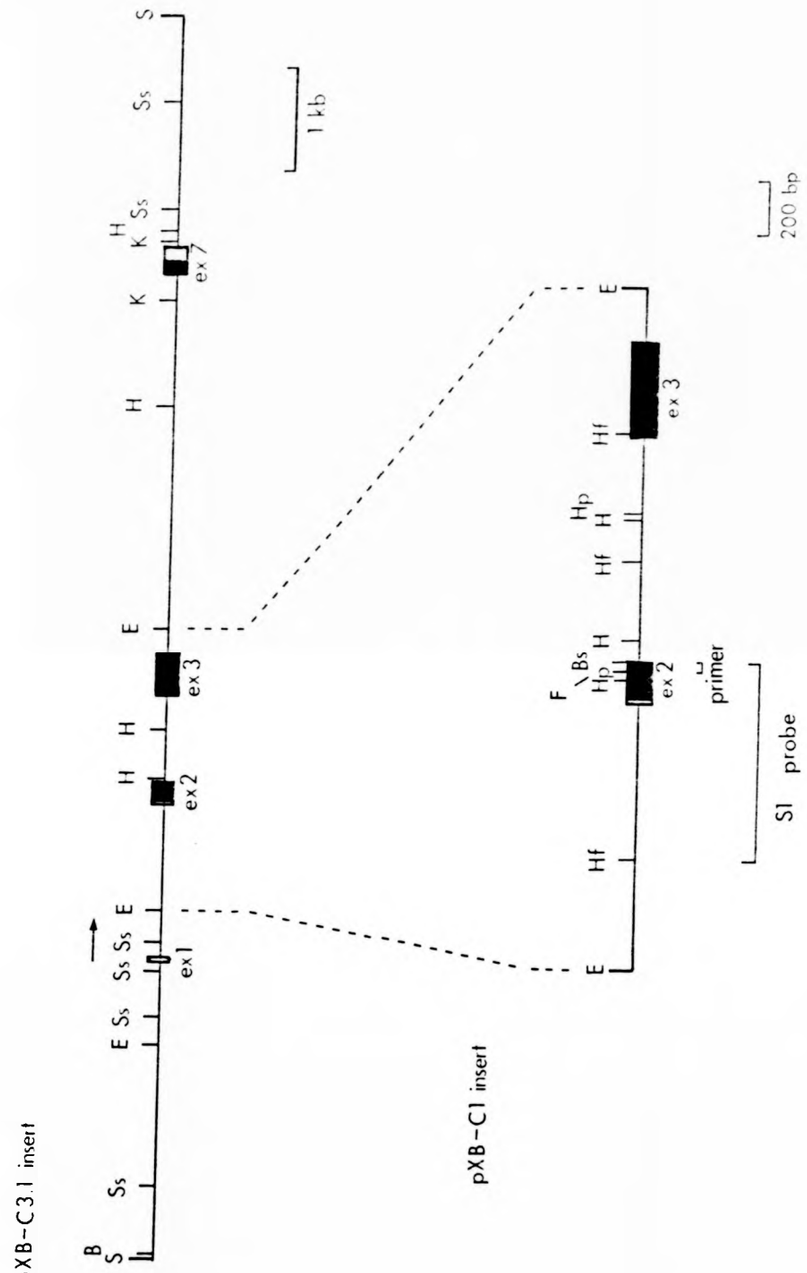


Figure 2.1 The Structure of the *X. borealis* Cardiac Actin Gene, pXB-C3.1, and Probes and Primers to Detect its Expression

The organisation of the plasmid clone pXB-C3.1, is discussed in Section 4.2.3. The 12 kb genomic insert contains the complete transcribed sequence of a *X. borealis* cardiac actin gene, with 2.8 kb upstream of the start of transcription and 2.2 kb downstream of the 3′ end. Empty bars represent 5′ and 3′ untranslated regions, while solid bars represent coding regions. The arrow indicates the start and direction of transcription.

The *Bst*NI/*Hinf*I fragment used for S1 analysis is shown, together with the 36 b *Fok*I/*Bst*NI fragment used for primer extension. Under normal S1 digestion conditions, a stretch of 147 b in the 600 b probe is predominantly protected by cardiac actin mRNA (see Section 5.2.3). The length of the major primer-extended product was 192 b (see Section 5.2.1). Not all the *Bst*NI or *Fok*I sites are marked on the map. Restriction sites: B = *Bam*HI; Bs = *Bst*NI; E = *Eco*RI; F = *Fok*I; H = *Hind*III; Hf = *Hinf*I; Hp = *Hpa*II; K = *Kpn*I; S = *Sal*I; Ss = *Sst*I.

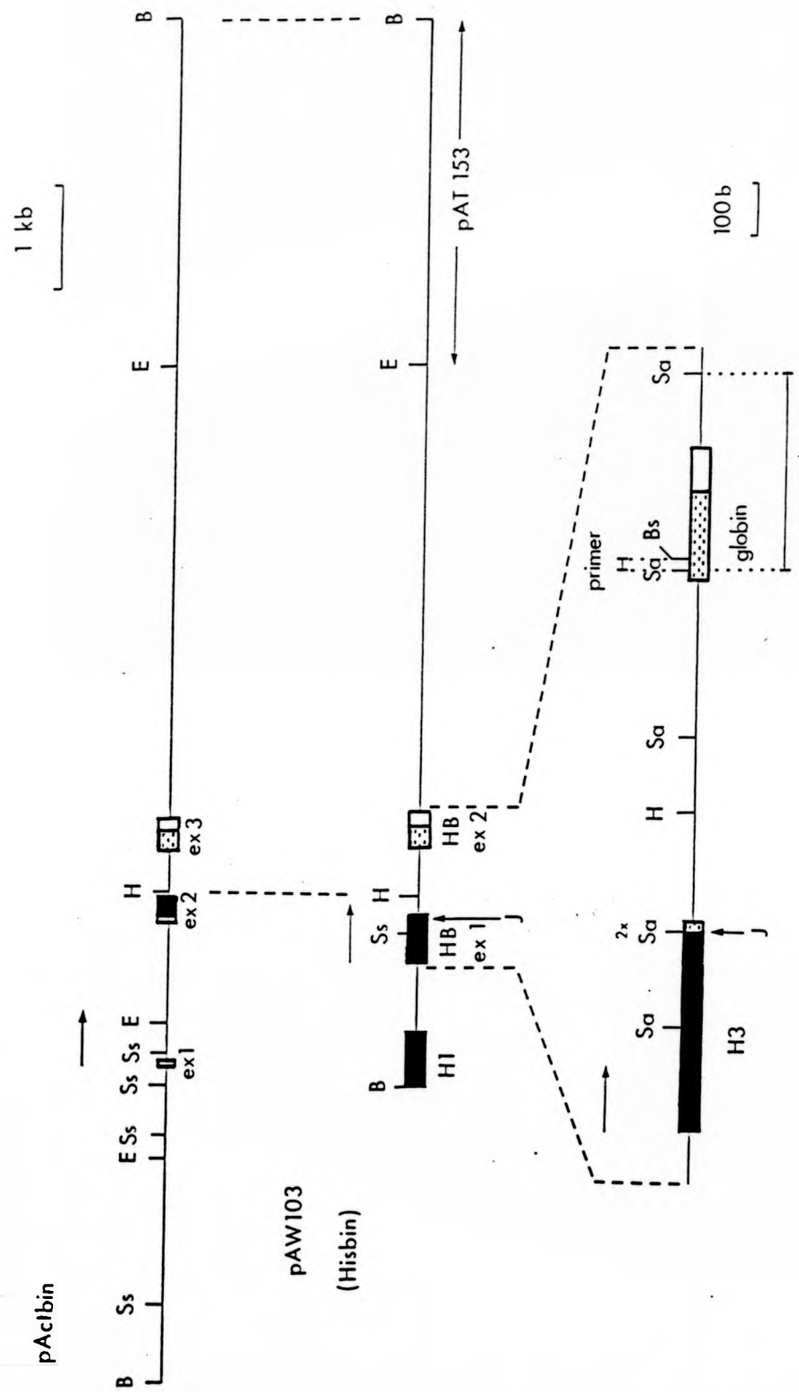


Figure 2.2 The Structure of the Actbin and Hisbin Clones, and Probes and Primers to Detect their Expression

The actbin clone is shown at the top and the hisbin clone below it. Their construction is discussed in Section 6.2.3. The final mouse β-globin exon is shared by both constructs. Empty bars represent 5′ and 3′ untranslated regions, the stippled bars represent globin coding region, and the solid bars represent both actin and histone coding sequences. The arrow indicates the start and direction of transcription of the actbin and hisbin (HB) genes. J marks the boundary between *Xenopus* histone H3 and mouse globin sequence in the hisbin. The 3′ part of a *Xenopus* histone H1 gene is also present in the hisbin construct.

An enlarged portion of the hisbin is illustrated, showing the *Sau*3A fragment used for S1 analysis of transcription from both actbin and hisbin clones. S1 nuclease digestion of a hybrid between this 392 b probe and correctly processed mouse β-globin mRNA leaves a 243 b protected product (see figure 7.9) The 24 b *Bst*NI/*Sau*3A primer is also indicated. When hybridised to actbin RNA, it is extended to a major product of 235 b in length by reverse transcriptase (see Figure 7.11). Not all the *Bst*NI sites are marked. Restriction sites: B = *Bam*HI; Bs = *Bst*NI; E = *Eco*RI; H = *Hind*III; Sa = *Sau*3A; Ss = *Sst*I.

2.5.2 Primers for Primer Extension

Preparation of the actin primer from the second exon (see Section 5.2.1) initially involved all the steps required for the synthesis of the S1 probe from the *Hinf*I fragment of pXB-C1. The kinased mixture of fragments was further restricted with *Fok*I (see Figure 2.1), and the digested DNA was electrophoresed on an 8% polyacrylamide-urea gel. A 36 base *Fok*I/*Bst*N1 labelled product could be clearly separated from its unlabelled 39 base complementary strand by this method. The former was excised from the gel and used as a primer.

The oligonucleotide usually used to primer extend from cardiac actin mRNA is complementary to the last fifteen bases of RNA coded by the first exon in pXB-C3.1 (see Section 6.2.2); another oligonucleotide complementary to the last twenty-two bases has also recently been synthesized. The preparations of oligonucleotides contain a number of smaller DNAs as well as the full-length product. These contaminants were removed by electrophoresis of several micrograms of the mixture on a 20% polyacrylamide gel, and excision of the required product after staining with ethidium bromide.

The purified oligonucleotide, which does not carry a phosphate group at its 5′ terminus, was 5′ end-labelled with T4 polynucleotide kinase and [γ-^{32}P] ATP. Kinased material was separated from unlabelled DNA by electrophoresis on a 20% polyacrylamide-urea gel. The excised labelled oligonucleotide was used for primer extension.

The primer from the globin exon (see Section 6.2.4) was made from the same *Sau*3A fragment used for S1 analysis (see Figure 2.2) by additional

digestion with *Bst*N1. After treatment with alkaline phosphatase, the products were 5' end-labelled with T4 polynucleotide kinase and [γ-^{32}P] ATP, and separated on an 8% polyacrylamide-urea gel. The smallest fragment (24 bases) was clearly fractionated from its complementary strand (27 bases), and was excised for use as the primer.

CHAPTER 3

CONSTRUCTION OF *XENOPUS BOREALIS* GENOMIC LIBRARIES

SECTION 3.1 INTRODUCTION

The construction of genomic libraries that represent most of a eukaryotic genome and can be easily screened for specific sequences, has been made relatively straightforward by the use of high capacity cloning vehicles, particularly λ replacement vectors and cosmids (Murray, 1983; Collins and Hohn, 1978). Although λ replacement vectors cannot carry such large inserts as cosmids, library construction and screening are more convenient and efficient with the former (for review, see Murray, 1983), and so they are frequently selected when maximal insert size is not at a premium.

In order to be infective, the phage genome must usually be between 80% (40 kb) and 100% (50 kb) of its wild-type length. As 20 kb of wild-type DNA is not essential for productive phage infection, it may be removed and replaced by other sequences. In λ replacement vectors, the phage genome has been manipulated so that the non-essential regions are clustered in a central 'stuffer' segment, which can be excised from the phage arms by a single restriction digest. The intact arms may then be purified by separation on a sucrose density gradient. To construct a library, lengths of digested eukaryotic or prokaryotic genomic DNA, which have the same cohesive ends as the restricted phage DNA, are ligated between the two arms. The non-essential region is therefore replaced and potentially viable transducing phage are formed, provided the total DNA length is within the limited size range. The ligation mixture is packaged into phage particles *in vitro* (Hohn and Murray, 1977; Sternberg *et al.*, 1977), and these are then used to infect a bacterial host. For such large DNA molecules, this method of

introducing recombinants is approximately 100 times more efficient than transfection of bacterial cells. The recombinant phage may be screened on agar plates by direct transfer of DNA from plaques to nitrocellulose filters and subsequent hybridisation to radioactively labelled probes (Benton and Davis, 1977).

The development of λ replacement vectors, *in vitro* packaging systems and efficient screening methods has therefore allowed extremely rapid cloning of specific DNA fragments up to a theoretical maximum of about 23 kb in length. In a preliminary screen, I isolated an 11 kb genomic fragment (from a λ library that was available in the laboratory at the time), which contained all of the cardiac actin gene except the first exon and its flanking regions. Thereafter, my aim was to use a replacement vector to clone the entire actin gene by screening a library with a highly specific probe derived from another previously isolated, incomplete cardiac actin clone (see Chapter 4). The following paragraphs discuss the structure and suitability of the vehicle EMBL3 for this purpose, while the remainder of the chapter considers the techniques employed for library construction, and the problems involved in obtaining an appropriate clone.

3.1.1 The Replacement λ Vector, EMBL3

A map of EMBL3 is shown in Figure 3.1, together with λ1059, the vector from which it was derived, and wild-type λ. Basically λ1059 was constructed by moving the essential genes (i.e. the group of functions including *N*, *O* and *P*) in the middle of the wild-type bacteriophage towards the right end, and manipulating and deleting the *Bam*HI sites, so

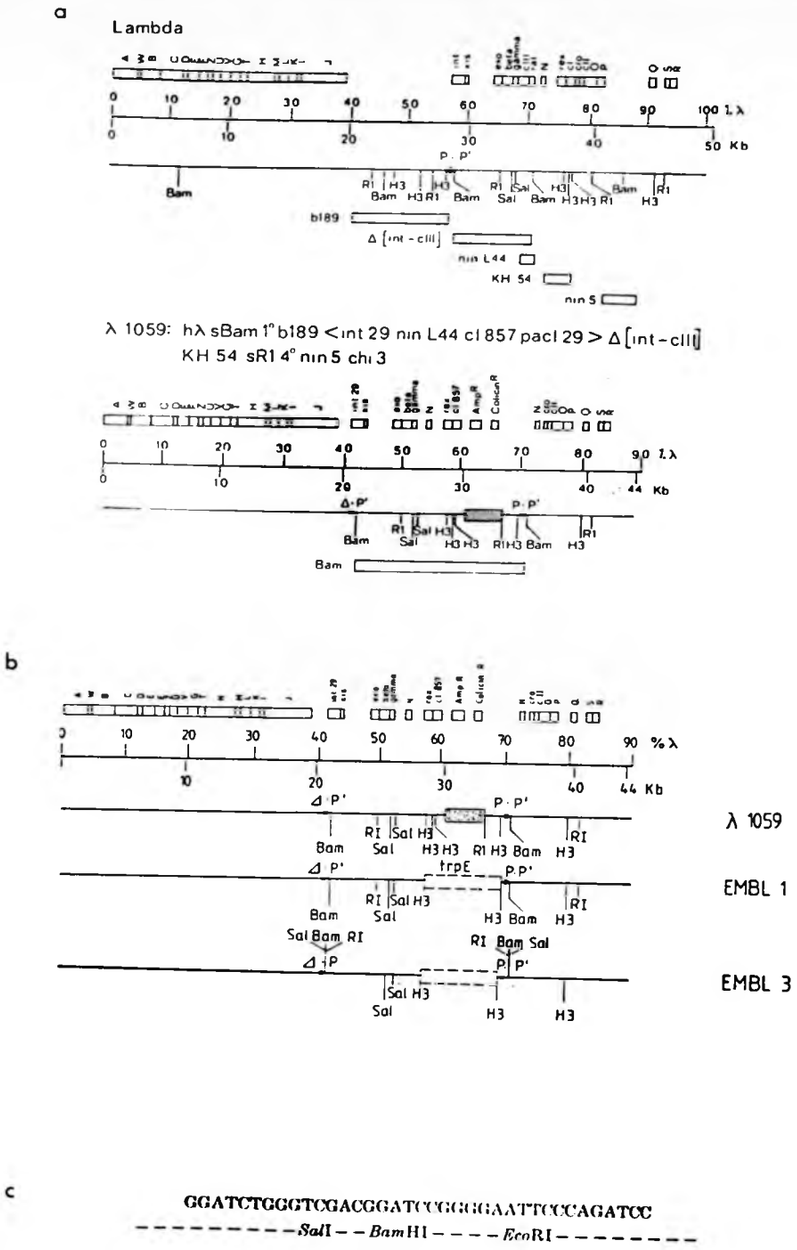


Figure 3.1 The Structure of Wild-Type λ Bacteriophage, and the Cloning Vectors, λ1059 and EMBL3

(a) Comparison of wild-type λ with λ1059 (from Karn *et al.*, 1980). Basically the latter has been modified by displacing the essential functions in the centre of the wild-type phage (including *N*, *O* and *P*) to the right hand side of the genome. The closely linked *gamma* and *red* functions remain in the stuffer fragment and are transcribed from the pL promoter, which is also in this segment. Deletion of some of the *Bam*HI restriction sites allows this enzyme to be used to excise the central region in preparation for cloning. Restriction sites: Bam = *Bam*HI; H3 = *Hind*III; R1 = *Eco*RI; Sal = *Sal*I.

(b) Comparison of λ1059 with EMBL3 (from Frischauf *et al.*, 1983). λ1059 was converted to EMBL1 by replacing the ColE1 plasmid sequence in the stuffer fragment with the *trp*E locus of *E. coli*. A further modification of the two *Bam*HI sites, by which a short polylinker was inserted in each, led to the construction of EMBL3.

(c) Sequence of the short polylinker in EMBL3 (starting and ending at the positions of the original *Bam*HI half-sites). The additional restriction sites can be used in a biochemical selection of recombinants and to excise the whole insert from genomic clones (see text).

that the two that remained were flanking the non-essential central region (Karn *et al.*, 1980; Figure 3.1a). The restriction enzyme *Bam*HI is ideal for removing the stuffer fragment, because the short cohesive ends it generates are also produced by the enzymes *Bgl*II, *Mbo*I and *Sau*3A. *Mbo*I and *Sau*3A recognise tetranucleotide sequences, so when they are employed to partially restrict genomic DNA, a more random population of high-molecular-weight fragments is formed than when enzymes recognising hexanucleotide sequences are used. The capacity of the vector is close to the theoretical limit (approximately 23 kb).

A further feature of λ1059 is that the central region contains the *red* and *gamma* genes, and so the vector is *Spi*$^+$ (sensitive to P2 interference). It cannot form plaques on strains of *E. coli* which are lysogenic for phage P2, because the combined *red* and *gamma* functions inhibit the host's RecBC nuclease. This enzyme must be active in the lysogenic cells for them to survive. On replacement of the central region with other DNA, the bacteriophage becomes *red*$^-$ *gam*$^-$ and can produce plaques on a *recA*$^+$ host lysogenic for P2 (Zissler *et al.*, 1971). Recombinants may therefore be selected by infecting such a bacterial strain with the packaged ligation mixture.

EMBL3 (Frischauf *et al.*, 1983; Figure 3.1b) has been modified further to incorporate a small polylinker sequence in place of the *Bam*HI site. The *Bam*HI recognition sequence is flanked by *Eco*RI and *Sal*I sites (Figure 3.1c). Inserts in clones from partial *Sau*3A libraries may therefore be excised by *Sal*I. Further, when the vector is prepared for library construction, it can be restricted with *Eco*RI, as well as *Bam*HI, and then precipitated with isopropanol to remove the short *Bam*HI/*Eco*RI linker fragment, which remains in solution (see Section 2.3.4). After

double digestion, the cohesive ends of the central fragment are no longer the same as those of the arms, and so the complete mixture can be ligated to genomic DNA without prior purification of the bacteriophage arms. The prevention of direct reincorporation of the stuffer by this method provides a biochemical enrichment for recombinants of up to two orders of magnitude (Frischauf *et al.*, 1983). The ColEI plasmid sequence within the central fragment of λ1059 has also been replaced in EMBL3 by a region of *E. coli* DNA (Figure 3.1b), a modification first made in the precursor of the vector called EMBL1. When libraries are screened with probes made from plasmids, this avoids possible cross-hybridisation of DNA from reconstituted λ vector with contaminating plasmid vector sequences in the probe.

Therefore, EMBL3 is one of the most versatile, high-capacity replacement λ vectors available, and it was selected for the construction of the *X. borealis* genomic libraries used in this project.

SECTION 3.2 RESULTS

3.2.1 Preparation of Genomic DNA for Library Construction

The restriction enzyme *Sau*3A, which recognises a tetranucleotide sequence and generates the same cohesive ends as *Bam*HI, was selected to partially digest *X. borealis* genomic DNA. Two methods were employed to produce populations of fragments with different susceptibilities to *Sau*3A digestion (see Figure 3.2). In the first (Method I), the chromosomal DNA was partially restricted, so that only the smaller fragments were in the range of 15-20 kb, and the entire unfractionated digest was used in the ligation reaction. However, although this procedure is extremely rapid and simple, in order to maximise the insert size, a digest must be chosen where most of the material is too large to clone, and this will select against infrequently cut genomic regions. Therefore, in the alternative technique (Method II), the digestion was allowed to proceed much further, until almost all the fragments were less than 10 kb in length. The DNA was fractionated on a sucrose gradient, and DNA longer than 10 kb was purified for cloning. Thus, regions containing few or poorly susceptible *Sau*3A sites were selected.

Method I

In this procedure, 1 mg of genomic DNA was digested with 20 units of *Sau*3A in a total volume of 8 ml of core buffer at 37°C. Aliquots of 0.48 ml (60 μg of DNA) were removed at 5 minute intervals from 0-50 minutes, and at 10 minute intervals thereafter. They were immediately added to 16 μl of 0.5 M EDTA (pH 8.0), and the enzyme was heat-inactivated by incubating at 70°C for 10 minutes. For selected samples,

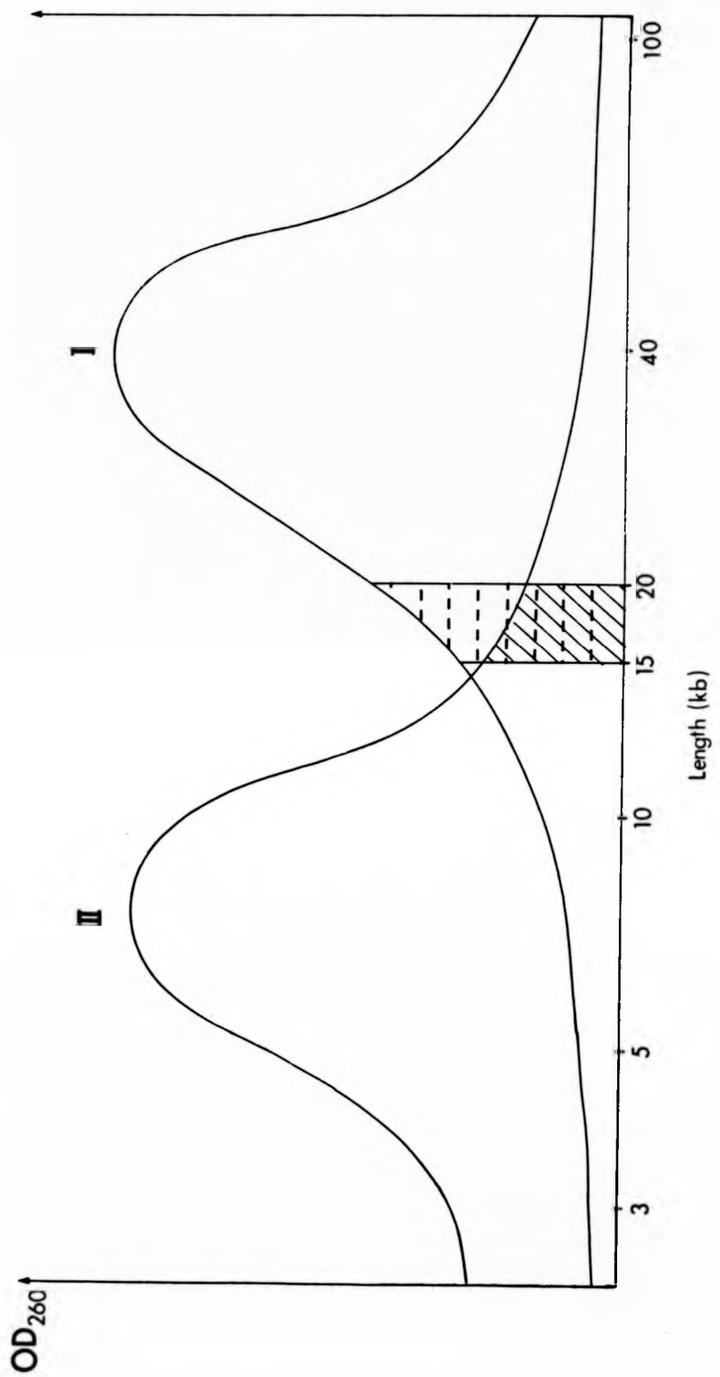


Figure 3.2 Diagram of Distribution of DNA Size in Different Partial Digests of Genomic DNA

This hypothetical graph plots OD_{260} v. fragment length, and represents two extreme size distributions in partial *Sau*3A digests which could be used to make genomic libraries containing inserts of 15-20 kb. The fragments in the appropriate size range for cloning are shaded. In one digestion, most of the DNA is too long to be successfully cloned, and the complete mixture of restricted fragments can be added to the ligation mix (Method I). In the other, most of the DNA is shorter than the required length and must be removed beforehand on a sucrose density gradient (Method II). As *Sau*3A does not cut the genome randomly, the genomic regions in each library will differ, according to their susceptibility to the enzyme. A large library made from DNA prepared by Method I did not include any clones, which hybridised with the cardiac actin gene-specific probe. However, even a library prepared by Method II contained fewer than the expected number of cardiac actin clones, suggesting that there is an unusual distribution of *Sau*3A sites at this gene locus (see Section 3.3).

8 μl aliquots (containing 1 μg of DNA) were taken, and electrophoresed overnight on a 0.4% agarose gel (Figure 3.3). Fractions containing material with a number average (median) length slightly greater than 20 kb were chosen. It has previously been estimated that such material produces a peak intensity of ethidium staining at approximately 40 kb on agarose gels, because ethidium stains longer molecules more strongly (Frischauf *et al.*, 1983). Three samples from consecutive time points, which produced material in the appropriate size range, were pooled (see Figure 3.3). Two separate sets of digested DNA samples (G-a and G-b) were pooled in this way. These were extracted with phenol and chloroform, and then ethanol precipitated. The pellets were redissolved in 0.1 x TE, and the DNA was phosphatased and prepared for library construction, as described in Section 2.3.5.

Method II

In order to effect a more complete digestion, 350 μg of genomic DNA were restricted with 20 units of *Sau*3A for 50 min in a total volume of 2.7 ml of core buffer. The reaction was stopped and the mixture fractionated on a sucrose density gradient following the procedure discussed in Section 2.3.5. The OD_{260} of the fractions was measured, and a graph of the distribution of DNA on the gradient is presented (Figure 3.4). Some of the samples were also immediately electrophoresed on a 0.6% agarose gel (Figure 3.5a), as described in Section 2.3.5.

In the experiment shown, the highest OD_{260} was observed in fraction 29. Fraction 27 produced a peak intensity on the gel at approximately 2.5 kb, and so the number average length of DNA in the sample was shorter than this value for the reason discussed in the previous section. At a rough estimate, the number average length of DNA in fraction 29 was

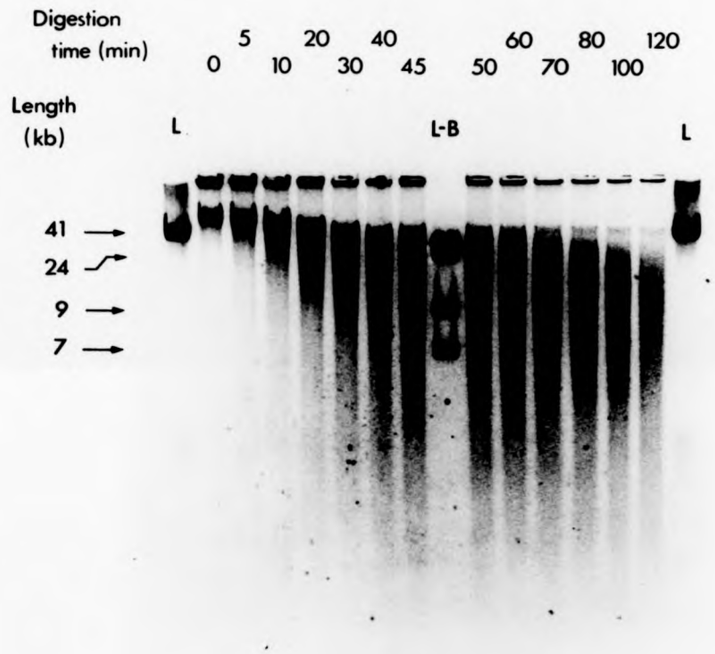


Figure 3.3 Selection of Partially *Sau* 3A-Digested Material for Cloning by Method I

In this method, 1 mg of genomic DNA was restricted with 20 units of *Sau*3A in a total volume of 8 ml at 37°C. Aliquots of 8 μl from different time points in the digest were electrophoresed overnight on a 0.4% agarose gel, together with lambda markers (L = uncut λ47.1; L-B = *Bam*HI-restricted λ47.1). The ethidium-stained gel was photographed, and the figure presents a picture of the Polaroid negative. The digestion times are indicated above each track, and the lengths of the markers are also shown on the left.

On the basis of this gel, samples from 20, 25 and 30 min (G-a), and 35, 40 and 45 min (G-b) were pooled for subsequent cloning, as the peak intensities in their tracks ranged from over 40 kb to approximately 25 kb. Only the shortest fragments, particularly in G-a, will be small enough to be cloned in EMBL3, and so the insert size in the genomic library will be maximised (see text).

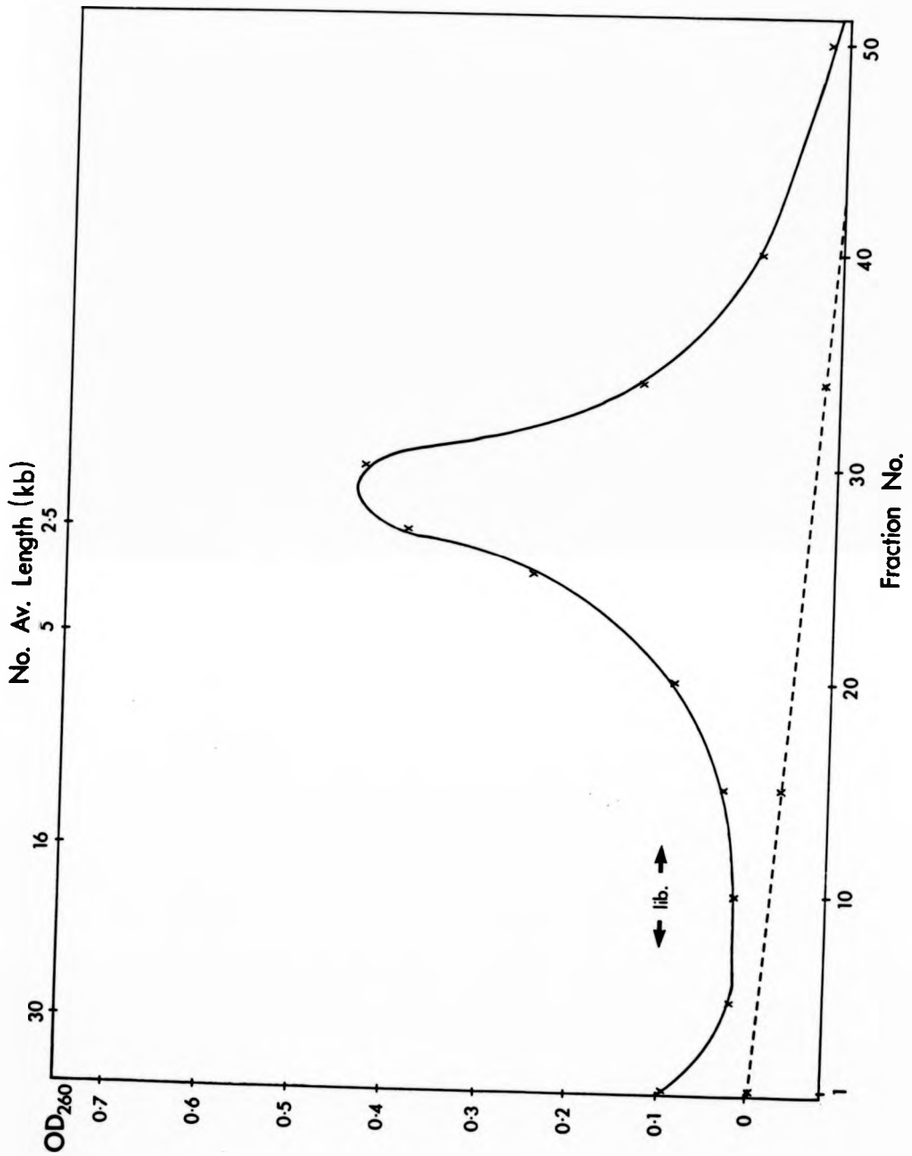
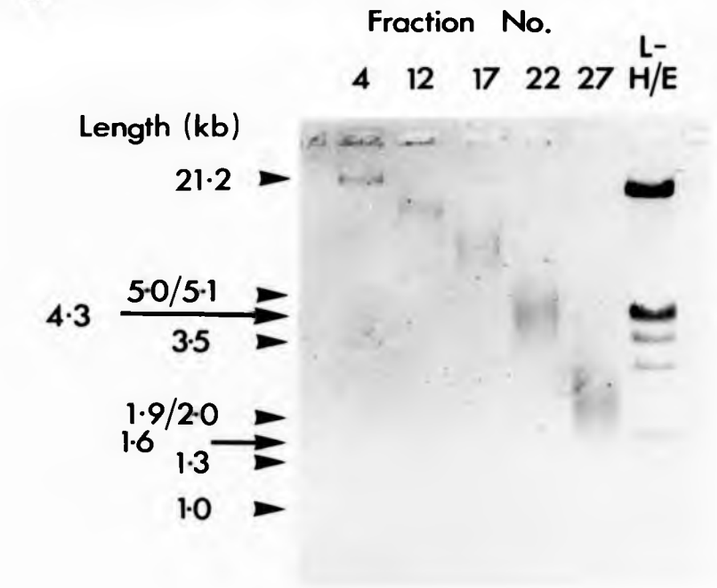


Figure 3.4 Graph of OD_{260} v. Fraction Number from a Sucrose Density Gradient Separating Fragments in a Partial *Sau*3A Digest (Method II)

A total of 350 μg of genomic DNA was restricted with 20 units of *Sau*3A for 50 min in a volume of 2.7 ml. The resulting products were separated on a 10-40% sucrose density gradient, which was then fractionated into fifty aliquots, numbered from the bottom of the gradient to the top. The OD_{260} of some of the fractions was measured and plotted on a graph (full line), together with values from a gradient containing no DNA (dashed line).

A few of the samples were also electrophoresed on an agarose gel (see Figure 3.5a) and a rough estimate of the number average length of their contents has been made from the peak intensity of ethidium staining in each track (top scale). As ethidium stains longer DNA more strongly, these values will be overestimates, and the number average length of fragments in the whole digest is probably less than 2 kb (see text). The range of fractions selected for library construction (fractions 8, 10 and 12) is marked (lib.).



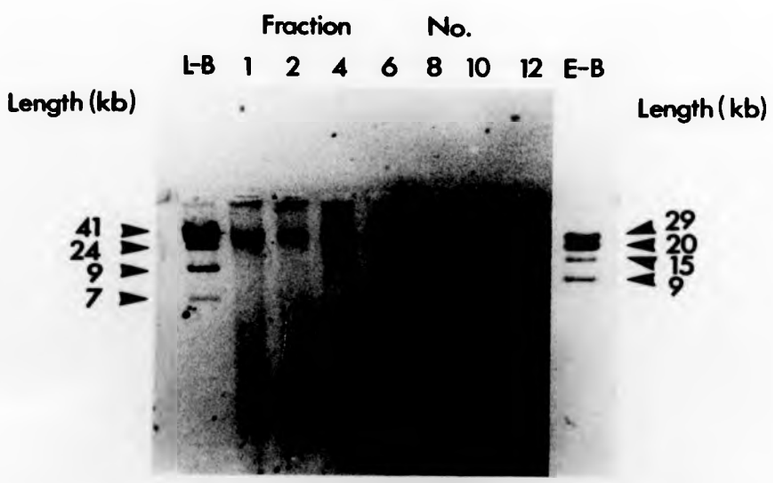


Figure 3.5 Analysis of DNA Fragment Size in Fractions Isolated from a Sucrose Density Gradient of Partially *Sau*3A-Digested Genomic DNA (Method II)

The figure shows two pictures of Polaroid negatives from photographs of ethidium-stained gels, viewed under ultraviolet illumination.

a. Aliquots from a small selection of fractions from a 10-40% sucrose gradient (see Figure 3.4) were electrophoresed on a 0.6% 3 x TAE agarose gel (see Section 2.3.5). The gel was also loaded with a sample of wild-type λ DNA that had been restricted with *Hind*III and *Eco*RI (L-H/E), and had subsequently been dissolved in the same high-salt solution as the fractions from the gradient. The lengths of the bands from the λ digest are indicated, and rough estimates of the fragment size at which ethidium staining is at its highest level have been made for each sample (see Figure 3.4). On the evidence of this gel, samples 1-12 were selected for further analysis.

b. Aliquots from selected, dialysed fractions were electrophoresed on a 0.6% agarose gel next to two marker lanes. The first marker track (L-B) contained λ47.1 DNA partially digested with *Bam*HI, while the second (E-B) contained EMBL3 DNA fully digested with *Bam*HI. In the latter, some reannealing of the *cos* ends in the phage arms has probably occurred to give a 29 kb band. Although the resolution on this gel is relatively poor, DNA fragments from fractions 8 and 10 (G-c), and 12 (G-d) were considered to be in the most appropriate size range for cloning, and were therefore processed further. It would appear that the higher levels of DNA in sample 12 are the result of more efficient precipitation during its processing, as samples 4 and 12 were stained similarly with ethidium, before this processing took place (see Figure 3.5a).

consequently 2 kb. Furthermore, like the peak intensity on an agarose gel, the maximum value on the graph of OD_{260} v fraction number is skewed towards longer DNA fragments. Therefore, given the relative symmetry of the distribution of fragment size on the graph, the median length of DNA in the complete digest was almost certainly less than the estimated median length for fraction 29 (i.e. less than 2 kb).

Selected fractions (1-12), containing DNA over 10 kb in length, were dialysed and processed as described in Section 2.3.5. From the graph in Figure 3.4, all these samples represent in total only approximately 6% by weight of the original material digested, and largely contain genomic regions with the fewest or least susceptible *Sau*3A sites. Aliquots from most of these samples were electrophoresed on a 0.6% agarose gel (Figure 3.5b). Fractions 8, 10 and 12 appeared to span the appropriate size range for cloning. Fractions 8 and 10 were pooled (G-c), while fraction 12 was kept separately (G-d). The latter contained a relatively high level of DNA compared to the fractions in G-c, probably because of more efficient ethanol precipitation during its processing.[5]

Both G-c and G-d were phosphatased in preparation for library construction, following the procedure of Section 2.3.5. Such treatment largely eliminates the complication of cloning multiple inserts in a single recombinant. The ethanol precipitates were redissolved in 0.1 x TE at an estimated concentration of 0.1 mg/ml.

3.2.2 The Construction of Genomic Libraries

In their original report, Frischauf *et al.* (1983) used EMBL3 to clone

genomic DNA from mouse liver. The bacterial strain Q358 and its P2 lysogenic counterpart Q359 (Karn *et al.*, 1980) were infected with the packaged ligation mix to construct the library. Using the doubly cleaved vector for biochemical selection, a similar number of plaques was observed with both strains (indicating that most clones were recombinants) and approximately 2×10^6 plaques were formed for every microgram of mouse DNA (prepared by Method I).

When the same bacterial strains were infected to construct the *Xenopus* libraries, the number of plaques on Q358 and Q359 were also approximately equal (Table 3.1). However, the yield of plaques per microgram of genomic DNA was surprisingly low for both G-a and G-b, considering that the packaging efficiency was similar to that in the experiments of Frischauf and his colleagues. Aliquots from some of the ligation mixes were electrophoresed on an agarose gel (Figure 3.6). Clearly the addition of restricted EMBL3 to phosphatased genomic DNA led to ligation between the different molecular species to give products in the predicted size range, indicating that there was no apparent problem with the ligation itself.

An alternative explanation for the low yield is that methylation of the eukaryotic genomic material was preventing normal phage propagation. It has been found with plant DNA, which is highly methylated, that an alternative bacterial strain, K803 (Wood, 1966), is much more suitable for cloning than Q358 and Q359 (T. Gallacher, per. comm.). It is not a P2 lysogen, and so the biochemical selection of doubly digested EMBL3 must be relied upon to reduce the level of reconstituted phage. However, unlike Q358 and Q359, it will allow the persistence and replication of DNA, which is initially heavily methylated, leading to

Table 3.1 Characterisation of Suitable Bacterial Strains, and Target: Vector Ratios for Efficient Library Construction

Ligations	Components of ligation			pfu/μg vector (x 10^{-5})			minimum number of recombinants (x10^{-5})	
	Vector (μg)	Insert (μg)		Q358	Q359	K803	/μg vector	/μg genomic insert
A	0.7	-		3.0	0.5	2.3	-	-
B	0.7	0.7	G-a	0.1	0.1	2.4	2.3	2.3
C	0.7	0.25	G-a	1.0	0.7	5.2	4.2	11.8
D	-	0.7	G-a	0	0	0	-	0
E	0.7	0.7	G-b	0.3	0.3	1.7	1.4	1.4
F	0.7	0.25	G-b	1.0	0.9	5.3	4.3	12.0
G	-	0.7	G-b	0	0	0	-	0

Table 3.1 Characterisation of Suitable Bacterial Strains, and Target: Vector Ratios for Efficient Library Construction

Genomic DNA was prepared by Method I (G-a and G-b; see Section 3.2.1) and ligated to EMBL3 vector in a fixed volume of 14 μl, containing 50 μg/ml of EMBL3 and 0.5 units of T4 DNA ligase. The components of seven different ligations (A to G) are listed in the table. Aliquots of 4 μl from ligations A and B were electrophoresed on an agarose gel (see Figure 3.6) to test the ligation step. Then 5 μl samples from ligations A to G were packaged and small aliquots of the resulting phage were used to infect three different bacterial strains, Q358, Q359 and K803 (see text). The number of plaques formed on each strain is listed. The packaging efficiency was about 3 x 10^7 pfu/μg for uncut lambda vector in this experiment. Uncut EMBL3 infected Q359 at a level of less than 0.1% of its infectivity in Q358, because of the P2 selection in the former strain. Restricted EMBL3, ligated by itself (ligation A), produced a significant number of plaques, some of which were also formed in a P2 lysogen. It is unclear if these result from rearrangements or the cloning of contaminating DNAs, but they are largely lost on addition of the genomic target (compare A and B in Q358). Not surprisingly, genomic DNA on its own did not produce any plaques (ligations D and G).

As can be clearly seen, the libraries from ligations B, C, E and F all gave the best titres in K803, probably because of the high level of methylation in the *Xenopus* genome (see text). The estimated minimum number of recombinants is derived from subtracting the yield of plaques in Q358 from that in K803. However, as there are nearly as many plaques in Q359 as in Q358, it would seem likely that most of the infective phage in Q358 contain inserts derived from a subset of genomic DNAs that can be cloned in this strain, and so the calculated value is an underestimate.

The results show that a molar ratio of roughly 2:1:2 for left arm:insert:right arm (ligations C and F) gives a higher yield than with excess genomic insert, an observation that is theoretically predicted (Maniatis *et al.*, 1982).

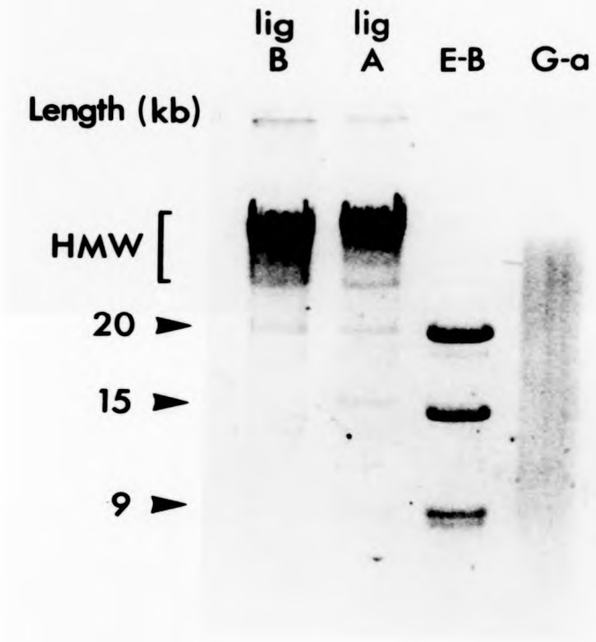


Figure 3.6 Test of the Ligation Step in the Synthesis of a Genomic Library

EMBL3 DNA, prepared by the procedure described in Section 2.3.4, was added to a ligation reaction with and without the genomic DNA preparation G-a (ligations B and A respectively; see Table 3.1). Aliquots of these mixtures were electrophoresed on a 0.4% agarose gel, together with samples containing equal amounts of the original components of the ligations, the EMBL3 vector restricted with *Bam*HI (E-B) and the genomic insert (G-a). In ligation A, much of the genomic DNA migrated with high-molecular-weight material (HMW), some of which consists of phage arms and insert in the appropriate arrangement to form viable recombinant phage. Other tests have shown that ligation reactions containing G-a alone do not produce significant levels of high-molecular-weight DNA, confirming that this material has been efficiently phosphatased. The upper bands in ligation A are formed by concatenation of the phage arms or of the stuffer fragment.

the synthesis of non-methylated daughter molecules.

Like plants, the genome of *Xenopus* is also highly methylated (see Burdon and Adams, 1980). Approximately 8% of its cytosines are modified, compared to only about 3% in most tissues of the mouse. The packaged ligation mixes, that gave low yields in Q358 and Q359, were therefore used to infect K803, and the results are presented in Table 3.1. Clearly many more plaques were obtained, and the minimum number of recombinants per microgram of vector and target insert in each ligation is estimated by comparison with the yield using Q358. This value only represents a rough approximation, but it confirms that the cloning efficiency is similar to that reported by Frischauf *et al.* (1983).

If the unsuitability of Q358 and Q359 was due to methylation, then phage recovered from plaques formed on a lawn of K803 in the first library amplification should subsequently be infective in the former strains, as the modified bases are lost during replication in bacteria. Indeed, 7 out of 9 randomly selected clones from a library containing the genomic DNA pools G-a and G-b (library 1), which was originally plated in K803, subsequently infected Q359. Before amplification, the number of phage from this library, which could replicate in Q359, was only about 6% of the number in K803. Therefore, if amplification in K803 had no effect on the infectivity in Q359, there would only be roughly a 5×10^{-6}% chance of picking 7 infectious clones in 9 selections. Hence, there can be no doubt that a single amplification in K803 alters the properties of the cloned genomic inserts in such a way that they can then be replicated in Q359, and this is presumably the result of demethylation.

The data in Table 3.1 also reveal that both pools of digested genomic

DNA, G-a and G-b, produced high yields of clones. The efficiency was decreased by adding excess quantities of insert (compare B and E with C and F). Although it was difficult to estimate the molar quantity of inserts in the ligation, as the size of genomic DNA in the fractions varied over such a wide range, the most productive molar ratio of left arm:insert:right arm was roughly 2:1:2, which is the theoretical optimum (see Maniatis *et al.*, 1982).

Table 3.2 summarises the components and sizes of all the libraries constructed and indicates the theoretical probabilities of isolating a single copy locus from each of them. The latter values are calculated from a rough estimate of the average insert size, made from the known lengths of a few characterised clones in each library. However, despite the high probabilities, only in library 3, in which DNA prepared by Method II was used, was a clone containing the cardiac actin gene found (see Section 4.2.3). Even in this library there was only a single positive clone, despite the fact that approximately five genome lengths of DNA are represented.

Of course, one obvious explanation of this phenomenon is that the libraries were comprised of far fewer genomic clones than had been estimated. The only test to eliminate this possibility was to screen them for other genes. A part of library 1 has been probed with a labelled fragment from the *X. borealis* histone gene cluster, and the number of positive clones found (E. Bagenal, per. comm.) was exactly the figure estimated from its known copy number in the genome (approximately 70 repeats per haploid genome; Turner and Woodland, 1983). Furthermore, preliminary evidence suggests that the same library in its amplified form has been successfully screened for genes expressed specifically in

Table 3.2 Size of the EMBL3 Libraries Constructed to Clone the *X. borealis* Cardiac Actin Gene

Library	Genomic DNA	Minimum Number of Recombinants	Estimated Average Insert Size (kb)	Theoretical Probability of Isolating a Sequence Found as a Single Copy in the Genome
1	G-a & G-b	6.8×10^5	18	98.5%
2	G-c	3.1×10^5	15	79.0%
3	G-d	11.8×10^5	12	99.1%

Table 3.2 Size of the EMBL3 Libraries Constructed to Clone the *Xenopus borealis* Cardiac Actin Gene

Three different libraries were screened for a complete copy of the *X. borealis* cardiac actin gene, and the details relevant to calculating their sizes are given in this table. Library 1 contains two pools of genomic DNA isolated by Method I, and libraries 2 and 3 contain pools of DNA isolated by Method II. The minimum number of recombinants is determined by the calculation discussed in Table 3.1. The average insert length is roughly estimated from the sizes of the small number of clones which have been characterised and, for G-c and G-d, also from the apparent average length of the DNA in the samples before cloning (see Figure 3.5b).

The probability (P) of isolating a clone containing part of a single copy gene from a genomic library is calculated from the formula $N = \ln(1-P)/\ln(1-f)$, where N = the number of clones in the library and f = the fraction of the haploid genome that is represented by a single recombinant (Clarke and Carbon, 1976).

Despite the high theoretical probabilities, especially in libraries 1 and 3, only library 3 contained a suitable clone. Even in this case, as there are about five genome lengths represented (approximately 1.4×10^7 kb), the gene is not found at such a high frequency as would be expected if fragments were cloned entirely randomly. This anomaly is discussed in Section 3.3 of this chapter.

Xenopus liver (R. W. Old, per. comm.), for two genes expressed in early *Xenopus* development (A. Mohammed, per comm.) and for homeo-box-containing genes (P. C. Turner, per. comm.), all of which are likely to exist at low copy number in the genome. It can therefore be concluded that the poor recovery of genomic clones containing the cardiac actin gene is due to a property of the gene itself and not the general representativeness of the library. A possible explanation for this bias is discussed in the final section of the chapter.

The two lambda clones of the cardiac actin gene which I have isolated, λXB-C2 (cloned in λ47.1) and λXB-C3 (cloned in EMBL3), both share the common property that they form small plaques with a titre approximately 50 times lower than most other plaques. This property may be associated with specific sequences in the genomic insert. All other clones I have isolated, whether they contain different actin genes or completely unrelated inserts, do not behave in this fashion. Once either λXB-C2 or λXB-C3 were purified, large quantities of DNA could still be made using standard methods (see Sections 2.3.2 and 2.3.3) merely by adjusting the ratio of phage that was added to host bacteria. However, initial isolation of a clone of this type is made difficult for three reasons. Firstly, it will be gradually lost from a library on repeated amplification, and so it is advisable to screen the library after the first amplification. Secondly, if the plaques are overlapping when the library is plated, phage which grow poorly can be competed out. Confluent plaques should therefore be avoided. Finally, the phage from a plug of agar containing the clone, which was picked after the first amplification, must be plated at relatively low plaque density on numerous plates, in order to increase the probability of selecting a single positive plaque on the second screen.

The reason for the unique behaviour of these actin clones is unknown. However, no gross rearrangements were observed in the inserts of either λXB-C2 or λXB-C3, as judged by comparison with the restriction map derived from genomic Southern blots (see Table 4.1). There is therefore no evidence that any regions in the gene have been removed to increase the titres of plaques derived from the clones to an acceptable (albeit low) level.

As λXB-C2 was already known to replicate poorly, all the necessary precautions were taken in the screening of the EMBL3 libraries to avoid competition, and so it is unlikely that the unusual amplification properties can in any way account for the low recovery of appropriate cardiac actin clones in these experiments.

SECTION 3.3 DISCUSSION

The successful isolation of a complete cloned cardiac actin gene suitable for expression studies proved to be more difficult than first expected. Although other evidence suggests that the constructed libraries were generally representative, they did not contain the predicted number of cardiac actin clones.

The most convenient method for the preparation of partially digested genomic DNA for construction of libraries (Method I) involves the restriction of DNA to a level where many or most of the products are still too large to be cloned in lambda. The entire mixture of fragments can therefore be ligated to the vector without any purification step and only the smaller DNAs will be cloned. In order to maximise the insert size, I used partial digests in which only the shortest material was less than 20 kb, and this is reflected in the average length of genomic DNA in library 1. However, this approach selects against the cloning of regions containing few or poorly susceptible *Sau*3A sites. In retrospect, it might have been more advisable to place less emphasis on optimising the insert size, so that a wider range of fragments could be cloned. However, at the time, the lengch of the actin gene transcription unit was unknown. Furthermore, screening for other genes has suggested that the cardiac actin gene is exceptional in its absence from the resulting library.

The required clone was finally isolated using DNA prepared by Method II. A simple explanation for such a result would be that the region has only relatively few or poorly susceptible *Sau*3A sites. If this were the case

though, the gene should be enriched in libraries 2 and 3, when clearly it is present at less than the expected frequency. Therefore a more acceptable interpretation is that the *Sau*3A sites are unevenly distributed in the region of interest in such a way that it is poorly represented in all the pools of DNA used for library construction. In this regard, the initial isolation of an incomplete copy of the gene (λXB-C2) from a small partial *Sau*3A library (see Section 4.2.2), before I constructed the EMBL3 libraries, either may be the result of the gene being better represented in the pool of DNA cloned or it may have merely been a fortuitous cloning event from a non-representative pool.

The proposed hypothesis for the unusual behaviour of the cardiac actin gene is difficult to substantiate, because, even if the entire insert was sequenced, it would also be necessary to sequence the flanking regions in order to determine the pattern of *Sau*3A sites in the whole region. However, as the other tests of library 1 have led to the cloning of many other genes, it would seem likely that it is the actin gene and not the library which is responsible for the problems encountered. Interestingly, the putative related cardiac actin gene (see Section 4.2.3 and Appendix II) was also only isolated from library 3, again at lower than the expected frequency.

CHAPTER 4

ISOLATION OF A COMPLETE *XENOPUS BOREALIS* CARDIAC ACTIN GENE

SECTION 4.1 INTRODUCTION

Cross (1984) previously isolated two actin clones from a partial *Eco*RI-digested *X. borealis* genomic library. One of these (λ35A) was shown to contain the first four exons of a type 1 (β-) cytoskeletal actin gene, on the basis of its amino acid sequence, intron positions, and the expression of the chromosomal gene in non-muscle tissues (also see Appendix I). The other (λ5AP, referred to here as λXB-C1; see Figure 4.1) contained the first two coding exons of a different actin gene. It encoded an amino acid sequence identical to that of mammalian cardiac actin (Vandekerckhove and Weber, 1978a; see Chapter 1), if it was assumed that the N-terminal methionine and cysteine were cleaved off after translation. This latter process takes place in all other known vertebrate muscle actin genes (e.g. Hamada *et al.*, 1982; Zakut *et al.*, 1982; Fornwald *et al.*, 1982). It was suggested, but not demonstrated, that the gene contained an intron in its 5' untranslated sequence, like all other characterised vertebrate actin genes (e.g. Zakut *et al.*, 1982; Fornwald *et al.*, 1982).

Sequencing of primer extension products, described in Section 5.2.1, has since clearly demonstrated that an intron does exist in the 5' untranslated region, and a sequence for part of the putative first exon was deduced by this technique (see Section 5.2.2). The whole insert of λXB-C1 was sub-cloned into pBR325 (pXB-C1). It was completely sequenced (Cross, 1984 and unpublished results), and shown not to contain the putative first exon sequence. It was therefore concluded that λXB-C1 lacked the first exon of the actin gene, as well as those coding sequences downstream of the third exon.

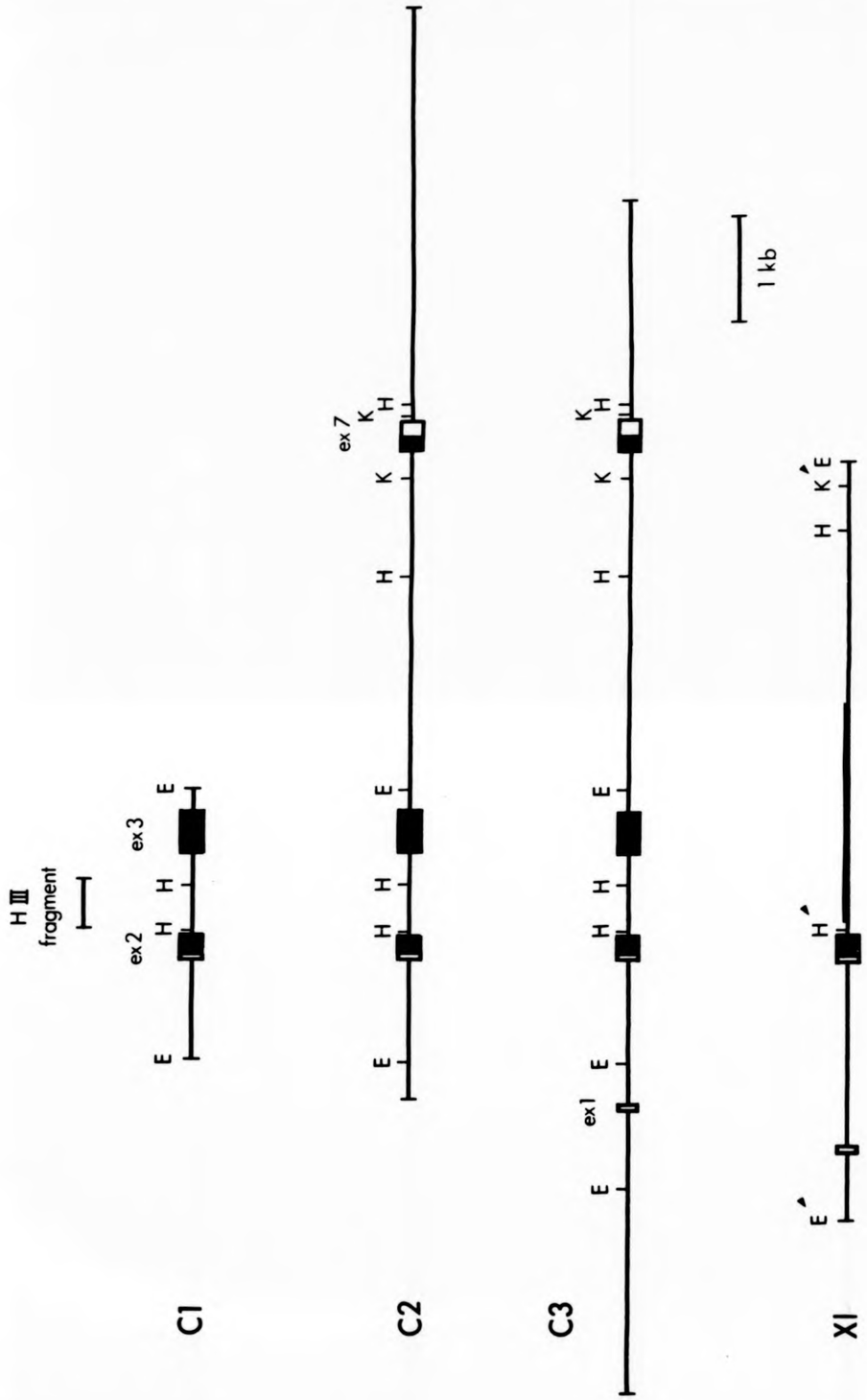


Figure 4.1 Restriction Maps of *X. borealis* and *X. laevis* Cardiac Actin Gene Clones

The upper three restriction maps show the inserts of the three *X. borealis* clones, λXB-C1, λXB-C2 and λXB-C3, derived from the locus encoding the cardiac actin gene. The position of the *Hind*III fragment used for screening genomic libraries is also indicated. Below the *X. borealis* maps is the restriction map (X1) of the cloned *X. laevis* cardiac actin gene (α1) described by Stutz and Spohr (1986). Empty bars represent 5' and 3' untranslated regions, while solid bars represent coding regions. In the *X. borealis* clones, not all the exons have yet been sequenced, but the final exon is designated the seventh, because of the highly conserved intron positions in fully sequenced, homologous genes of other vertebrates (e.g. Hamada *et al.*, 1982; Chang *et al.*, 1985). A region in the *X. laevis* gene, part of which appears to cross-hybridise with the *Hind*III probe (see Table 4.1), is marked in red. Additional restriction site data would be required to delimit this region more accurately. It is clear that only λXB-C3 contains the entire transcription unit of the *X. borealis* cardiac actin gene. A few of the restriction sites in the introns and promoter of the *X. laevis* clone (marked with an arrowhead) may be in the same position as those in the *X. borealis* gene, suggesting that the genes share homologies even in intervening and flanking sequences.

Restriction sites: E = *Eco*RI; H = *Hind*III; K = *Kpn*I.

If the regulation of expression from this gene was to be investigated in transformed *Xenopus* embryos, it was clear that a clone, at least containing the first exon and a few hundred nucleotides of upstream sequence, was required. A number of genomic libraries were therefore screened with a probe, which would not cross-react with other actin genes. This chapter discusses the selection of a probe and the characterisation of the relevant clones obtained.

SECTION 4.2 RESULTS

4.2.1 Identification of a Highly Specific Probe for the Putative Cardiac Actin Gene

In order to screen genomic libraries specifically for clones which contained the actin gene sequences of pXB-C1, it was necessary to find a probe, which would not cross-hybridise with other regions in the genome.

As the coding sequences of different actin isoforms are so highly conserved, a region was selected in an intron of the gene. A 451 bp *Hind*III fragment (see Figure 4.1) was separated from the other products of a *Hind*III digest of pXB-C1 on an agarose gel, and labelled by nick-translation. This probe only contains sequences in the second intron.

The filters from Southern blots of digested *X. borealis* genomic DNA were hybridised to the fragment (e.g. Figure 4.2). For most digests, the probe hybridised to two separate bands, one of which was more intense than the other. However, for a few enzymes, more than two bands were visible. The sizes of some of the major bands are listed in Table 4.1.

In the *Eco*RI digest, a 2.6 kb fragment reacted most strongly with the probe. This is the same length as the insert of pXB-C1, and bands in the *Hind*III and *Pvu*II digests also correlated with internal fragments in the clone (see Table 4.1). When the probe hybridised to digested genomic fragments with flanking restriction sites that were not both internal to the pXB-C1 insert, as expected, no size correlation with the lengths of positive fragments from pXB-C1 was observed. These results

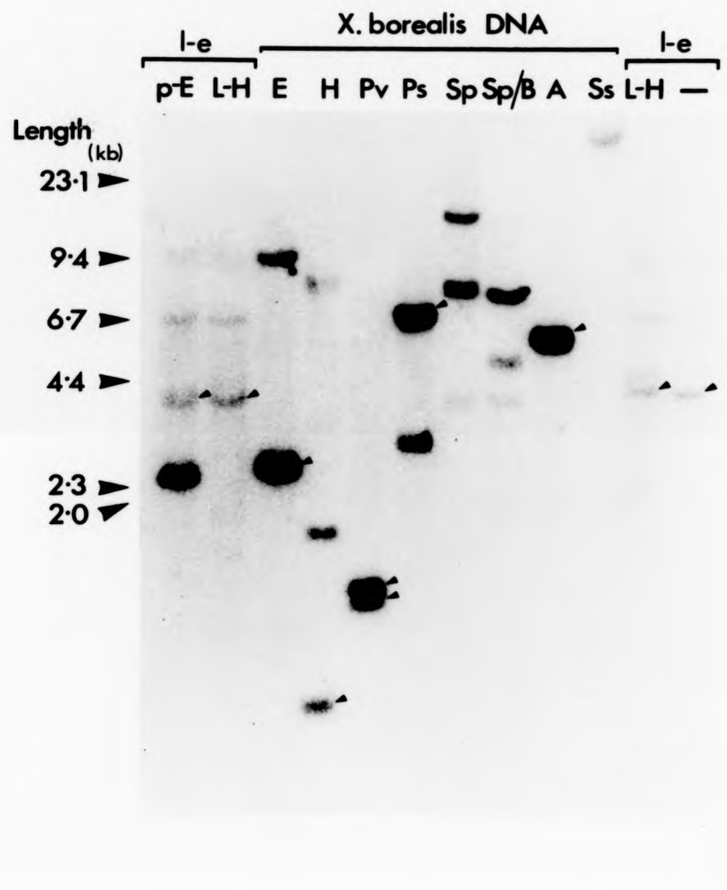


Figure 4.2 Southern Blot of Restricted Genomic DNA from *X. borealis* and *X. laevis* Probed with the *Hind*III Fragment

For each digest, 8 μg of genomic DNA was added to approximately 20 units of the appropriate enzyme in a total volume of 30 μl. After incubation at 37°C for 3 hours, the samples were immediately loaded on a 0.6% agarose gel and electrophoresed overnight. The DNA was transferred to nitrocellulose, which was subsequently hybridised to the nick-translated *Hind*III fragment under standard conditions (see Section 2.2.8). The filter was washed twice in 3 x SSC at 50°C for 10 min.

Tracks containing *X. borealis* genomic DNA: E = *Eco* RI-digested; H = *Hind*III; Pv = *Pvu*II; Ps = *Pst*I; Sp = *Sph*I; Sp/B = *Sph*I/*Bam*HI; A = *Acc*I; Ss = *Sst*I.

Tracks containing *Eco*RI-digested *X. laevis* genomic DNA (l-e) are supplemented with λDNA markers (L-H) and *Eco*RI-digested pXB-C3.1 (p-E).

The central eight lanes contain *X. borealis* DNA restricted with different enzymes. In most digests two bands are observed, one of which is more intense (marked with an arrowhead) and is derived from the locus represented by pXB-C1 (see Table 4.1). For the *Pvu*II digest, there are two strongly hybridising fragments derived from the pXB-C1 locus. Furthermore, there are four positive bands in the tracks containing *Sph*I-digested DNA (S and S/B). In pXB-C1 the *Hind*III fragment is cut at a single site by *Sph*I, and so, to account for the pattern observed, it seems likely that the enzyme also attacks the homologous region in the locus represented by λXB-Crl (see Appendix II). *Sst*I failed to restrict the DNA properly and so no conclusions can be drawn from this particular test, although the digest has since been repeated successfully.

The outer lanes (labelled l-e) all contain 8 μg of an *Eco*RI-restricted preparation of *X. laevis* genomic DNA (a kind gift from Edward Bagenal). The undigested DNA was sheared considerably more than the *X. borealis* material, accounting, at least to some extent, for the relatively weak signals. In two of these tracks (marked L-H), *Hind*III-digested, ^{32}P-labelled lambda DNA markers have been added, although the bands are too faint to see clearly at this exposure. Their sizes and migrations are indicated. In the left-hand track (marked p-E), 120 pg of pXB-C3.1 (a larger clone including the λXB-C1 insert), digested with *Eco*RI, has been mixed with the *X. laevis* genomic DNA. This digest of the clone contains the same level of the hybridising 2.6 kb *Eco*RI fragment as 3 μg of *Eco*RI-digested *X. borealis* genomic DNA (in lane E), if it is assumed that there is only a single copy of the actin gene in the haploid genome. The blot therefore confirms that the cardiac actin allele is only represented at low copy number, although a series of accurately quantitated controls would be required to make a more precise estimate. The cloned DNAs were electrophoresed with genomic DNA in order to equalise the quantity of nucleic acids in each track. Fragments of equivalent length should therefore migrate at the same speed in all lanes.

Table 4.1 Correlation of Lengths of Bands Hybridising to the *Hind*III Intron Probe in Southern Blots of Genomic and Cloned DNA

Digest	from Southern blots			from Restriction Map
	length of positive genomic fragments (kb)	length of positive fragments in pXB-C1 (kb)	length of positive fragments in pXB-C3.1 (kb)	length of putative positive fragments in α1 (kb)
	X. borealis (major band)			
*Eco*RI	2.6	2.6	2.6	-
*Hind*III	0.43	0.45	0.45	-
*Pvu*II	1.0	1.0	1.0	-
	1.1	not internal	1.1	-
*Pst*I	6.9	not internal	6.5	-
*Sst*I	7.2	not internal	7.0	-
	X. laevis			
*Eco*RI	7.5	-	-	7.0
*Hind*III	4.0	-	-	4.0
*Pst*I	3.4	-	-	3.2

Table 4.1 Correlation of Lengths of Bands Hybridising to the *Hind*III Intron Probe in Southern Blots of Genomic and Cloned DNA

Restricted genomic and cloned DNAs were electrophoresed on 0.6% agarose gels. For this comparison, only enzymes that produce a completely internal hybridising fragment (i.e. where both restriction sites are in the insert) in the *X. borealis* clones pXB-C1 and pXB-C3.1 or the *X. laevis* clone α1 (Stutz and Spohr, 1986) are considered. Particularly in the genomic Southern blots, it is difficult to accurately estimate the lengths of bands that migrate slowly, leading to the larger discrepancies in the size correlations above 6 kb.

For *X. borealis* genomic DNA, the most intense band in each digest has been selected (e.g. see Figure 4.2) for comparison. In all cases, its length is essentially identical to the positive fragment in pXB-C3.1. This observation confirms that the insert of the clone is an unrearranged and more extensive copy of the insert in the original recombinant pXB-C1. The latter also shares the same positive bands in those digests of it that delineate an entirely internal hybridising fragment.

For *X. laevis* genomic DNA, there are no clear major hybridising bands. The size of the hybridising fragment from α1 is estimated from the restriction map of Stutz and Spohr (1986), assuming that the homologous region to the *Hind*III fragment probe, lies in the same position as in the *X. borealis* gene. For all three digests, a positive band of comparable size can be found in the genomic Southern, suggesting that the α1 clone represents one of a pair of cardiac actin gene alleles sharing close homology in their intervening sequences, as well as their coding regions, with pXB-C3.1.

indicate that the major cross-hybridising band in each track is derived from the putative cardiac actin gene, of which pXB-C1 is a part. Hybridisation signals from genomic blots were compared to measured amounts of cloned actin DNA (e.g. Figure 4.2). This suggested that the locus is at low copy number, and it seems probable that it is present at a single site in the haploid genome, as is found in other vertebrates.

The probe for the blots was not only selected for its likely specificity, but also because of its small size, which reduces the number of hybridising restriction fragments from a single locus to a minimum. In most tracks containing digested genomic DNA two positive bands were observed, an observation consistent with the hypothesis that a second genomic locus contains sequences closely related to the intron probe. This region has now been cloned (λXB-Cr1; see Section 4.2.3 and Appendix II), and shown to encode an actin gene. Other evidence considered in the discussion section of this chapter suggests that it represents a second cardiac actin allele.

The *Hind*III probe also hybridised to at least two bands in genomic Southern blots of restricted *X. laevis* DNA. The lengths of one of these fragments in the *Eco*RI, *Pst*I and *Hind*III digests are consistent with the predicted sizes from the limited restriction map (see Figure 4.1) of a *X. laevis* cardiac actin gene (Stutz and Spohr, 1986; Table 4.1), if it is assumed that the probe hybridises in approximately the same position in the genes of both species. Indeed, although sequence information would be needed to confirm this, the upstream *Hind*III site, which demarcates the probe in *X. borealis*, appears to be maintained in *X. laevis*, and this also may apply to a few other sites in the introns (see Figure 4.1). This evidence, in conjunction with the sequence data and

expression experiments discussed later, shows that pXB-C1 does contain part of a functional *X. borealis* cardiac actin gene. This gene shares homologies in intervening sequences with a transcribed *X. laevis* cardiac actin gene, even though the species are thought to have diverged about 10 million years ago (Bisbee *et al.*, 1977). However, the correct regulation of expression from a cardiac actin gene construct, which does not contain the *Hind*III fragment (see Section 7.2.3), suggests that at least this sequence is not necessary for proper control, despite its conservation between different *Xenopus* species.

The *Hind*III fragment was selected as a suitable probe for the screening of genomic libraries for clones with longer inserts, which included the pXB-C1 genomic sequences.

4.2.2 Screening of a Partial *Sau*3A Genomic Library Cloned in λ47.1

An amplified partially *Sau*3A-digested *X. borealis* genomic library in λ47.1 (a kind gift from Dr. Phil Turner) was screened with the *Hind*III fragment. In 2.5×10^6 clones analysed, a single cross-reacting phage (λXB-C2) was recovered twice. The original library probably only contained $2\text{-}3 \times 10^5$ recombinants (with average insert size of about 12 kb), and so this clone had been relatively poorly replicated in the amplification process. Indeed, the plaque size observed was very small, and the titre of purified phage stocks from individual plaques was particularly low (see Sections 3.2.2 and 4.2.3 for further discussion).

A restriction map of the 11 kb insert in λXB-C2 is shown in Figure 4.1.

It clearly was derived from the same gene as λXB-C1, but most of the additional sequence was downstream of the original clone. The extra 400 bases, which were upstream of λXB-C1, were sequenced, but the predicted sequence of the first exon, derived from primer extension, was not present. To confirm this observation, a 15 base oligonucleotide, complementary to the RNA encoded by the last 15 bases of the first exon (see Figure 5.3), was used to screen the filter from a Southern blot of restricted λXB-C2 DNA. No signal was detected (data not shown).

As the number of clones in the original λ47.1 library only represented about one haploid genome length of DNA, it was not surprising that only one copy of the required single-copy gene was present. Some of the filters were rescreened with the complete insert of λXB-C1, which contains coding regions that should share a relatively high degree of homology with all actin genes. Although this probe hybridised most strongly to λXB-C2, it also reacted with other clones. Some of these have been characterised and shown to contain actin genes. In λ3H1 (Cross *et al.*, 1986; Appendix I), a virtually complete type 1 (β-) cytoskeletal actin gene has been found. The clone λ3B contains an entire skeletal actin gene (M. Boardman, per. comm.).

4.2.3 Screening of other Genomic Libraries Cloned in EMBL3

A detailed account of the construction of further partial *Sau*3A libraries is given in Chapter 3, where it is proposed that an unusual distribution of *Sau*3A sites in the region of the cardiac actin gene led to difficulty in isolating a suitable clone. Finally, a library (library 3), in which the longest products (of about 10-20 kb in length)

from a *Sau*3A genomic digest had been cloned, was shown to contain three recombinants, which hybridised to the *Hind*III fragment probe.

Two of these clones (λXB-Cr1 and λXB-Cr2) appear to be derived from the second putative cardiac actin locus, and are discussed elsewhere (Appendix II). However, analysis of the other clone (λXB-C3) demonstrated that it shared regions in common with λXB-C1 and λXB-C2 (Figure 4.1), and furthermore that it also shared additinal restriction fragments in common with the genomic locus (see Table 4.1). It spanned a few kilobases upstream of the first coding exon. The phage produced only very small plaques, just like λXB-C2, and it seems likely that a specific sequence present in the insert of both of these clones must be responsible for this phenomenon (also see Section 3.2.2).

The filter from a Southern blot of electrophoresed digests of λXB-C3 was hybridised to the end-labelled 15-base oligonucleotide complementary to the 3' portion of the first exon, under conditions known to produce hybrids with RNA (see Section 6.2.2). The resulting autoradiograph (Figure 4.3) showed that the first exon sequence was present in λXB-C3. The insert was subcloned into the plasmid vector pEMBL8, using the EMBL3 *Sal*I polylinker sites, which flanked it, and it was subsequently mapped for restriction sites (pXB-C3.1; Figure 4.4). This revealed that there was approximately 2.8 kb upstream of the putative start of transcription.

The two small *Sst*I fragments at the 5' end of the gene, one of which contains the homologous sequence to the first exon probe (Figure 4.3), were subcloned in the plasmid vector pEMBL18. They have since been largely sequenced (G. S. Cross, per. comm.). By sequencing of primer

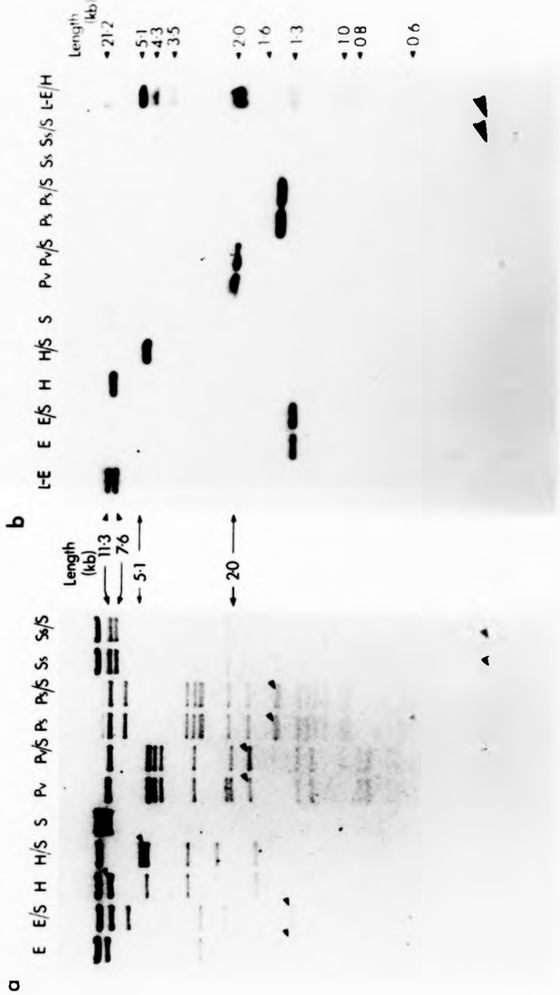


Figure 4.3 Demonstration that λXB-C3 Contains the First Exon of the Cardiac Actin Gene

Aliquots containing one microgram of λXB-C3 DNA were digested with a number of enzymes, both with and without added *Sal*I, which excises the insert of the clone. The resulting samples were electrophoresed on a 1% agarose gel, the gel was photographed (a) and the DNA was transferred to nitrocellulose. The Southern blot was probed with an end-labelled oligonucleotide that is complementary to the last 15 bases of RNA encoded by the first exon of the cardiac actin gene (b).

(a) A Polaroid negative from a photograph of an ethidium-stained gel, which was illuminated under ultraviolet light. The hybridising bands seen in (b) are marked with arrowheads.

(b) Southern blot of the same gel, including two tracks containing lambda DNA markers. Their sizes are indicated in the figure. The hybridisation and washing conditions for the filters (37°C in 3 x SSC for 2 days, and room temperature in 3 x SSC for 2 x 15 min, respectively) were already known to be suitable for stable hybridisation of the oligonucleotide to RNA (see Section 6.2.2). The short *Sst*I-digested hybridising band is too faint to see at this exposure, but is marked with a large arrowhead.

Tracks: E = *Eco*RI-digested pXB-C3.1 DNA; E/S = *Eco*RI/*Sal*I; H = *Hind*III; H/S = *Hind*III/*Sal*I; S = *Sal*I; Pv = *Pvu*II; Pv/S = *Pvu*II/*Sal*I; Ps = *Pst*I; Ps/S = *Pst*I/*Sal*I; Ss = *Sst*I; Ss/S = *Sst*I/*Sal*I; L-E = λ47.1 digested with *Eco*RI; L-E/H = wild-type λ digested with *Eco*RI and *Hind*III.

Figure 4.4 Restriction Map of pXB-C3.1

The 12 kb insert of λXB-C3 was subcloned into pEMBL8 at the *Sal*I site (pXB-C3.1), using the *Sal*I sites in the polylinkers of the EMBL3 arms. The clone contains the complete transcription unit of the *X. borealis* cardiac actin gene, with 2.8 kb upstream of the start of transcription and 2.2 kb downstream of the 3' end. Empty bars represent 5' and 3' untranslated regions, while solid bars represent coding regions. Not all exons have been sequenced, but the final exon, found in a short *Kpn*I fragment, is designated the seventh by comparison with the homologous actin genes of other vertebrates. The arrow shows the start and direction of transcription at a site internal to the smallest *Sst*I fragment. The *Bam*HI site was created from a *Sau*3A site in the genomic insert on ligation to the *Bam*HI-digested right arm of EMBL3.

Restriction enzymes: B = *Bam*HI; E = *Eco*RI; H = *Hind*III; K = *Kpn*I; S = *Sal*I; Sp = *Sph*I; Ss = *Sst*I.

extension products made from mRNA, it had been possible to deduce about the last 35 bases at the 3' end of the first exon (see Section 3.2.2). When compared to the proposed first exon region in the genomic clone, it is absolutely identical. In addition, the neighbouring promoter sequences contain a 'TATA' box and a 'CAAT' box in the appropriate positions (see Section 1.4.2) and share extensive homologies with the *X. laevis* cardiac actin gene promoter, and more limited homologies with other vertebrate cardiac actin gene promoters (see Chapter 8). This information is consistent with the proposal that the gene represented by pXB-C3.1 is transcribed *in vivo* (for further discussion, see Chapter 5), and this transcription is initiated from a site in the smallest *Sst*I fragment of the clone (see Figure 4.4).

An *Eco*RI fragment, containing the coding sequence of the penultimate and final exons from the type 1 cytoskeletal actin clone, λ3H1 (see Appendix I), was labelled by nick-translation and hybridised to filters from Southern blots of pXB-C3.1. Due to the high level of sequence conservation in the coding regions of all vertebrate actin genes, cross-hybridising bands were identified (data not shown). In particular, a *Kpn*I fragment of approximately 700 b reacted with the probe. The ends of this fragment were sequenced, using the method of Maxam and Gilbert (1978). Gunning *et al.* (1984a) have published the sequence of the 3' untranslated region of the *X. laevis* cardiac actin mRNA, which they derived from a cDNA clone. Sequence comparison demonstrates that the 3' untranslated region of the *X. borealis* gene is highly conserved and is internal to the *Kpn*I fragment of pXB-C3.1, as are the rest of the final exon and part of the penultimate exon (G. S. Cross, per. comm.; see Figure 4.4). It is therefore concluded that pXB-C3.1 contains the entire transcription unit of the cardiac actin gene.

SECTION 4.3 DISCUSSION

Since the original proposal that λXB-C1 was derived from a cardiac actin gene, further sequence data has accumulated to support this hypothesis. Vandekerckhove and Weber (1984) found an actin protein in *X. laevis* with an identical N-terminus to mammalian cardiac actin. It was only expressed in striated muscle. Furthermore, Mohun *et al.* (1984) isolated a *X. laevis* actin cDNA, encoding the same N-terminal amino acids as λXB-C1. In the adult, the RNA from which it was synthesised, was only found in heart tissue. Sequence comparison of 3' untranslated regions suggests that this message (Gunning *et al.*, 1984a) is encoded by the *X. laevis* cardiac actin gene cloned by Stutz and Spohr (1986), if it is assumed that the two mismatches in 138 bp are due to sequencing error or polymorphism.

More extensive clones, derived from the same *X. borealis* gene as λXB-C1, have been shown to share some similarities in their restriction maps to the cloned *X. laevis* cardiac actin gene. As well as the known sequence of the encoded proteins being identical in both genes, the 5' and 3' untranslated regions, and the promoter sequences are all highly conserved (G. S. Cross, per. comm.). The 3' untranslated regions of warm-blooded vertebrate cardiac actin genes also share some homologies with those of the *X. laevis* and *X. borealis* genes (see Gunning *et al.*, 1984a). The sequences conserved in the actin promoters of all vertebrate species are discussed briefly in Chapter 8.

The highly specific intron probe used in these studies hybridised to a second locus, which has now been cloned and shown to contain an actin

gene (see Appendix II). Stutz and Spohr (1986), using different specific probes for their genomic Southern blots, have recently suggested that there are pairs of alleles with homology to a cardiac actin gene, a skeletal actin gene and a processed skeletal actin gene in *X. laevis*. The results in this chapter and Appendix II show that the same applies for the *X. borealis* cardiac actin gene and confirm that, in this case, both alleles encode actins. The whole genome duplication, which has been proposed to have occurred in an ancestor of *X. laevis* and *X. borealis* about 30 million years ago (Bisbee *et al.*, 1977), is likely to be responsible for these paired loci (see Figure 6.4b). It is currently unknown if the gene encoded by λXB-Cr1 is expressed.

The following chapter presents evidence from sequencing of primer extension products that the gene encoded by λXB-C3 is transcribed, and that, in the adult, these transcripts are localised to the heart. All the available data therefore suggest that the clone contains a cardiac actin gene, and the mapping described in the preceding section confirms that the complete transcription unit is present, with 2.8 kb upstream of the cap site and approximately 2.2 kb downstream of the poly(A) addition site.

CHAPTER 5

EXPRESSION OF THE CHROMOSOMAL *XENOPUS BOREALIS* CARDIAC ACTIN GENE DURING NORMAL DEVELOPMENT

SECTION 5.1 INTRODUCTION

The sequence data and restriction mapping presented in the last chapter indicates that λXB-C3.1 encodes a cardiac actin gene. The experiments described in the following sections are designed to answer two important questions about expression at this gene locus. Firstly, is the gene transcribed at all? Secondly, if it is expressed, are transcripts localised to the heart of the adult, as has been demonstrated for the homologous gene in other vertebrates? Because of the possible presence of an additional closely related cardiac actin allele in the *X. borealis* genome, the detection of a specific product by transcriptional assays like S1 analysis or primer extension, using probes or primers from λXB-C3.1, cannot exclude the possibility that it is the other allele which is being transcribed. Therefore, the early sections of this chapter describe the sequencing of part of the mRNA detected by primer extension, in order to provide additional evidence that the message is derived from the gene in pXB-C3.1.

The remainder of the chapter deals with the regulation of expression of this gene in embryos, tadpoles and adult frogs. I also discuss some of the recent experiments that other workers have performed on the closely related species *X. laevis*, using the cardiac actin gene as a marker for early muscle cell differentiation.

SECTION 5.2 RESULTS

5.2.1 Primer Extension Analysis of Transcripts from the Cardiac Actin Gene

The experiments in this section were all undertaken by Dr. Gareth Cross and provide important background information on the expression of the cardiac actin gene *in vivo*. A primer was prepared from the first coding exon of the actin gene, as described in Section 2.5.2. To do this, the fragments from a *Bst*NI digest of the *Hinf*I fragment of pXB-C1, which contained the first coding exon of the cardiac actin gene, were labelled at their 5' ends with T4 polynucleotide kinase. They were then digested with *Fok*I and the resulting mixture was fractionated on an 8% polyacrylamide-urea gel. The required single-stranded primer, bounded by a *Fok*I and a *Bst*NI site (Figure 5.1) was 36 bases long and hybridised to the 3' part of the RNA encoded by the first coding exon.

The primer was hybridised in aqueous solution to *X. borealis* total RNA from stage 40 swimming tadpoles at various temperatures. It was then extended with reverse transcriptase before electrophoresis on an 8% polyacrylamide-urea gel (Figure 5.2). A major product of 192 bases was detected, but other shorter bands were also visible. Within the resolution of the gel, a major band of identical size was observed with *X. laevis* tadpole RNA, although sequence data reveals that it is, in fact, one nucleotide shorter. The highest hybridisation temperature at which strong signals were still produced (70^{o}C) was selected for all further analyses.

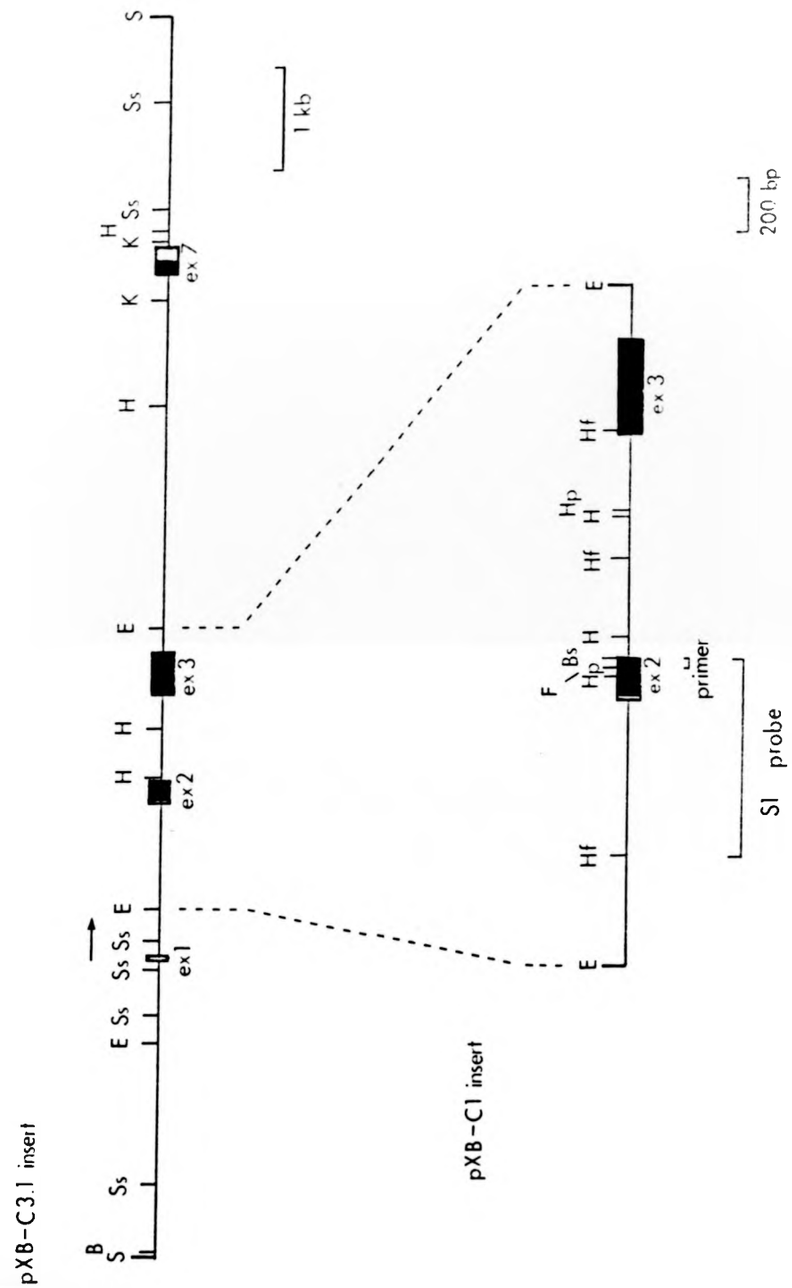


Figure 5.1 Probes and Primers for the Detection of Transcipts from the Cardiac Actin Gene

The structures of the clones pXB-C1 and pXB-C3.1 are discussed in Chapter 4. Empty bars represent 5' and 3' untranslated regions, and solid bars represent coding regions. The arrow indicates the start and direction of transcription.

The *Bst*NI/*Hinf*I fragment used for S1 analysis (Section 5.2.3) is shown. It overlaps the upstream end of the first coding exon. The 36 b *Fok*I/*Bst*NI primer is derived from this fragment. When the primer was hybridised to cardiac actin mRNA and extended with reverse transcriptase, the major product was 192 b in length (see Figure 5.2). Under normal digestion conditions, 147 b of the 600 b S1 probe were protected by the same message (see Figure 5.4).

Not all the *Bst*NI or *Fok*I sites are indicated on the map of pXB-C1. Restriction sites: B = *Bam*HI; Bs = *Bst*NI; E = *Eco*RI; F = *Fok*I; H = *Hind*III; Hf = *Hinf*I; Hp = *Hpa*II; K = *Kpn*I; S = *Sal*I; Ss = *Sst*I.

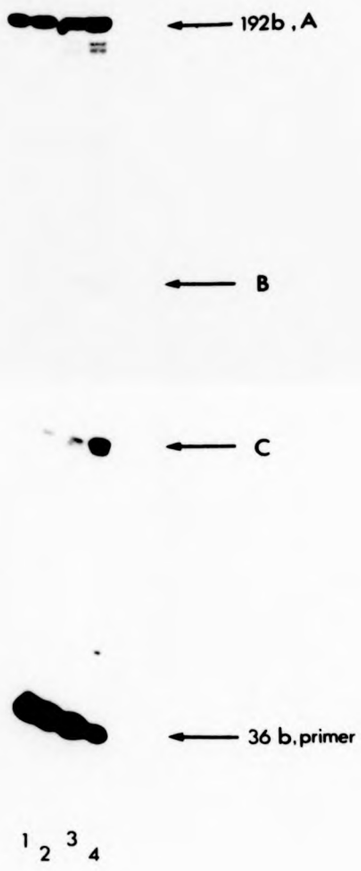


Figure 5.2 Characterisation of the Primer Extension Reaction with a Second Exon Primer Specific for the *Xenopus* Cardiac Actin Gene

Aliquots of the 36 b *Fok*I/*Bst*NI end-labelled primer were hybridised to 12 μg of total RNA from *X. borealis* stage 40 swimming tadpoles at 55°C (track 1), 60°C (track 2), 65°C (track 3) and 70°C (track 4) for 3 hours. After primer extension, the products were separated on an 8% polyacrylamide-urea gel. As well as the major 192 b full-length product (A), many shorter bands were observed. On the basis of this experiment, subsequent hybridisations were all performed at 70°C.

In a large-scale primer extension reaction, two of the minor fragments (B and C), together with band A, were gel isolated and it was attempted to sequence them by the method of Maxam and Gilbert (1978). Only A and C contained sufficient labelled material to produce readable sequence. They were synthesised from identical RNA templates, except product A extended further upstream. Therefore, the shorter fragment appears to result from extension off degraded RNA or from premature termination of DNA synthesis (see Sections 5.2.1 and 5.2.2 for further details).

The figure is shown with the kind permission of Dr. Gareth Cross.

A large-scale primer extension reaction was set up using *X. borealis* tadpole RNA, and, after electrophoresis, the major product (A) and two of the smaller fragments (B and C) were excised from the gel (see Figure 5.2). These end-labelled DNAs were sequenced by the method of Maxam and Gilbert (1978). One of the shorter bands (B) contained insufficient labelled material to obtain readable sequence. However, the other two products gave totally unambiguous sequences, which were derived from the same RNA template, except A extended further upstream than C. The shorter product contained only actin coding sequence, and, therefore, appears to be the result of extension from degraded RNA or of premature termination of DNA synthesis by the enzyme. At a given hybridisation temperature, all the other minor truncated products in the assay were found at the same relative ratios to each other in different samples, even using RNAs extracted at different stages of development. It would therefore seem quite likely that the RNAs, which they are derived from, are all synthesised by the same gene and probably are formed by the same mechanism that produces the shortened band C.

5.2.2 The Sequence of the Primer-Extended Product from *X. borealis* Cardiac Actin Transcripts: Synthesis of Oligonucleotides Complementary to Parts of the RNA Encoded by the First Exon

The complementary sequence to the full-length primer-extended product (i.e. the sequence of the mRNA) is given in Figure 5.3, together with the corresponding transcribed sequence in pXB-C1 (Cross, 1984). The sequences are identical throughout the coding region and 19 bases of 5′ untranslated region before they diverge completely. This divergence is

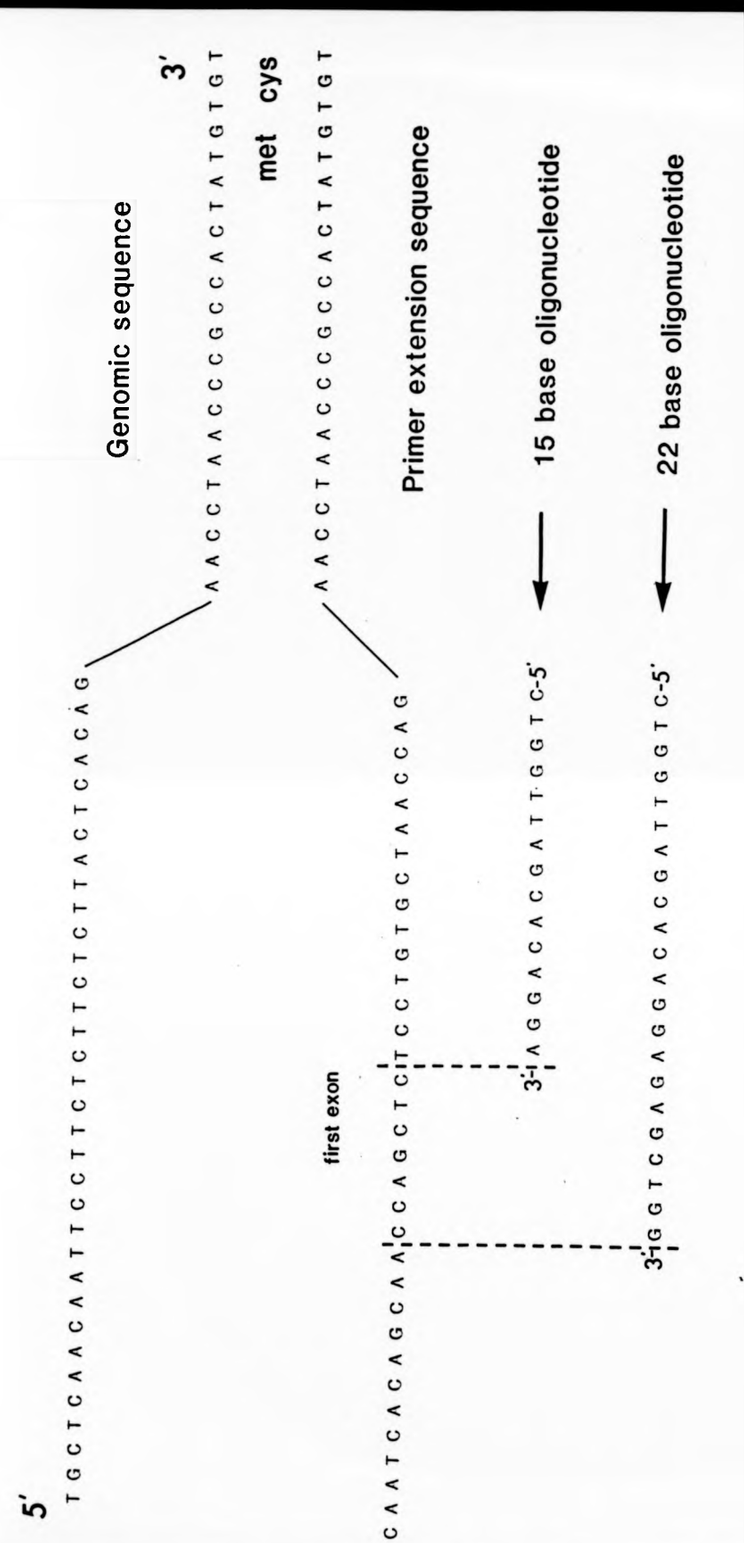


Figure 5.3 Comparison of the Genomic Sequence from pXB-C1 with the Sequence Derived from Primer Extension Products

At the top, the figure shows the transcribed sequence at the 5' end of the first coding exon of the cardiac actin gene, derived from pXB-C1. Below it, is the cDNA sequence of the mRNA to which the *Fok*I/*Bst*NI primer hybridises (Figure 5.2). It has been deduced by sequencing the major primer extension product using the chemical cleavage method (see Section 5.2.1). The sequences are identical throughout the coding region and for 19 bases of 5' untranslated region, suggesting that the message is transcribed from the gene encoded by pXB-C1. However, like all other characterised vertebrate actin genes (e.g. Zakut *et al.*, 1982; Fornwald *et al.*, 1982), the 5' untranslated region is interrupted by an intron. If the GT--AG splice site consensus is adhered to (Breathnach and Chambon, 1981), splicing leaves only 16 bases of 5' untranslated region from the second exon. The first exon, containing the remainder of sequence derived from primer extension (including the 5'-CAG-3' at the 3' end) and a further 15 bases at the 5' end of the message that could not be easily resolved from the primer extension sequence gel, is not found in pXB-C1, but is present in pXB-C3.1 (see Chapter 4).

The sequences of two oligonucleotides, which are complementary to the 3' portion of the RNA encoded by the first exon, are also presented.

consistent with the proposed existence of an intron in the 5' untranslated sequence. Splicing of the primary transcript is likely to leave only 16 bases of 5' untranslated region from the first coding exon, if the GT--AG splice site consensus is followed (Breathnach and Chambon, 1981), implying that the additional three conserved nucleotides (5'-CAG-3') are also present in the upstream exon.

Two oligonucleotides were synthesised, which are complementary to the predicted 3' end of the RNA transcribed from the first exon (Figure 5.3). The first (a kind gift from G. S. Cross) was 15 bases in length, the second (a kind gift from M. A. McCrae) was 22 bases in length. The former was end-labelled, and used as a probe in a Southern blot analysis to show that λXB-C3 contained the first exon of the cardiac actin gene (see Section 4.2.3). This genomic region has since been sequenced (G. S. Cross, per. comm.) and shown to encode an identical RNA sequence to that predicted from the sequencing of primer extension products, confirming that the first coding exon contains only 16 bases of 5' untranslated region. These results demonstrate that a cardiac actin transcript is expressed in *Xenopus* tadpoles, which has an identical sequence (at least at the 5' end) to the message encoded by the gene in λXB-C3. Furthermore, like all other sequenced vertebrate actin genes (e.g. Zakut *et al.*, 1982; Fornwald *et al.*, 1982), the *X. borealis* cardiac actin gene contains an intron in its 5' untranslated region (see also Section 4.2.3). The elucidation of the exact position of the cap site by primer extension from the 15 base oligonucleotide is discussed in Section 6.2.2.

In summary, taking into account other relevant data (see Section 5.3), it can be concluded from this experiment that λXB-C3 encodes a gene that

is transcribed *in vivo*. The localisation of the high levels of transcripts found in embryos is discussed in Section 5.2.4.

5.2.3 Expression in Adult Frogs: Transcripts from the Cardiac Actin Gene are Found Solely in the Heart

An S1 assay was devised by Dr. Gareth Cross to analyse transcription from the cardiac actin gene. The probe is one of the products formed in a *Bst*NI restriction digest of the *Hinf* fragment of pXB-C1, which contains the first coding exon. The whole mixture of products was kinase-labelled and used directly for S1 analysis (see Section 2.5.1). The relevant fragment spans the 5' intron-exon boundary of the first coding exon (see Figure 5.1). It was hybridised to *X. borealis* total RNA from stage 40 swimming tadpoles at various temperatures, and then digested with S1 nuclease under mild conditions. After electrophoresis on an 8% polyacrylamide-urea gel (Figure 5.4), a series of bands was observed. The strongest of these was measured as 147 bases in length, using labelled fragments from a pBR322 *Hinf*I digest as size markers. Of course, the widely spaced fragments in the marker track of this gel will not necessarily allow an estimate that is accurate to the base. However, the predicted length of the protected probe from sequence data is also 147 bases, and the patterns of bands observed under more stringent digestion conditions (see Figures 6.1 and 6.2) are also consistent with the major fragment from mild digestion being this length. The optimum hybridisation temperature in formamide was found to be 55^{o}C, and all samples were hybridised at this temperature in subsequent analyses.

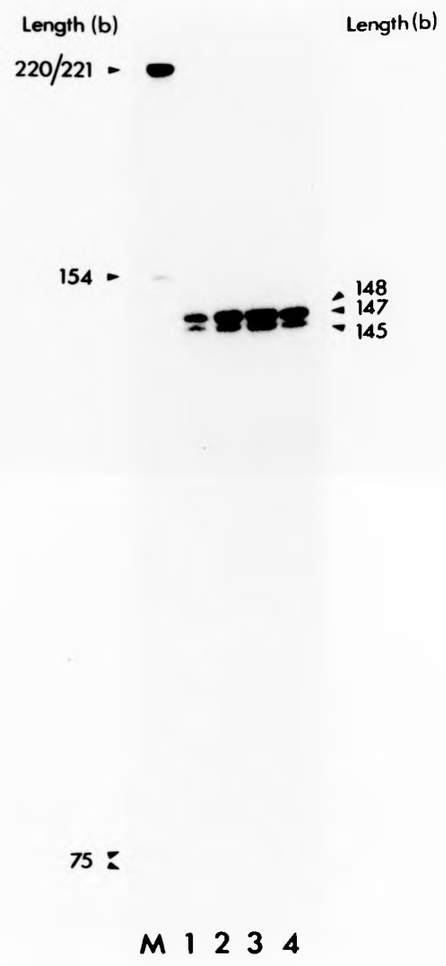


Figure 5.4 Characterisation of an S1 Nuclease Assay with a Second Exon Probe Specific for the *Xenopus* Cardiac Actin Gene

Aliquots of the *Hinf*I/*Bst*NI end-labelled S1 probe were mixed with 8 μg of *X. borealis* stage 40 tadpole total RNA and hybridised at 45°C (track 1), 50°C (track 2), 55°C (track 3) and 60°C (track 4) in 80% formamide for 3 hours. After S1 digestion with 100 units of S1 at 30°C for 30 min, the products were separated on an 8% polyacrylamide-urea gel. In the photograph, the full-length probe, produced from DNA-DNA hybrids (the probe is not strand-separated), has migrated too slowly to be seen. The sizes of fragments protected by mRNA were estimated using a marker track (M), containing labelled fragments from a *Hinf* I digestion of pBR322. The two strands of the 75 bp fragment are denatured and have migrated at slightly different speeds.

The major band in the S1 analysis is approximately 147 bases in length; subsequent experiments (see Section 6.2.1) and predictions from sequence data (Cross, 1984) are consistent with this figure. Two other fainter bands of 148 b and 145 b are also observed, presumably resulting from the slightly imprecise nature of S1 digestion.

The optimum hybridisation temperature is 55°C, and this temperature was used for all subsequent experiments.

S1 nuclease does not always cut precisely at the site where complementary sequences diverge, and this leads to the presence of two other fainter bands (148 and 145 bases in length). The S1 probe appears to be entirely specific for cardiac actin gene transcripts. Shorter products, which might result from reaction with RNAs that diverge from the cardiac actin mRNA, especially in their 5' untranslated regions (e.g. skeletal actin mRNA), are not observed (see Section 5.3).

The probe also cross-hybridises with RNA from *X. laevis* tadpoles, which is presumably produced from the homologous gene in this species (see Section 4.2.1). S1 digestion of the RNA-DNA hybrids under mild conditions gives protected fragments of identical length to the *X. borealis* cardiac actin mRNA, although frequently additional shorter fragments are also observed. The fact that the extra products are only formed in samples containing cardiac actin mRNA, coupled with evidence from further studies using more stringent S1 conditions (see Section 6.2.1), suggests that even these shorter fragments are protected by the same cardiac actin message. In fact, they result from minor sequence differences between the 5' untranslated regions of the transcripts from the two species (see Section 6.2.1).

In Figure 5.5, S1 analyses of *X. borealis* tadpole RNA and total RNA from different tissues of an adult *X. laevis* frog are presented. A *X. laevis* frog was used in this study, as these animals are sacrificed relatively frequently in the laboratory, particularly for the artificial fertilisation procedure. The only cardiac actin transcripts in adult organs were found in the heart. The skeletal muscle preparation has since been shown to contain skeletal actin transcripts, but the latter are not observed in the heart RNA sample (G. S. Cross, per. comm.).

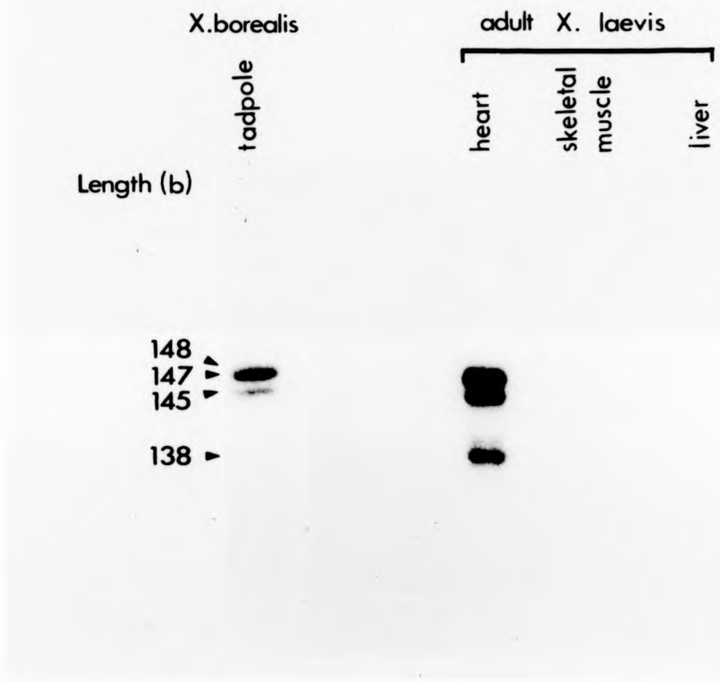


Figure 5.5 Expression of the Cardiac Actin Gene in Adult *X. laevis* Frogs

Aliquots of the *Hinf*I/*Bst*NI end-labelled S1 probe were hybridised to 8 μg of total RNA from *X. borealis* stage 40 tadpoles and 8 μg of total RNA from adult *X. laevis* heart, skeletal muscle (from the leg) and liver. The S1 nuclease digestion was performed at 22^{o}C with 300 units of enzyme in an attempt to maximise the resulting signal.

The same protected fragments are observed from *X. laevis* RNA as *X. borealis* RNA, although additional shorter bands are formed because of the minor sequence differences between the two messages (see Section 6.2.1). Even after longer exposures, no signal was detected in adult tissues other than the heart. Subsequent analysis has shown that the skeletal muscle preparation contains skeletal actin mRNA (G. S. Cross, per. comm.).

Thus, I have shown that expression of the *X. laevis* homologue of the cardiac actin gene is localised to the heart in the adult, and presumably this applies for the *X. borealis* gene as well. As discussed in Chapter 4, the *X. laevis* homologue almost certainly synthesises the RNA cloned as a cDNA by Mohun *et al.* (1984). The latter was shown to be localised specifically in the heart of the adult, in agreement with my results.

5.2.4 Expression in Embryos: The Cardiac Actin Gene is Activated Early in Development and its Transcripts are Localised in a Tissue-Specific Fashion

It has already been demonstrated that the cardiac actin gene is expressed strongly in stage 40 swimming tadpoles (see Sections 5.2.1 and 5.2.3). Numerous experiments have shown that transcripts are also present in neurulae and tailbud tadpoles (e.g. see Section 7.2.6), but, as expected, no signal has been detected in oocytes.

The relevant details of early development in *Xenopus* have been discussed in the first chapter (Section 1.6). Figure 5.6 shows an S1 analysis of RNA samples from single *X. laevis* embryos at different developmental stages. To ensure correct staging, carefully selected embryos were examined individually in order to avoid contamination with any rare abnormally developing embryos in a batch. Transcripts were first detected in the late gastrula (stage 12½), slightly earlier than has previously been suggested by Mohun *et al.* (1984) and 8 hours before the reported appearance of myofibrillae at stage 21 (Nieuwkoop and Faber, 1956). However, the inability to detect transcripts before the late

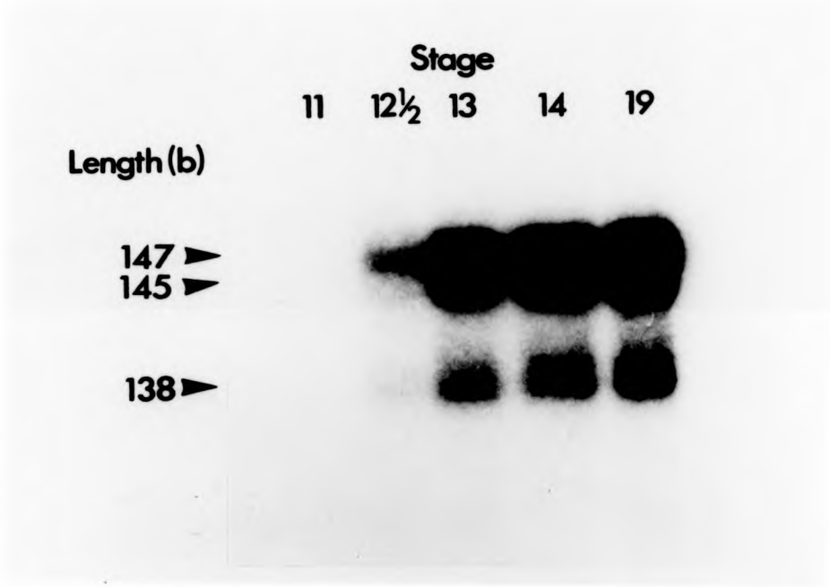


Figure 5.6 Determination of the Time of Cardiac Actin Gene Activation in *X. laevis* Embryos by S1 Analysis

Total RNA samples were prepared from single *X. laevis* embryos (approximately 5 μg of RNA per embryo), which had been carefully staged. They were then examined by S1 analysis with the *HinfI/Bst*NI fragment. S1 nuclease digestion was performed at 37°C with 120 units of enzyme. Like the example in Figure 5.5, additional shorter products are observed, as well as the major protected fragments of between 145 b and 148 b in length. The samples were taken at stage 11 (mid-gastrula), stages 12½ and 13 (late gastrula), stage 14 (early neurula), and stage 19 (mid-neurula). Compared to other experiments, the signal from stage 19 total RNA is unusually low.

gastrula could merely result from insufficient sensitivity of the assay. To test this hypothesis, poly(A)$^+$ RNA from 80 pooled *X. laevis* embryos at the late blastula stage (a kind gift from Prof. Hugh Woodland) was assayed by S1 analysis. No signal was observed (data not shown), and so it must be concluded that the cardiac actin gene is not even expressed weakly before gastrulation. A number of other prominent tadpole mRNAs are also first detected at the late gastrula and early neurula stages (Dworkin *et al.*, 1984), perhaps reflecting the requirement for new gene activities involved in the specialised cell differentiation that starts to take place at this time.

The localisation of transcripts has also been examined in neurulae and tadpoles. Typical data are shown in Figure 7.6. In both *X. laevis* and *X. borealis*, expression is almost entirely localised to the myotome region of the neurula, and no detectable transcription is observed in other mesodermal derivatives, including the notochord, which lies directly adjacent to the myotomes. In the tailbud tadpole, transcripts are localised to a dissected region containing the somites. These results are in close agreement with previous findings for the *X. laevis* cardiac actin gene (Mohun *et al.*, 1984), and strongly suggest that the gene is expressed only in regions in which skeletal muscle cells will be formed during early development. At later tadpole stages (from stage 40 onwards), transcripts have also been detected in the newly formed heart of *X. laevis* (Mohun *et al.*, 1984; G. S. Cross, per. comm.).

Therefore, the cardiac actin gene is activated at a very early stage in *Xenopus* development (stage 12½), only 14 hours after the egg has been fertilised. In contrast to the adult frog, it is expressed in skeletal muscle as well as cardiac muscle during the embryonic and tadpole stages.

SECTION 5.3 DISCUSSION

In Chapter 4, sequence and restriction mapping data led to the proposal that the insert of pXB-C3.1 encodes a cardiac actin gene. The results described in this chapter provide clear evidence that the *X. borealis* gene from which pXB-C3.1 is derived is expressed. In the adult frog its transcripts are found only in the heart. Of course, it is conceivable, though extremely unlikely, that the two putative *X. borealis* cardiac actin alleles (see Section 4.2.1) potentially encode RNAs with identical 5' untranslated regions, but that the gene represented by pXB-C3.1 is not expressed. However, two lines of evidence argue against this hypothesis. First, the presence of a 'TATA' box, a 'CAAT' box and other actin gene-specific conserved sequences upstream of the cap site in pXB-C3.1 (see Section 8.2) suggests that the gene has a functional promoter. Second, in Chapter 7, it is shown that, when pXB-C3.1 is injected into embryos, expression from the clone is regulated in a very similar way to the functional chromosomal gene.

Both the primer extension and S1 assays described in this chapter appear to be totally specific for cardiac actin mRNA. They do not produce a detectable signal with the high levels of skeletal actin mRNA that are present in embryos and adult skeletal muscle (Mohun *et al.*, 1984; Cross, per. comm.), or with cytoskeletal actin mRNAs. The S1 probe shares 84% homology in the coding sequence with the skeletal actin mRNA of *X. borealis* (M. Boardman, per. comm.). The first coding exon of the *X. borealis* skeletal actin gene has a virtually identical 5' untranslated region (one mismatch in 19 bases) and has the same 5' splice site as that of the *X. laevis* gene (Stutz and Spohr, 1986; G. S. Cross, per.

comm.). There are 14 and 13 mismatches respectively between the encoded transcripts and the S1 probe in the first 19 bases of 5' untranslated region. Therefore, any hybrids with the skeletal actin transcripts of either species would be expected to protect a shorter product on S1 digestion than those hybrids containing the cardiac actin message. The primer also shares 83% homology with the corresponding region in the skeletal actin gene, but again this is apparently insufficient to allow significant cross-reaction.

At first sight, this high degree of specificity is rather surprising. We have also reported a similar phenomenon for a primer made from a cytoskeletal gene (Cross *et al.*, 1986). It hybridises to other cytoskeletal actin mRNAs in *X. borealis* and *X. laevis*, all of which presumably share very high levels of sequence homology. But there is no evidence that the primer reacts with muscle-specific actin transcripts. One explanation is that the optimal temperatures used for hybridisation of both S1 probes and primers, which are determined experimentally, are significantly higher than the T_m for totally complementary strands. For S1 analysis, the higher temperature reduces the level of DNA-DNA probe reannealing, while, for primer extension, it may be necessary to melt secondary structures in the mRNA or primer (see Cross *et al.*, 1986).

Many of the expression experiments have been performed on *X. laevis*, which expresses a cardiac actin gene that is highly homologous to the one in *X. borealis*. The frogs and embryos of the former species are more readily available in the laboratory. Furthermore, most of the work on the expression of the *X. borealis* clone has been performed in *X. laevis*, and so it has been important to analyse precisely the regulation of the endogenous gene in this species.

The strict localisation of cardiac actin mRNA in the heart of the adult has also been observed by Mohun *et al.* (1984), who, like us, could not detect any transcripts in leg muscle. However, Vandekerckhove and Weber (1984), using their N-terminal tryptic peptide characterisation procedure, have demonstrated the presence of the cardiac actin isoform in adult skeletal muscle of the leg, as well as in the heart. In the former tissue, the ratio of cardiac:skeletal isoforms was 1:4, while in the latter it was 3:1. If the equivalent messages were present at a similar ratio, cardiac actin mRNA would be easily detectable in skeletal muscles by the S1 assay, so poor sensitivity of transcript analysis cannot explain the anomaly. Furthermore, it seems unlikely that a greater stability of the cardiac actin protein compared to the RNA can alone account for these results, as such a differential stability is not observed in other organisms, like the chicken and mouse, where the two sarcomeric actin messages also become more strictly localised to either the heart or skeletal muscle during development (Vandekerckhove *et al.*, 1986). However, it is plausible that younger frogs were used in the protein study, and that these might still have been expressing the cardiac actin gene in skeletal muscle. Alternatively, another gene, which encodes a protein with the same N-terminus as the cardiac actin isoform, but which is sufficiently diverged in nucleotide sequence for the RNA to be undetectable with the probes and primers used in this study, could be transcribed in skeletal muscle cells.

Like Mohun *et al.* (1984), I have found that the cardiac actin gene is strongly expressed in embryonic myotomes. As discussed in Section 1.5.2, expression in fetal skeletal muscle is also a common feature of the homologous gene in warm-blooded vertebrates (e.g. Minty *et al.*, 1982; Paterson and Eldridge, 1984). Commonly, as in *Xenopus*, the

transcripts are subsequently lost from this tissue during development into the adult. The embryonic coexpression of the two sarcomeric actin genes, followed by their differential expression in the adult, makes them particularly suitable for the study of shared and unique control mechanisms responsible for developmental changes in gene regulation.

The other interesting feature of cardiac actin gene expression is that it is activated at stage 12½ (the late gastrula), long before the first myofibrillae are seen in the somites (at stage 21, the mid-neurula). The transcripts are therefore a very early marker for commitment to muscle development. Consequently, molecules that regulate this gene may well be involved in the earliest stages of muscle differentiation, and so a study of the expression of cardiac actin mRNA by transformation of embryos with the cloned gene may be of particular developmental interest. Furthermore, Gurdon and his colleagues have exploited the early expression to reexamine a number of embryological experiments in *Xenopus*, in an attempt to investigate if cell interactions are required for the formation of mesodermal structures *in vivo*. These studies are briefly considered in the final part of the chapter.

5.3.1 The Formation of Muscle in Early *Xenopus* Development

Basically two mechanisms can be envisaged by which cells become committed to form mesodermal tissue, and subsequently muscle and other mesodermal cell types, in the embryo. In one, determinants, which are localised in the cytoplasm of the fertilised egg before the first cell division, might become located in a specific region of cells as the embryo divides. This region would then be committed to a mesodermal

fate. In the other mechanism, mesodermal cells might be derived from presumptive non-mesodermal cells by interactions with another cell type. These two models are clearly not mutually exclusive and are discussed from a more general point of view in Section 1.2.

Evidence that the latter process can take place in *X. laevis* was presented by Ôgi (1969) and Nieuwkoop (1969), who grafted the animal caps of blastulae on to the vegetal regions of blastulae and allowed the conjugates to develop. By histological criteria, neither region on its own forms muscle or indeed any other mesodermal structure in culture, but, in the conjugate, many animal pole cells were induced by the endodermal cells to become muscle. These mesodermal induction experiments have been repeated by Gurdon *et al.* (1985b), using cardiac actin gene expression as a marker for muscle tissue. Their results were in close agreement with the earlier studies and the report characterised the minimum contact time required and the minimal time needed for gene activation to occur after induction.

Gurdon *et al.* (1985a) have also performed a series of experiments in which they separated regions of a fertilised egg with a hair loop, and then, after incubation, analysed actin gene expression in the parts that retained the nucleus. They also divided 8-cell and 32-cell embryos by a similar method. They found that only regions containing cytoplasm from the lower half of the egg (mostly on the dorsal side) produced transcripts after a 24 hr period of culture. Their interpretation of these results was that the fate of some 'primary' mesoderm cells was fixed by the inherited localisation of a cytoplasmic determinant and these cells subsequently induced the formation of more 'secondary' mesoderm. However, it is impossible to exclude the hypothesis that

induction was the only process taking place and that cell division led to the establishment of inducing and responsive cells in an isolated vegetal half.

A 32-cell embryo can be divided into four tiers of cells roughly perpendicular to the animal-vegetal axis, the first at the animal pole and the fourth at the vegetal pole. Studies, using a vital dye to label cells in individual tiers, have led to the conclusion that the third tier is the major contributor to muscle *in vivo*, with only a minor role for the second tier (e.g. Nakamura *et al.*, 1978). However, a more recent set of carefully controlled experiments, with an injected fluorescent tracer as the cell marker, has demonstrated that the second tier (from the animal half) is the major source of muscle cells (Cooke and Webber, 1985). Provided this is true, then even if some cells in the vegetal half of the embryo are committed to form muscle due to cytoplasmic localisations *in vivo*, they can only be equivalent to a small fraction of those cells, which are subsequently induced to differentiate into muscle and which are largely in the animal half. Therefore, it is currently believed that mesodermal induction is the major mechanism for mesoderm formation in *Xenopus*, and it can be envisaged that further studies with cardiac actin mRNA as a marker will help to elucidate the details of this process further.

Two groups have analysed the expression of the chromosomal cardiac actin gene, when cell interactions are disrupted in the *X. laevis* embryo by the removal of Ca^{2+} and Mg^{2+} ions (Gurdon *et al.*, 1984; Sargent *et al.*, 1985). This treatment, which dissociates the cells of the embryo, leads to inactivation of gene expression if it is performed during the time at which the gene is normally switched on. However, Gurdon and his

colleagues found that dissociation prior to this time (i.e. over the cleavage and blastula stages), followed by reaggregation, had no effect on the total level of actin transcripts.

By contrast, Sargent *et al.* (1985) reported that expression was severely repressed by this procedure if the cells were agitated during dissociation in order to keep them properly separated. But, if the dissociated cells were retained in the vitelline membrane of the embryo (a protocol suggested to be similar to dissociation without agitation, where cells aggregate in heaps), expression was unaffected. This demonstrated that the absence of Ca^{2+} and Mg^{2+} *per se* was not responsible for the inhibition with agitated cells. In conclusion, these experiments would appear to demonstrate that cell contacts during cleavage and blastula stages are required for muscle formation, although tight cell adhesion is not necessary. The cell interactions that take place may well include those which are responsible for mesodermal induction, but they may also be involved in contacts that are merely permissive for mesodermal differentiation. Hence, these experiments do not directly bear on the localisation/induction controversy.

In summary, the use of cardiac actin mRNA as an early marker for muscle development has led to some progress in elucidating the embryological mechanisms that are responsible for the determination of mesodermal cell fate. I have concentrated my studies on developing a transformation system for cloned genes in *Xenopus* embryos. This has allowed the regulation of expression of the cardiac actin gene to be investigated, and eventually the molecules responsible for its control may be isolated (see Section 8.3). The results from these experiments to date are described in the following two chapters.

CHAPTER 6

DETECTION OF TRANSCRIPTS FROM THE CLONED CARDIAC ACTIN GENE IN INJECTED *XENOPUS* EMBRYOS

SECTION 6.1 INTRODUCTION

In previous chapters the isolation and characterisation of a clone, encoding the entire transcription unit of a functional *X. borealis* cardiac actin gene, has been described. The next aim of the project was to investigate the temporal and spatial regulation of its expression in injected *Xenopus* embryos, and to compare this regulation with that of its chromosomal counterpart. However, to accomplish this, it was obviously necessary to distinguish the transcripts of the injected genes from endogenous *Xenopus* cardiac actin mRNA.

Two alternative approaches could be used. In the first, the clone could be injected into a closely related *Xenopus* species, which expresses a message that is detectably different from *X. borealis* cardiac actin transcripts. In the second, a construct could be designed from pXB-C3.1, which would produce message that is readily distinguished from normal actin mRNAs. The design of such a construct and the development of transcriptional assays, which allowed both of these approaches to be used, are described in this chapter. Initially the injection of unaltered genes into the embryos of a related frog was most appropriate, as it was unclear which sequences in the clone could be removed without affecting the regulation of expression of the cardiac actin gene.

SECTION 6.2 RESULTS

6.2.1 Differentiation of *X. borealis* and *X. laevis* Cardiac Actin Transcripts by S1 Analysis

In Section 5.2.3, an end-labelled *HinfI/BstNI* fragment, spanning the 5' intron/exon boundary of the first coding exon in the cardiac actin gene, was used as an S1 probe. It gives products of identical size (the major one being 147 b in length), when used for S1 analysis of *X. borealis* and *X. laevis* message, under normal S1 digestion conditions (see Figure 5.5). However, it was first shown by Dr. Gareth Cross that, with increasing amounts of S1 nuclease at higher temperatures, the end-labelled fragments from *X. laevis* mRNA were preferentially truncated (Figure 6.1). Consequently, if the assay was performed at 50°C, it became possible to clearly distinguish products from the mRNAs of the two species.

In Figure 6.2a, a wider range of S1 concentrations was used at 50°C. In the most stringent conditions, even the products from *X. borealis* cardiac actin transcripts were shortened, and the resolution of the gel was sufficiently high to separate all the resulting bands. A diagrammatic representation of these results is given in Figure 6.2b. Since the characterisation of the assay, the partial sequence of the *X. laevis* cardiac actin gene has been published (Stutz and Spohr, 1986), and the relevant portion is compared with the *X. borealis* gene sequence (Cross, 1984) in Figure 6.2b. In the part of the coding region that is complementary to the S1 probe, the transcript sequences in the two species are identical, except for a single mismatch near the 3' end.

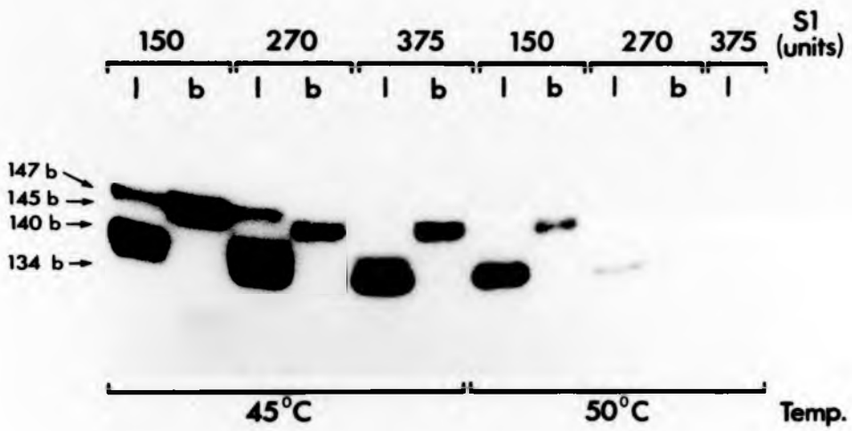
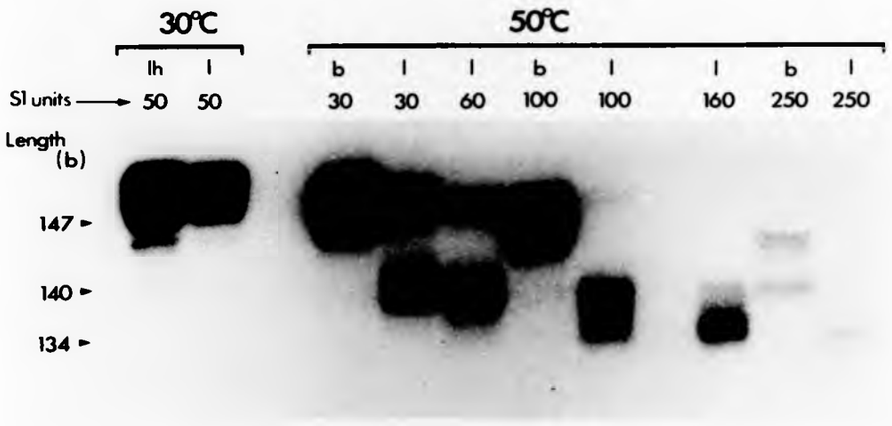


Figure 6.1 Differential S1 Analysis of *X. laevis* and *X. borealis* Cardiac Actin mRNAs

Total RNA from stage 30 tadpoles of *X. laevis* (l) and *X. borealis* (b) was hybridised to the 5'-end labelled *Hinf*I/*Bst*NI second exon probe, and 20 μg aliquots were digested with different levels of S1 at 45°C or 50°C for 30 min. The resulting fragments were separated on an 8% polyacrylamide-urea gel. The longest product observed at these high temperatures was 147 bases for RNA from both species, and it represents precise full-length protection of the RNA-DNA hybrid. However, particularly at a temperature of 50°C, the longer bands were preferentially lost in the reaction with *X. laevis* RNA.

This figure is shown with the kind permission of Dr. Gareth Cross.

a



b

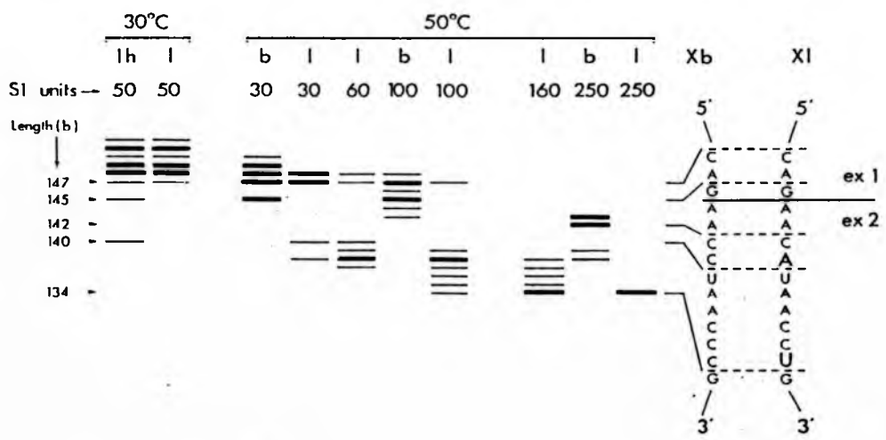


Figure 6.2 Further Characterisation of the Differential S1 Assay

(a) Total RNA from stage 30 tadpoles of *X. laevis* (l) and *X. borealis* (b) was hybridised to the *Hin*fI/*Bst*NI second exon probe, and 20 μg aliquots were digested with a wide range of S1 concentrations at 50°C for 30 min. The resulting fragments were separated on an 8% polyacrylamide-urea gel. On the left are two S1 assays of 5 μg of RNA from *X. laevis* adult heart (lh) and tadpoles (l), digested under very mild conditions (50 units of S1 at 30°C), so that most of the products are still longer than 147 bases. With various exposures of the same gel, fragments differing by only a single base are clearly separated in all tracks.

(b) Diagrammatic representation of the pattern of bands in Figure 6.2a. Major bands are indicated by broad lines, and minor bands by narrow lines. The sizes of some important fragments are marked on the left. The sequence on the right is the relevant hybridising portion of the 5' untranslated region (largely from exon 2) in both *X. borealis* (Xb; G. S. Cross, per comm.) and *X. laevis* (Xl; Stutz and Spohr, 1986) cardiac actin mRNAs (for a more complete *X. borealis* sequence, see Figure 5.3). The points of cleavage, which produce the marked fragments, are indicated (dashed lines). The two mismatches between the *X. borealis* probe and the *X. laevis* message are shown in bold type.

Basically, under stringent conditions, the hybrids are attacked at the mismatches, but the temperature also appears to be sufficiently high to allow partial melting at A-U and T-A base pairs, leading to a complex pattern of *X. laevis* products in some cases and the progressive truncation of *X. borealis* fragments. Furthermore, the bands in close proximity to the 134 b *X. laevis* fragment, which is produced by mismatch, are all longer than 134 b, suggesting that digestion has occurred by the melting of hybridised strands from the end of the hybrid, rather than by local internal melting. However, such a mechanism seems less likely to be responsible for the production of the 140 b fragment at the other mismatched site, where neighbouring bands are shorter and have been cut in an A/T-rich region.

However, in the 19 bases of 5' untranslated region that also hybridise to the probe, two separate base differences are found. The S1 digestion pattern with *X. laevis* message at high stringencies correlates extremely well with the positions of these mismatches. Furthermore, the shorter products from *X. borealis* mRNA and additional products from *X. laevis* mRNA appear to result from cleavage at A-U or T-A base pairs (see Figure 6.2b).

The intensity of the signal in this assay decreases significantly with higher stringency, presumably due to S1 attack at other A-U or T-A base pairs in the DNA-RNA duplex, particularly at the ends of the hybrids. Additional minor products less than 134 bases in length are frequently visible, but many others are probably too small to remain on the gel, because they have been cut near the labelled end of the probe. Digestion temperatures above 50°C (52°C and 54°C) were also used for the assay (data not shown), but the bands were merely reduced in intensity, and no clearer differentiation was achieved.

Thus, an S1 assay was designed which could distinguish two highly homologous messages. The minimum level of S1 nuclease required to separate the protected fragments from the two species was approximately 150 units. As the enzyme's activity declines on storage at 4°C, the optimal quantity of S1 for maximum sensitivity was usually predetermined in a set of pilot assays, before the test samples were analysed. The lengths of digested products were independent of RNA concentration over a range of 1-20 μg of total tadpole RNA per assay (Figure 6.3). This is possibly because the S1 digestion buffer contains an additional fixed amount (3 μg) of salmon sperm DNA in all reactions, which partly compensates for the lowest quantities of RNA substrate.

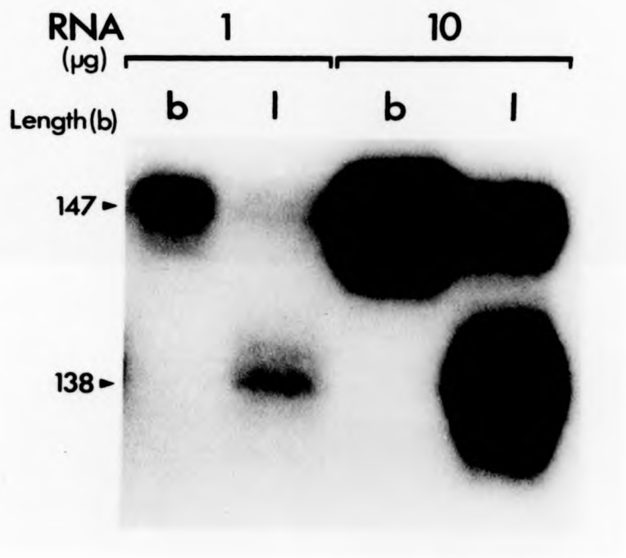


Figure 6.3 Effect of RNA Concentration on the Size of Protected Fragments in the Differential S1 Assay

Aliquots of the *Hinf*I/*Bst*NI probe were hybridised to 1 or 10 μg of total RNA from stage 30 tadpoles of *X. borealis* (b) or *X. laevis* (l). Subsequently, the samples were digested with 60 units of S1 nuclease at 50°C for 30 min, and the products were separated on an 8% polyacrylamide-urea gel. Clearly, a ten-fold increase in the amount of RNA analysed has no effect on the sizes or relative quantities of the different fragments formed, and the same result has been demonstrated with all RNA levels that have been assayed (1-20 μg of total tadpole RNA).

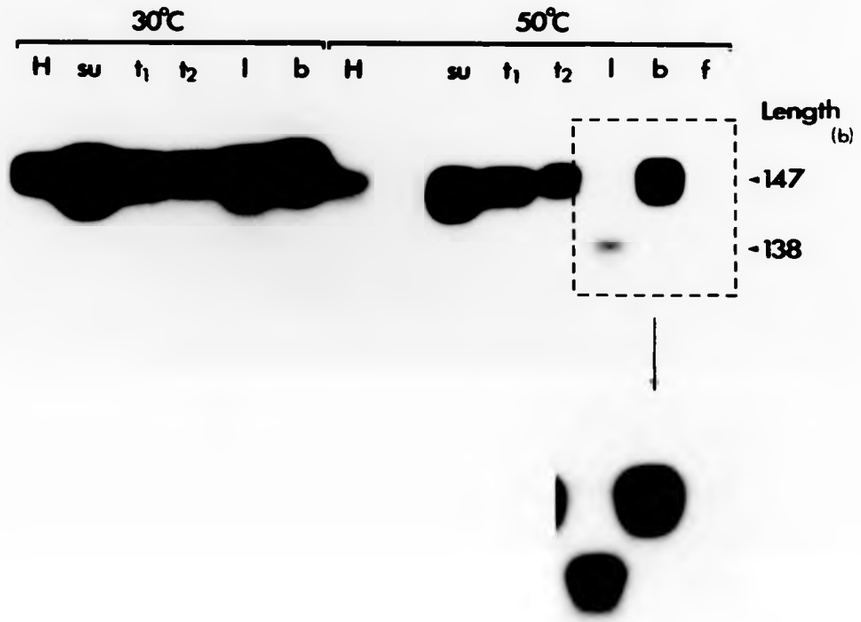
The species *X. borealis* and *X. laevis* are thought to have diverged only about 10 million years ago (Bisbee *et al.*, 1977). The S1 assay was therefore performed on tadpole RNA from other frogs, including two species, *X. tropicalis* and *Hymenochirus*, that are much more distantly related to *X. borealis* (Figure 6.4a). The phylogenetic tree, derived by Bisbee *et al.* (1977) from immunological distances between the albumins of these frogs, is presented in Figure 6.4b. It reveals that *X. tropicalis* diverged from *X. borealis* some 30 million years ago, while the *Hymenochirus* line branched about 120 million years ago. However, very surprisingly, RNA from embryos of both of these species gives the same pattern of bands as RNA from *X. borealis*, even at high stringencies. Indeed, from those frogs examined, only *X .fraseri* transcripts protected different products when digested at 50°C, and these were identical to the fragments protected by *X. laevis* cardiac actin message. A species of unknown origin (*Xenopus* s.u. [species unknown]), which is kept in the laboratory, was tested, and its transcripts also behaved like those of *X. borealis*.

Therefore, during evolution, the cardiac actin mRNAs of *X. laevis* and *X. fraseri* frogs have recently developed sequence changes in their 5' untranslated regions, which make only them sufficiently diverged from *X. borealis* transcripts to be differentiated in the S1 assay.

6.2.2 Differentiation of *X. borealis* and *X. laevis* Cardiac Actin Transcripts by Primer Extension

Although the differential S1 analysis provided a reproducible and unequivocal assay to distinguish *X. borealis* and *X. laevis* cardiac actin

a



b

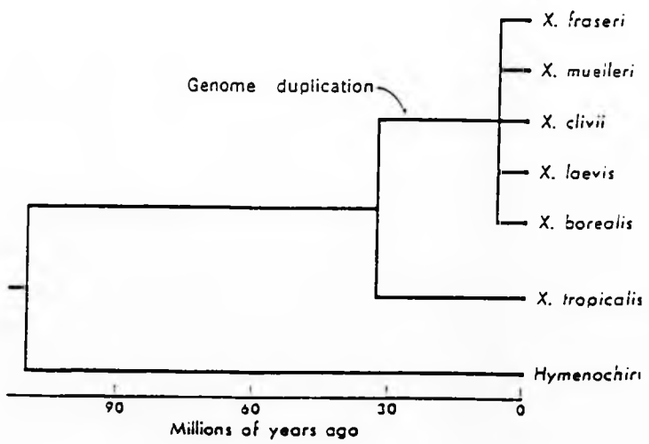


Figure 6.4 Differential S1 Assay on RNA from Several Frog Species

(a) Aliquots of total RNA extracted from four tadpoles of *Hymenochirus* (H), *Xenopus* s.u. (Su), *X. tropicalis* (t_1 and t_2), *X. laevis* (l) and *X. borealis* (b) were hybridised with the *HinfI/Bst*NI probe. The embryos used for samples t_1 and t_2 were produced by crossing pairs of frogs that were imported from different sources in Africa. All the hybridised mixtures were then divided into two, and each sample was digested with 120 units of S1 nuclease for 30 min at 30°C or 50°C. A sample containing the RNA from a single *X. fraseri* tadpole (f) was also hybridised in the same way and digested at 50°C. The resulting products were separated on an 8% polyacrylamide-urea gel.

Most of the RNAs gave the same pattern of bands as *X. borealis* RNA at both digestion temperatures. However, a longer exposure of this gel revealed that *X. fraseri* RNA behaves like *X. laevis* RNA at 50°C(see inset).

(b) Phylogenetic tree of *Xenopus* and *Hymenochirus* species based on immunological relatedness of albumins (from Bisbee *et al.*, 1977). *Hymenochirus* branched from the *Xenopus* line about 120 million years ago. *X. tropicalis* and the *X. laevis* species group (also including *X. borealis*, *X. fraseri* and *Xenopus* s.u.) diverged approximately 30 million years ago, before the duplication of the genome in the latter.

transcripts, it did not demonstrate that these transcripts were initiated from the correct promoter. To do this, a primer extension assay was designed using the 15 base oligonucleotide complementary to the last 15 bases of *X. borealis* cardiac actin mRNA encoded by the first exon (see Figure 5.3). In Figure 6.5, the extension reaction was performed after hybridisation of the end-labelled oligonucleotide to either *X. borealis* or *X. laevis* total tadpole RNA at different temperatures. Hybridisation temperatures between 37°C and 45°C were found to produce maximum levels of products. In subsequent experiments, hybridisation was performed at 42°C.

Although changes in hybridisation temperature did not differentiate between the two cardiac actin messages, the major product with *X. borealis* RNA was one base longer than that with *X. laevis* RNA. The length of the latter was measured against M13 sequencing markers as 47 or 48 bases. If it is assumed that transcription begins at an adenine residue, as is commonly the case (Breathnach and Chambon, 1981), then the actual length is 48 bases (see Figure 6.6). This position is seven bases upstream of the initiation site suggested by Stutz and Spohr (1986) from the sequence of their 'full-length' cDNA clone. Their proposed start site is the second adenine residue within a sequence 5'-CACCCAG-3', which is also conserved in mouse, human and chicken cardiac actin genes (Minty and Kedes,1986). The same nucleotide in this sequence has been proposed as the initiation site for the human gene, and therefore, by extension, for the genes of the other species. However, the length of the primer extension product in my experiment is sufficiently small to make it highly unlikely that I have overestimated its length by seven bases, and so it must be concluded, perhaps surprisingly, that transcription from the *X. borealis* cardiac actin gene

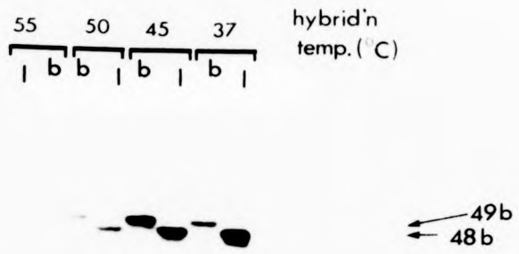


Figure 6.5 First Exon Primer Extension Analysis of *X. laevis* and *X. borealis* Cardiac Actin mRNAs

Total RNA from stage 30 tadpoles of *X. laevis* (l) and *X. borealis* (b) was hybridised at different temperatures to the 15 base oligonucleotide, which is homologous to the region of cardiac actin mRNA encoded by the 3' end of the first exon. After primer extension, the products were fractionated on a 15% polyacrylamide-urea gel.

Clearly hybridisation temperature has no effect on the relative signals from the two RNA samples (the weak reaction with *X. borealis* message at 37°C was not reproducible). Temperatures between 45°C and 37°C allowed optimal hybrid formation, and 42°C was used for further analyses. The major product from *X. borealis* cardiac actin transcripts was one base longer than that from *X. laevis* RNA. The size of the *X. laevis* major band was measured on another gel by comparison with markers from an M13 sequencing reaction. It is either 47 or 48 bases in length. If it is assumed that the transcriptional initiation site is at an adenine residue (Breathnach and Chambon, 1981), it can be concluded that the product length is 48 bases (as indicated in the figure).

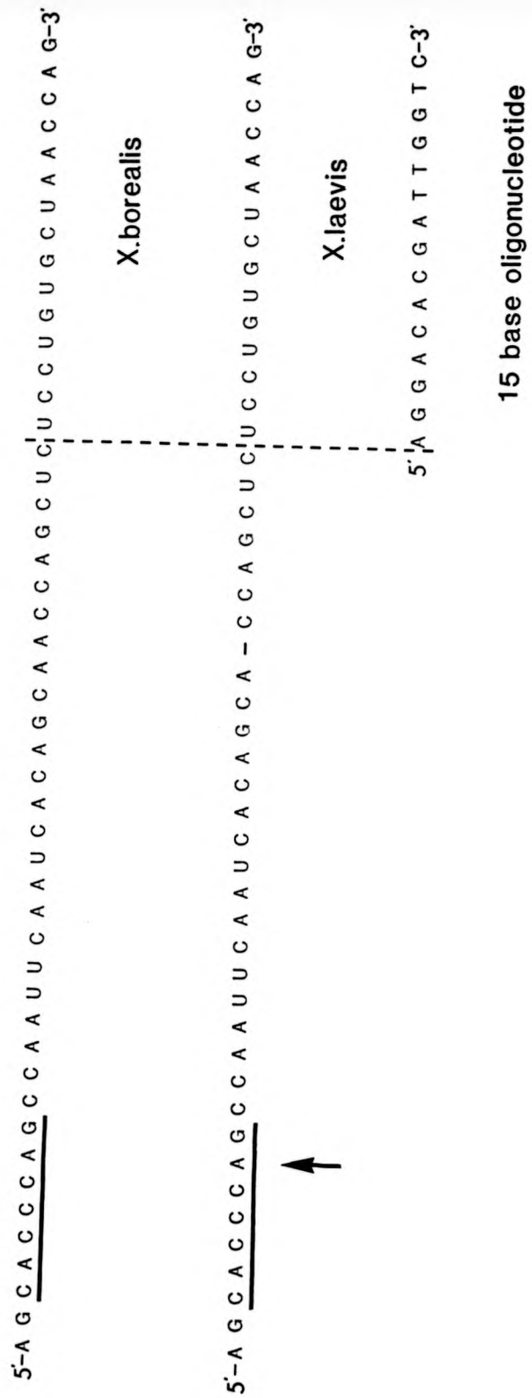


Figure 6.6 Comparison of First Exon Sequences in the Cardiac Actin mRNAs of *X. laevis* and *X. borealis*

The sequences of the first exons from *X. borealis* cardiac actin mRNA (Cross, per. comm.) and the *X. laevis* homologue (Stutz and Spohr, 1986) are presented with the sequence of the 15 base oligonucleotide primer. The two RNAs are identical, except for the loss of a single adenine residue in *X. laevis*. The transcriptional initiation site proposed by Stutz and Spohr (1986) for the *X. laevis* gene is indicated with an arrow. This is found in a short sequence (underlined) that is totally conserved in other known vertebrate cardiac actin genes (Minty and Kedes, 1986). The equivalent adenine residue in the human cardiac actin gene is its transcriptional initiation site. However, my results clearly show that this is not the start site for the *X. borealis* gene (see Section 6.2.2).

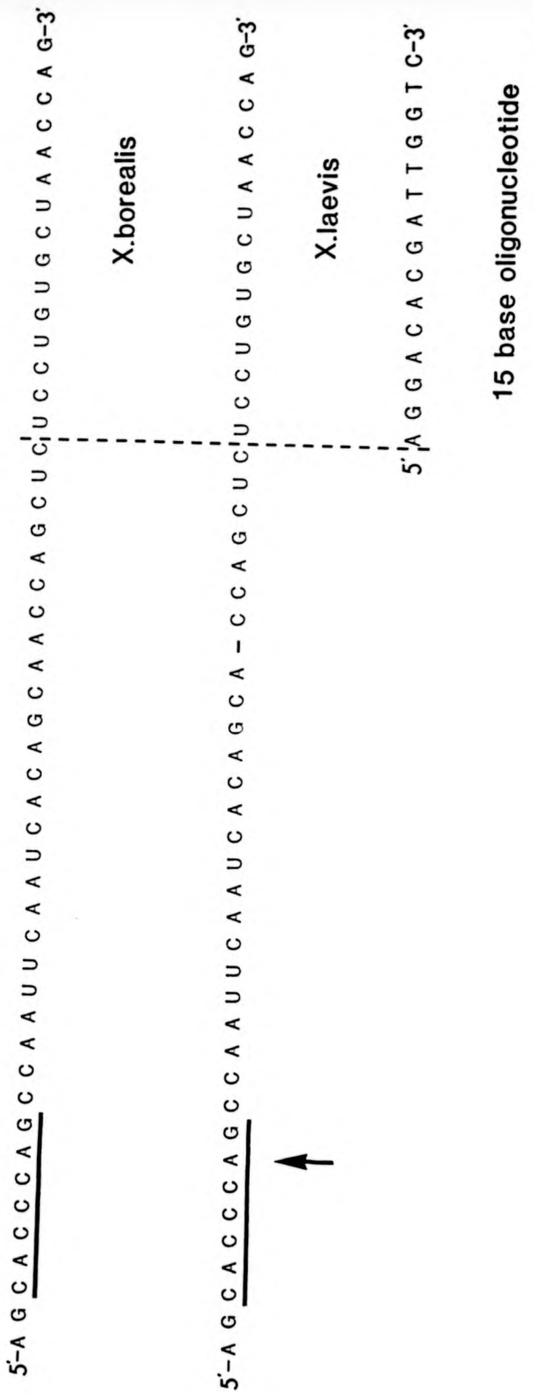


Figure 6.6 Comparison of First Exon Sequences in the Cardiac Actin mRNAs of *X. laevis* and *X. borealis*

The sequences of the first exons from *X. borealis* cardiac actin mRNA (Cross, per. comm.) and the *X. laevis* homologue (Stutz and Spohr, 1986) are presented with the sequence of the 15 base oligonucleotide primer. The two RNAs are identical, except for the loss of a single adenine residue in *X. laevis*. The transcriptional initiation site proposed by Stutz and Spohr (1986) for the *X. laevis* gene is indicated with an arrow. This is found in a short sequence (underlined) that is totally conserved in other known vertebrate cardiac actin genes (Minty and Kedes, 1986). The equivalent adenine residue in the human cardiac actin gene is its transcriptional initiation site. However, my results clearly show that this is not the start site for the *X. borealis* gene (see Section 6.2.2).

is not started within this highly conserved region.

A minor band, one base shorter in length than the major one, is observed with primer-extended RNA from both species. This is almost certainly produced by the same mRNA species, as similar doublets from a single RNA have been reported in other primer extension studies (Cross *et al.*, 1986; Luse *et al.*, 1981).

Now that the sequence of the *X. laevis* cardiac actin first exon has been published (Stutz and Spohr, 1986), it has been possible to compare it to the sequence in *X. borealis* (G. S. Cross, per comm.). Figure 6.6 shows that, as expected from the hybridisation evidence, the regions which hybridise to the primer are identical. Indeed the sequences encoded by the first exons are entirely the same except for the insertion of a single adenine residue in the *X. borealis* mRNA, accounting for the extra nucleotide in the primer extension product.

The extension reaction was also performed using tadpole RNA from other species of frog (Figure 6.7). In this case, a number of significant differences were observed. The *X. tropicalis* and *Hymenochirus* RNAs appeared to hybridise weakly to the primer at 42^{o}C, and may therefore contain a mismatch in the complementary region. Although the products with *X. tropicalis* and *Xenopus* s.u. RNA appeared to be the same length as with *X. borealis* RNA, the product from *Hymenochirus* RNA was one base longer. Therefore, the first exons of cardiac actin genes from these different species typically show subtle variations in their length and sequence.

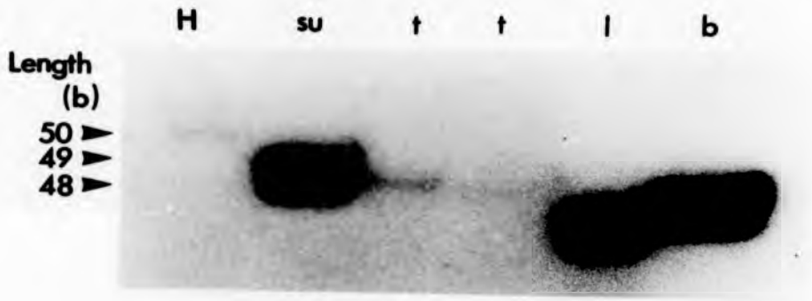


Figure 6.7 Primer Extension Assay of RNA from Other Frog Species

Aliquots of total RNA extracted from two tadpoles of *Hymenochirus* (H), *Xenopus* s.u. (Su), *X. tropicalis* (t_1 and t_2, see Figure 6.4), *X. laevis* (l) and *X. borealis* (b) were hybridised to the 15 base oligonucleotide primer at 42°C. After primer extension, the products were separated on a 15% polyacrylamide-urea gel.

The intensity of the bands from *Hymenochirus* and *X. tropicalis* RNAs is relatively low. S1 analysis of the same samples has shown that this difference is not a direct consequence of a depleted cardiac actin mRNA concentration. It is therefore likely that there is a mismatch in the RNA sequence that hybridises to the oligonucleotide, which leads to a reduced level of hybrids.

The major product from *X. tropicalis* transcripts is the same length as that from *X. borealis* mRNA (49 bases). This also appears to apply for *Xenopus* s.u. RNA, although the unusual curvature of migration on the gel results in a slight misalignment of the bands in this track with the bands in the neighbouring *X. tropicalis* track. However, the product with *Hymenochirus* cardiac actin message is clearly one base longer (50 bases), while the *X. laevis* product is one base shorter (48 bases; see Figure 6.5). Therefore, the data demonstrate that both the length and sequence of the first exon in the cardiac actin gene have undergone minor changes throughout the evolutionary period covered by these different species.

6.2.3 Construction of an Actin-Globin Fusion Gene

The differential S1 assay and oligonucleotide primer extension reaction proved to be invaluable in the early studies of expression from the injected actin clone. However, pXB-C3.1 was manipulated to produce a fusion gene in order to increase the sensitivity of transcript assays and to allow a direct comparison with the expression of a gene which was not regulated in a tissue-specific fashion.

The fusion contained the first two exons of the actin gene and the final exon of the mouse β-globin gene, joined in intron sequences at *Hind*III sites in both genes (Figure 6.8). The highly conserved 3' untranslated region (G. S. Cross, per. comm., Gunning *et al.*, 1984a) and coding sequence for the last 234 amino acids of the actin protein were therefore removed, as well as the associated introns. In a correctly spliced message, the globin region would be translated from this "actbin" clone in the proper frame.

The globin part of the actbin gene was obtained from another fusion construct, in which the front part of a *X. laevis* histone H3 gene had been joined to the penultimate mouse β-globin exon at *Bam*HI sites in both genes (the "hisbin", a kind gift from A. Wilson and S. Sheikh; see Figure 6.8). The junction had been manipulated by end-filling of the *Bam*HI 5' overhangs to bring the globin coding sequence into frame.

Therefore, these two constructs, the actbin and the hisbin, contained identical final exons, but the upstream exons and transcriptional promoter regions were different. In subsequent injection experiments, the hisbin clone acted as an important control gene, because, unlike the

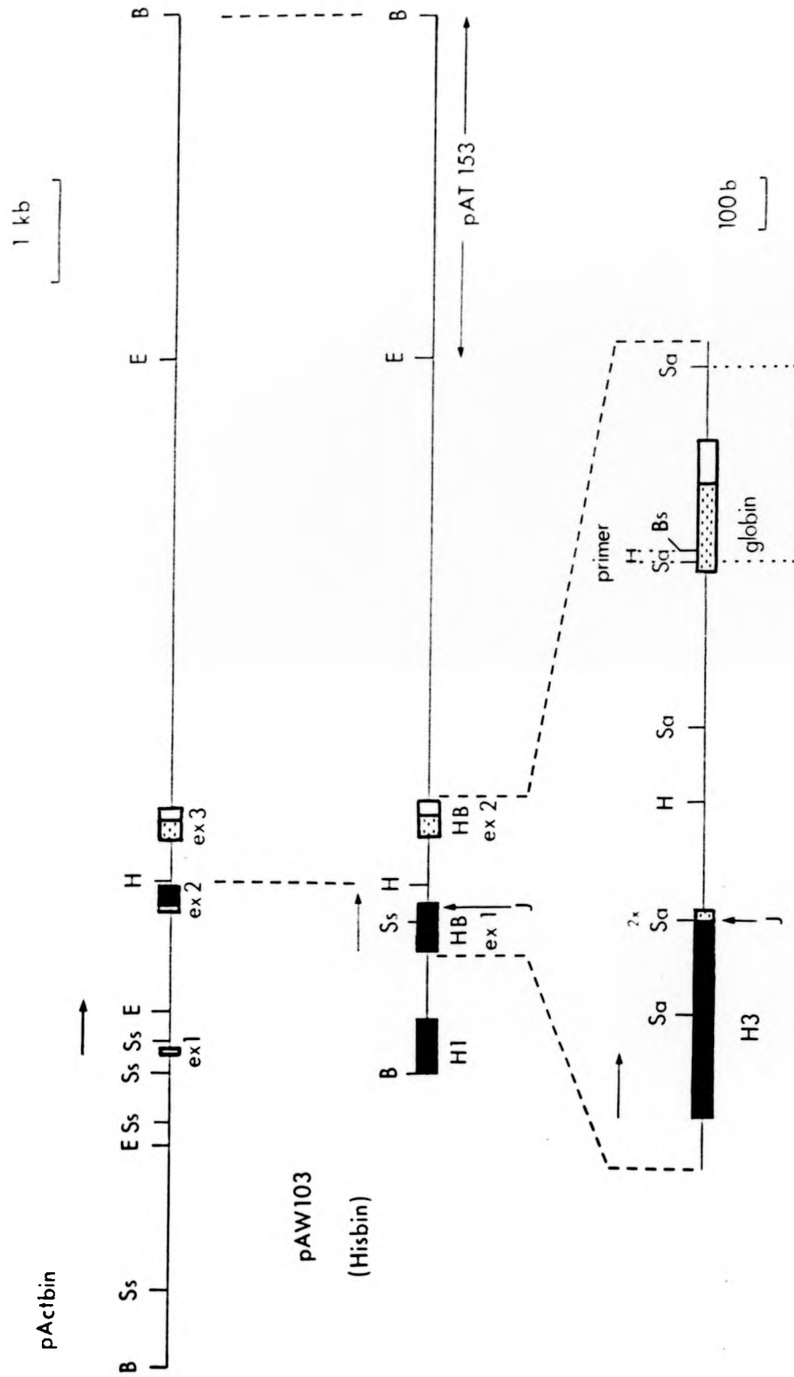


Figure 6.8 The Structure of the Actbin and Hisbin Clones

The actbin clone is shown at the top and the hisbin clone below it. The *Hind*III/*Bam*HI fragment containing the last globin exon is shared. Empty bars represent 5′ and 3′ untranslated regions, the stippled bars represent the globin coding region, and the solid bars represent both actin and histone coding regions. The arrow shows the start and direction of transcription. J marks the boundary between the *Xenopus* histone H3 gene and the mouse β-globin sequence in the hisbin (HB). This clone also contains the 3′ end of a *Xenopus* histone H1 gene, bounded by a unique *Bam*HI site. *Bam*HI was used to linearise both actbin and hisbin clones before injection.

An enlarged portion of the common globin region is also illustrated. The *Sau*3A fragment used for S1 analysis of transcripts from both clones is indicated, together with the *Bst*NI/*Sau*3A fragment used for primer extension. These assays are discussed in Section 6.2.4. Not all the *Bst*NI sites are shown on this map. Restriction sites: B = *Bam*HI; Bs = *Bst*NI; E = *Eco*RI; H = *Hind*III; Sa = *Sau*3A; Ss = *Sst*I.

actin promoter, the histone promoter would not be expected to be expressed in a tissue-specific fashion in transformed embryos.

6.2.4 Detection of Expression from the Actbin and Hisbin Clones

It had already been established that, when the hisbin clone was injected into oocytes, transcripts with correctly processed and polyadenylated 3' ends were observed (A. Chambers and R. W. Old, per. comm.). Expression from both actbin and hisbin constructs was, therefore, routinely studied by S1 analysis using a 3'-end-labelled *Sau*3A fragment, which overlaps the 3' portion of the final globin exon (Figure 6.8). Correctly processed mouse globin transcripts protected a region of 243 bases in the 392 base probe from digestion (see Figure 7.9). The signal intensity was maximised when actbin or hisbin RNAs were hybridised at 60°C in formamide (data not shown). The probe did not cross-hybridise detectably with *Xenopus* globin RNA in tadpoles (e.g. see Figure 7.14b).

An alternative assay, which could demonstrate correct transcriptional initiation and splicing at both introns of the actbin gene, used a 24 base 5'-end-labelled *Bst*N1/*Sau*3A fragment from the upstream end of the *Sau*3A fragment discussed above (see Figure 6.8). This primer hybridised optimally at a temperature of 60°C in aqueous solution (data not shown), and did not cross-react with any *Xenopus* RNAs (see Figure 7.11). From sequence data, the length of a fully extended product from correctly initiated and spliced actbin message is predicted to be 235 bases, and this was indeed the length of the major extended band (see Section 7.2.3).

SECTION 6.3 DISCUSSION

In this chapter, I have discussed the development of assays to differentiate between two highly homologous mRNAs from species which diverged only about 10 million years ago (Bisbee *et al.*, 1977). These assays have allowed the expression of a *X. borealis* cloned cardiac actin gene to be investigated in embryos of *X. laevis*, a species that is presently used for most experiments in our laboratory. Examination of RNA from frogs, which separated from the *X. borealis* lineage much earlier than *X. laevis*, has suggested that the differences between *X. laevis* and *X. borealis* cardiac actin transcripts may be rather fortuitous. Both the S1 and primer extension assays on *X. tropicalis* tadpole RNA give identical length products to *X. borealis* RNA, even though the species diverged about 30 million years ago. Indeed *Hymenochirus*, which diverged at least 130 million years ago, synthesises cardiac actin message that is indistinguishable from *X. borealis* RNA by the S1 assay, and gives a primer extension product that is only one base longer. The region which hybridises to the first exon oligonucleotide primer may, however, also contain at least one mismatched base in *X. tropicalis* and *Hymenochirus*. This would most easily be explained if the hybridising sequence in the RNAs of the *X. laevis* species group mutated after these frogs diverged from *X. tropicalis*.

In the most likely scenario to account for the results from S1 analysis, it would appear that the sequence of the second exon in the cardiac actin gene has generally been highly conserved over the past 120 million years of frog evolution. Not only does a probe synthesised from the *X. borealis* gene hybridise to RNA from all species analysed, but the RNA

usually protects fragments of identical size, even under conditions where some single base mismatches can be recognised. However, probably before the branching of the *X. laevis* and *X. fraseri* lines, but after their divergence from *X. borealis* and *Xenopus* s.u., two single base changes took place in the 5' untranslated sequence. These relatively recent alterations form the basis of the differential S1 assay. The results of Bisbee *et al.* (1977) also support the hypothesis that *X. fraseri* and *X. laevis* are more closely related to each other than to *X. borealis*.

There is a greater, albeit limited, degree of evoutionary change in the first exons of the cardiac actin gene from the different frog species. *Xenopus* s.u. and *X. borealis* RNAs give a product of identical length in this assay as well as in the S1 assay, suggesting that they are relatively closely related. *X. fraseri* RNA has not yet been examined to test if it behaves more like *X. laevis* RNA than the RNA from any other species, as has been found with the S1 assay. A summary of the phylogenetic tree, which can be derived from the differences in the cardiac actin message of the frogs examined, is presented in Figure 6.9.

Although the differential S1 assay separates the products from *X. laevis* and *X. borealis* cardiac actin transcripts more clearly than the primer extension analysis, only the latter demonstrates correct initiation of transcription, and so it has been used to confirm any conclusions made from S1 assays. The pattern of primer extension products (both full-length and minor shorter bands) is identical for RNA from both *X. laevis* tadpoles and adult heart. This result implies that the same transcription initiation site is probably used in the two tissues (embryonic skeletal muscle and adult cardiac muscle), as has been

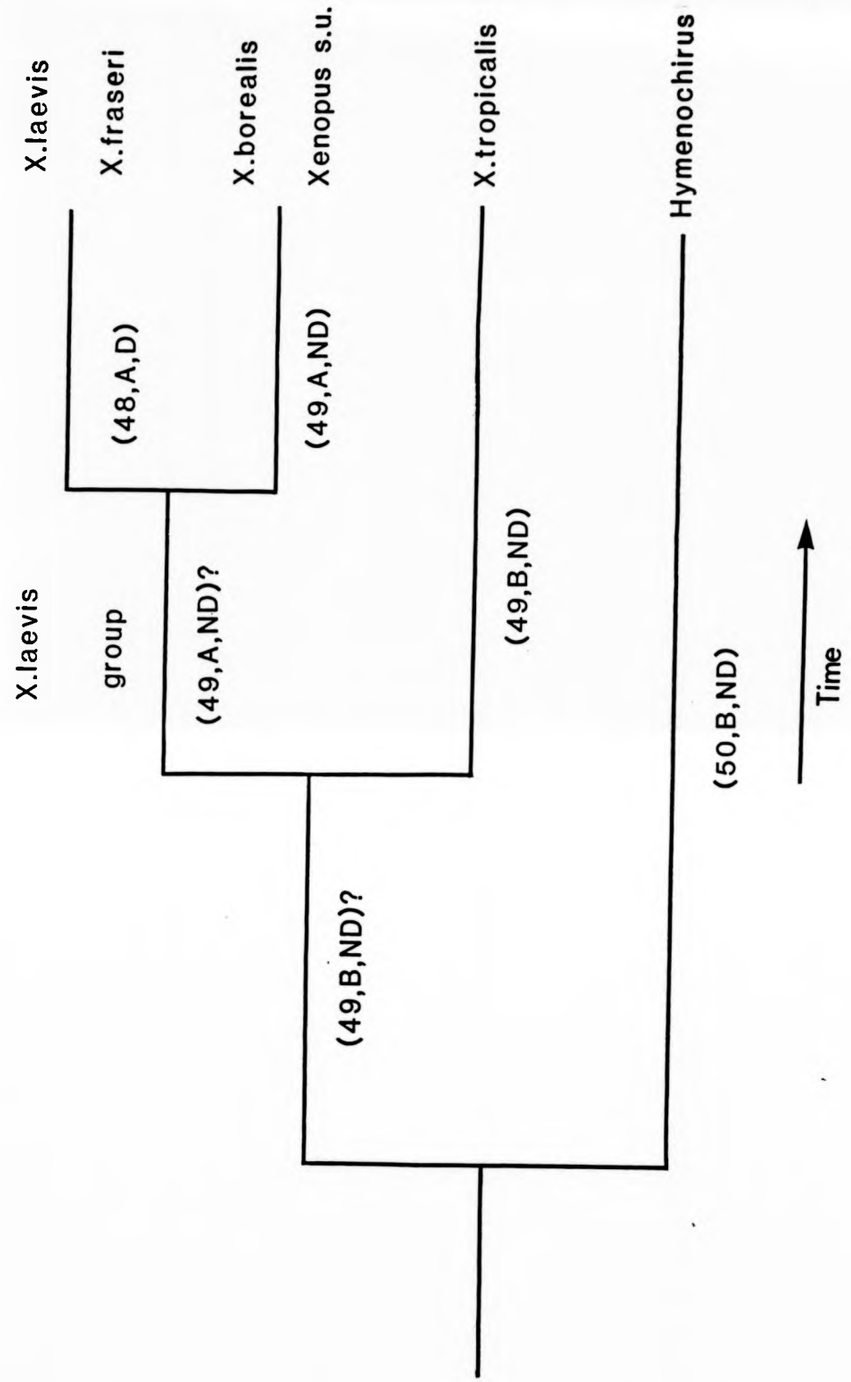


Figure 6.9 Phylogenetic Tree of *Xenopus* and *Hymenochirus* Species Based on Cardiac Actin mRNA Divergence

Cardiac actin transcripts from a number of frog species have been shown to display subtle differences in their reactions in differential S1 and primer extension assays (see Figures 6.4a and 6.7). As all the messages are still very highly conserved, it is extremely unlikely that any specific change that occurred during evolution reverted at a later date. Hence, an evolutionary tree describing species-relatedness can be deduced from the limited data available on the cardiac actin mRNAs. However, it is impossible from this evidence to elucidate the order in which these evolutionary events took place, so the position along the time axis (not to scale) on this tree is derived from Bisbee *et al.*, (1977: see Figure 6.4b).

The characteristics recognised by the transcription assays are as follows:

(i) the length of the oligonucleotide primer extension product - 48-50 bases
(ii) the divergence of the RNA sequence which hybridises to the oligonucleotide: A - totally complementary, B - mismatched.
(iii) the divergence of 5' untranslated sequence encoded by the second exon, as detected by the S1 assay under stringent conditions: D - different from *X. borealis* products, ND - not different.

For each branch on the tree, all three of these characteristics are listed in brackets. Two branches represent ancestral species and the likely characteristics of their cardiac actin messages are given. For *X. fraseri*, only the S1 analysis has currently been performed, but this suggests that it is more closely related to *X. laevis* than any of the other frogs examined (see Figure 6.4). This hypothesis is partly supported by the data of Bisbee *et al.* (1977).

recently proposed for the human cardiac actin gene (Minty and Kedes, 1986). Therefore the developmental regulation of tissue-specific expression for the *Xenopus* and human genes must operate at a single promoter. It does not involve the differential activity of two promoters, which has been demonstrated in other muscle-specific genes like those encoding the alkali light chain of myosin (Daubas *et al.*, 1985).

Although the major reason for developing the differential assays was to allow expression from the unaltered *X. borealis* clone to be detected in injected *X. laevis* embryos, there are additional advantages in this approach. As products are also derived from the transcripts of the endogenous *X. laevis* chromosomal genes, the accuracy of the dissections used to test for localised expression in embryos can be carefully controlled. Furthermore, a direct comparison of transcription from the injected genes relative to the chromosomal genes can be made (see Section 7.3).

However, the S1 assay is relatively insensitive, because the highly stringent S1 digestion conditions dramatically reduce the quantity of hybrids that are not nicked in base-paired regions. Furthermore, because of its small size, excessively large amounts of end-labelled oligonucleotide primer are required to saturate the primer extension hybridisation reaction. Therefore, the sensitivity of this assay was usually reduced, because the primer was not used in vast excess. The actbin clone and the associated test for correct 3' processing of actbin mRNA were developed to overcome these sensitivity problems. The related hisbin construct was also suitable as a control gene, because the tissue-specific promoter had been replaced by the promoter of a gene normally expressed in all cell types.

CHAPTER 7

THE INJECTION OF ACTIN, ACTBIN AND HISBIN CLONES INTO *XENOPUS* EMBRYOS AND OOCYTES

SECTION 7.1 INTRODUCTION

In Section 1.3, I discussed the introduction of cloned genomic DNA into whole animals and its merits as a method for analysing gene regulation. Although proven systems are already established for this approach in *Drosophila* (Spradling and Rubin, 1983; Goldberg *et al.*, 1983) and the mouse (Palmiter and Brinster, 1985), there are potential advantages in using *Xenopus* to study gene control in early development (see Sections 1.3 and 7.3).

Busby and Reeder (1983) have previously demonstrated correct temporal activation of injected Xenopus rRNA genes in embryos. Furthermore, Brown and Schlissel (1985) have recently injected mixtures of somatic and oocyte 5S RNA genes to examine their transcriptional regulation. It has been postulated that these genes compete for transcription factors *in vivo* (Wormington and Brown, 1983). The chromosomal oocyte 5S genes are transcribed in the oocyte and blastula, but not after gastrulation of the embryo. Consequently, only somatic 5S transcripts are found in later embryonic and adult somatic cells. It was proposed that this might be entirely explained if the somatic genes had a much higher affinity for the progressively limiting amounts of transcription factors, which are gradually sequestered by the rising number of genes formed as the embryo divides. However, only a moderate preference for the somatic genes over oocyte genes (approximately four fold) has been demonstrated with cell-free extracts from somatic cells (Pelham *et al.*, 1981), and this is insufficient to explain the effect *in vivo*. By contrast, Brown and Schlissel demonstrated a much more significant preference for injected 5S somatic genes over oocyte genes in embryos,

showing that it was comparable to the relative preference found for the chromosomal genes. As predicted, the injection of more of the limiting transcription factor, TFIIIA (see Section 1.4.2), decreased the degree of competition between the 5S genes for this factor and therefore raised the relative level of transcription from the oocyte 5S genes.

Despite these successes with a few specific genes, experiments with injected genes transcribed by RNA polymerase II have proved less encouraging. In many cases, cloned DNA from other organisms has been used (e.g. Rusconi and Schaffner, 1981; Etkin *et al.*, 1984). But even when homologous *Xenopus* adult and larval globin genes (Bendig and Williams, 1983, 1984), or *Xenopus* vitellogenin gene constructs (Andres *et al.*, 1984) were injected, correct developmental regulation of expression was not observed. Most, if not all, injected genes appear to be activated at the end of the phase of rapid cleavage in the fertilised egg (stage 7½-8). At this time, known as the mid-blastula transition (MBT; see Section 1.6), cell cycles become asynchronous, and transcription is detected for the first time since fertilisation, leading to the activation or reactivation of many genes (Gerhart, 1980; Newport and Kirschner, 1982a). The expression of injected genes that is observed at about this time may therefore be a consequence of these general gene activation events occurring at the MBT (see Section 7.3).

In contrast to these relatively disappointing results with introduced genes, Krieg and Melton (1985) have recently reported the correct temporal regulation of an injected *Xenopus* gene, GS17. It is normally expressed at high levels from the mid-blastula stage until mid-gastrulation. By injecting constructs containing only the front part of the GS17 gene, Krieg and Melton have demonstrated that the regulation is

probably at the transcriptional level.

The major aim in my project was to investigate the regulation of expression from an injected cardiac actin gene, and, in particular, to find out if transcription from such a gene could be localised only to the regions which express the endogenous chromosomal counterpart. The first part of this chapter is devoted to experiments that characterise the optimal conditions for injection and investigate the fate of injected molecules. The quantities of DNA injected in previous experiments have varied considerably from approximately 50 pg (Bendig and Williams, 1984; Krieg and Melton, 1985) to over 1 ng (Etkin *et al.*, 1984). More significantly, both supercoiled and linearised plasmids have been used. Etkin *et al.* (1984) found that injected linearised DNAs persisted better and replicated more efficiently than circular molecules, but most of the other reports discussed above have described the injection of closed circles. It was therefore necessary to test both supercoiled and linear DNAs to assess which form would be most suitable for these expression studies and to estimate the maximum quantities of plasmid that could be injected without affecting embryonic development.

The rest of the chapter discusses the expression of the injected actin, actbin and hisbin clones in embryos and oocytes. The distribution of both injected DNAs and their transcripts is analysed at a number of developmental stages. Correct tissue-specific regulation of the actin gene is observed, demonstrating for the first time that this system is suitable for the study of tissue-specific expression in early development. The major results from the work will also be described in a forthcoming report (Wilson *et al.*, 1986).

All the injections and almost all of the dissections in this project were kindly performed by Prof. Hugh Woodland. Generally, *X. laevis* embryos have been used, because of their suitability in the differential transcription analyses (see Sections 6.2.1 and 6.2.2) and their general availability in the laboratory. The details of the probes and primers for expression assays have been summarised in Section 2.5, where restriction maps of the clones are presented.

SECTION 7.2 RESULTS

7.2.1 Characterisation of the Injection Technique and the Survival of Injected Molecules

Injected linear DNA survives and replicates better than supercoiled DNA

Approximately 0.3-0.6 ng (10-20 nl of a 30 μg/ml solution) of DNA from samples containing either supercoiled or linear pXB-C3.1, was injected into the cytoplasm of *X. laevis* fertilised eggs, and the embryos were allowed to develop. The clone was linearised by digestion with *Sal*I, which cuts at sites flanking the insert. Both the mixture of vector and insert, and the insert alone, which had been isolated from the vector DNA on an agarose gel, were injected. Total nucleic acid preparations were made from the embryos at different stages of development. The DNA from these samples was restricted with *Eco*RI and the resulting fragments were separated electrophoretically on an agarose gel. They were then transferred from the gel to nitrocellulose by the method of Southern (1975) and the filter was probed with the nick-translated 2.6 kb insert of pXB-C1. This probe should hybridise to a 2.6 kb *Eco*RI fragment in pXB-C3.1 (see Figure 2.1). This fragment contains no *Sal*I sites, and so is unaffected by the *Sal*I digestion used for linearisation.

In Figure 7.1, an autoradiograph of the blot is presented, showing that a 2.6 kb digestion product specifically reacted with the probe. Furthermore, the persistence of supercoiled plasmid was clearly considerably poorer than that of linear molecules. The quantity of supercoiled DNA in the embryo generally declined between the time of injection and stage 11 (mid-gastrula stage), but the level increased for

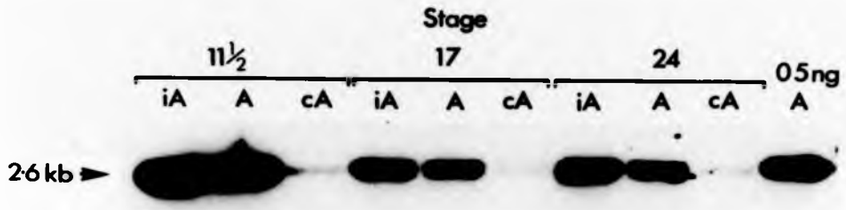


Figure 7.1 Persistence of Circular and Linearised pXB-C3.1 Injected into *X. laevis* Embryos

Fertilised eggs were injected with solutions containing circular pXB-C3.1 (cA), the mixture of fragments from a *Sal*I digestion of pXB-C3.1 (A) and the gel-isolated insert from the products of a *Sal*I digestion of pXB-C3.1 (iA). The concentration of the insert DNA was approximately the same in each of these solutions. Nucleic acids were extracted from pools of six embryos at stage 11½ (mid-gastrula), stage 17 (mid-neurula) and stage 24 (early tailbud tadpole). The DNA from one embryo equivalent of each nucleic acid preparation was restricted with *Eco*RI, and the resulting fragments were electrophoresed on a 0.8% agarose gel, together with 0.5 ng of uninjected, linearised pXB-C3.1 that had been digested with *Eco*RI in the same way. The latter is roughly equal to the amount of DNA originally injected into a single embryo. As for all agarose gels discussed in this chapter, the tracks were examined under ultraviolet illumination after electrophoresis to ensure that all samples contained digested DNA. This DNA was then transferred to nitrocellulose and the filters were probed with the 2.6 kb nick-translated insert of pXB-C1.

As expected, the probe hybridised to a 2.6 kb fragment which is also present in the sample containing uninjected actin plasmid. At much longer exposures, additional bands were observed in all tracks containing nucleic acids from tailbud tadpoles, including those from uninjected embryos. They are cross-hybridising fragments from actin genes in the *X. laevis* genome, particularly the highly homologous cardiac actin genes.

the linear DNAs. In fact, the replication of linear molecules observed in this experiment was relatively low compared to many other experiments (see Figure 7.3). The presence of unattached prokaryotic vector sequences in the injected mixture had no apparent effect on the survival of the insert in embryos (even though, as will be shown later, the vector fragments religate to the insert in random orientations during the early cleavage stage; see Figure 7.3d).

Linear actin, actbin and hisbin plasmid clones have all been shown to be highly replicated in embryos up to the gastrula stage (see Figures 7.3 and 7.16). In most cases, at least a 10-fold increase is observed. By contrast, in the supercoiled form, these same plasmids survive poorly. Indeed, in one experiment described later (see Figure 7.6c), all the injected circular actin molecules were lost before the tailbud tadpole stage. The linearised λ actin clone (λXB-C3) has also shown improved persistence compared to the undigested clone, although even the persistence of the former was much lower than for linearised plasmid containing the same insert (see Figure 7.7).

As a consequence of these results, almost all the other experiments described in this chapter have been performed with linear plasmids.

Higher amounts of injected DNAs become increasingly toxic to embryos

A wide range of dilutions of a solution, which contained the mixture of fragments from a *Sal*I digestion of pXB-C3.1 (at concentrations of 10-200 μg/ml), was injected into fertilised eggs. Increasing concentrations of DNA (100 μg/ml or more) led to rising levels of exogastrulation and embryo death. However, some embryos still survived to the early tailbud tadpole stage, even when they were injected with the 200 μg/ml sample.

Although this experiment was basically designed to examine expression (see Figure 7.8), it was concluded that plasmid concentrations of up to 50 μg/ml were suitable for injection, because, compared to uninjected controls, a high proportion of embryos injected with dilutions in this range survived through neurulation. Therefore, all further injections were performed with solutions at 30-50 μg/ml of DNA (equivalent to injecting 0.3-1.0 ng/embryo if the volume of injection is assumed to be 10-20 nl). Generally, injected DNAs had been previously purified on caesium chloride gradients. Small-scale preparations of λ clones from plates were usually found to be toxic, even at low concentrations. However, DNA fragments isolated from gels by the paper/dialysis membrane technique (Section 2.2.5) did not appear to affect embryonic development significantly when injected.

DNA persists and replicates similarly in most injected embryos

The actbin clone was restricted at the unique *Bam*HI site, found at the 5′ end of the insert. It was injected into fertilised eggs, which were then allowed to develop to the mid-neurula stage (stage 18). Nucleic acids were prepared from individual embryos and the DNA was restricted with *Eco*RI. The products were separated on an agarose gel, the DNA was blotted on to nitrocellulose, and the filter was hybridised to the labelled 5.1 kb *Hind*III/*Eco*RI fragment of the hisbin clone (see Figure 2.2), which contains the final mouse β-globin exon. This probe detects a 6.3 kb region in the *Eco*RI-digested actbin, and does not cross-react with *Xenopus* genomic DNA (see Figures 7.14 and 7.16).

In Figure 7.2, samples from five separate embryos were examined. Four of them contained very similar levels of plasmid DNA, which had been replicated since injection, but the other barely produced a signal.

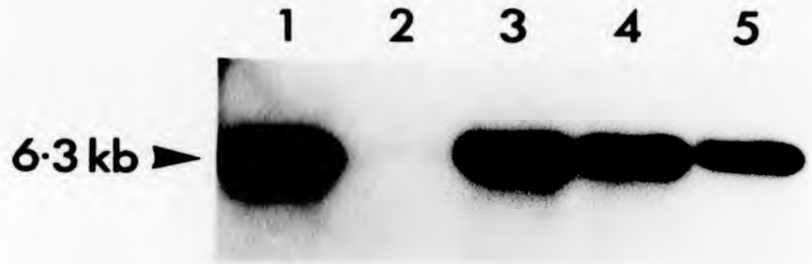


Figure 7.2 Persistence of Injected DNA in Different Embryos

Fertilised *X. laevis* eggs were injected with the linearised actbin clone. Nucleic acids were prepared from individual embryos at the mid-neurula stage (stage 18). The DNA from five different embryos was digested with *Eco*RI, and the products were separated on a 0.8% agarose gel, blotted on to nitrocellulose, and then probed with the nick-translated 5.1 kb *Hind*III/*Eco*RI fragment of the hisbin clone. The latter contains the final mouse β-globin exon, and does not cross-hybridise with *Xenopus* genomic DNA (see Figures 7.14 and 7.16).

As expected, a 6.3 kb fragment reacted with the probe. Generally, similar amounts of the actbin plasmid were found in each embryo. The sample from embryo 2, however, contains a very low level of DNA, which may well have spilled over from adjacent tracks during loading. It is therefore likely that this embryo was not injected. Alternatively, it is conceivable that only a small volume of the DNA solution was initially injected, or the result may represent an unusual but genuine example of poor persistence in this specific embryo.

Indeed, the signal in the latter was so low that it would seem highly likely that it was caused by spillage from an adjacent track during loading, and that the embryo had not been injected in the first place. Alternatively, this result could be explained if either a very small volume was initially introduced into the fertilised egg, or the DNA had persisted very poorly after injection, although both of these would seem relatively unlikely.

The experiment therefore demonstrates that injected DNAs generally persist in a highly consistent fashion in different embryos. This conclusion is strongly supported by many of the other experiments described in the chapter. Small pools of injected embryos (between 4 and 10 in number) have usually been examined, and the consistency of expression and DNA levels observed over a time course, for example, indicates that most of the embryos in each pool must be replicating and expressing the introduced gene in a similar fashion to embryos at the same developmental stage in other pools (e.g. see Figure 7.16). Although eggs from different matings can give quantitatively different results, in any one batch of embryos the observations are usually highly reproducible.

The fate of injected linear and supercoiled plasmids

Etkin *et al.* (1984) have previously shown that injected DNAs form high-molecular-weight concatenates during the cleavage stage of development, particularly when the molecules have been linearised prior to injection. The actbin clone was linearised with *Bam*HI and injected into fertilised two-cell eggs. Nucleic acid preparations were made from pooled embryos at early developmental stages. Both undigested and *Hin*dIII-digested aliquots of each sample were electrophoresed and the DNA was blotted on

to nitrocellulose (Figures 7.3a and 7.3b respectively). The filters were probed with the 5.1 kb globin-specific *Hind*III/*Eco*RI fragment of the hisbin clone.

Examination of the migration of undigested DNAs (Figure 7.3a) showed that most of the plasmid in the sample, which had been taken only a few minutes after injection, was still in the linear form. However, some DNA migrated to the same position as the relaxed circular configuration of the clone and a very small proportion behaved like supercoiled plasmid. A band moving even more slowly than the relaxed form was also observed. As development progressed through cleavage, this high-molecular-weight band, which comigrated with *Xenopus* chromosomal DNA, became increasingly broad and intense, continuing to replicate until the mid-gastrula stage.

The large size of the plasmid did not allow resolution of relaxed circles from the resulting smear after stage 6½, and so it cannot be concluded from this evidence alone that circular molecules of this kind were entirely absent. Indeed multimeric relaxed circles containing only two or three copies of the clone, would probably migrate at the same speed as the most intense part of the smear. However, the studies of Etkin *et al.* (1984) with smaller plasmids have shown that these multimeric molecules are not formed. Furthermore, they found that even relaxed monomers were usually rapidly lost in embryos. Indeed, in the experiment described here, it would appear that relaxed circles have disappeared within 40 min of injection, a time at which the smear of hybridising high-molecular-weight DNA is still insufficiently intense to mask any signal from such molecules. It is therefore concluded from these results that the heavily replicated plasmid DNA formed during the

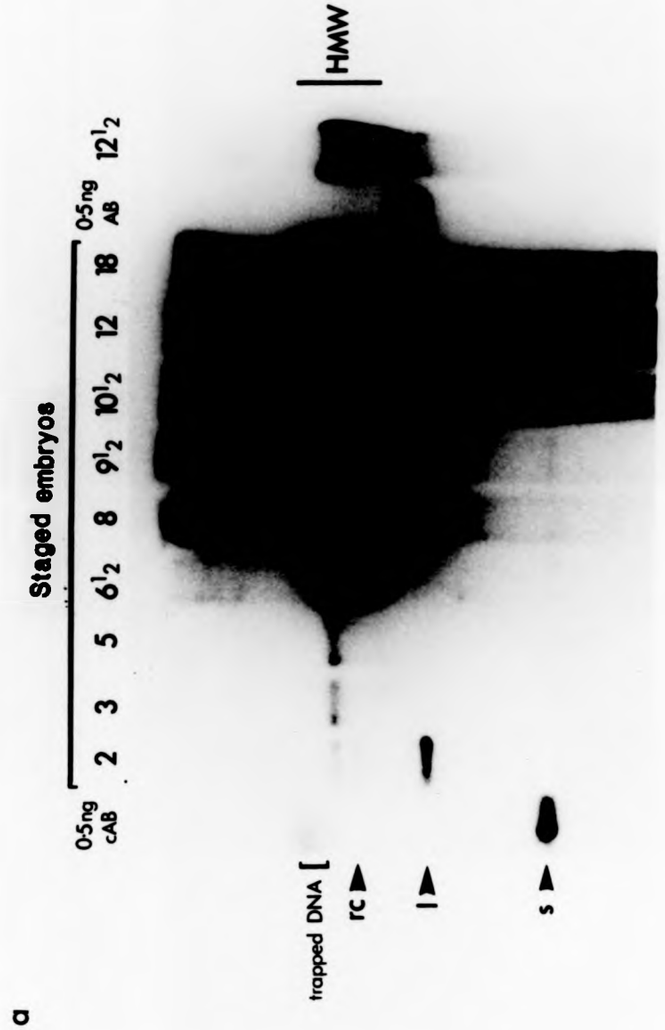


Figure 7.3a The Fate of DNAs Injected into *Xenopus* Embryos

(a) *Bam*HI-linearised actbin DNA was injected into fertilised eggs at the 2-cell stage. Nucleic acid preparations were made from pools of six embryos at different stages of development. One embryo equivalent of undigested material was electrophoresed on a 0.6% agarose gel and blotted on to nitrocellulose. The filter was probed with the globin-specific *Hind*III/*Eco*RI fragment. In addition to the injected embryo samples, 0.5 ng aliquots of undigested (cAB) and *Bam*HI-digested (AB) pure actbin DNAs were mixed with one embryo equivalent of nucleic acids from uninjected *X. laevis* stage 30 tadpoles and fractionated on the same gel. The right hand track contains a sample from stage 12½ injected embryos from another experiment, in which less replication was observed.

In the figure, the positions of supercoiled (s), linear (l) and relaxed circular (rc) plasmid are indicated; as expected, in addition to the former band, the latter two bands are observed in a longer exposure of the track marked cAB. However, this track also contains more slowly migrating hybridising fragments, which were not observed when the plasmid was electrophoresed without added *X. laevis* tadpole nucleic acids. It is therefore likely that a small proportion of the monomeric DNA becomes trapped in high-molecular-weight chromosomal material (HMW) present in this control track. However, in no way can this artifact account for the extremely intense signals seen in the same position in tracks containing nucleic acids from injected blastulae, gastrulae and neurulae. Most of the hybridising DNA in these tracks runs in the same position as the narrow band of high-molecular-weight hybridising material present at earlier stages (i.e. shorter exposures of these lanes appear similar to the track on the far right-hand side). However, a small fraction of the DNA migrates even more slowly, perhaps because the gel is overloaded with molecules that cannot diffuse easily through the agarose pores. Furthermore, there are some shorter DNAs, which may well be formed by shearing, during the preparation of nucleic acids.

Stages of development: stage 2 = 2 cells (within 5 minutes of injection); stage 3 = 4 cells; stage 5 = 16 cells; stage 6½ = late cleavage; stage 8 = mid-blastula; stage 9½ = late blastula; stage 10½ = early gastrula; stage 12 = late gastrula; stage 18 = mid-neurula.

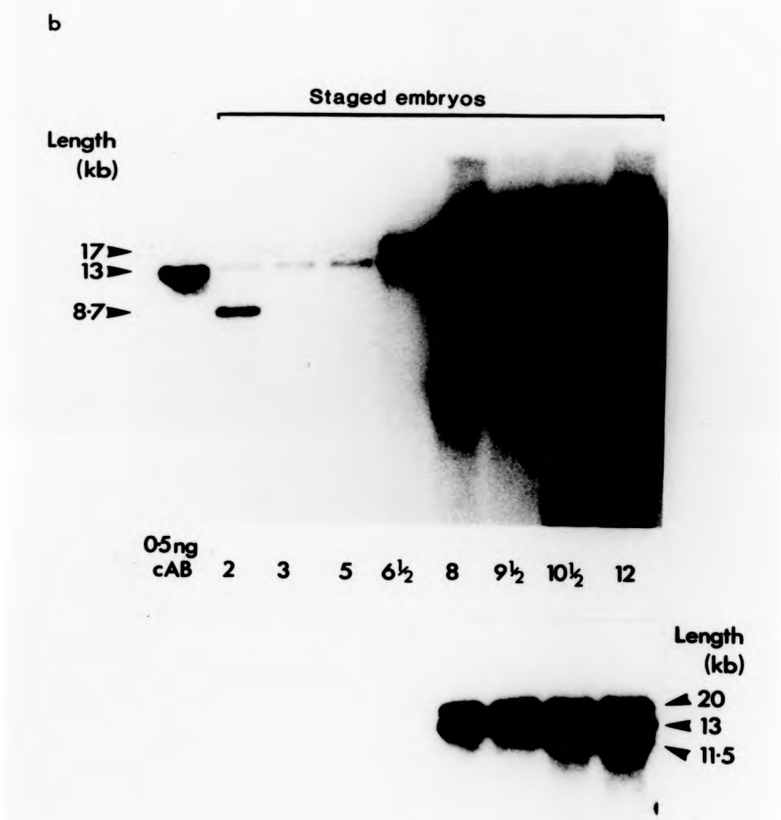


Figure 7.3b The Fate of DNAs Injected into *Xenopus* Embryos

(b) One embryo equivalent of each of the nucleic acid preparations examined in (a) was digested with *Hind*III and electrophoresed on a 0.8% agarose gel, together with a sample consisting of the products from a *Hind*III digest of a mixture containing 0.5 ng of the pure circular actbin plasmid and one embryo equivalent of DNA from uninjected *X. laevis* tadpoles. The DNA was blotted on to nitrocellulose and the filter was probed with the globin-specific *Hind*III/*Eco*RI fragment. Two exposures of the same gel are presented (the lower picture was exposed for a considerably shorter time period).

Immediately after injection, at the 2-cell stage, a band consistent with digestion of linear monomers (8.7 kb in length) is still observed, but this is rapidly lost, and two other major bands of 13 kb and approximately 17 kb (marked as 20 kb in the shorter exposure) become predominant. They appear to originate from the random concatenation of injected molecules (see Figure 7.4a). These bands remain the major species in the gastrula (see short exposure), but other hybridising DNAs, particularly one of approximately 11.5 kb in length, also become visible. They may represent rearranged contaminating DNAs in the injected plasmid preparation that replicate more efficiently in embryos, or they may result from rearrangements occurring very early on in the embryo (see text).

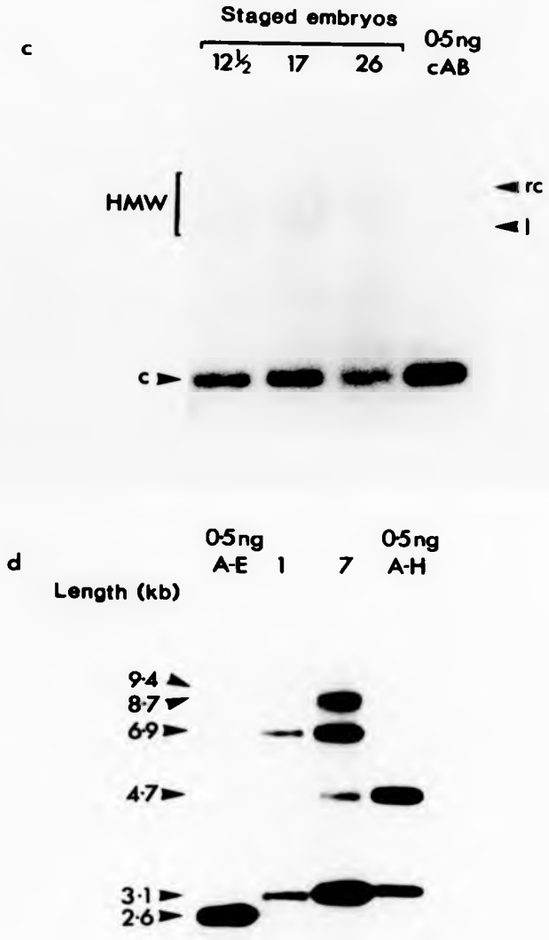


Figure 7.3cd The Fate of DNAs Injected into *Xenopus* Embryos

(c) Supercoiled actbin DNA was injected into *X. laevis* embryos, and nucleic acid preparations were made at different stages of development. One embryo equivalent of each was electrophoresed on a 0.8% agarose gel, along with 0.5 ng of undigested supercoiled DNA (cAB), which is approximately the quantity of DNA initially injected into a single fertilised egg. The DNA was blotted on to nitrocellulose and the filter was probed with the globin-specific *Hind*III/*Eco*RI fragment.

The migration position of supercoiled (c), linear (l) and relaxed circular (rc) forms of the actbin molecule are indicated. As was also shown in Figure 7.1 for pXB-C3.1, supercoiled actbin circles persist poorly in embryos. Particularly in the samples at stage 17 (mid-neurula) and stage 26 (tailbud tadpole), a minor proportion of the hybridising material comigrates with high-molecular-weight DNA, which is poorly resolved from relaxed circular actbin plasmid on this gel. However, these hybridising molecules may be plasmid monomers trapped in *Xenopus* genomic DNA at these later stages (see Figure 7.3a).

(d) The mixture of fragments from a *Sal*I digestion of pXB-C3.1 was injected into *X. laevis* embryos at the single-cell stage. Nucleic acids were prepared from embryos at about 15 min after injection (stage 1; approximately at the time of the first cell division), and at stage 7 (early blastula). One embryo equivalent of these samples was digested with *Hind*III and the resulting fragments were fractionated on a 0.8% agarose gel (in the central two tracks). One half nanogram of uninjected, *Sal*I-restricted pXB-C3.1 was also treated in the same way (A-H); and a similar quantity of the same purified preparation was digested with *Eco*RI and subsequently electrophoresed (A-E). The separated DNA fragments were blotted on to nitrocellulose and the filter was then hybridised to the nick-translated insert of pXB-C1.

The proposed origin of the bands detected is shown in Figure 8.4b. The hybridising 0.45 kb fragment has migrated too far to be observed on this photograph At the early time point, the pattern of bands is consistent with a model in which most of the plasmid has circularised, and few of the molecules are still in the linear form. However, by the blastula stage, more random concatenation between both fragments in the injected mixture appears to have occurred.

cleavage and blastula stages is almost all incorporated into high-molecular-weight material.

When supercoiled DNA has been injected into the germinal vesicles of *Xenopus* oocytes (Laskey *et al.*, 1977) or into fertilised eggs (Newport and Kirschner, 1982b), it has been shown that it is rapidly converted to the relaxed form. However, it is then converted back to supercoiled plasmid as the DNA is packaged into chromatin. An event analogous to the latter process may be taking place in the experiments described here, or alternatively the relaxed circles may be degraded or directly converted into high-molecular-weight DNA.

The small amount of supercoiled actin plasmid, which is formed on injection of linear molecules, is maintained during early development, but it is poorly replicated (see Figure 7.3a). This result is consistent with the data for injected circular pXB-C3.1 in Figure 7.1, and is further supported by the experiment shown in Figure 7.3c. In the latter example, the supercoiled actbin clone was injected into fertilised eggs, the embryos were incubated, and nucleic acids were isolated from them at different stages. The undigested DNA was electrophoresed, Southern blotted and probed with the globin-specific fragment. The results indicate that the injected plasmid largely persisted in a supercoiled form, but it survived poorly compared to linear molecules. A weak smear was also observed to comigrate with high-molecular-weight DNA, although a proportion of this moved at the same speed as relaxed circles on the gel. Furthermore, such a weak signal may really represent supercoiled plasmid monomers that have been trapped in the *Xenopus* genomic DNA (see Figure 7.3a). Even if it is not artifactual, the fact that the injection of supercoiled molecules does

not lead to the synthesis of large amounts of high- molecular- weight material implies that other DNAs, probably linear monomers, are required for the production of this material when linearised plasmids are introduced.

Examination of the blot of *Hind*III-digested samples from the embryos injected with linear actbin plasmid (Figure 7.3b) confirmed the conclusions already made, and showed that the high-molecular-weight DNA consisted largely of multimeric plasmid concatenates joined in random orientations. Immediately after injection, three bands were observed. The shortest (8.3 kb) represented the hybridising *Bam*HI/*Hind*III fragment, which would be expected from a *Hind*III digest of the *Bam*HI-linearised actbin clone. This fragment had disappeared within 40 minutes, supporting the hypothesis that all linear molecules were rapidly lost or converted to a different form during early cleavage. The middle band (approximately 13 kb in length) comigrated with the hybridising fragment in a *Hind*III digest of the circular actbin clone. Such a product could be formed if the opposite ends of either a single *Bam*HI-digested actbin molecule or of two different linear actbin molecules were ligated together. The length of the other band (approximately 17 kb) was consistent with the size of the hybridising product from a *Hind*III digest of a fusion between the same ends of two linear actbin molecules (see Figure 7.4a). During the cleavage and blastula stages, DNA represented by the middle and upper bands was strongly and faithfully replicated, so that these bands remained the major species throughout this period. Therefore, the data showed that a large proportion of the injected molecules comigrating with *Xenopus* chromosomal DNA were linked to each other in random orientations, and other experiments suggest that these conatenates persisted until the

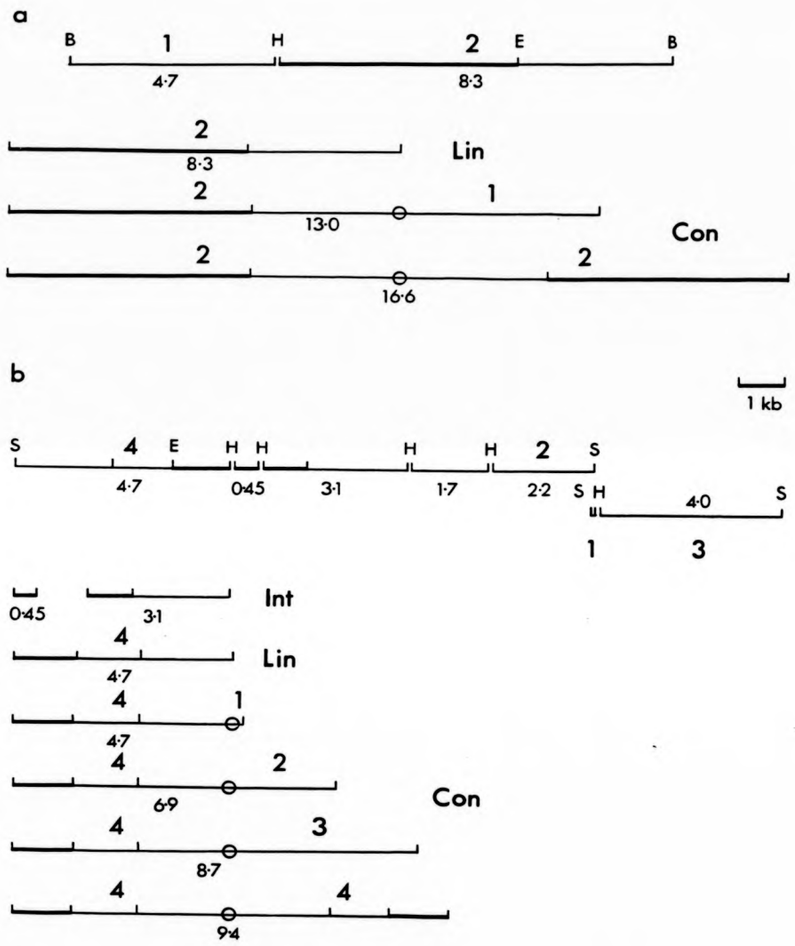


Figure 7.4 Origin of Specific Bands in Southern Blots of Digested DNA from Injected Embryos

At the top of both diagrams the fragments, which are produced on *Hind*III digestion of pure *Bam*HI-linearised actbin (a) and *Sal*I-linearised actin (b) plasmids, are indicated. Their sizes are given and the end-fragments are numbered in order of increasing size. The regions containing the front part of the actin gene in the actbin and actin are marked 1 and 4 respectively. Fragment 1 in the actin clone is a short segment of the polylinker in pEMBL8 of roughly 20 bases in length. In (a) and (b), the regions, which hybridise to the probes used in Figures 7.3b and 7.3d respectively, are drawn with a bold line.

Below the maps of the restricted plasmids are presented all the possible hybridising fragments that will be produced by the *Hind*III digestion of linear monomers or randomly concatenated multimers. These include fragments which are completely internal to the linearised clones (Int), those which are formed only in linear monomers (Lin), and those which can be created by concatenatation (Con). For the latter, the point of ligation in each concatenated molecule is marked with a circle. The sizes of the fragments are consistent with the bands observed in Figures 7.3b and 7.3d. In (b), the 0.45 kb and 3.1 kb fragments are internal to the *Sal*I-restricted actin insert, so they are unaltered by concatenation.

Restriction sites: B = *Bam*HI; E = *Eco*RI; H = *Hind*III; S = *Sal*I

tadpole stage at least (see Figure 7.16f). At the earliest time point, the middle band was at its greatest intensity relative to the top one, consistent with the observation that circular monomeric molecules (largely relaxed) were at their highest level relative to random concatenates at this stage.

By the blastula stage, a series of weaker bands are also frequently visible in tracks containing DNA from injected embryos. In the experiment shown here (Figure 7.3b), a hybridising fragment of 11.5 kb was particularly prominent. These bands almost certainly do not represent randomly formed junction fragments between plasmid and chromosomal sequences, because the copy number is much higher than one per haploid genome. Furthermore, the same integration event would need to have occurred in embryos from each time point. For the latter reason, random rearrangement of the injected DNA also cannot explain the additional hybridising fragments. They must either result from highly specific and reproducible rearrangements, or they could originate from minor rearranged contaminants in the plasmid preparation, which are replicated more efficiently in embryos than the unaltered actbin clone. Other experiments with dissected embryos have shown that the spurious bands are present at the same relative levels in all parts of the embryo, suggesting that if rearrangement after injection is responsible, it must have taken place very early in development.

It was impossible at such early stages to assess if any of the injected material had integrated into the *Xenopus* genome. This question is briefly addressed in other sections (e.g. Sections 7.2.6 and 7.3), and one major conclusion is that most of the DNA probably remains extrachromosomal.

In Figure 7.3d, nucleic acids prepared from embryos, which had been injected with the mixture of fragments from a *Sal*I digest of pXB-C3.1, were restricted with *Hind*III, electrophoresed and Southern blotted. When probed with the nick-translated insert of pXB-C1, a complex pattern of bands was observed. The proposed origin of these hybridising fragments is depicted in Figure 7.4b. Shortly after injection, the insert appeared to have mainly circularised, but, by the blastula stage, the pattern of fragments was no longer consistent with the presence of only circular monomers. Indeed the vector and insert had apparently ligated to each other in random orientations during the cleavage stages of embryogenesis. It has also been found that fragments from *Sal*I-linearised pXB-C3.1 and the *Bam*HI-linearised hisbin clone will concatenate with each other, when coinjected into fertilised eggs (see Figure 7.16f), so the cohesive ends of the DNAs do not need to be identical for ligation to occur.

In summary, the plasmids in this study all persist poorly in embryos if they are injected in the circular form. By contrast, linear molecules are generally replicated by at least 10-fold, and often considerably more, between the single-cell stage and the mid-gastrula. Specific hybridising plasmid bands are observed on Southern blots of restricted nucleic acids from injected embryos of all developmental stages, including later tadpole stages (see Section 7.2.5). Their sizes indicate that the replicated molecules are mostly unrearranged and have become ligated into large multimeric concatenates. The majority of these are probably extrachromosomal (see Sections 7.2.6 and 7.3), because they are gradually lost from the mid-gastrula stage onwards.

Basically, the data presented in this section confirm the findings of Etkin *et al.* (1984). In my study, however, the preferential replication of linear over circular molecules is much more marked. In addition, the experiments show that the resulting concatenated molecules are joined in random orientations. A similar process takes place in microinjected sea urchin embryos (Davidson *et al.*, 1985), whereas the concatenates of linear molecules that are formed in injected mouse eggs are organised only in head-to-tail arrays (see Palmiter and Brinster, 1985). In my experiments in *Xenopus*, different coinjected DNA fragments were all ligated to each other, even if they did not share the same cohesive ends.

In the expression experiments described in subsequent sections, all plasmids were therefore introduced into fertilised eggs in the linearised form at a concentration of between 30 and 50 μg/ml.

7.2.2 Expression from the Injected Cardiac Actin Clone is Localised to the Myotomes

As discussed previously, it was unclear if sequences downstream of the transcription initiation site would be required for proper regulation of expression from the cardiac actin gene, so the unaltered actin gene was initially used for injection experiments. Both λXB-C3 and pXB-C3.1 were injected into the animal pole of artifically fertilised *X. laevis* eggs at the single-cell stage. The highest levels of expression were obtained with the *Sal*I-linearised mixture of fragments from pXB-C3.1.

Dissections of injected mid-neurulae (approximately stage 18) and

tailbud tadpoles (stage 30) were performed (Figure 7.5) to separate the myotomes from other embryonic regions. Differential S1 analysis of the samples (Figure 7.6) confirmed that transcripts from the chromosomal *X. laevis* actin genes were almost entirely localised to the myotomes at both stages of development, as previously shown by Mohun *et al.* (1984). It was also found that *X. borealis* mRNA from the injected clone was distributed similarly, and was present at comparable levels to the *X. laevis* message at both these stages.

In Figure 7.6c, an experiment in which circular plasmid was injected is also presented. No expression was observed, indeed, all the DNA was lost by the tadpole stage (see Figure 7.6d). When small quantities of injected circular plasmid did persist into later development, correspondingly low levels of expression were detected. The undigested λ clone, λXB-C3, also survived poorly in the embryo. When a mixture of its insert and vector arms, produced by digestion with *Sal*I, was injected, persistence was improved. With this approach, correctly localised transcription was observed from the λ clone (Figure 7.7), although the relative level of expression was several fold lower than from the linearised plasmid, presumably because of the smaller numbers of DNA molecules present at the time of expression. This was both a consequence of the reduction in numbers of actin genes that were injected, due to the larger size of the vector, and of the fact that the DNA persisted less well during early development.

In all these experiments, each analysed sample contained RNA from dissected portions derived from the same number of embryos, so equal amounts of RNA were not used for each dissected region. This allowed direct comparison with analyses of samples from whole injected embryos,

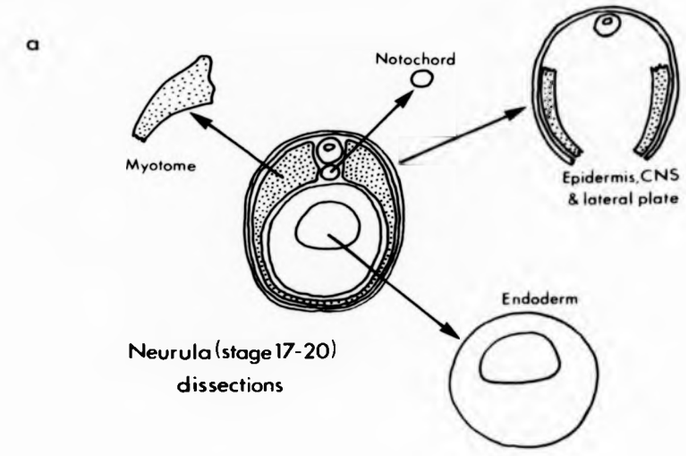
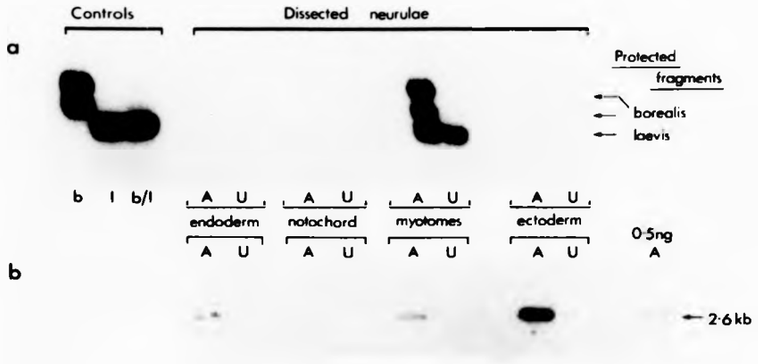


Figure 7.5 Diagram of Embryo Dissections

The diagrams show two of the different dissections used in this study. (a) Mid-neurula (stage 17-20) dissection. (b) Tailbud tadpole (stage 30) dissection. For the mid-neurula dissection, the fraction containing epidermis, CNS and lateral plate is referred to as the ectoderm fraction in the text. These diagrams were drawn by Professor Hugh Woodland.



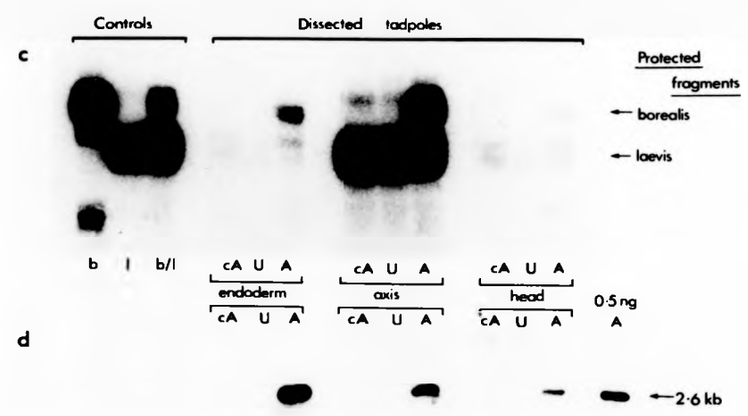


Figure 7.6 Correctly Localised Expression of the Injected *X. borealis* Cardiac Actin Clone, pXB-C3.1, in *Xenopus* Embryos

X. laevis embryos injected with the *Sal*I-linearised actin clone, pXB-C3.1 (A), and the uncut circular clone (cA), both origially dissolved at 50 μg/ml, as well as uninjected controls (U), were dissected at the mid-neurula stage (stage 18) or at the tailbud tadpole stage (stage 30). Total nucleic acids were extracted from pooled fragments of six to ten embryos. Differential S1 analysis was used to distinguish *X. borealis* and *X. laevis* cardiac actin mRNA. The RNAs from four tadpoles (approximately 20 μg) of *X. borealis* and *X. laevis* (tracks b and 1) were analysed as controls, together with a mixture of 20 μg of *X. laevis* tadpole RNA and 2 μg of *X. borealis* tadpole RNA (track b/1). The RNA from the dissected samples was derived from four embryo equivalents for each fraction. The S1-protected fragments were fractionated on an 8% polyacrylamide-urea gel.

*Eco*RI-digested DNA from one embryo equivalent of the same samples was electrophoresed on a 0.8% agarose gel and Southern blotted. Filters were probed with the 2.6 kb *Eco*RI insert of pXB-C1, which hybridised to a 2.6 kb internal fragment in the insert of pXB-C3.1. One half nanogram of *Eco*RI-digested pXB-C3.1 was run as a control. This is approximately equal to the amount of DNA originally injected into one embryo.

(a) S1 analysis of RNA from dissected fractions of mid-neurulae. (b) Southern blots of *Eco*RI-digested DNA from the same samples as in (a). With a slightly longer exposure, a signal can be observed in the injected notochord fraction. (c) S1 analysis of RNA from dissected fractions of tailbud tadpoles. The autoradiograph has been overexposed to show the low levels of *X. laevis* and *X. borealis* cardiac actin message in the head and endoderm fractions. (d) Southern blots of digested DNA from the same samples as in (c).

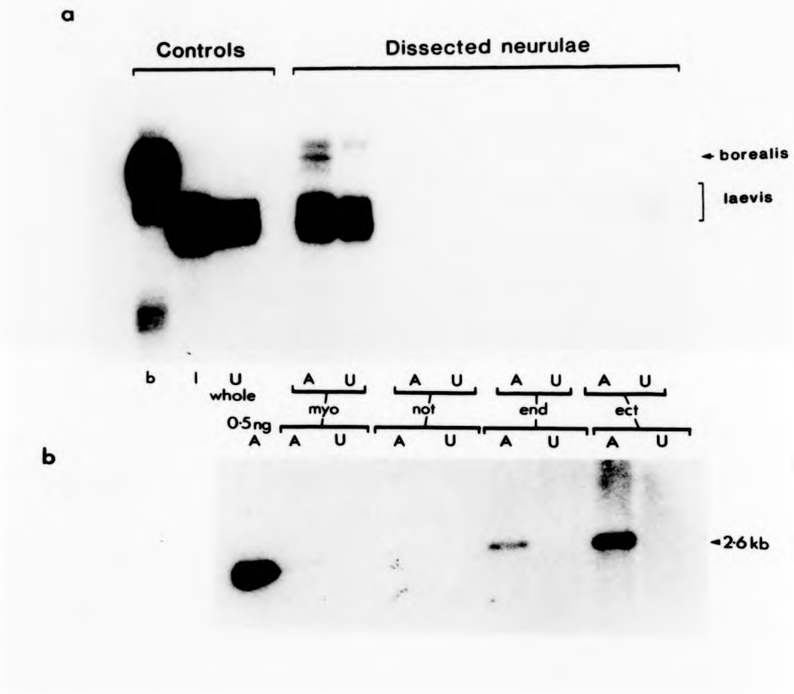


Figure 7.7 Correctly Localised Expression of the Injected *X. borealis* Cardiac Actin Clone, λXB-C3, in *Xenopus* Embryos

X. laevis embryos injected with a 40 μg/ml dilution of the *Sal*I-linearised actin clone, λXB-C3 (A), and uninjected controls (U) were dissected at the mid-neurula stage (stage 18). Total nucleic acids were extracted from pooled fragments of ten embryos. Differential S1 analysis was used to distinguish *X. borealis* and *X. laevis* cardiac actin mRNA. The RNAs from four *X. borealis* (b) and *X. laevis* (l) tadpoles were analysed as controls. The RNA from the dissected samples was derived from four embryo equivalents for each fraction.

*Eco*RI-digested DNA from one embryo equivalent of the same samples was electrophoresed and Southern blotted. Filters were probed with the 2.6 kb *Eco*RI insert of pXB-C1, as in Figure 7.6. One half nanogram of *Eco*RI-digested λXB-C3 was run as a control. This is approximately equal to the amount of DNA originally injected into one embryo. Abbreviations: myo = myotomes; not = notochord; end = endoderm fraction; ect = ectoderm fraction.

(a) S1 analysis of RNA from dissected fractions of mid-neurulae. A sample from whole, uninjected mid-neurulae was also tested. In this experiment, the RNAs were slightly underdigested, so there are two faint upper bands from *X. laevis* cardiac actin transcripts. However, the major band from *X. borealis* mRNA does not coincide with these. (b) Southern blots of *Eco*RI-digested DNA from the same samples as in (a).

and, since the signal from the myotome fraction was similar to that from intact embryos for both *X. laevis* and *X. borealis* mRNAs, it was clear that significant amounts of RNA were not being lost or degraded during dissection and processing. Furthermore, as the ectoderm and endoderm samples from mid-neurulae contain more RNA than the myotome fraction (A. Mohammed, per. comm.), this method of presentation was biased against showing myotome-specific expression from the clone. Indeed, even though the notochord contains less RNA, a signal would probably have been expected from this region, if the injected genes were expressed at the level found in the myotome fraction. Although in the tadpole dissection the axis sample contains somewhat more RNA than the other two fractions, significant expression from the clone should also have been easily detected in the latter.

In dissections at both mid-neurula and tadpole stages, occasionally a very low level of expression from linearised pXB-C3.1 was observed in non-somitic fractions (see Figure 7.6c), but in all cases this was accompanied by a roughly comparable signal from *X. laevis* actin mRNA. One simple explanation for the presence of the latter is that these fractions were contaminated with myotome tissue. Alternatively, other regions may normally express the cardiac actin gene at a low level. For example, in the ectoderm sample from mid-neurulae, which contains both ventral and lateral mesoderm (see Figure 7.5), a minor signal was usually detected, as was also observed by Mohun *et al.* (1984). Whatever the explanation for the chromosomal gene expression in other fractions, these studies could not strictly eliminate the possibility that a proportion of transcription from the injected clone arose from cells that were not expressing the endogenous genes. However, the consistent paired association of low levels of *X. borealis* and *X. laevis* actin gene

expression in approximately similar ratios in non-myotome fractions strongly suggests that this proportion could only be very small.

Primer extension from the 15 base oligonucleotide complementary to the first exon of the actin gene showed that transcripts from pXB-C3.1 were correctly initiated (see Figure 7.12c). Although the lower minor band from reaction with *X. borealis* cardiac actin mRNA is of identical length to the upper major band from *X. laevis* message, the ratio of the two correctly initiated transcripts could be roughly estimated with this technique. By comparison with the ratio estimated by differential S1 analysis, it was clear that most, if not all, the transcription from the injected clone was correctly initiated. Indeed, there was no evidence for aberrant initiation sites from the primer extension results.

There are two different ways in which the myotome-specific expression of the injected gene might not reflect appropriate tissue-specific control. Firstly, almost all the injected DNA may be localised to the myotomes. Secondly, because this system has not been used to examine localised expression before, the possibility exists that any injected gene might be preferentially transcribed in the myotomes. The latter explanation is excluded in the next section by injection of the hisbin fusion construct. To investigate the distribution of the injected clone, DNA from the same dissected samples used for S1 analysis was digested with *Eco*RI, and the resulting fragments separated on a 0.8% agarose gel. After Southern blotting, filters were probed with the insert of pXB-C1. In the neurula dissections, the plasmid was relatively evenly distributed throughout the embryo, and was never at highest levels in the myotomes. In the experiment shown (Figure 7.6b), the most DNA was found in the ectoderm fraction. Quite clearly the distribution of the

injected actin clone was not reflected in its expression. A similar conclusion could be made for the tadpole dissections (Figure 7.6d), where the endoderm sample usually contained most plasmid. Therefore the simple hypothesis that the observed tissue-specific expression was due to localisation of the injected DNA was eliminated.

In one experiment, the effect of concentration of the injected DNA was tested (Figure 7.8). Concentrations above about 50 μg/ml became progressively more toxic to developing embryos (see Section 7.2.1). Increasing levels of injected *Sal*I-linearised pXB-C3.1 were found to produce higher numbers of transcripts throughout the range of concentrations analysed, suggesting that the molecules, which induce expression, are not limiting, even though the number of exogenous genes are in vast excess over the chromosomal genes (see Section 7.3). Furthermore, the level of endogenous cardiac actin message is not noticeably affected by any concentration of injected genes.

In summary, it has been shown that injected actin genes initiate correctly and show appropriate spatial regulation of expression at the neurula stage. The cloned genes continue to be controlled like their chromosomal counterparts, at least up to the tadpole stage.

7.2.3 Expression of the Actbin Gene, but Not the Hisbin Gene, is also Localised to the Myotomes of Developing Embryos

Both hisbin and actbin clones were linearised by digestion with *Bam*HI, which cuts uniquely at the 5′ end of both inserts, and then the plasmids

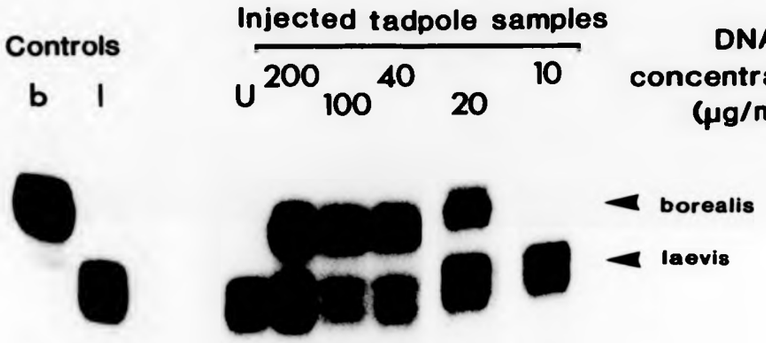


Figure 7.8 Effect of Increasing Concentration of Injected pXB-C3.1 on the Level of Expression in Early Tailbud Tadpoles

X. laevis embryos were injected with roughly the same volume of solutions containing different concentrations of *Sal*I-linearised pXB-C3.1 (10-200 µg/ml). They were allowed to develop to the early tailbud tadpole stage (stage 24). Those embryos injected with high concentrations of DNA (i.e. 100 and 200 µg/ml, and especially the latter) displayed a stronger tendency to exogastrulate than uninjected controls, but at least a few survived to the tailbud tadpole stage. However, no embryos injected with the 200 µg/ml solution developed to later tadpole stages.

Total nucleic acids were extracted from pools of tadpoles from each set of injections. RNA from four embryo equivalents of each sample was analysed by the differential S1 assay (the tracks are marked with the appropriate injected concentration). Four embryo equivalents of RNA from uninjected tadpoles at the same stage (U), and *X. borealis* (b) and *X. laevis* (l) stage 30 tadpoles were tested as controls.

Clearly the level of *X. borealis* cardiac actin transcripts rises as the injection concentration increases. The relatively low amount of endogenous *X. laevis* mRNA in some tracks, particularly the ones containing the 100 µg/ml and 40 µg/ml samples, suggests that some RNA or protected probe may have been lost, possibly because of inefficient precipitation. If this is assumed to be the case and is taken into account when comparing adjacent tracks, the quantity of *X. borealis* transcripts still appears to increase over the concentration range 10-100 µg/ml. However, the tadpoles injected with the 200 µg/ml solution do not perhaps produce as much mRNA as expected. This may be due to the toxic effects of DNA at this concentration. Alternatively, the analysis of only those embryos which survived to the tadpole stage may have selected for embryos in the batch that had been injected with the least DNA.

were injected separately into fertilised eggs. Mixtures of the linearised actin and hisbin clones were also used. Injected embryos were incubated to the mid-neurula stage, and total nucleic acids were prepared from pooled dissected regions. S1 analysis with the globin probe, which overlaps the 3' end of the message (see Section 6.2.4) demonstrated that the localisation of actbin transcripts was essentially identical to that of *X. borealis* actin mRNA described previously, whether the clone was originally injected in the animal pole or the vegetal pole of the egg (Figure 7.9). Furthermore, injection of the actbin construct into *X. borealis* embryos produced a similar level of message compared to *X. laevis* embryos. As the promoter region of the homologous cardiac actin gene in *X. laevis* is so similar to that of the gene encoded by pXB-C3.1 (Stutz and Spohr, 1986; G. S. Cross, per. comm.), this perhaps was not unexpected. However, the result demonstrates that the regulatory sequences of the cardiac actin gene and the molecules which control them have been conserved at a functional level, ever since the divergence of the species about 10 million years ago (Bisbee *et al.*, 1977).

In contrast to the actbin, hisbin transcripts were never at their highest levels in the myotomes, even when a coinjected actin clone was shown to be expressed in a tissue-specific fashion. In the experiment shown (Figure 7.9) the message was largely localised to the ectoderm fraction.

Southern blots of restricted DNA from all these samples probed with the globin-specific *Hind*III/*Eco*RI fragment revealed that the distribution of hisbin mRNA roughly paralleled the relative levels of DNA in each sample (Figure 7.9b). Although the distribution of both hisbin and actbin

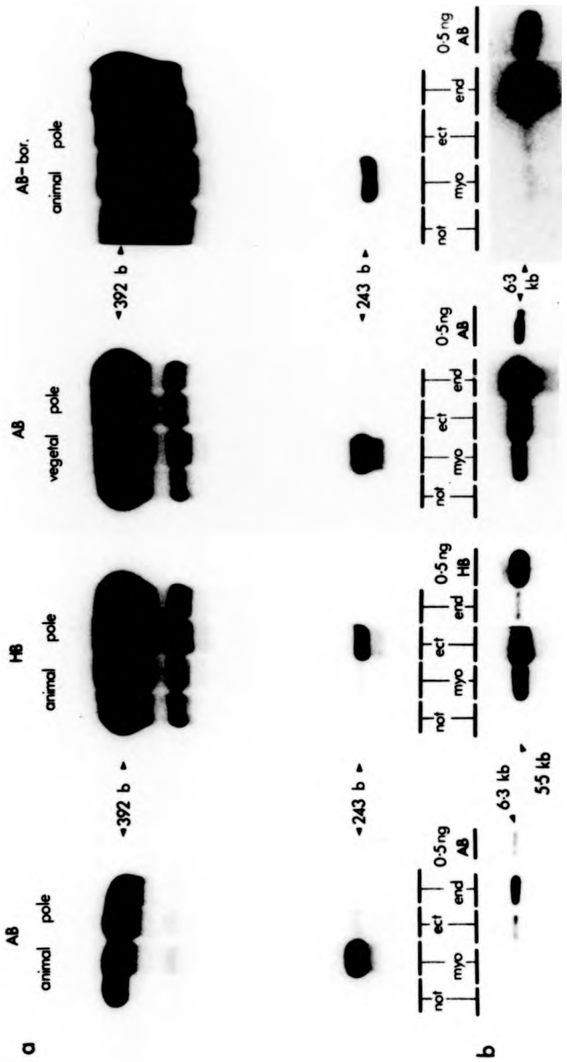


Figure 7.9 Localisation of Expression from Injected Actbin and Hisbin Clones in *Xenopus* Embryos

Single-cell *X. laevis* fertilised eggs were injected with the *Bam*HI- linearised actbin (AB) or hisbin (HB) clones at the animal pole, or with the linear actbin clone at the vegetal pole. *X. borealis* fertilised eggs were also injected with the linear actbin clone at the animal pole (AB-bor). Embryos were dissected at the neurula stage (stages 17-20) and total nucleic acids were extracted. S1 analysis was performed with the globin-specific *Sau*3A fragment (see Section 6.2.4), which detects correct 3' processing of globin mRNA. The protected fragments were fractionated on an 8% polyacrylamide-urea gel. For DNA analysis, one embryo equivalent of each actbin-injected sample was digested with *Eco*RI, while DNA from hisbin injections was restricted with *Eco*RI and *Sst*I. After electrophoresis and Southern blotting, filters were probed with the *Hind*III/*Eco*RI fragment of the hisbin clone, which contains the final globin exon, and hybridises with specific fragments from the restricted actbin (6.3 kb) and hisbin (5.5 kb) DNAs. As a control, 0.5 ng of the appropriately digested, uninjected subclones was also probed, an amount which is approximately equal to the quantity of DNA originally injected in one embryo. Abbreviations: not = notochord; myo = myotomes; ect = ectoderm fraction; end = endoderm fraction.

(a) S1 analysis of total nucleic acid preparations from dissected neurulae. Two embryo equivalents of RNA were hybridised with the 392 base *Sau*3A probe, spanning the 3' end of the last globin exon. The protected fragment in hybrids with mouse globin RNA is 243 bases long. A fully protected band is also observed, as both strands of the *Sau*3A fragment were present in the hybridisation mixture. Below it, a less intense 360 b band represents a minor contaminating fragment in the probe. (b) Southern blots of digested DNA from the same samples as in (a). Neither the S1 probe nor the probe for the Southern blot cross-react with nucleic acids from uninjected embryos (see Figures 7.14b and 7.16).

plasmids did vary somewhat between experiments, the myotome fraction never contained most DNA, irrespective of the pole of the egg into which the clones were injected.

To examine these results quantitatively, the bands from the S1 analyses were excised and the radioactivity was measured. The resulting data are presented in Figure 7.10, and confirm the conclusions made from visual inspection of the gels.

Primer extension using the 15 base oligonucleotide complementary to part of the first cardiac actin exon showed that transcription of the actbin in the myotomes was correctly initiated (e.g. see Figure 7.12c). The RNA samples, which were analysed in Figure 7.9, have also been examined by primer extension from a kinase-labelled *Bst*N1/*Sau*3A fragment complementary to the globin coding sequence (see Section 6.2.4). As the globin exon is the last exon in the hisbin construct, this assay tests not only for correct initiation of transcription, but also for appropriate processing at both the predicted splice sites. A product was observed of the length calculated (235 bases) for an actbin transcript, which has been processed properly at its two splice junctions (Figure 7.11). However, a smaller, minor product was also detectable. It may result from premature termination of extension by reverse transcriptase, or possibly even from extension off degraded RNA. Alternatively, it may be a full-length product from differently processed or aberrantly initiated transcripts. However, its length could not be accounted for by the removal of the entire second exon through a splicing event between the first and third exons.

Dr. Mark Boardman has constructed a fusion, known as the 'skatbin',

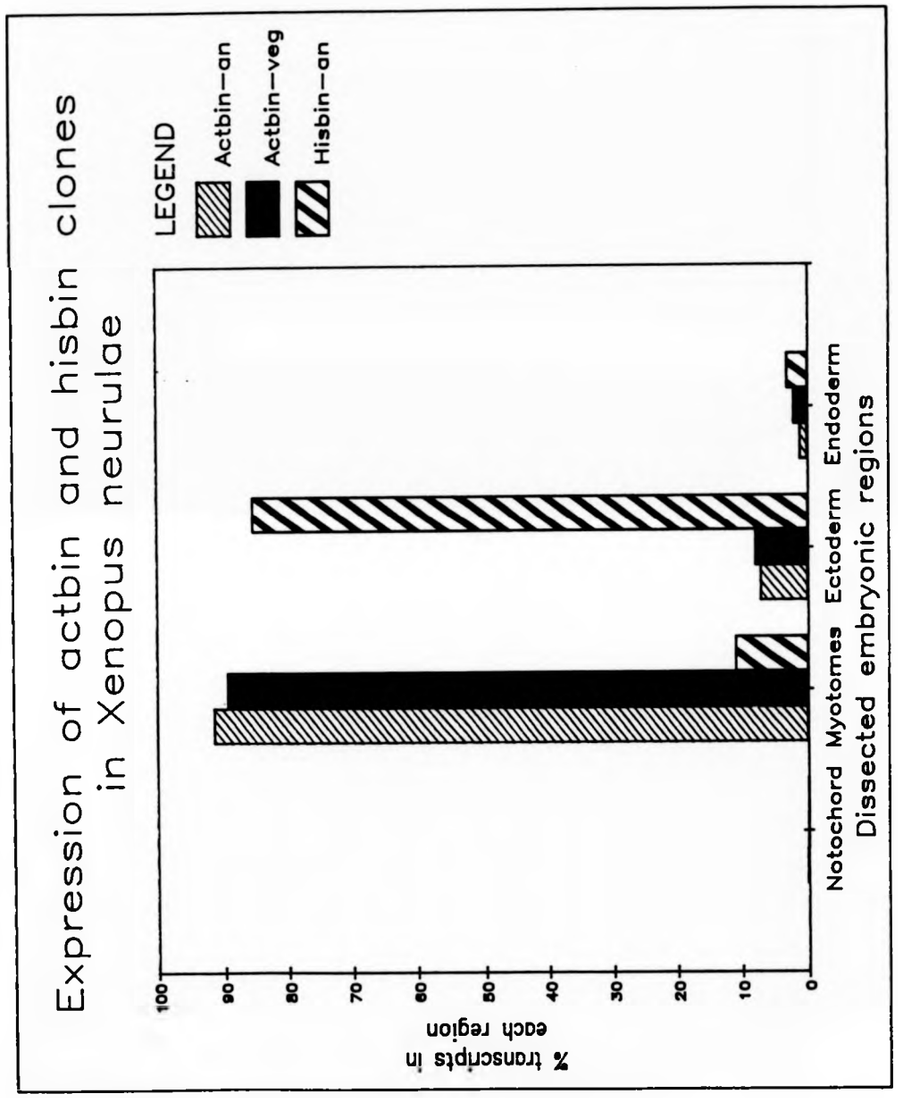


Figure 7.10 Histogram Showing Distribution of Transcripts from Injected Actbin and Hisbin Clones in *Xenopus* Neurulae

S1-protected fragments were excised from the gels in Figure 7.9 and the radioactivity measured. The data are presented in graphical form for injections of the actbin clone into both animal (an) and vegetal (veg) poles of *X. laevis* embryos, and injection of the hisbin clone into the animal pole of *X. laevis* embryos. The total number of transcripts from the actbin construct was up to ten times the number normally produced by the hisbin.

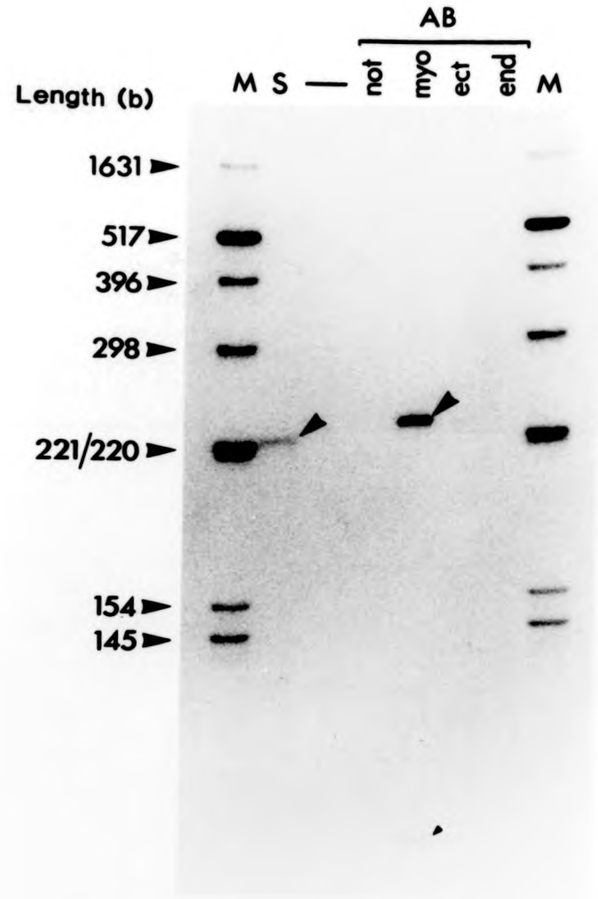


Figure 7.11 Analysis of RNA from Actbin- and Skatbin-Injected Embryos by Primer Extension with the Globin-Specific Primer

The nucleic acid samples from dissected neurulae that had been injected with linearised actbin DNA at the animal pole (AB), and that were examined by S1 analysis in Figure 7.9, were tested by primer extension with a globin-specific *Bst*NI/*Sau*3A primer (see Section 6.2.4). Two embryo equivalents of each preparation were used. In addition, a sample from whole neurulae injected with the skatbin clone (a kind gift from Dr. Mark Boardman) was also tested (S). The extended products were loaded on an 8% polyacrylamide-urea gel, adjacent to a mixture of labelled fragments from a *Hinf*I digest of pAT153 (M). The sizes of these marker bands are indicated. The 24 base unextended primer has migrated too far to be seen in this photograph.

The major product from actbin transcripts (marked by a large arrowhead) is approximately 235 bases in length, the predicted size for correctly spliced message. A minor band of greater than 90 bases is also observed (small arrowhead). As with the S1 analysis, most of the actbin mRNA is found in the myotomes, although a small proportion is detected in the ectoderm fraction at longer exposures (see Sections 7.2.2 and 7.2.4 for a general discussion of non-myotome expression). The length of the major skatbin product (approximately 225 bases; marked with a large arrowhead) also correlates well with the predicted size for correctly spliced message, calculated from the known length of the *X. borealis* skeletal actin 5′ untranslated region (G. S. Cross, per. comm.).

between the *X. borealis* skeletal actin gene and the mouse β-globin gene. It has been made by attaching intron sequences upstream of the third actin exon to the downstream portion of the globin gene at the identical *Hind*III site used to construct the actbin clone. Hence, an artificial intron is formed in the skatbin, which is very similar in organisation to the one in the actbin.

A preparation of nucleic acids from whole injected mid-neurulae (a kind gift from Dr. Mark Boardman) was analysed by the same primer extension assay (Figure 7.11). The size of the major band observed correlated well with the predicted length of correctly processed skatbin mRNA, calculated from the known length of the *X. borealis* skeletal actin 5' untranslated region (G. S. Cross, per. comm.). Therefore, the primary transcripts from both of the actin-globin fusion constructs are mainly correctly spliced at the artificial second intron which is formed, in both cases, by ligation of intron sequences from the two different genes.

From the experiments discussed in this section, it is concluded that the actbin clone is transcribed from the correct initiation site in embryos, and the resulting message is distributed in a similar pattern to that from both the cardiac actin clone and its chromosomal counterpart. The more uniform distribution of hisbin transcripts clearly illustrates that this particular tissue-specificity is not a general feature of any injected gene.

7.2.4 Inappropriate Expression of the Actbin Clone Outside the Myotomes May Be Incorrectly Initiated

Usually the expression of the actbin construct, like the actin clone, was very low in non-myotome fractions of mid-neurulae. In Section 7.2.2, I have already discussed the possibility that much of this transcription might arise in cells, which are also expressing the chromosomal gene. Frequently it was not possible to detect correctly initiated transcripts by primer extension from the first exon oligonucleotide (see Section 6.2.2) in these non-myotome fractions, although it was unclear if this was merely due to insufficient sensitivity of the assay. However, in a few cases (e.g. Figure 7.11), RNAs with normal 5' ends were observed by primer extension analysis, particularly in the ectoderm fraction of dissections of injected neurulae.

In one experiment, a different localisation of actbin mRNA was demonstrated by the globin S1 analysis. A relatively high level of transcripts was found in the ectoderm fraction, approximately half the amount in the myotome region (Figure 7.12a). Southern blots of digested DNA from these samples showed an unusally disproportionate quantity of the clone in the ectoderm region (Figure 7.12b), which could perhaps explain the surprisingly high content of mRNA here. However, primer extension (Figure 7.12c) produced a much lower signal in this fraction relative to the myotomes, suggesting that the transcription from the non-somitic cells in this experiment was mostly not initiated from the actin promoter. One explanation for this result is that cryptic initiation sites in the vector sequence were responsible for the aberrant transcripts. It is unclear if such RNAs are also synthesised

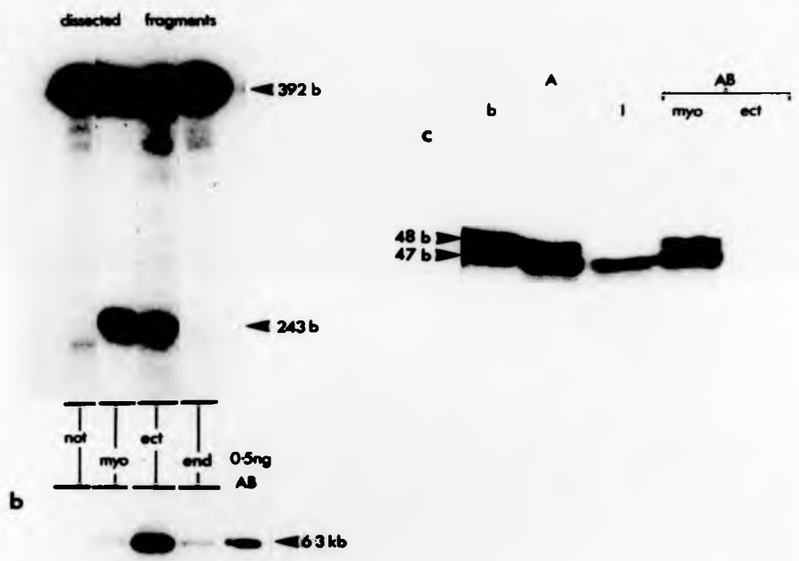


Figure 7.12 Initiation of Actbin Transcripts in Myotomes and Non-Myotome Tissues

(a) S1 analysis of two embryo equivalents of RNA from pooled regions of six actbin-injected embryos, dissected at the mid-neurula stage (stage 18). The high level of actbin transcripts in the ectoderm fraction has not been found in any other experiments.

(b) Southern blots of one embryo equivalent of *Eco*RI-digested DNA from the same samples as in (a) probed with the globin-specific fragment. One half nanogram of linearised, uninjected actbin plasmid was also analysed in the same way. This quantity is approximately equal to the amount of DNA originally injected into one embryo. Abbreviations: not = notochord; myo = myotomes; ect = ectoderm fraction; end = endoderm fraction.

(c) Primer extension from two embryo equivalents of RNA using the oligonucleotide complementary to the first exon of the cardiac actin gene (see Section 6.2.2). The 49 base major product from *X. borealis* tadpole RNA (b) is one base longer than the major product from *X. laevis* neurula RNA (1). RNAs from actin-injected neurulae (A), and from the actbin-injected myotome and ectoderm fractions (AB) used in (a) were also analysed.

at lower levels in all the other injection experiments with the actbin construct, or indeed with the actin clone.

7.2.5 Actbin Transcripts are Still Correctly Localised in Injected Embryos at Later Stages of Development

Tadpoles injected with both the actbin and hisbin clones have been dissected and nucleic acid samples tested by S1 analysis with the globin probe. In the tailbud tadpole (stage 30), the actbin was expressed in a comparable fashion to the expression from the actin clone (described in Section 7.2.2), while hisbin transcripts were found at low levels in all dissected fractions. Most DNA was usually detected in the endoderm fraction, and not in the axis fraction, which contains the somites. Typical expression data from one experiment are presented in the form of a histogram in Figure 7.13.

Preliminary investigation of the localisation of transcripts from injected constructs in stage 40 swimming tadpoles has also been performed. Relatively straightforward dissection can separate such tadpoles into a number of major structural parts (see Figure 7.14a). For example, the embryonic heart is clearly visible. When the heart region was analysed for actbin expression, no signal was detected (Figure 7.14b). However, this fraction contained very little RNA, and I have not been able to detect even endogenous cardiac actin mRNA in the heart at this early stage, although I have yet to examine RNA from pools of more than two embryos. Dr. Gareth Cross has detected such transcripts in the hearts of tadpoles at later stages of development. Indeed Mohun *et al*. (1984) have reported the presence of endogenous

at lower levels in all the other injection experiments with the actbin construct, or indeed with the actin clone.

7.2.5 Actbin Transcripts are Still Correctly Localised in Injected Embryos at Later Stages of Development

Tadpoles injected with both the actbin and hisbin clones have been dissected and nucleic acid samples tested by S1 analysis with the globin probe. In the tailbud tadpole (stage 30), the actbin was expressed in a comparable fashion to the expression from the actin clone (described in Section 7.2.2), while hisbin transcripts were found at low levels in all dissected fractions. Most DNA was usually detected in the endoderm fraction, and not in the axis fraction, which contains the somites. Typical expression data from one experiment are presented in the form of a histogram in Figure 7.13.

Preliminary investigation of the localisation of transcripts from injected constructs in stage 40 swimming tadpoles has also been performed. Relatively straightforward dissection can separate such tadpoles into a number of major structural parts (see Figure 7.14a). For example, the embryonic heart is clearly visible. When the heart region was analysed for actbin expression, no signal was detected (Figure 7.14b). However, this fraction contained very little RNA, and I have not been able to detect even endogenous cardiac actin mRNA in the heart at this early stage, although I have yet to examine RNA from pools of more than two embryos. Dr. Gareth Cross has detected such transcripts in the hearts of tadpoles at later stages of development. Indeed Mohun *et al.* (1984) have reported the presence of endogenous

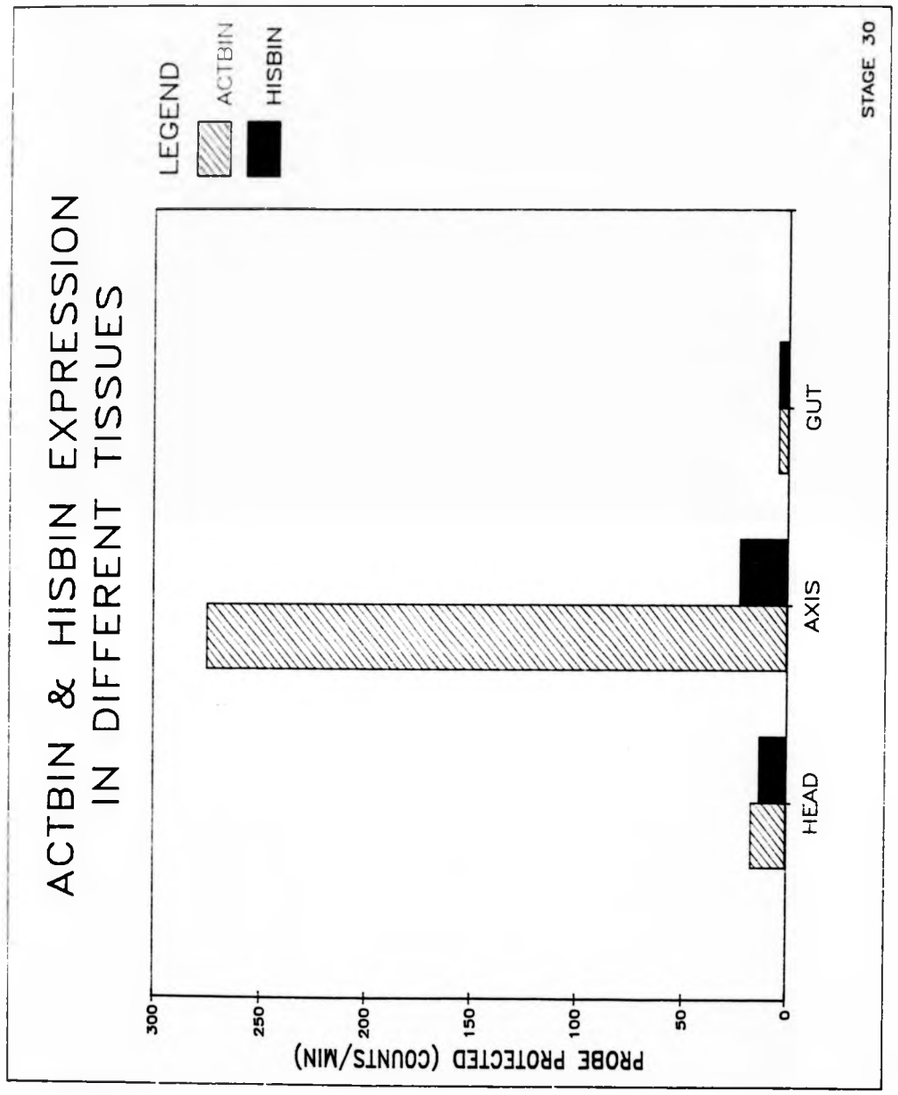


Figure 7.13 Histogram Showing Distribution of Transcripts from Injected Actbin and Hisbin Clones in *Xenopus* Tailbud Tadpoles

Tailbud tadpoles injected with either the actbin or hisbin clones were dissected as shown in Figure 7.5. Nucleic acids were extracted from the pooled dissected fragments of six tadpoles. The samples were analysed with the globin-specific S1 probe and the resulting products were fractionated on an 8% polyacrylamide-urea gel. The 243 b protected fragment in each track was excised and the radioactivity measured. The data are presented in graphical form, and show a marked preference for expression of the actbin gene in the axis sample, which contains the somites and most of the endogenous cardiac actin mRNA (see Figure 7.6). By contrast, the actbin DNA was at highest levels in the gut (endoderm) fraction. The hisbin gene was poorly expressed in all regions (though slightly more transcripts were found in the axis fraction compared to the other two fractions), probably because most of the injected hisbin DNA was lost by this stage of development.

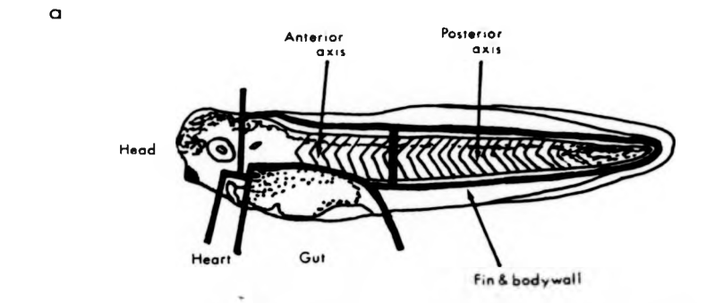


Figure 7.14 Localisation of Actbin Transcripts and DNA in Injected Stage 40 Swimming Tadpoles

(a) Diagram of the dissection of stage 40 swimming tadpoles used in this study. In Figure 7.14b, the axis fraction contains nucleic acids from both the anterior and posterior axis regions.

This diagram was drawn by Professor Hugh Woodland.

(b) *X. laevis* embryos were injected with the *Bam*HI-linearised actbin clone, and allowed to develop into swimming tadpoles, before dissection. Nucleic acid preparations from both injected (AB) and uninjected tadpoles were made. RNA from two embryo equivalents of each sample was analysed, using the globin-specific *Sau*3A fragment as an S1 probe. The faint band just below the 243 base RNA-protected fragment, which is present in injected and uninjected samples, is partly produced by a minor contaminant in the probe, although this contaminant also comigrates with a secondary product in the S1 analysis.

No actbin mRNA is detectable in the heart fraction. However, endogenous cardiac actin mRNA has not been detected in heart RNA from the same number of tadpoles at this stage.

(c) Southern blot of *Eco*RI-digested DNA from one embryo equivalent of the same samples as in (b). The fragments were separated on a 0.8% agarose gel, before transfer to nitrocellulose. One half nanogram of *Eco*RI-restricted actbin plasmid was also analysed as a control which is roughly the quantity of DNA injected into a single embryo. The filter was probed with the globin-specific *Hind*III/*Eco*RI fragment. The parts represented by the dissected fractions do not constitute a complete embryo, and so the total amount of actbin DNA in these samples is considerably less than the level in the whole embryo preparation.

cardiac actin message in the hearts of stage 42 tadpoles, but they probably analysed RNA from a larger pool of tadpoles, as their assay would not appear to be significantly more sensitive than the one used in this study.

In Figure 7.15, other dissected fractions from swimming tadpoles, injected with either the actbin or the hisbin clones, have been analysed. Like the experiment in Figure 7.14b, most of the actbin transcripts were localised to the axis. The body wall and fin fraction also revealed significant expression, but it has since been found that this sample generally contains some endogenous cardiac actin transcripts as well. It has not yet been established whether the actbin RNA, which is present at low levels in the dissected regions that do not express the endogenous gene, is correctly initiated.

Southern blots showed that the distribution of the injected DNA was much more variable at this later stage of development (compare the gut samples in Figures 7.14c and 7.15b). It was never specifically localised to the axis, however, and seemed to be continuing to persist at relatively high levels in at least some of the other regions.

However, the results using the hisbin gene are much less suitable as controls to allow a valid judgement of the data on actbin expression in the swimming tadpole. Although the hisbin DNA persisted at a relatively similar level to the actbin DNA up to the neurula stage, it was preferentially lost during tadpole development. Therefore, hisbin DNA and transcripts were poorly represented in all tissues of the tadpole. Furthermore, the distribution of the remaining plasmid was, like the actbin, much more variable (compare distributions of hisbin and actbin

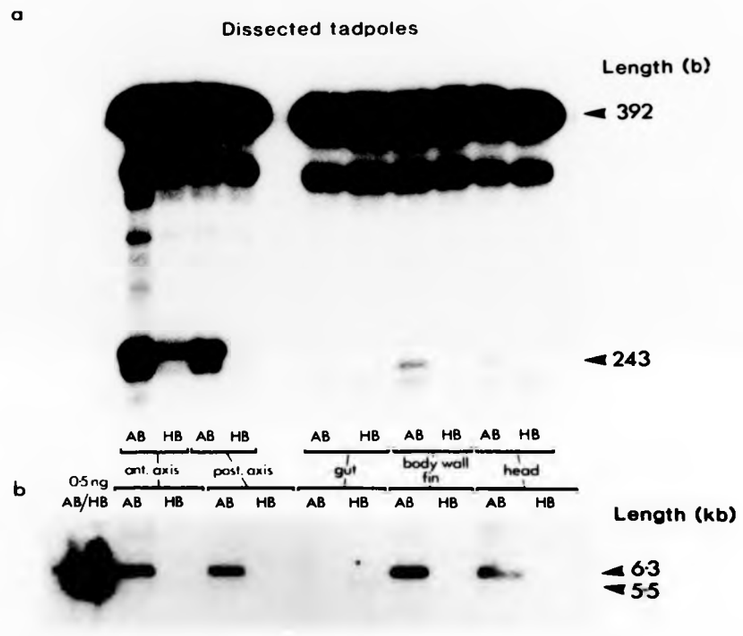


Figure 7.15 Localisation of Actbin and Hisbin Transcripts and DNA in Injected Stage 40 Swimming Tadpoles

X. laevis embryos were injected with the *Bam*HI-linearised actbin or hisbin clones, and allowed to develop into swimming tadpoles, before dissection. Nucleic acids were extracted from both actbin- (AB) and hisbin-injected (HB) tadpoles. RNA from two embryo equivalents of each sample was analysed using the globin-specific *Sau*3A fragment as an S1 probe. For DNA analysis, one embryo equivalent of each actbin-injected sample was digested with *Eco*RI, while DNA from hisbin-injected tadpoles was restricted with *Eco*RI and *Sst*I. A mixture containing 0.5 ng of linearised actbin plasmid and 0.5 ng of linearised hisbin plasmid was also digested with *Eco*RI and *Sst*I, as a control. These quantitites represent the approximate amounts of each clone injected into a single embryo. The resulting fragments were fractionated on a 0.8% agarose gel and then transferred to nitrocellulose. The filter was hybridised to the globin-specific *Hind*III/*Eco*RI fragment. For the actbin clone, a 6.3 kb fragment hybridises to the probe, while, for the hisbin, a 5.5 kb hybridising fragment is produced.

(a) S1 analysis of samples from dissected embryos. The 243 base protected fragment was not observed when RNA from uninjected tadpoles was analysed. (b) Southern blot of digested DNA from the same samples. The only signal from the hisbin-injected embryos is detected in the gut fraction.

DNA in Figure 7.15b). Consequently, a group of fertilised eggs injected with the hisbin clone at the same time as the actbin injections served as an inadequate control by the swimming tadpole stage. However, the continued localisation of actbin transcripts, whatever the distribution of DNA, suggests that expression is still being prohibited in non-muscle tissue.

As the actbin mRNA may well be relatively stable and its total level in injected embryos did not rise significantly after the neurula stage, the results on tadpoles discussed in this section might really reflect the persistence of correctly localised message synthesised at an earlier time in development. However, in Figure 7.15, it is clear that the posterior axis was expressing the actbin gene at a similar level to the anterior axis. The former develops from a very small group of cells in the tailbud tadpole, which is derived from the neural plate of earlier stages (Pasteels, 1939), a region that would be contained in the ectoderm fraction of a neurula dissection. Thus, most of the actbin transcripts of the posterior axis must have been synthesised at least after the neurula stage, and, almost certainly, mainly after the tailbud tadpole stage. Furthermore, the actbin DNA content of the region also appeared to be high, considering the small number of cells from which the tail originates. After the tailbud tadpole stage, these muscle cells are clearly synthesising new cellular components extremely rapidly, and are dividing repeatedly relative to many other cells in the tadpole. These processes may be reflected in the preferential accumulation of both actbin transcripts and DNA that seems to take place in the tail during its development. In other regions like the anterior axis, where such a significant change in transcriptional activity does not occur, it has not been possible to determine if the injected genes

are still being expressed in the tadpole.

7.2.6 The Temporal Control of Expression of the Actin and Actbin Clones Differs from that of the Hisbin

The expression of the injected clones was studied at different stages of development by S1 analysis. Low levels of transcripts from the unaltered actin gene were first detected at the mid-blastula transition (Figure 7.16a), which is 8 hours before the endogenous chromosomal actin gene is first expressed (see Section 5.2.4) and approximately the time at which other injected DNAs have been shown to be prematurely activated (Bendig and Williams, 1983, 1984; Etkin *et al.*, 1984). However, the total level of *X. borealis* actin message rose sharply from the late gastrula stage (stage 12½) in parallel with the appearance of endogenous *X. laevis* transcripts. The pattern of expression from the actbin clone was essentially identical to that of the actin clone (Figure 7.16c).

Low levels of transcripts from the hisbin gene were also observed from the MBT onwards (Figure 7.16c). The total amount of message showed a secondary rapid increase, but this occurred consistently by stage 10½ (the early gastrula), at least 3 hours before the surge of actin and actbin exrpession. By stage 12½, the time of the acceleration of transcription from the actin-derived clones, the level of hisbin mRNA was usually approaching its maximum.

Southern blots of restricted DNA from these temporal samples (Figures 7.16b and 7.16d) demonstrated that there was an initial increase in the amount of all the injected clones up to the mid-gastrula stage, as

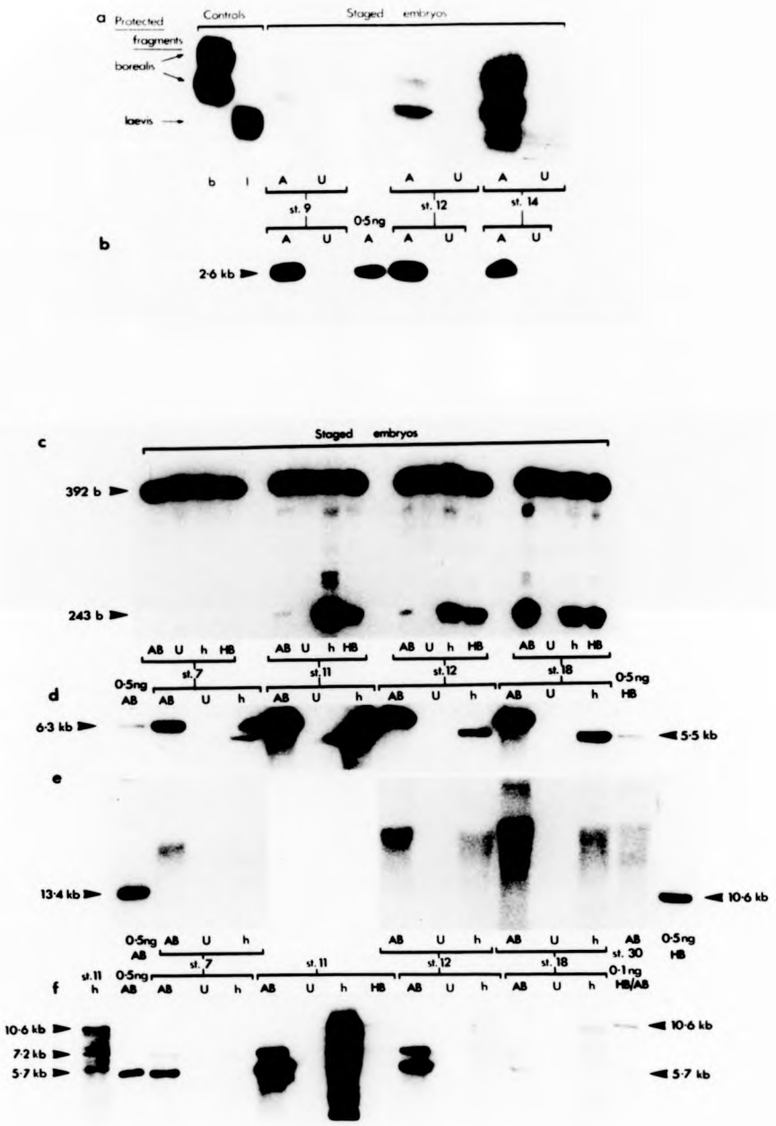


Figure 7.16 Temporal Control of Expression from the Injected Actin, Actbin and Hisbin Clones in *Xenopus* Embryos

(a) Differential S1 analysis of two embryo equivalents of RNA from actin-injected (A) and uninjected (U) *X. laevis* embryos at different stages of development. Stage 9 = late blastula; stage 12 = late gastrula; stage 14 = early neurula. Two embryo equivalents of *X. borealis* and *X. laevis* tadpole RNAs (tracks b and 1) were used as controls. *X. laevis* cardiac actin mRNA was first detected after stage 12 (see Section 5.2.4).
(b) Southern blot of one embryo equivalent of *Eco*RI-digested DNA from the same samples as in (a). The filter was probed with the insert of pXB-C1. One half nanogram of linearised actin DNA was digested with *Eco*RI and analysed as a control.

(c) S1 analysis of two embryo equivalents of RNA from embryos injected with the actbin clone (AB), the hisbin clone (HB), or a mixture of actin and hisbin clones (h). The globin-specific *Sau*3A fragment was used to assay RNA from these embryos at different stages of development. Stage 7 = mid-blastula; stage 11 = mid-gastrula; stage 12 = late gastrula; stage 18 = mid-neurula. Uninjected control embryos (U) were also used.
(d) Southern blots of digested DNA from the same samples as in (c). DNA from actbin-injected and uninjected embryos was digested with *Eco*RI, while DNA from hisbin-injected embryos was restricted with *Eco*RI and *Sst*I. The blot was hybridised to the globin-specific *Eco*RI/*Hind*III probe, which reacts with fragments from the actbin and hisbin clones of 6.3 kb and 5.5 kb respectively. One half nanogram of appropriately digested, uninjected actbin and hisbin plasmid DNA was blotted as a control.
(e) Southern blots of undigested DNA from the same samples as in (c), using the same probe as in (d). DNA from actbin-injected embryos at stage 30 was also examined. Samples containing 0.5 ng of *Bam*HI-linearised, undigested actbin and hisbin clones were blotted as controls.
(f) Southern blots of *Eco*RI-digested DNA from the same samples as in (c), using pAT 153 as a probe. One half nanogram of uninjected circular actbin, and a mixture of 0.1 ng of circular actbin and 0.1 ng of circular hisbin were digested and blotted as controls. Two hybridising fragments of 5.7 kb and 7.2 kb were produced by DNA from actbin-injected embryos. For hisbin-injected embryos the hybridising bands were 7.2 kb and 10.6 kb in length. As expected, the 5.7 kb and 10.6 kb bands migrated to the same position as the hybridising fragment from an *Eco*RI digest of the circular actbin and hisbin clones respectively. The other bands represent products from molecules in which the same ends of two different actbin or hisbin DNAs had been ligated together. DNA from embryos injected with hisbin and actin produced a more complex pattern, particularly as pAT153 cross-hybridises with the pEMBL vector of the actin clone. This pattern is more clearly seen on the left hand side, which shows a shorter exposure of the stage 11 track. The lengths of hybridising fragments are predictable from the maps of the two plasmids involved, if it is assumed that the three different molecules in the mixture have joined to each other in randomly orientated concatenates.

DNA from one embryo equivalent of each sample was used for the Southern blots in (d), (e) and (f).

previously discussed (see Section 7.2.1). From that time on, including the period when actin and actbin expression surged at the late gastrula stage, the levels of DNA slowly declined although significant amounts still usually remained at stage 40 (see Figure 7.14c). Therefore, although the changes in the total amount of plasmid present at different developmental stages could at least partly account for the profile of hisbin expression, they could not explain the main late gastrula surge seen with the actin and actbin clones. As the vast majority of plasmid molecules appear to be lost between the mid-gastrula and tadpole stages, it seems likely that most of the copies present in the gastrula are extrachromosomal (see Section 7.3).

The S1 analysis of blastula poly(A)$^+$ RNA from eightly pooled uninjected embryos (see Section 5.2.4) showed clearly that the chromosomal cardiac actin gene is not even expressed weakly before gastrulation. Therefore, the expression of injected actin and actbin clones from the mid-blastula transition until the late gastrula represents inappropriate regulation of the transformed genes. Primer extension has demonstrated that at least some of the prematurely synthesised transcripts are correctly initiated. Therefore, unlike some of the actbin mRNA in non-myotome tissues at later stages (see Section 7.2.4), these transcripts cannot have been formed by improper initiation from cryptic start sites in the clones.

Figure 7.16 also shows Southern blots of unrestricted DNA, and *Eco*RI-digested DNA. The blot of digested samples was probed with nick-translated pAT153. This probe hybridised to fragments incorporating the *Bam*HI-restricted end of the linearised plasmids that contained pAT153 vector sequence (Figure 7.16f). Essentially the gels confirmed the

results discussed in Section 7.2.1, showing that the injected plasmid had concatenated in random orientations, forming high-molecular-weight material from the cleavage stage onwards. These concatenates were still present in the tadpole (data not shown). The complex pattern of bands observed for embryos coinjected with actin and hisbin clones can only be accounted for if it is assumed that the three DNA molecules involved all become randomly linked with each other. Thus, even DNAs with different cohesive ends (the actin and hisbin molecules) are ligated together in embryos. It is known that, when single-stranded DNAs are injected into oocytes and embryos, the complementary strand is rapidly synthesised (H. R. Woodland, per. comm.). It is therefore quite possible that the cohesive ends of enzyme-digested molecules are filled in very quickly on injection and that the resulting blunt-ended duplexes participate in the ligation reaction.

During development, the relative level of background hybridisation rises in tracks containing DNA from injected embryos. This observation suggests that, although unrearranged concatenates remain the major molecular species, other products are formed, which result from rearrangement, deletion, inaccurate replication, or ligation to other DNAs. In some cases, specific bands that do not correlate with the size of any fragments from unaltered concatenates also appear with time (see Section 7.2.1). The persistence of unrearranged concatenates in different embryos becomes increasingly variable, particularly by the swimming tadpole stage. As discussed in Sections 7.2.5 and 7.3, this non-uniformity in behaviour may limit the usefulness of transformed *Xenopus* for gene expression studies at later stages of development.

7.2.7 Regions of the Embryo which Do Not Normally Express Injected Actbin Genes can be Induced to Express the Genes in Grafting Experiments

Gurdon *et al.* (1985b) have previously investigated the expression of chromosomal cardiac actin genes in isolated fragments of the late blastula, and in conjugates between the animal and vegetal regions from such dissected embryos (see Section 5.3). If these embryos were dissected into animal, equatorial and vegetal parts (see Figure 7.17e), and the parts were cultured in isolation, then only the equatorial region was shown to express the actin gene, when whole embryo controls had reached the neurula stage. However, if a conjugate was formed by placing an animal cap on top of a vegetal region, actin message was detected, because the ectodermal cells of the animal pole were induced to form mesoderm by the endodermal cells of the vegetal pole.

We repeated these experiments using embryos injected with the actbin and hisbin clones. When dissected regions were cultured in isolation (Figure 7.17a), actbin transcripts were only observed in the equatorial fraction, while the hisbin construct was expressed in both animal and equatorial regions. Southern blots showed that DNA localisation in these experiments was relatively variable between experiments (compare the distribution of actbin plasmid in Figures 7.17b and 7.17d). However, even when the plasmids were heavily concentrated in the equatorial region, the hisbin gene, unlike the actbin gene, was always expressed in other regions, particularly the animal cap.

When animal caps from injected embryos were grafted on to uninjected vegetal regions, they then expressed the actbin clone (Figure 7.17c).

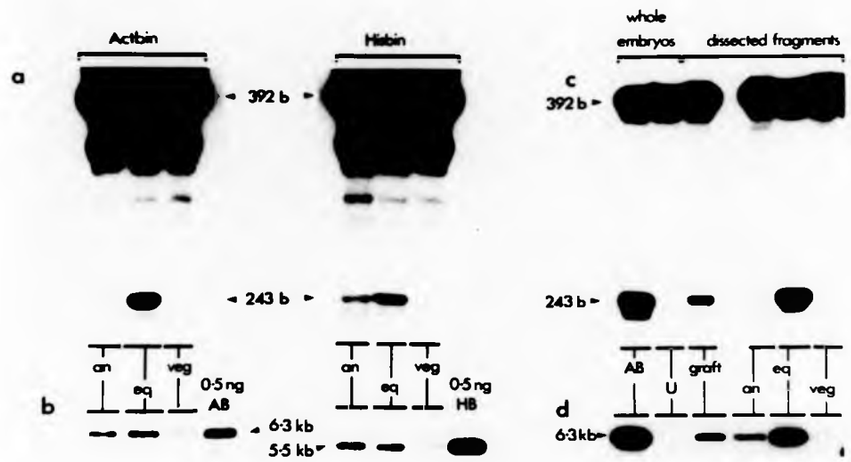


Figure 7.17 Actbin Expression is Activated in Animal Caps When They are Induced to Form Mesoderm

(a) Nucleic acids were prepared from isolated dissected fragments of actbin- and hisbin-injected blastulae, after they had been cultured until whole embryo controls were at stage 18. They were assayed by S1 analysis with the globin-specific *Sau*3A fragment.

(b) Southern blots of digested DNA from the same samples as in (a). DNA from actbin-injected samples was digested with *Eco*RI, and DNA from hisbin-injected samples was digested with *Eco*RI and *Sst*I. The filter was hybridised to the globin-specific *Hind*III/*Eco*RI fragment. One half nanogram of both the actbin and hisbin clones was treated in the same way and used for a control.

(c) S1 analysis of RNA from isolated dissected fragments of actbin-injected blastulae, as in (a). Actbin-injected (AB) and uninjected (U) whole embryo RNA is also included, as well as RNA from a graft between animal cells of an injected embryo and vegetal cells of an uninjected embryo. (d) Southern blots of *Eco*RI-digested DNA from the same samples as in (c) probed as described in (b).

RNA from two embryo equivalents of each sample was used for all S1 analyses, and DNA from one embryo equivalent was used for Southern blots. Abbreviations: an = animal pole cells; eq = equatorial region; veg = vegetal pole cells.

(e) Diagram of the blastula dissection and the animal-vegetal graft. This was drawn by Professor Hugh Woodland.

These results closely correlate with those of Gurdon *et al.* (1985b), and provide further strong evidence that expression of the injected clone is being regulated in an identical fashion to its endogenous chromosomal counterpart. They exclude the possibility that the injected DNA in presumptive non-mesodermal regions exists in some irreversibly non-expressible state.

7.2.8 Expression of the Actbin and Hisbin Clones in *Xenopus* Oocytes

Many workers have injected cloned genes into *Xenopus* oocytes in order to investigate the sequences, which are required for accurate and efficient transcriptional initiation. Several have examined genes that are not expressed in a tissue-specific fashion, like the herpes virus TK gene (McKnight and Gavis, 1980; McKnight and Kingsbury, 1982), sea urchin histone genes (Probst *et al.*, 1979) and the SV40 late genes (Wickens and Gurdon, 1983). In other cases, tissue-specific genes have been injected. For example, Bendig and Williams (1984) have examined the expression of *Xenopus* globin genes, while Bergsma *et al.* (1986) have recently investigated transcription from a chicken skeletal actin gene. Particularly when such tissue-specific genes are derived from *Xenopus*, it might be expected that they would not be expressed in the oocyte, in line with their chromosomal counterparts. Nevertheless, perhaps because the injected clones are extrachromosomal, or because they are in vast numerical excess in the oocyte, accurate and efficient transcription initiation is sometimes observed. However, the promoter sequences necessary for such genes to be expressed do not always include all the regions required for tissue-specific regulation (e.g. Bergsma *et al.*,

1986). In other examples, genes encoding rabbit β- globin (Rungger *et al.*, 1981), the *Xenopus* globins (Bendig and Williams, 1984), and chick ovalbumin (Wickens *et al.*, 1980) have been found to be transcribed, but, for the most part, not from the normal initiation site.

As the *Xenopus* cardiac actin gene had been demonstrated to be controlled properly in embryos, unlike the globin genes from the same species (Bendig and Williams, 1983, 1984), and generally no expression had been detected from aberrant initiation sites, it seemed plausible that it might not be transcribed in oocytes. To test this hypothesis, approximately 5-10 ng of circular or linearised actin and actbin clones were injected into *X. laevis* oocytes, in a preliminary study. In previous experiments, Harland *et al.* (1983) have found that the TK gene is expressed much more (i.e. several hundred-fold more) efficiently in a circular configuration than in a linearised form in oocytes (see Section 1.4.1), in direct contrast to the results in embryos that I have described. However, the size of this preference appears to be dependent on the injected gene. Probst *et al.* (1979), for example, have demontrated that the level of transcription from circular sea urchin histone genes in *Xenopus* oocytes is only about three times higher than that from linear molecules. No concatenation of linear molecules occurs in oocytes, and the DNA is generally unstable compared to closed circular material, perhaps accounting for the minor differences in expression of the two molecular species of the histone clone, but certainly not explaining the effects on the TK gene. Neither linear nor circular DNAs appear to be replicated in the oocyte.

In control experiments with the hisbin clone, we detected significant expression from linear as well as supercoiled DNA, even though most of

the former had become degraded within 24 hours. By S1 analysis, using the *HinfI/BstNI* probe from the first coding exon of the cardiac actin gene, I was unable to detect any transcription from coinjected circular or linear pXB-C3.1 (data not shown). Primer extension with the 15 base oligonucleotide complementary to RNA encoded by the first exon has also failed to show any expression. However, recently Dr. Gareth Cross has repeated the primer extension work with a 22 base oligonucleotide complementary to a longer stretch of RNA encoded by this exon (see Figure 5.3). Using this more sensitive assay, there is some evidence that a basal level of actin transcripts is present in oocytes injected with circular pXB-C3.1 Further studies will be required to confirm this observation.

Although no expression has been detected from *Bam*HI-linearised actbin molecules in oocytes, analysis with the globin-specific S1 probe, which assays for correct 3' processing of transcripts, has demonstrated that circular actbin genes produce an abundant level of RNAs with normal 3' ends. However, primer extension, using the globin-specific primer, indicated that these RNAs were extremely heterogeneous, and did not appear to be initiated from the actin promoter. Therefore, it seems likely that cryptic initiation sites, perhaps in the vector sequences of the clone, may be responsible. These might then be separated from the actin gene by linearisation with *Bam*HI, accounting for the lack of transcription from the linear form. Such sites may be related to those which appear to be active in the non-myotome regions of injected embryos in at least one experiment (see Section 7.2.4).

SECTION 7.3 DISCUSSION

In this chapter, I have described for the first time the correct spatial regulation of expression from a cloned gene, when it is injected into fertilised *Xenopus* eggs. In the discussion section, I firstly consider the general applicability of using micro-injected *Xenopus* embryos for studies of early developmental gene control. The important characteristics of this transformation system are discussed. I then concentrate on the specific results for the actin gene, outlining future avenues for further investigation.

The *Xenopus* transformation system has a number of advantages, especially for the study of early developmental gene regulation, compared to the established transgenic systems in the mouse and *Drosophila*. The biological part of the experiments, involving injection of embryos and their development to the mid-neurula stage takes less than one day. The size of the embryos allows relatively rapid and accurate dissection into the germ layers, and each embryo contains sufficient FNA to be analysed individually if necessary. Thus, once made, new constructs and deletions of the original gene can be tested for localised expression within a few days.

In this transformation system, the cytoplasm of *Xenopus* embryos is micro-injected with DNA, but clearly the correct transcriptional regulation of introduced genes suggests that some of the DNA must later reach the nucleus. How does this translocation occur? Two studies have considered the subsequent cellular localisation of injected DNAs in *Xenopus*. Forbes *et al*. (1983) investigated the fate of injected lambda

DNA in unfertilised eggs by autoradiography, coupled with fluorescence and electron microscopy. They found that a double bilayer membrane, which contained pores and was associated with nuclear lamin proteins, formed around the DNA. These nucleus-like structures have recently been observed in embryos injected with lambda (Shiokawa *et al.*, 1986), where they surround the cleavage nuclei. It seems likely that they fuse with the cleavage nuclei during early development, as the latter have been shown to contain injected DNAs (see Gurdon and Melton, 1981). However, it is not yet known if the miniature nuclei must fuse in this way if their contents are to persist to later developmental stages, like the neurula and tadpole.

In the experiments described in this thesis, linear DNAs have usually been injected. They are ligated into long, randomly orientated concatenates, which replicate without significant rearrangement, increasing in numbers until the gastrula stage. By contrast, the supercoiled forms of the same plasmids persisted poorly in the embryo, and no net accumulation of DNA was observed at any stage of development. Studies by Etkin *et al.* (1984), Krieg and Melton (1985) and Busby and Reeder (1983) have also reported DNAs that do not replicate efficiently as circular molecules in *Xenopus* embryos. By contrast, other supercoiled DNAs have been amplified 50- to 200-fold following injection (Bendig and Williams, 1983; Etkin *et al.*, 1984; Rusconi and Schaffner, 1981). The reason for these differences is not clear, but, at least in my work, as the same molecules can be highly replicated if they are first linearised, the phenomenon cannot be explained by the existence of "poison" sequences in the plasmids, which inhibit replication. Such sequences have been postulated by Lusky and Botchan (1981) to explain the poor replication of recombinants containing pBR322 in transformed

eukaryotic cells.

All the studies of transcription from the actin, actbin and hisbin clones have been accompanied by analyses of DNA content in the same samples. Although the distribution of DNA in developing embryos does vary somewhat between experiments, frequently it is remarkably consistent within the same batch of embryos. A particularly important point is that the injected plasmids become widely distributed in virtually all cases, whether they are injected in the animal or vegetal poles of fertilised eggs. This system would therefore appear to be suitable for the study of genes expressed specifically in other *Xenopus* tissues during early development. A good candidate would be a gene like the recently isolated epidermal keratin gene (Jonas *et al.*, 1985), which is an early terminal product of epidermal differentiation. In any injection experiments of this kind, it cannot, of course, be estimated to what extent DNA content varies between different cells in the same tissue, but the success of the work on the cardiac actin genes suggests that this is not a critical factor.

In Section 1.2.2, I briefly discussed the possible use of anti-sense transcripts to block the expression of regulatory genes in vertebrates in order to ascertain the function of their products. This is one approach available for investigating developmental regulation in an organism which cannot realistically be extensively searched for developmental mutants. As injected DNA is distributed in all parts of the embryo, one feasible method for producing anti-sense transcripts in all regions is to introduce a construct in which the anti-sense transcription unit is fused to a highly active non-tissue-specific promoter. Such a technique might overcome any problem of poor RNA

stability, which would make direct injection of anti-sense transcripts inviable.

As yet, it has not been demonstrated that the injected genes in this system become integrated into chromosomal DNA. It is probable that most of the copies present at the mid-gastrula stage are associated in long unintegrated concatenates, because they are slowly lost throughout development to the swimming tadpole stage. However, an alternative, but less likely, explanation could be that integrated concatenates are excised by recombination events over the period from stage 12 to stage 40, during which only about five cell cycles take place.

A number of previous reports have suggested that some injected DNA does indeed become integrated in *Xenopus* embryos (Rusconi and Schaffner, 1981; Etkin *et al.*, 1984). The main lines of evidence are the persistence of injected DNA as high-molecular-weight material in adult frogs and the generation of putative junction fragments, containing both plasmid and chromosomal sequences, upon digestion with restriction enzymes. The persistence of unintegrated episome-like injected molecules in the adult frog in a study of injected *Xenopus* vitellogenin gene derivatives (Andres *et al.*, 1984) shows that the retention of plasmids in adulthood does not necessarily require attachment to the chromosomes. Putative junction fragments could well represent rare rearranged molecules and they must be cloned to provide definitive proof of integration, as has been done in injected sea urchin embryos (Flytzanis *et al.*, 1985). In the studies which have considered the fate of DNA in adult frogs it has been found that only a few copies per haploid genome remain after metamorphosis, even though about the same number of molecules were initially injected as in the experiments

performed in my project. As there are at least two thousand copies of the actin or actbin clones for every endogenous chromosomal copy of the actin gene at the early neurula stage, this further supports the proposal that most are unintegrated, and are gradually lost.

The disappearance of injected DNAs appears to occur at variable rates in different embryos, particularly during later tadpole development. Furthermore, the cell types that retain the highest levels of the injected plasmid in each pool of tadpoles are not necessarily the same (see Figures 7.14 and 7.15). Consequently the resulting mosaic tadpoles are not so well suited for studies of gene expression as embryos at earlier stages. Similar investigations to the ones described in this thesis may therefore be less plausible if the gene is only expressed after neurulation. Indeed, if more extensive studies of actin gene regulation were to be considered in swimming tadpoles, it would probably be advisable to coinject a non-tissue-specific gene (which would form concatenates with the actin clone) in order to control more rigorously for the non-uniform and unpredictable distribution of DNA.

The *Xenopus* transformation system described here is therefore quite unlike the transgenic systems of *Drosophila* and the mouse. In the latter, the exogenous DNAs are integrated into the genome and extrachromosomal material does not persist, so only expression of chromosomal copies is investigated. Indeed, in *Drosophila*, transgenic offspring usually contain only a single inserted copy in each cell (see Section 1.3), compared to the thousands of extrachromosomal copies in *Xenopus* embryos.

In the mouse, tandem arrays of the injected genes become integrated into

the genome, usually at a single site before first cleavage. The resulting transgenic individuals frequently express the introduced gene at very variable levels (see Palmiter and Brinster, 1985), almost certainly because of the position effects of different adjacent chromosomal sequences in each animal. Generally the *Xenopus* system does not appear to suffer from this problem of variability, and this suggests that expression from the vast excess of unintegrated gene copies may be the major source of transcripts during the early stages of development examined. Therefore transformed *Xenopus* embryos may be analagous to transient expression systems in cell lines, where the expression of non-chromosomal genes is also investigated (see Section 8.2).

A similar system to that in the frog has been developed in sea urchins. Their embryos, which also develop rapidly and independently, are injected in the cytoplasm with linear DNA molecules Long extrachromosomal concatenates are formed in early development (see Davidson *et al.*, 1985), and a few copies become integrated into the genome. However, although correct temporal expression of an injected actin gene has been demonstrated in sea urchins, correct tissue-specific expression has yet to be shown in these embryos, which are less easily dissected, because of their small size.

Why has the work on actin gene regulation described in this chapter been so fruitful in comparison to previous studies of tissue-specific gene expression in injected *Xenopus* embryos? In one other successful report, Krieg and Melton (1985) have recently shown correct temporal regulation of the injected gastrula-specific gene, GS17, (see Section 7.1) in *Xenopus*. They suggested that only genes normally expressed very early

in embryonic development, like GS17, might be correctly controlled in these experiments, while genes which are transcribed much later would not necessarily behave properly, as has been found for globin (Bendig and Williams, 1983, 1984). Certainly, in our experiments, the increasingly uneven distribution of DNA in the tadpole may make it more difficult to interpret expression data from these later stages. However, the evidence for localisation of large numbers of actbin transcripts to the tail of the swimming tadpole, a region which is largely formed between stage 30 and stage 40, suggests that specific expression can still take place at this time (see Section 7.2.5).

The relatively low amounts of globin transcripts detected in the gastrula and neurula by Bendig and Williams (1983, 1984), long before the endogenous gene is activated, could well represent a type of 'leaky' expression. As discussed in this chapter, basal levels of correctly initiated *X. borealis* actin and actbin transcripts are also found after the mid-blastula transition, before the *X. laevis* chromosomal cardiac actin gene is first transcribed in the late gastrula. However, preliminary evidence has suggested that, unlike transcripts from the hisbin clone, these actin-related RNAs may be preferentially localised to the equatorial region of the late blastula. This region includes all the cells which will normally form the myotomes, and so this result may indicate that mesodermal cells already contain some muscle-specific gene regulators by this stage of development. These would be able to weakly activate extrachromosomal (but not chromosomal) muscle-specific cardiac actin genes. We are presently repeating this experiment on localisation, but it is particularly difficult to investigate the phenomenon, as the total amount of transcription from injected genes in the blastula is so low that maximal assay sensitivity is required to

demonstrate a significant level of cell-type-specificity.

The simple explanation that the abnormal behaviour of globin genes in *Xenopus* embryos and tadpoles is related to the fact that it is only expressed later in development, unlike GS17, is not therefore fully supported by the data from the injected actin gene and its construct. Although the actin-derived clones may be expressed inappropriately in the blastula and early gastrula, just like the globin genes, they are regulated properly at later stages. An alternative hypothesis, which might account for the incorrect globin expression at these later times, is that the vector sequences in the clone must be removed prior to injection before expression can be regulated correctly. Such a requirement has been observed for globin (Chada *et al.*, 1985) and a number of other genes (Shani, 1986; Krumlauf *et al.*, 1985) in transgenic mice.

However, in all the experiments with actin described in this chapter, vector DNA was still present, and removal of this prokaryotic sequence has recently been shown to have no effect on the spatial distribution or quantity of transcripts from the injected clone (data not shown). In contrast, Shani (1986) has shown that vector sequences alter the distribution and quantity of message from a skeletal actin gene construct in adult transgenic mice. It may therefore be the case that prokaryotic DNA does not generally affect expression of transformed genes in *Xenopus* embryos in the same way as in mice.

A direct comparison of relative transcription rates from the injected and chromosomal actin genes can be made, if it is assumed that the *X. borealis* and *X. laevis* cardiac actin mRNAs have similar stabilities.

The injected clone is expressed at highest relative levels just after the chromosomal genes are activated in the late gastrula (see Figure 7.16a). At this time there are up to ten times more transcripts from the clone than from the endogenous genes. However, the injected genes are in at least a two thousand fold excess in the future myotome region, if it is assumed that the ratio of injected to chromosomal genes is the same in this region as it is in the whole embryo. The relative amount of actin mRNA made from the plasmid falls after this period, but this decline is accompanied by decreasing levels of persisting injected DNA and increasing amounts of chromosomal DNA. However, even in the stage 30 tadpole, the injected genes clearly outnumber the chromosomal actin genes in the somites by a considerable margin, although the two contribute roughly equal amounts of RNA. Mohun *et al.* (1984) calculated that between 10^7 and 10^8 cardiac actin transcripts are present in a normal *X. laevis* neurula. It would appear that a similar number of *X. borealis* mRNA molecules are generally produced from the injected genes by this stage.

The use of the actbin and hisbin constructs allowed direct comparisons of localised expression from a muscle-specific and a non-tissue-specific gene. Typically less than 10% of transcripts from the hisbin gene were found in the myotome fraction of neurulae, partially reflecting the preferential distribution of injected DNA to other regions. However, at least 90% of actbin transcripts were localised to the myotomes, thus showing roughly a 100-fold preference for expression in this tissue.

Despite the vast excess of injected genes, we have never found any reduction in expression from chromosomal cardiac actin genes in these experiments. The rate of transcription per gene is much lower for the

injected DNA than for the chromosomal genes, but expression from the former increases when more DNA is injected (see Section 7.2.2). Injections start to become toxic when more than 1 ng is injected per embryo. It is unclear if all the injected genes are being transcribed at low levels in these experiments, or if only a small proportion are selectively activated. Possibly other genomic sequences flanking the insert of pXB-C3.1 are required to raise the level of expression to that of the chromosomal genes, even though transcripts from the clone are properly localised without these sequences. For example, a nuclear matrix association region and perhaps an adjacent enhancer (see Section 1.4.2) may be absent from the injected gene. If the expressed genes are not integrated, this might also affect the relative transcription rate.

The use of primer extension, coupled with a comparison to S1 analysis, has demonstrated that properly localised actin and actbin message is correctly initiated in these experiments. In the one case where there was sufficient actbin RNA present to be easily detected in other regions, it was clear that most of it was incorrectly initiated. Aberrant initiation of this kind may limit the usefulness of techniques like *in situ* hybridisation to further characterise the localised expression of injected genes in *Xenopus* embryos, and it may always be necessary to use assays, which detect the position of the 5' end of the message.

In addition to demonstrating proper localisation, I have also shown that the surge of expression at the late gastrula stage coincides with the temporal activation of the chromosomal genes. As discussed previously in this section, the anomalous 'leaky' transcription before this time may in fact still be localised and is currently under further

investigation.

At present, transformed froglets are developing in the laboratory and these will be examined in due course. It is hoped that specific expression may be detected in the hearts of such animals, which contain high levels of endogenous cardiac actin transcripts. The adult heart, unlike the embryonic heart, does not express the skeletal actin gene, while adult skeletal muscle does not express the endogenous cardiac actin gene (see Section 5.2.3). Dr. Mark Boardman has recently constructed a skeletal actin/β-globin fusion (the 'skatbin'; See Section 7.2.3), which has a very similar structural organisation to the actbin clone. The expression of the skatbin is regulated in a tissue-specific fashion in *Xenopus* embryos. It is therefore envisaged that, if injected DNA does persist in the hearts or skeletal muscle of metamorphosised frogs, cardiac actin and skeletal actin constructs could be coinjected in order to ascertain if they are differentially regulated in these tissues during later development.

It will also be interesting to discover if any of the injected material is stably integrated and therefore transferred through the germ line. However, the long period of time required for frogs to reach maturity (about two years) means that F1 and F2 transgenic progeny could not realistically be considered for the detailed study of gene regulation in the same way as in transgenic flies and mice.

In this discussion section, I have considered the general properties of the DNA injection system in *Xenopus* embryos and its suitability for studies of genes activated early in development. I have also

highlighted those areas in which further work on the regulation of the actin and actbin clones is still required. These include the studies of expression in the blastula, at later stages of tadpole development and after morphogenesis.

However, there is considerable scope to extend the investigations further. It should be possible to delimit the sequences necessary for correct developmental regulation of expression from the cardiac actin gene fairly rapidly. In addition, it is planned to use the oocyte as an alternative to *in vitro* transcription systems in order to study the molecules which specifically stimulate the expression of the actin gene in muscle cells. Both of these topics are considered in more detail in the final chapter.

CHAPTER 8

PROSPECTS

SECTION 8.1 INTRODUCTION

This discussion concentrates on two lines of research, which will form the basis for future work on the developmental regulation of the *X. borealis* cardiac actin gene. The first is the delineation of gene sequences, which are necessary for correct tissue-specific regulation. Similar experiments have already been performed with other muscle-specific actin genes in transformed myogenic cell lines and the results from these studies are considered. In the second, it is hoped that the *trans*-acting factors, which interact with such regulatory sequences, might be identified and cloned. Many groups are currently attempting to isolate cell-type-specific factors that control expression of a number of well-characterised genes. A few factors that regulate specific genes transcribed by RNA polymerase II have been identified and partially purified (Wu, 1984b; Emerson and Felsenfeld, 1984; see Section 1.4.2). However, up until now, it has not been possible to clone them, partly because they are present at such low concentrations. The approach described here may be able to circumvent the problem by directly assaying for the RNA, which encodes the actin-specific factors.

SECTION 8.2 THE DELINEATION OF SEQUENCES RESPONSIBLE FOR THE DEVELOPMENTAL REGULATION OF THE SARCOMERIC ACTIN GENES

A number of vertebrate skeletal and cardiac actin genes have already been cloned and sequenced. References for these studies are given in Section 1.5. The sequences of the 5' flanking regions of these genes have been compared in an attempt to identify motifs, which might be important in transcriptional regulation and would therefore be conserved. Man and mouse, and, to a lesser extent, chick, all share significant homologies in the 5' flanking sequences of their cardiac actin genes (Minty and Kedes, 1986). Similarly, elements have been identified that are conserved between the upstream regions of chicken and rodent skeletal actin genes (Ordahl and Cooper, 1983; Hu *et al.*, 1986; Bergsma *et al.*, 1986). However, in general, the conserved sequences of the cardiac and skeletal genes are not common to them both.

Further experiments are clearly required to show that regions of homology actually have a functional role in regulation. Recent studies, mainly with transformed myogenic cell lines and cell cultures, have started to address this question. These myoblast cells can be induced to differentiate into fused myotubes *in vitro*, leading to the activation of chromosomal muscle-specific genes like the sarcomeric actin genes. The results of the transformation experiments are discussed briefly in this section and relevant sequence comparisons to putative regulatory elements are made for the *X. borealis* cardiac actin gene.

Most of the reports have described transient expression assays, in which the introduced genes remain extrachromosomal and are therefore not

subjected to the position effects of adjacent genomic sequences. The advantages of this approach over stable transformation systems were demonstrated in one of the earliest studies, when Melloul *et al.* (1984) used rat L8 cells that had been stably transfected with a rat skeletal actin fusion construct. Although gene expression was induced on differentiation by up to 50 fold, there was considerable variation between clones, presumably because of position effects. Such inconsistencies make it difficult to interpret data from deletion analyses, and so subsequent investigations of this nature have usually been performed in transiently transformed cells.

A number of studies of this kind have described tissue-specific control of skeletal or cardiac actin gene expression. They have demonstrated that transformed, differentiated myogenic cell lines and cell cultures, but not other cell types, contain high levels of transcripts from the introduced genes (e.g. Seiler-Tuyns *et al.*, 1984; Grichnik *et al.*, 1986; Minty and Kedes, 1986). Furthermore, sarcomeric actin genes have been shown to be controlled normally in heterologous systems (e.g. Seiler-Tuyns *et al.*, 1984; Minty and Kedes, 1986) suggesting, perhaps not surprisingly, that regulatory mechanisms are conserved between different vertebrates. As expected, non-muscle-specific genes like the β-cytoskeletal actin gene were expressed at similar levels in transformed myogenic and non-myogenic cells (e.g. Melloul *et al.*, 1984; Grichnik *et al.*, 1986.

However, in some examples, particularly in those using the mouse myogenic C2C12 cell line transformed with cardiac actin genes, there was already strong expression before the cells had differentiated and activated their endogenous sarcomeric actin genes. A recent study by

Minty *et al.* (1986) has confirmed that it is the C2C12 cell line and not the transformed gene, which is responsible for this effect, as the rat L8 muscle cell line did not behave in this fashion when the same constructs were introduced into it.

There are two important conclusions from these results. Firstly, transformed genes, which do not have the same history as endogenous chromosomal genes, are not always treated in the same way as the latter by the transformed cell. Of course, this is possibly because the cloned genes lack some of the sequences required for normal regulation. However, an alternative hypothesis is that either their extrachromosmal state (and consequent association with different neighbouring sequences to the endogenous genes) or their different historical background leads to their escape from a regulatory pathway, which inhibits the chromosomal actin genes. The second conclusion is that, as some 'undifferentiated' cell lines can activate transformed muscle-specific genes, they may in fact have already undergone some of the early muscle differentiation processes, although these changes alone are insufficient to lead to chromosomal actin gene expression.

Therefore, it may be difficult to identify precisely the developmental stages of muscle differentiation that myogenic lines represent. This is an additional reason for using transgenic animals to confirm any results derived from cell lines, aside from the increased stringency of testing all the cell types in the whole organism. If expression from injected *Xenopus* cardiac actin genes is localised to the presumptive muscle tissue of the blastula (see Section 7.3), this effect would be comparable to the premature activation of cardiac actin genes in C2C12 cells, and would serve as a useful marker for the earliest stages of

muscle differentiation *in vivo*.

Some investigations in the past year have started to delimit important functional sequences to specific regions in the striated muscle actin genes. Grichnik *et al*. (1986) found that a 411 bp fragment flanking the 5' end of the chicken skeletal actin gene, but not including the transcription initiation site, could amplify expression from a fused bacterial CAT (chloramphenicol acetyltransferase) gene by 9-15 fold on differentiation of transformed chicken primary myoblast cultures. The same group has since delineated the regulatory sequences to approximately the first 200 nucleotides of the upstream region (Bergsma *et al*., 1986).

The sequence of this upstream region in the chicken skeletal actin gene was compared with other available skeletal actin sequences by this group, in order to search for possible conserved regulatory elements. In addition to the 'TATA' box and 'CAAT' box, which are common features of many genes transcribed by RNA polymerase II (see Section 1.4.2), they recognised an inverted 13 bp repeat spanning between positions -80 and -139 which might play a regulatory role. The downstream repeat is part of the 20 bp sequence, which had previously been reported to be highly conserved in the chicken and in rodents (Ordahl and Cooper, 1983; Hu *et al*., 1986). Furthermore, it was suggested by Bergsma *et al*. (1986) that the upstream repeat in the chicken gene is also conserved in rodents and that, in addition, both repeats are present in the 5' flanking regions of the smooth and cardiac actin genes of the chicken. Jaynes *et al*. (1986) have recognised another different repeat sequence that is not only found in multiple positions in the 5' flanking regions of the rat

and chicken skeletal actin genes, but is also homologous to other copies in the mouse muscle creatine kinase gene. All of these motifs are potential regulatory elements for skeletal actin genes, but further studies must be performed in order to demonstrate directly that they are functionally important.

Bergsma *et al.* (1986) also showed that the chicken skeletal actin gene was transcribed at relatively high levels in *X. laevis* oocytes. Only 107 bp of 5' flanking sequence were required, suggesting that this region contains elements that are necessary for non-regulatory, transcriptional control. However, such evidence does not exclude the strong possibility that tissue-specific regulatory sequences are also present in the same region.

The regulation of human cardiac actin gene expression has also recently been studied in transformed cell lines (Minty and Kedes, 1986). It was found that 485 bp of upstream sequence from this gene (as well as the first exon and 24 bp of the first intron) directed high-level transient expression (40-50 fold higher than deleted constructs) of a CAT fusion gene in differentiated mouse C2C12 cells, but not in a mouse fibroblast line. Minty and Kedes proceeded to delete the 5' flanking sequence and identified two upstream regions, a distal one starting between -443 and -395, and a proximal one starting between -177 and -118, which appeared to be largely responsible for the tissue-specific effect. However, a fragment containing sequences from -440 to -47 only stimulated expression from a fused SV40 promoter by 4 to 5 fold in C2C12 myotubes. It is therefore clear that these regions alone will not fully activate any nearby promoter sequence in muscle cells, although this may only be because the promoter must be positioned at a precisely defined distance

from the regulatory elements.

In sequence comparisons between the human, mouse and chicken cardiac actin genes, Minty and Kedes (1986) found many homologies in the 5' flanking region. In particular, the sequence 5'-CC--A-rich--GG-3' (designated CCArGG) is present in a number of conserved positions in each gene, and the CCArGG element at any one position has a highly conserved internal A-rich sequence. Furthermore, similar regions were identified within 50 to 150 nucleotides upstream of the cap site in skeletal actin genes (one of which is part of the upstream repeat recognised by Bergsma *et al.* [1986]), and also in other muscle-specific genes, like the myosin heavy chain gene and the troponin T gene. Not surprisingly, however, considering the small size of this relatively non-specific consensus sequence, it has also been found in positions flanking non-muscle-specific genes. For example, the human, rat and chicken β-cytoskeletal actin genes all contain such a sequence in their 5' flanking regions. Nevertheless, this does not exclude CCArGG sequences as important regulatory elements in muscle, it merely suggests that other different regions are also required to account for strict tissue-specificity. Indeed, in a similar example, the regulatory octamer sequence of immunoglobulins (see Section 1.4.2) has also been found in other genes that are not specifically expressed in lymphoid cells.

Dr. Gareth Cross has sequenced the first 170 bases upstream of the cap site in the *X. borealis* cardiac actin gene. Generally the sequence shows less homology with the equivalent regions in the cardiac actin genes of warm-blooded vertebrates than the latter do with each other. A putative 'TATA' box and 'CAAT' box are found in roughly the correct

location relative to the cap site. More importantly, the two CCArGG elements, which should be in this region are highly conserved (see Figure 8.1), implying that they have indeed been maintained over considerable evolutionary distances.

In summary, although deletion analysis has shown that the upstream regions of the striated muscle actin genes can direct tissue-specific expression, it has not yet been possible to identify the exact sequences responsible at a functional level. Sequences, which are conserved from frogs to man, may well indicate the regions in which to look, but the results so far suggest that multiple elements probably play a role, which may make their elucidation much more difficult.

The first report of actin gene regulation in transgenic mice was made this year by Shani (1986). He introduced the rat skeletal actin/human globin gene fusion construct which Melloul *et al.* (1984) had already used successfully in rat L8 cells. It still contains the front two-thirds of the actin gene coding sequence. Tissue-specific expression was observed, which was dependent on the removal of vector sequences from the clone. Furthermore, expression from the construct was correctly developmentally regulated, as it was reduced in the adult heart relative to the neonatal heart, in the same way as the expression from the endogenous skeletal actin gene in rodents (Mayer *et al.*, 1984; Minty *et al.*, 1982).

In the studies described in this thesis, sequences upstream of the third exon have been shown to be sufficient to direct proper tissue-specific expression from the cardiac actin gene. Using the subtle differences

Figure 8.1 Comparison of CCArGG Elements in the Cardiac Actin Genes of the Chicken, Man and *X. borealis*

```
Element 3   -124  CCTTGG--CCATTCATGGCC          chicken
                  ******  ****  ******
            -157  CCTTGGCTCCATGAATGGCC          human
                  ***   ****** ******
            -141  CCT--ACTCCATTAATGGCT          X. borealis

Element 4    -98  GCGCCGAGCCGGCCAAATAAG-AGAAGGTGGC  chicken
                  ***   **  * *********   ********
            -119  GCGA--AGGGGACCAAATAAGG-CAAGGTGGC  human
                   * *  **   ********** * ***  **
            -105  CCCA-AAGCT-ACCAAATAAGGGC-AGGAAGC  X. borealis
```

The 5' flanking sequences from the chicken (Eldridge *et al.*, 1985), human (Minty and Kedes, 1986) and *X. borealis* (G. S. Cross, per. comm.) cardiac actin genes have been compared. Although the *X. borealis* gene shows only limited overall homology to its chicken and human counterparts, two CCArGG sequences (see text) described by Minty and Kedes (elements 3 and 4) are found in *X. borealis*. Identical nucleotides are marked with asterisks and the CCArGG elements are underlined. The position of the most upstream nucleotide in each sequence is given on the left. The other conserved elements that have been recognised in chicken and man are probably too far upstream to be included in the 170 bases of *X. borealis* upstream region that has currently been sequenced.

Like Minty and Kedes, I have also compared the sequences directly adjacent to the CCArGG elements. Although the CCArGG sequences themselves are very highly conserved, neighbouring regions display many more mismatches, suggesting that only the specific sequences within the CCArGG motifs are important.

It should be noted that Chang *et al.* (1985) have also published the sequence of the chicken cardiac actin gene. It differs somewhat from the sequence of Eldridge *et al.* (1985) in element 4, particularly upstream of the CCArGG sequence, where it shows increased homology to the human gene.

between the cardiac actin transcripts of two very closely related *Xenopus* species, it has been possible to make an initial test of the complete cloned gene. Thus, it has been subsequently demonstrated that removal of regions downstream of the second intron has no effect on the distribution or total level of the resulting transcripts. Of course, it cannot be excluded, however, that other sequences outside the original clone might influence regulation.

Further localisation of the elements which activate the actbin construct in embryonic myotomes will involve standard techniques like deletion analysis. It will be important to remove the remainder of the actin gene transcription unit from future constructs to ascertain whether the regulation occurs solely at the transcriptional level, or whether post-transcriptional events also play a role. One possible problem with these experiments is that sequence deletions might lead to altered replication or localisation of injected extrachromosomal DNAs. Of course, Southern blots of nucleic acid preparations will partly control for these changes. However, if relatively subtle differences in gene activity are to be recognised, it may be necessary to coinject the actbin gene, for example, as well as the manipulated construct. As the two molecules will form concatenates with each other, any changes in DNA replication or distribution should affect both genes, and hence the actbin will serve as a useful internal control.

If the sequences responsible for proper regulation are found to be complex, an alternative approach, which has been suggested recently, would be to coinject polymers containing tandem copies of putative regulatory elements. Xiao and Lis (1986) have reported that such a polymer, which they constructed from forty copies of the pair of

regulatory elements in the *Drosophila hsp*70 gene, specifically inhibits the transient expression of cloned heat shock genes in cotransfected cell lines. Such a competition assay for putative factor-binding regions might therefore be suitable to study the expression of the extrachromosomal transformed genes in *Xenopus* embryos.

Even if it is not possible to localise the precise position of important regulatory sequences in the cardiac actin gene, it may still be plausible to identify factors, which interact with these sequences. A novel approach to such studies, which is based on the use of the *Xenopus* oocyte as an expression system, is described in the following section. By making use of both the transcriptional and translational capacities of this cell type, it is conceivable that the RNAs encoding the *trans*-acting regulatory factors will be cloned, without previous purification of the factors themselves.

SECTION 8.3 THE ISOLATION OF *TRANS*-ACTING FACTORS, WHICH INFLUENCE CARDIAC ACTIN GENE EXPRESSION

The *Xenopus* oocyte is a particularly useful type of cell in which to study the processes of gene expression. Its large size makes microinjection of both DNA and RNA relatively straightforward. In addition, the oocyte contains a large maternal reserve of transcriptional and translational apparatus, in preparation for the rapid cell division that will take place after fertilisation. Therefore, large numbers of injected genes can be packaged into chromatin in the oocyte and may be transcribed in this form (Gurdon and Melton, 1981). Furthermore, numerous groups have injected RNAs into oocytes, since the first experiments by Lane *et al.* (1971), and have found that they are efficiently translated relative to *in vitro* translation systems. The resulting products are often correctly processed and biologically active (e.g. Reynolds *et al.*, 1975).

Preliminary studies on the expression of the cardiac actin gene have suggested that it is transcribed weakly, if at all, in oocytes. Although this might be expected as the endogenous chromosomal gene is not active in this cell type, such correlation between chromosomal and extrachromosomal gene expression has not always been observed in the oocyte (Bendig and Williams, 1984). If the results for the cardiac actin gene are confirmed, then it should be possible to coinject extracts from cells that normally express the gene (i.e. from the neurula or the adult heart) in order to activate transcription. Such a complementation assay has been successfully used to show that the Ela

gene product synthesised in bacteria can stimulate the adenovirus E3 promoter in *Xenopus* oocytes (Ferguson *et al.*, 1984). In addition, if a nuclear extract from chicken erythroid cells is added to a histone-supplemented extract from the germinal vesicles of *Xenopus* oocytes, the mixture will package a cloned β-globin gene into chromatin that contains a DNase I hypersensitive region also found in active genes *in vivo* (Emerson and Felsenfeld, 1984; see Section 1.4.2). Thus, the chromatin packaging apparatus of the oocyte can be influenced by the introduction of cell-type-specific nuclear factors from other cells.

Although nuclear extracts consisting of muscle proteins might be tested in this way for actin-specific regulators, an alternative approach is to inject RNA from the same cells and to allow the oocyte to synthesise the factors. Of course, potential problems can be envisaged. For example, the relevant transcript or its product could be unstable in oocytes, or some vital post-translational processing step may only take place in muscle cells. Indeed, the factors may not even be proteins. However, it is relatively straightforward to test different RNA preparations in order to verify if this will be a plausible method.

Preliminary investigations have produced potentially promising, but inconsistent results (G. S. Cross, per. comm.). Injected *X. laevis* poly(A)$^+$ RNA from whole neurulae has, on some occasions, produced a stimulation of expression from pXB-C3.1 in oocytes, as judged by the oligonucleotide primer extension assay. This assay is required to distinguish between pXB-C3.1 transcripts and the *X. laevis* cardiac actin transcripts in the injected RNA preparation. As poly(A)$^+$ RNA is used, the RNA content of a whole embryo can be injected into a single oocyte for these studies. The primary objective at the moment is to develop

another relatively sensitive differential primer extension assay, which gives a lower background signal than the one incorporating the 15 base first exon oligonucleotide that is presently in use (see Figure 7.12c). This should help to determine whether the injected gene is expressed at a basal level in oocytes or if it is totally inactive, a question that is yet to be resolved.

The inconsistencies that have so far been observed in the RNA experiments may reflect variable treatment of micro-injected material by different batches of oocytes. Such an effect has been reported for the oocyte-type 5S RNA genes from transplanted *Xenopus* somatic cell nuclei (Korn and Gurdon, 1981). In this example, only oocytes from some frogs were able to reactivate the oocyte 5S genes. Presumably they alone contained constituents, which were able to remove chromatin fixed in an inactive configuration, so that it could then be replaced by a different structural arrangement which was accessible to the appropriate transcription-regulating factors. If oocyte variability is a problem in our experiments, it may be necessary to first remove a small proportion of oocytes from different individuals surgically, in order to determine which are most suitable for expression experiments.

If a system can be characterised in which specific injected RNA preparations stimulate expression from injected cardiac actin genes, we would then propose to use this functional assay to clone the transcripts responsible. Such an approach has been made possible by the development of *in vitro* transcription systems (e.g. Melton *et al.*, 1984; Krieg and Melton, 1984), which can produce high levels of artificially synthesised RNAs. Thus, if the RNA preparations that activate cardiac actin gene

expression are cloned as cDNAs in an appropriate vector, artificial transcripts could subsequently be injected in an attempt to reproduce the effect observed with the original natural mRNA. Initially, material synthesised from large pools of different clones would be tested in a single oocyte, but then, if a stimulation was detected, these pools would be divided up in a search for the RNA species involved. A recent report by Noma *et al.* (1986), in which they described the cloning of the mouse IgG1 induction factor, was based on this approach. This group found that RNA isolated from T cells, which synthesise the induction factor, could direct the production of a secreted, biologically active product on injection into *Xenopus* oocytes. They proceeded to clone the T cell RNA and demonstrated that artificial RNAs made from the resulting cDNA library also had the same effect. They then gradually decreased the pool size of clones used to make each RNA preparation, until they identified a single clone, which produced RNA encoding the induction factor.

One possible problem with this approach for identifying transcription factors is that it may be difficult to recognise multiple factors that affect the same gene, particularly if they operate synergistically. However, artificially synthesised anti-sense RNA coinjected with the original embryonic mRNA preparation may provide an alternative assay for these. Although the suitability of anti-sense RNA for inhibiting endogenous gene expression remains in doubt, Melton (1985) has demonstrated that translation of globin RNA introduced into oocytes can be blocked by the complementary RNA molecule, so this technique does appear to be feasible for inhibiting translation from injected sense RNA in the oocyte at least. It also has the advantage that it is not necessarily dependent on the synthesis of full-length RNA.

Hence, clearly these proposed experiments in *Xenopus* oocytes could potentially lead to the cloning of the RNAs which encode actin gene-specific regulatory factors. This would represent an important development, as other strategies for cloning specific transcription factors require the initial isolation and purification of significant levels of these molecules, a task that has yet to be achieved.

Undoubtedly, it is easy to conceive of reasons why our RNA injection experiments might always fail to give positive results or might even lead to false positives. However, only after we have clearly demonstrated the correct functioning of RNA prepared from muscle cells in this assay, will we consider the more time-consuming cloning work. If the results of the preliminary RNA injections appear promising, further controls could be performed by injecting altered cardiac actin genes that are deleted in their 5' flanking region, and have been found to be expressed poorly in embryos as a consequence. Moreover, if adult heart poly(A)$^+$ RNA stimulates pXB-C3.1 expression, it will be interesting to investigate its effect on the skeletal actin gene, which would not be expected to be activated.

If the oocyte assay system for *trans*-acting factors is to be viable, it seems as likely that it would work with the *Xenopus* cardiac actin gene as with any other gene. The extrachromosomal copies of pXB-C3.1 have already been shown to be correctly regulated in *Xenopus* embryos, even though they are in vast excess compared to their chromosomal counterparts. This result has two important implications. Firstly, large numbers of unintegrated cardiac actin genes in a single cell can be almost entirely inactivated if no muscle-specific factors are present, an observation that is further supported by the preliminary

data from oocyte injections. Secondly, the muscle-specific factors appear to be at sufficiently high concentrations in the myotomes to activate at least some of the multiple copies of the injected genes. However, it could be suggested that the distribution of the injected plasmid in different cells is so non-uniform that some nuclei only contain a few copies, and that only these nuclei are expressing the transformed genes. If this was true, insufficient synthesis might be expected from injected myotome RNA in the oocyte to activate transcription from many gene copies, and so the number of injected genes would need to be carefully controlled. However, the fact that, when more pXB-C3.1 DNA is injected in embryos, rising levels of expression are observed, argues against the hypothesis that the muscle-specific factors are limiting.

Clearly the use of oocytes to assay for *trans*-acting factors may be an important pathway in the cloning of the genes that synthesise them. In the near future, it is therefore anticipated that considerable efforts will be concentrated on assessing the feasibility of this experimental approach.

SECTION 8.4 CONCLUDING REMARKS

The primary aim of this project was to investigate the regulation of expression from a cloned cardiac actin gene on injection into fertilised *Xenopus* eggs. The results have demonstrated that the gene is controlled in a very similar fashion to its chromosomal counterpart. Not only has this study indicated that further experiments to delineate the sequences responsible for such regulation are viable, but it also suggests that the *Xenopus* embryo system will be suitable for the investigation of other genes expressed during early development.

Transformed *Xenopus* can therefore be used to study gene expression at stages of development that are not so readily accessible to scrutiny in transgenic vertebrates, like the mouse. The speed and relative simplicity of the biological part of the experiments in *Xenopus* are additional advantages that make this system particularly useful.

It had also been considered at the outset of this project that, if the cloned cardiac actin gene was regulated properly in embryos and oocytes, it might then be possible to use the oocyte as an assay system for *trans*-acting factors. The results discussed in this thesis indicate that this is a feasible approach, and recent reports by other workers have suggested ways in which we might clone such factors. Within the next few months we hope to have clarified the RNA injection experiments, so that we will be able to assess if this line of study is worthy of further investigation. As the expression of the cardiac actin gene occurs long before the formation of myofibrillae in *Xenopus*, it is hoped that any factors that might be identified will be associated with some of the earliest stages of muscle differentiation in this organism.

APPENDIX I

Cytoskeletal Actin Gene Families of *Xenopus borealis* and *Xenopus laevis*

Gareth S. Cross[1], Clive Wilson[1], Harry P. Erba[2] and Hugh R. Woodland[1]

MRC Animal Development Group, Department of Biological Sciences, University of Warwick, Coventry CV4 7AL, England[1]; The MEDIGEN Project, Department of Medicine, Stanford University School of Medicine and Veterans Administration Medical Center, Palo Alto, California 94304, USA[2]

Proofs should be sent to the former address

SUMMARY

We have sequenced the coding and leader regions, as well as part of the 3' untranslated region, of a X. borealis type 1 cytoskeletal actin gene (defined according to the arrangement of acidic residues at the N-terminus; Vandekerckhove et al., J. Mol. Biol. 152: 413-426). The encoded amino acid sequence is the same as the avian and mammalian β (type 1) cytoskeletal actins, except for an isoleucine at position 10 (as found in the mammalian γ cytoskeletal actins), and an extra amino acid, alanine, after the N-terminal methionine. Five introns were found, in the same positions as those of the rat and chicken β actin genes. The 5' and 3' untranslated regions resemble those of the human γ (type 8) cytoskeletal actin gene more closely than the mammalian β genes.

Primer extension showed that this type 1 gene is transcribed in ovary and tadpole. Sequencing of primer extension products demonstrated two additional mRNA species in X. borealis, encoding type 7 and 8 isoforms. This contrasts with the closely related species X. laevis, where type 4,5 and 8 isoforms have been found. The type 7 isoform has not previously been found in any other species. The mRNAs of the X. borealis type 1 and 8 and X. laevis type 5 and 8 isoforms contain highly homologous leaders. The X. borealis type 7 mRNA has no leader homology with the other mRNA species and, unlike them, has no extra N-terminal alanine codon. The evolutionary implications of these data are discussed.

KEY WORDS: Actin genes - Evolution - 5' and 3' untranslated regions

INTRODUCTION

Actins are a family of highly conserved contractile proteins. Some isoforms are muscle-specific, while the cytoskeletal actins are found in probably all eukaryotic cells. Of the four mammalian muscle-specific isoforms (Vandekerckhove and Weber 1978a), two are specific to striated muscles (the heart and skeletal muscle isoforms, differing in their relative abundance in the heart and in skeletal muscle), and two are specific to smooth muscle (the aortic and stomach isoforms, differing in their relative abundance in the aorta and the stomach). In mammals there are two different major cytoskeletal actins, the β and γ isoforms (Vandekerckhove and Weber 1978b), although there have been unconfirmed reports of additional isoforms in the nuclei of non-muscle cells (Bremer et al. 1981), and in brain synaptosomes (Marotta et al. 1978). The β and γ isoforms are also found in chickens (Vandekerckhove and Weber 1981), but, in addition, there has recently been a report that chickens have a gene encoding a type 5 (see Table 1) actin (Bergsma et al. 1985), although such a protein has so far not been detected *in vivo*. The cytoskeletal actins differ from the muscle-specific isoforms in 22 to 25 amino positions (Vandekerckhove and Weber 1979). For a given isoform, however, such as the β cytoskeletal actin, the amino acid sequence has been found to be identical in humans, rats and chickens (Kost et al. 1983; Nudel et al. 1983; Ponte et al. 1984). The β and γ isoforms of mammals and chickens differ between each other in only four amino acids: the three acidic residues at the N-terminus and the amino acid at position 10 (Vandekerckhove and Weber, 1978b).

The conservation of these different cytoskeletal isoforms in mammals and birds has suggested separate functions for each of them, although there is little evidence for this so far. It is therefore interesting that this conservation does not extend to amphibians, where Vandekerckhove et al. (1981) have partially sequenced the N-terminal tryptic peptides of cytoskeletal

Table 1. Distribution of the different vertebrate cytoskeletal actin isoforms

Actin type[a]	N-terminal amino acids	Mammals	R. pipiens	T. cristatus	X. laevis	X. borealis
1	asp-asp-asp	+ (= β)		+		+
3	asp-glu-asp			+		
4	glu-asp-asp		+		+	
5	asp-glu-glu		+		+	
7	glu-glu-asp					+
8	glu-glu-glu	+ (= γ)		+	+	+

[a] Nomenclature as suggested by Vandekerckhove et al. (1981).

actins from several different species (Xenopus laevis, Rana pipiens, and Triturus cristatus) and found new isoforms which differ in their content and order of glutamate and aspartate residues in the first three amino acid positions. As Table 1 shows, different amphibian species have different sets of cytoskeletal actins. These differences may have occurred because of selective advantages of certain isoforms to species in particular environments, or they may be merely neutral mutations (ie there may be no functional difference between the isoforms). In this paper we show, by the sequencing of primer extension products, that Xenopus borealis, a species thought to have diverged from Xenopus laevis only 8 million years ago (Bisbee et al. 1977), contains mRNA species encoding two cytoskeletal isoforms, types 1 and 7 (see Table 1), which have not been found in X. laevis, as well as a type 8 isoform which is present in both species.

We also present the structure of a X. borealis type 1 cytoskeletal actin gene, demonstrating the conservation of its introns and amino acid sequence when compared with mammalian and avian β (type 1; see Table 1) cytoskeletal genes. This gene, however, shows greater similarities in its leader and 3' untranslated region sequences, as well as at amino acid position 10, to a human γ (type 8) cytoskeletal gene. We speculate on the evolutionary relationships of the Xenopus genes we have detected to the cytoskeletal actin genes of other organisms.

MATERIALS AND METHODS

Materials

Clone λ35A was isolated from a partial EcoRI genomic library cloned in λgtWES (W. Bains, PhD thesis, University of Warwick, Coventry, 1982). Clone λ3H1 was isolated from a partial Sau 3A1 library in λ47.1 (a generous gift from Dr. P.C. Turner). Kenyan Xenopus borealis frogs, used to make DNA for these libraries, as well as RNA for primer extension analysis, were bred in

this laboratory. M. laevis, used to make RNA, were obtained from the South African Snake Farm, Fishhoek, South Africa. RNAs were kindly supplied by Ms J.E.M. Ballantine and Dr. P.C. Turner, and were made as described previously (Woodland et al. 1984).

α-^{32}P deoxynucleoside triphosphates (> 2000Ci/mmole) and [γ-^{32}P] ATP (> 5000Ci/mmole) were from Amersham International. Restriction enzymes were from Boehringer-Mannheim Ltd, Bethesda Research Laboratories, and New England Biolabs. Escherichia coli DNA polymerase I (Klenow fragment and the Kornberg enzyme) and calf intestinal alkaline phosphatase were from Boehringer-Mannheim. T4 DNA polymerase I, DNA ligase and polynucleotide kinase were from Bethesda Research Laboratories. AMV reverse transcriptase was obtained from Life Sciences Inc.

Isolation of actin clones

Initially a M. borealis partial Eco RI library was screened by the procedure of Benton and Davis (1977). The probe was a nick-translated Dictyostelium actin cDNA clone. This was hybridized to the nitrocellulose in 4 x SSC (1 x SSC = 0.15M NaCl, 0.015M sodium citrate), 1 x Denhart's additives (0.2% Ficoll, 0.2% polyvinyl-pyrrolidone, 0.2% bovine serum albumin), 60mg/l sonicated E. coli DNA, 0.1 mg/ml poly(A) at 60°C for 48 hours, and washed at 55°C in 4 x SSC for 1 hour. Plaques that showed a positive signal were picked and screened a second time. These included the recombinant λ35A.
Subsequently a second library (a partial Sau 3A1 digest cloned in the vector λ47.1) was screened, resulting in the isolation of clone λ3H1, containing the 3.6kb Eco RI fragment of λ35A in addition to surrounding M. borealis sequence. For this screen, the hybridization was in 3 x SSC, 0.1 x Denharts additives, 0.1mg/ml poly(A) at 65°C for 16 hours and the filters were washed in 3 x SSC at 50°C. The probe used was the nick-translated insert of λ5AP, a clone obtained on screening the first library which contains sequence encoding the

first 150 amino acids of a muscle-specific actin. The clone λ3H1 was purified by further plating and rescreening.

Mapping the clones

We used double digests of nick-translated DNA and partial digests of DNA fragments labelled at one end with T4 polynucleotide kinase or T4 polymerase I (Smith and Birnstiel 1976). The resulting fragments were separated on agarose gels. To locate the coding regions, unlabelled restriction fragments were separated on agarose gels, blotted onto nitrocellulose (Southern 1975), and hybridized with the nick-translated Dictyostelium actin probe under the same conditions as used to isolate λ35A.

Sequencing of DNA

For this, as indicated in Fig. 1, we used both the chemical base modification procedure of Maxam and Gilbert (1978), and the enzymic method of Sanger et al. (1980) applied to single-stranded M13 subclones. For the former procedure, DNA restriction fragments were end-labelled, either with T4 polynucleotide kinase (after treatment with alkaline phosphatase) or T4 DNA polymerase I, prior to asymmetrical cleavage with a second restriction enzyme. The products were separated on 5% non-denaturing acrylamide gels and extracted from the gel as described by Maxam and Gilbert (1978). Five reactions were set up for each end-labelled sample (G, A+G, A>C, C, C+T). In those regions where the Sanger method alone has been used the sequence of both strands has, for the most part, been obtained.

Primer extension

The primer was prepared from a 450bp Bst NI fragment, one end of which was situated in the first coding exon (= exon 2 in Fig. 1), whilst the other was situated 5' to this exon. This Bst NI fragment was isolated from the

3.6kb Eco RI fragment of λ35A (recloned into the plasmid pBR325 for easy preparation), treated with alkaline phosphatase, 5' end-labelled with T4 polynucleotide kinase and [γ-^{32}P] ATP, cut with Hha I and fractionated on an 8% acrylamide, 7M urea sequencing-type gel. The 103 nucleotide fragment of the non-coding strand corresponding to residues 7 to 41 of the coding region was extracted from the gel.

Hybridization of this primer to RNA, and its subsequent extension, were as described previously (Woodland et al. 1984), except that 8% sequencing-type gels were used to separate the primer extension products. In preliminary experiments, to optimize hybridization temperatures for example, 0.01-0.02mg of total RNA was used in 10μl hybridizations. To obtain enough extended material to sequence, however, larger amounts of RNA were used. Thus, to sequence X. borealis primer extension products A and B (see Fig. 3), 0.75mg of ovary RNA was hybridized to primer overnight in 0.35ml hybridization solution (0.4M NaCl, 10mM Pipes pH 6.4) containing 50% formamide at 45°C (equivalent to a hybridization at 70°C without formamide). To sequence X. borealis product C, 0.2mg of tadpole RNA was hybridized to primer for 3 hours at 65°C in 0.13ml of hybridization solution without formamide, but containing 1% sodium dodecyl sulphate. To sequence X. laevis products A and B, 0.4mg of XTC cell (Pudney et al. 1973) RNA was hybridized in 0.2ml of 1% SDS-containing hybridization buffer at 70°C for 3 hours.

RESULTS AND DISCUSSION

The structure of a X. borealis cytoskeletal actin gene

We have isolated a clone, λ35A, from a λgtWES X. borealis genomic library. The 3.6kb Eco RI insert was found to contain sequence encoding the first three exons of an actin gene. Subsequently another clone, λ3H1, was isolated from a λ47.1 library (made from the DNA of a different individual). This clone contained an extra 1.4kb of sequence in the 3' portion of the gene

and an extra 10kb of sequence upstream from the 3.6kb Eco RI fragment. This enabled us to determine the rest of the coding sequence, as well as 318 bases of the 3' flanking sequence. The structure of this gene and the sequencing strategy are shown in Fig. 1.

Fig. 2 shows the sequence of regions 2 and 3 from Fig. 1. As customary, we have numbered the amino acids according to the nomenclature suggested by Elzinga and Lu (1976) and Vandekerckhove and Weber (1978b) for rabbit skeletal muscle actin. Except for two positions, the amino acid sequence encoded by this gene (376 amino acids in total) is identical to that of the mammalian and avian β (type 1) cytoskeletal actin isoforms (Kost et al. 1983; Nudel et al. 1983; Ponte et al. 1984). This is a remarkable conservation of protein sequence, as the divergence of the mammalian and amphibian ancestral lines is estimated to have occurred 300 to 400 million years ago (Wilson et al. 1977). One of the differences occurs at amino acid position 10 where, instead of a valine codon, an isoleucine codon was found. However, an isoleucine occurs at this position in the mammalian γ (type 8) isoform (Vandekerckhove and Weber 1978c). The second difference is the presence of an extra codon, for alanine, after the N-terminal methionine codon of the frog gene. This codon is not found in the human, rat or chicken β actin (Kost et al. 1983; Nudel et al. 1983; Ponte et al. 1984), or human γ actin (H. Erba, submitted for publication) genes that have been sequenced, although it has been found in a type 5 chicken gene (Bergsma et al. 1985). The significance of this is discussed below.

Introns interrupt the coding sequence of the Xenopus gene at positions 41/42 (ie between the codons for these amino acids), 121/122, 267 (ie within the codon for this amino acid) and 327/328. These are in the same positions as the introns in the rat and chicken β actin genes (Kost et al. 1983; Nudel et al. 1983). No genomic sequences of avian or mammalian γ cytoskeletal actin genes have been published so far.

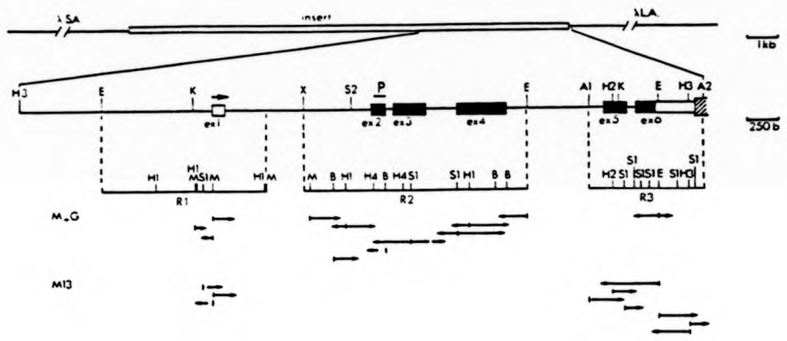


Fig. 1 Structural map and sequencing strategy for the X. borealis genomic clone λ3H1. λL.A and λS.A mark the long and short arms, respectively, of the vector. 3 regions, R1, R2 and R3 have been mapped in greatest detail. Symbols for restriction enzymes: A1, Acc I; A2, Ava I; H1, Hinf I; H2, Hinc II; H3, Hind III; H4, Hha I; B, Bst NI; M, Msp I; E, Eco RI; K, Kpn I; X, Xba I; S1, Sau 3A1; S2, Sst I. Sequencing was in the region and direction represented by the arrows, and was by the method of Maxam and Gilbert ('M+G', Maxam and Gilbert 1978) or of Sanger ('M13', Sanger et al. 1980). The solid blocks represent coding exons, the open bars, untranslated sections of the mRNA. The hatched region represents a portion of the λ cloning vector. The horizontal arrow over exon 1 indicates the start and direction of transcription. The region P is the Hha I/Bst NI fragment used for primer extension.

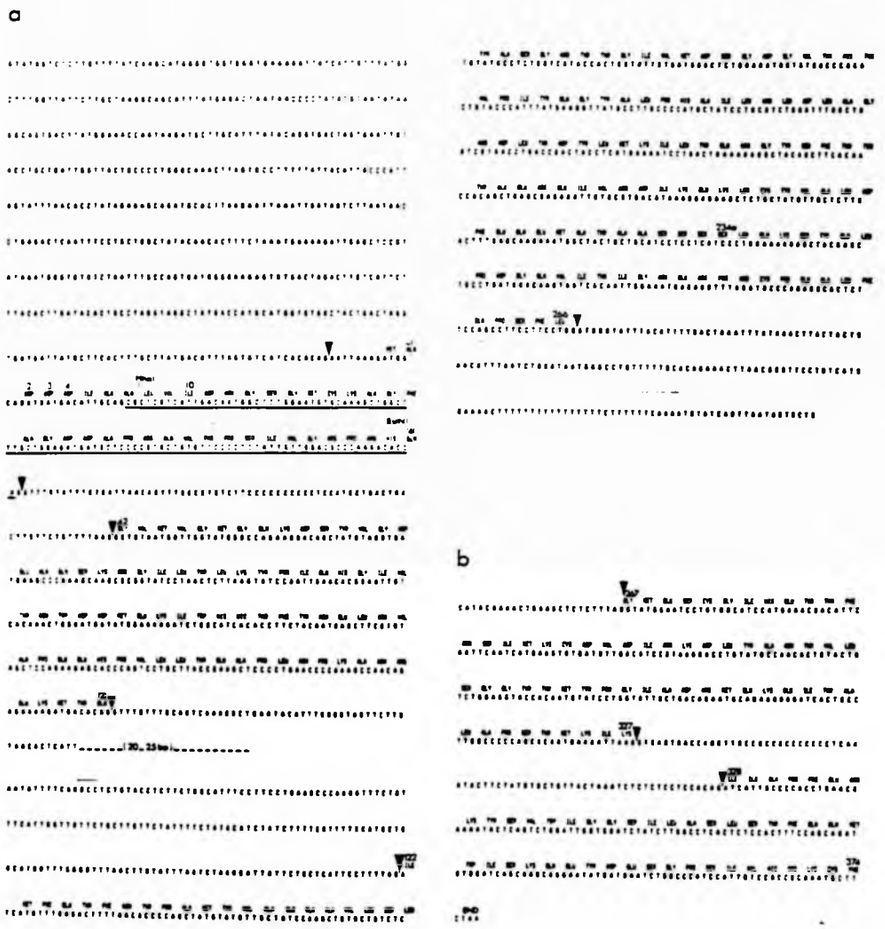


Fig. 2 The sequence of regions R2 (a) and R3 (b) from Fig. 1. There are 60 bases on each line. Amino acids of the encoded actin protein have been indicated above their respective codons. Selected amino acids have been numbered above the sequence according to the numbering system derived from rabbit skeletal muscle actin (Elzinga and Lu 1976; Vandekerckhove and Weber 1978b). As the preceding amino acid is probably post-translationally removed it is numbered -1. The arrowheads indicate probable splice sites, following the GT/AG rule of Breathnach and Chambon (1981). The underlined sequence indicates the position of the Hha I-Bst N1 fragment's non-coding strand used for primer extension.

Confirmation that this gene truly encodes a cytoskeletal actin and not a muscle-specific isoform was provided by the primer extension experiments described below. These demonstrated the existence of the gene's transcripts in the oocyte. We have also found its transcripts in RNA from embryos at all stages of development up to the tadpole, although the abundance of this mRNA species decreases relative to the other X. borealis cytoskeletal actin mRNAs (from the type 7 and type 8 genes - see Figs. 3 and 4) between the oocyte and the stage 12 embryo (unpublished data).

5' transcriptional mapping

To map the start of transcription, a primer was made from a Hha I - Bst NI fragment in the first coding exon (see Figs. 1 and 2, and Methods). The 103 base, 5' end-labelled, single stranded primer was hybridized to X. borealis ovary or tadpole RNA at various temperatures, extended with reverse transcriptase, and the products separated on a sequencing-type polyacrylamide gel. As shown in Fig. 3, three major products were visualized by autoradiography of such gels. Surprisingly, two of these were obtained in greater amounts when the primer hybridization was at the highest temperature, 70°C in this case, while for the smallest product the optimum temperature was 65°C (data not shown). A similar phenomenon was noted by Woodland et al. (1984) while using a histone H4 primer. It is possible that higher temperatures can promote hybridization by the destruction of mRNA secondary structure. As with histone mRNAs, there are, in fact, leader and coding regions which could base pair with the primer-hybridizing part of the mRNA. The primer extension products from such experiments were eluted from the gels and sequenced by the method of Maxam and Gilbert (1978). The material used for such sequencing typically consisted of a doublet of bands. We believe these doublets to result from extension on only one mRNA species, because the sequence obtained was ambiguous, and because such a phenomenon has previously

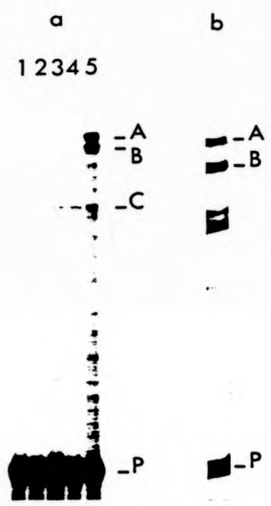


Fig 3 Primer extension products resulting from hybridizing the primer from λ3H1 to (a) X. borealis ovary RNA, and (b) X. laevis XTC cell RNA. Hybridizations were in 10μl of hybridization solution containing 1% sodium dodecyl sulphate. In (a) 10μg of RNA was hybridized overnight at 50°C (track 2), 55°C (track 3), 60°C (track 4) and 70°C (track 5). In (b) 20μg of RNA was hybridized for 3 hours at 70°C. The primer was then extended using reverse transcriptase and the products analyzed on 8% sequencing-type polyacrylamide gels. Electrophoresis was from top to bottom. In track 1 of (a) unhybridized primer was run. The extension products we have sequenced have been labelled. The primer is labelled P.

been reported for a single species of globin mRNA (Luse et al. 1981).

The sequenced portions of the three primer extension products are shown in Fig. 4. Part of the sequence of the largest product exactly matches the coding sequence of the type 1 gene, and the first 8 bases 5' to this. The remaining 50 bases of product A sequence are identical to a region approximately 1200 bases upstream in the genomic clone. The 5' untranslated leaders of all other sequenced vertebrate actin genes are also interrupted by an intron (Chang et al. 1985; Fornwald et al. 1982; Kost et al. 1983; Nudel et al. 1983; Zakut et al. 1982). To determine the start of transcription, primer extension product A was sized by running it next to a set of sequencing tracks. We estimate the accuracy to be within 2 bases. As shown in Fig. 5, 29 to 30 bases upstream from the start of transcription (the 'cap-site') lies the beginning of the sequence 5'-TTAAATA-3'. This resembles the 'TATA box' promoter element (with consensus sequence ($TATA^{T}_{A}A^{T}_{A}A$), found in this position in many eukaryotic polymerase II genes (Breathnach and Chambon 1981). Two potential 'CAAT boxes' can also be found. One sequence, 5'-GCCCATA-3', lies approximately 90 bases upstream from the start of transcription, and hence in a similar position to that found in many other genes (Breathnach and Chambon 1981)). However, another sequence, 5'-GGCCAATCG-3', more closely resembles the consensus sequence for this promoter element, GG^{C}CAATCT (Benoist et al. 1980; Efstratiadis et al. 1980), but is situated 30 bases further away from the cap site.

Previously, searches for regions of homology in the 5' upstream regions of similar genes in different species have been made in order to identify sequences which may be involved in transcriptional control. Thus, a region of homology 20 nucleotides long has been found in the rat and chicken skeletal muscle actin genes (Ordahl and Cooper 1983) around, and including, the CAAT box. A block of 25 nucleotides inbetween the CAAT and TATA boxes is very similar in the rat and chicken β actin genes (Kost et al. 1983; Nudel et al.

```
                                                                met ala asp asp asp ile ala
--CTTGCTAGCTTTGCTTTTTTTTTCCCCGTAAAGGAAACTTCCAGCACCAAGAATTAAAG ATG GCA GAT GAT GAC ATT GCA G     X.bor A = Type 1

  ..............    ...... ..... ........ .  .. ......... .... met ala glu glu glu ile ala
--CTTGCTAGCTTTGC---TTTTTTTCCCGTTAAGGAAACCTTAAGTACCAAGAATAAAAG ATG GCA GAA GAA GAG ATT GCA G     X.bor B = Type 8

                  .. .......  .. ..  .............. met ala glu glu glu ile ala
               --GTTAAGGAAATCTCTAGTTCCAAGAATTAAAG ATG GCA GAA GAA GAG ATT GCA G     X.lae A = Type 8

                  .. .............. .............. met ala asp glu glu ile ala
               --GTCAAGGAAACTTCCAGTACCAAGAATTAAAG ATG GCA GAT GAA GAG ATT GCA G     X.lae B = Type 5

                .     .          ...    .  . .   met glu glu asn ile ala
            --GGAGACAATTCGTGTTTGCGTCCTACCCTCAGATCACA ATG GAA GAA GAT ATT GCT G       X.bor C = Type 7
```

Fig. 4 Partial sequence of X. borealis and X. laevis mRNAs derived from sequencing primer extension products. The sequences of X. borealis products A and B (see Fig. 3) were obtained using ovary RNA while tadpole RNA was used to obtain product C sequence. XTC cell RNA was used to obtain the sequence of X. laevis products A and B. In each case the sequence of the leader region is incomplete. We have indicated the type of actin encoded by each sequence, using nomenclature suggested by Vandekerckhove et al. (1981). We have compared the leader of the X. borealis type 1 mRNA with the other 4 mRNAs. An asterisk above a base indicates that it is the same as the equivalent base of the X. borealis type 1 leader. A gap has been introduced into the X. borealis type 8 mRNA sequence to optimize the upstream homology between this leader and that of the type 1 gene.

1983). We have compared the X. borealis type 1 gene with the rat and chicken β (type 1) genes, and found several small regions of homology in the leader and 5' flanking regions. Of the 16 most distal bases in the 25 nucleotide homology block mentioned above, 11 bases of the rat sequence and 10 bases of the chicken sequence are the same within the region, 5'_GTTTCTGAAAGATGCC_3', between the TATA box and most distal putative CAAT box of the Xenopus gene. The sequence 5'-CTGTGPuCG-3" occurs just upstream from the frog and rat TATA boxes, but not that of the chicken. The sequence 5'- GCGAGGC-3' is found in all three genes, however, at approximately the same distance from the cap site (100-110 nucleotides), and if the most proximal potential CAAT box is used for the frog gene, the same distance (approximately 20 nucleotides) upstream from the CAAT boxes. In the leader region, a 12 base segment was found to have 11 bases in common with part of the rat β gene leader (see Fig. 5), but this sequence was not in the chicken leader. However, a comparison of the frog leader with that of a human γ (type 8) gene (H. Erba, submitted for publication), revealed a high degree of homology in the first 20 bases (see Fig. 6b), with little homology elsewhere. It seems likely that this region has been conserved for functional reasons, but further comparisons with other species are required to ascertain if the smaller length homologies described above are meaningful or merely the result of chance.

3' untranslated region (3'UTR)

It is probable that all of the 318 bases of cloned DNA 3' to the termination codon of this X. borealis gene are transcribed, as two very similar genes of the closely related species X. laevis (see Fig. 4) both have 3' UTRs approximately 800 bases in length (T. Mohun, personal communication). Fig. 6(a) shows a sequence comparison between this 3' region and the 3' UTR of a human γ (type 8) cytoskeletal actin cDNA clone (H. Erba, submitted for publication). There are many blocks of complete homology in the first 120

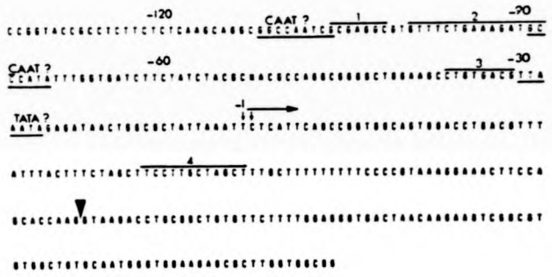


Fig. 5 Sequenced portion of the region R1 from Fig. 1, including exon 1 and the region immediately upstream from the start of transcription. Each complete line consists of 60 bases. The large vertical arrowhead indicates the 5' end of the leader intron. The horizontal arrow indicates the direction and start of transcription. The exact point of the latter is not known; the two small vertical arrows indicate possible positions. The bases 5' to the start of transcription have been given negative numbers. Sequences representing possible promoter elements have been underlined. Sequences which are wholly or partly homologous to similar regions in other cytoskeletal genes have been overlined. Sequence 1 is found 100-110 nucleotides upstream from the start of transcription of the rat and chicken β actin genes (Kost et al. 1983; Nudel et al. 1983). Sequence 2 has homology to the first 16 bases of the 25 nucleotide homology block found between the CAAT and TATA boxes of the rat and chicken β actin genes (Kost et al. 1983; Nudel et al. 1983). The rat sequence is the same in 11 positions and the chicken sequence is the same in 10 positions. 7 of the 8 bases of sequence 3 are found immediately upstream from the TATA box of the rat β gene (Nudel et al. 1983). A sequence in the leader of the latter gene is the same as sequence 4 in 11 out of 12 positions.

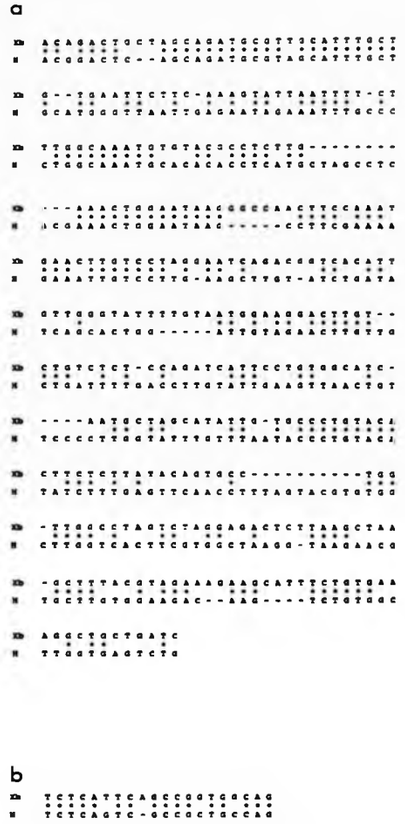


Fig. 6 Sequence comparisons between (a) the 3' UTRs, and (b), the 5' UTRs of the X. borealis (Xb) type 1 gene and a human (H) γ cytoskeletal cDNA clone (H. Erba, submitted for publication). In (a) the comparison begins immediately after the termination codons and proceeds for the full 318 bases of downstream Xenopus sequence present in λ3H1. In (b), a portion of the X. borealis gene's leader, beginning with the start of transcription, is compared with a similar sequence in the human gene. Outside this region there is little homology between the two leaders. Asterisks indicate positions where bases are the same. Gaps have been introduced in both sequences to maximize homology.

bases with a few regions of homology in the remainder. Apart from the first 15 bases, there is far less homology when the frog type 1 gene is compared with the rat, human, or chicken β actin 3'UTRs (Kost et al. 1983; Nudel et al. 1983; Ponte et al. 1984). This is a remarkable example of non-coding sequence conservation, and it is presumably maintained for functional reasons. It may be relevant to its function that the region of 21 bases from bases 11 to 31 of the X. borealis gene's 3' UTR is capable of forming a hairpin loop structure (see Ffg. 7). In this region there is only 1 base difference between the Xenopus and human genes. This phenomenon of 3' UTR sequence conservation appears to be a general one for actins, as it is for several other types of gene (Yaffe et al. 1985), though the conservation in this actin seems extreme. Regions of high homology have been reported between the 3'UTRs of the human and X. laevis cardiac muscle actin genes (Gunning et al. 1984), between the 3'UTRs of human, rat and chicken skeletal muscle actin genes (Gunning et al. 1984), and between the 3'UTRs of the human, rat and chicken β cytoskeletal actin genes (Kost et al. 1983; Nudel et al. 1983; Ponte et al. 1984). In contrast, apart from two small regions of homology between the rat cardiac and skeletal muscle gene 3' UTRs (Mayer et al. 1984), there is little homology between the 3'UTRs of the different isoforms within a species. This conservation of 3'UTR sequence, therefore, together with the 5'UTR conservation shown in Fig. 6(b), and the presence of an isoleucine codon at position 10, suggests that the gene we have cloned is functionally, and presumably evolutionarily, related to the mammalian and avian γ cytoskeletal genes, rather than the β genes (even though the three N-terminal amino acids are in the β configuration). The function of the 3'UTR conservation in certain genes is not known. Yaffe et al. (1985) have suggested that in some cases, such as the c-fos gene (Miller et al. 1984), it is because this region is involved in translational control. Evidence has been found suggesting that the translational control of cytoskeletal actins takes place in the early

Fig. 7 Predicted hairpin loop structure formed from bases 11 to 31 of the X. borealis type 1 mRNA's 3' UTR.

embryonic stages of X. laevis (Ballantine et al. 1979). Alternative explanations are possible, however. For example, this region may be important in message stability.

Expression of other cytoskeletal actins in Xenopus

The sequences of the other X. borealis primer extension products (see Fig. 3) are shown in Fig. 4. Product B represents a type 8 isoform, and like product A has an alanine after the N-terminal methionine, while product C represents a type 7 isoform with no alanine in this position (position -1). The type 7 isoform has not previously been found in any other species. Thus in total, we have found mRNAs encoding isoform types 1, 7 and 8 in X. borealis.

By the partial sequencing of N-terminal tryptic peptides, Vandekerckhove et al. (1981) found the type 4, 5 and 8 isoforms in several non muscle tissues of X. laevis. Shown in Fig. 4 are the sequences of the two largest primer extension products produced as a result of hybridizing our X. borealis primer to XTC cell (a X. laevis fibroblast-like cell line; Pudney et al. 1973) RNA (see Fig. 3b). These sequences encode type 5 and type 8 actin isoforms; two of the isoforms found by Vandekerckhove et al. (1981). Primer extension bands probably representing at least one more mRNA species were also detected using X. laevis RNA, but it has not yet proved possible to sequence these (although we would predict the existence of an mRNA species encoding a type 4 isoform). We have, therefore, partially confirmed the results of Vandekerckhove et al. (1981). This difference in actin isoform content between X. borealis and X. laevis is surprising, as these species are thought to have diverged only 8 million years ago (Bisbee et al. 1977). Our ability to detect the type 5 isoform of X. laevis argues against the possibility that we were unable to detect a X. borealis type 5 mRNA because it hybridized poorly to our primer. Although more complex possibilities exist, it seems likely that we have

detected all of the major cytoskeletal actin mRNAs in the X. borealis ovary and X. laevis XTC cells. We have also found the same primer extension bands using X. borealis tadpole and X. laevis ovary RNA, so these mRNA species are unlikely to be tissue-specific.

Despite encoding different isoforms the X. laevis type 5 and X. borealis type 1 mRNAs have very similar leader sequences. As shown in Fig. 4, these are also closely related to the X. laevis and X. borealis type 8 mRNA leaders. It is possible that these genes have all arisen from a common ancestral gene relatively recently. Another, and not necessarily alternative, possibility is that the common ancestor to these two species possessed genes for all three of these isoforms, the type 1 being later silenced in X. laevis, and the type 5 in X. borealis. Within one species, the similarity of the leaders of the two genes may be the result of a gene conversion event or may reflect their formation by a recent duplication event. In this latter respect it is very relevant that X. borealis and X. laevis are believed to be tetraploid species, and are indeed more-or-less tetraploid relative to X. tropicalis which diverged from the other two species approximately 30 million years ago (Theibaud and Fischberg 1977; Tymowska and Fischberg 1977). It is not known if the duplication of the genome preceded the divergence of the two tetraploid species.

As can be seen in Fig. 4, the X. laevis type 5 and type 8 mRNAs also have the extra alanine codon (at position-1) found in the X. borealis type 1 and type 8 mRNAs. Vandekerckhove et al. (1981) determined the N-terminal protein sequences of these X. laevis isoforms and found that they began with the acidic residues, blocked by an acetyl group. Therefore, the alanine must be post-translationally removed, along with the first methionine.

Evolution of actin genes

It will be difficult to be certain how vertebrate actin genes are related

until more sequence data is available from a much wider range of animals. Nevertheless, we believe it to be useful to make a few speculations based on these new data, as a context for future observations.

The X. borealis cytoskeletal actin gene we have cloned contains introns at the same positions in the coding region as the β cytoskeletal actin genes of the rat and chicken (Kost et al. 1983; Nudel et al. 1983): at positions 41/42, 121/122, 267 and 327/328. The four mammalian and avian striated muscle-specific actin genes that have been sequenced have, however, all contained introns at a different set of positions; at codons 41/42, 150, 204, 267 and 327/328 (Chang et al. 1985; Fornwald et al. 1982; Hamada et al. 1982; Zakut et al. 1982). A human aortic-type smooth muscle actin gene has been sequenced and found to contain a set of introns which is the same as those of the striated muscle-specific genes, but with additional introns at positions 84/85 and 121/122 (Ueyama et al. 1984). Our results suggest, therefore, that the cytoskeletal and muscle-specific genes became separate before the divergence of the amphibians from the rest of the vertebrates. In support of this, we have sequenced (unpublished data) part of a X. borealis muscle-specific actin gene, encoding amino acids from the N-terminus to position 150, and found it to contain introns in the same positions as the mammalian striated muscle-specific genes (ie at positions 41/42 and 150). Further, the amino acid sequence of this frog actin was identical to that of the chicken and human cardiac actins (Chang et al. 1985; Hamada et al. 1982). These results with amphibian genes support the evolutionary proposals of Vandekerckhove and Weber (1984). They found that the muscle-specific actins of invertebrates did not vary greatly in amino acid sequence from the cytoskeletal isoforms of vertebrates and invertebrates. However, they obtained the complete amino acid sequence of a lamprey muscle-specific actin and found it to resemble the sequence of mammalian and avian muscle-specific isoforms. This suggested that this type of muscle-specific actin gene arose

prior to the divergence of jawless fish from the rest of the craniates.

We have tried to construct a scheme for the evolution of vertebrate cytoskeletal actin genes which accounts for the presence or absence of the extra alanine codon at position -1 in various genes. The information available can be summarized as follows: we have found in X. borealis two ala(+) mRNA species (ie possessing an alanine codon at position-1), and one ala(-) mRNA species, with no alanine at position -1 (X. laevis also possesses two ala+ species and the primer extension product we have been unable to sequence may correspond to the ala- species). We have shown that one of these ala(+) genes, although a type 1 gene (see Table 1) from its N-terminal amino acid sequence, appears to be more closely related to the human γ (type 8) gene than the human β (type 1) gene, as judged by extensive 3' UTR homology, as well as some 5' UTR homology, and the presence of an isoleucine codon at position 10. The mammalian and avian β, and mammalian γ, cytoskeletal genes are both ala(-), however.

Have these different types of actin gene arisen from an ancestral gene by deletion or addition of this extra alanine codon? The mammalian and avian muscle-specific genes sequenced so far (Chang et al. 1985; Fornwald et al. 1982; Hamada et al. 1982; Hanauer et al. 1983; Ueyama et al. 1984; Zakut et al. 1982) have a cysteine codon following the N-terminal methionine codon (ie they are cys+), as do all the actin genes (including cytoskeletal) analyzed from sea urchins (Schuler et al. 1983), Drosophila (Fyrberg et al. 1981), and nematodes (Files et al. 1983). It seems likely, therefore, that the ancestral prechordate actin gene was cys(+). The cysteine codon was presumably retained in invertebrates as well as vertebrate muscle-specific genes (an extra acidic residue being inserted at the N-terminus of the latter), but has been mutated to an alanine in some vertebrate cytoskeletal genes, and deleted in others. It is interesting to note that, unless two point mutations occurred simultaneously, this change from TGPy to GCN, would have involved an

intermediate codon for a third type of amino acid.

At what time, in the course of evolution, did such changes at position -1 occur? Both mammals and birds possess ala(-) β and γ cytoskeletal genes (Kost et al. 1983; Nudel et al. 1983; Ponte et al 1984; H. Erba, submitted for publication), suggesting that if this ala(-) deletion occurred only once, it occurred before their divergence from a common ancestor, and before the divergence of the β (type 1) and γ (type 8) genes. The presence of an ala(-) gene in X. borealis (the leader sequence of which shows small regions of homology to both β and γ mammalian genes) suggests that such a deletion occurred before the divergence of amphibians from the rest of the vertebrates. In X. borealis, however, there is also an ala(+) gene, with striking homology in its 5' and 3' UTR to the ala(-) mammalian γ (type 8) gene. How has this occurred, if the ala(-) deletion arose before the divergence of the β- and γ-like genes? One of the simpler explanations is that the ancestral cytoskeletal actin gene (cys+ or ala+) contained UTRs similar to the mammalian γ cytoskeletal gene, the mammalian β actin being the gene to have modified its 5' and 3' UTRs since the β/γ divergence from the ancestral gene. According to this scheme, gene duplication would have given rise to two cytoskeletal genes, one of which became ala(+), the other becoming ala(-). Both types persisted in amphibians, while only the ala(-) is currently detectable in mammals. Gene duplication of the ala(-) gene type at some stage presumably gave rise to the mammalian and avian β and γ cytoskeletal genes. We cannot say if this β/γ divergence occurred before or after the divergence of amphibians from the rest of the vertebrates as we do not yet have any 3' UTR sequence information from the X. borealis type 7 and 8 genes.

Alternative scenarios to that presented above are certainly possible, however. For example, there may have been more than one deletion event giving rise to ala(-) genes during the course of evolution. It is also possible that gene conversion events may have caused ala(+) genes to gain 5' or 3' UTR

sequences from ala(-) genes or vice versa.

Recently, Bergsma et al. (1985) have identified an ala(+) cytoskeletal actin gene in chickens encoding a type 5 (see Table 1) isoform. They found that this gene possessed just a single intron, in its leader, yet could find no evidence for it being a pseudogene. This contrasts with the 5 introns found in all other vertebrate cytoskeletal actin genes (Kost et al. 1983; Nudel et al. 1983), including the X. borealis ala(+) gene described here. Further information is needed to fit this gene into an evolutionary scheme. However, this observation is consistent with the view that the ala(+) mutatio arose early in evolution but, alternatively, the ala(+) mutation may have arisen several times. The presence of an ala(+) actin gene in the soybean (Shah et al. 1982)) may be an indication of the latter case, and suggests a functional advantage in having such a codon in this position.

In all four Xenopus ala(+) genes examined here (see Fig. 4), as well as the chicken type 5 and soybean genes, the same alanine codon, GCA, is found after the N-terminal methionine. Other synonymous alanine codons are more frequently used in the rest of the X. borealis type 1, chicken type 5 and soybean actin genes, however. GCA is used in only 4 out of 29 positions for each of the frog and chicken genes, and in 9 out of 28 positions in the soybean. This suggests that the role of the alanine codon following the N-terminal methionine may be in its contribution to the DNA or RNA sequence in this region, rather than simply encoding alanine. It is interesting in this respect that the 6 bases of leader sequence 5' to the N-terminal methionine codon of the soybean and X. borealis type 8 genes are identical, giving a total run of 16 identical bases.

It seems likely that the cysteine codon at position -1 of cys(+) genes has a function as it is present in all six Drosophila genes (Fyrberg et al. 1981), all four genes of the nematode Caenorhabditis elegans (Files et al. 1983), both sea urchin genes that have been sequenced so far (Schuler et al.

1983) and all six sequenced vertebrate muscle-specific genes (Kost et al. 1983; Fornwald et al. 1982; Hamada et al. 1982; Hanauer et al. 1983; Ueyama et al. 1984; Zakut et al. 1982). What is more, in all of these genes, except the human skeletal muscle gene (Hanauer et al. 1983), the more rarely used synonymous codon TGT is found (this includes the X. borealis muscle-specific actin gene mentioned above). This is the only position in which this codon is used out of the six cysteine codons of the chicken skeletal muscle gene (Fornwald et al. 1982), six cysteine codons of the sea urchin pSpG17 gene (Cooper and Crain 1982), four cysteine codons of the sea urchin gene 1 (Schuler et al. 1983), and the seven cysteine codons each of the Drosophila 79B and 88F genes (Sanchez et al. 1983). Thus, like the extra alanine codon of ala(+) genes, the role of this cysteine codon at position-1 may be in its contribution to the DNA or RNA sequence. It is possible that structural demands favour certain sequences around the start of translation of actin mRNAs.

If conservation of such a sequence takes place, it does not seem to extend, at least in amphibians, to the acidic amino acid codons at positions 2 to 4. The difference in amino acids in these positions between the cytoskeletal actins of two such closely related species as X. laevis and X. borealis (which can form viable hybrids), as well as the differences found in other amphibians (Vandekerckhove et al. 1981), appears to indicate that there are no functional differences between such isoforms. It is also possible, though on the face of it less likely, that such differences are adaptive, as the actin genes may be subject to different selective pressures even in such similar animals as X. laevis and X. borealis.

REFERENCES

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.

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

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

Bergsma, D.J., Chang, K.S. and Schwartz, R.J. (1985). Novel chicken actin gene: third cytoplasmic isoform. Mol. Cell Biol. 5, 1151-1162.

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.

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.

Chang, K.S., Rothblum, K.N. and Schwartz, R.J. (1985). The complete sequence of the chicken alpha-cardiac gene: a highly conserved vertebrate gene. Nucleic Acids Res. 13, 1223-1237.

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

Efstratiadis, A., Posakony, J.W., Maniatis, T., Lawn, R.M., O'Connell, C., Spritz, R.A., DeRiel, J.K., Forget, B.G., Weissman, S.M.,

Slightom, J.L., Blechl, A.E., Smithies, O., Baralle, F.E., Shoulders, C.C. and Proudfoot, N.J. (1980). The structure and evolution of the human beta-globin gene family. Cell 21: 653-668.

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

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

Fornwald, J.A., Kuncid, 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. Nucleic Acids Res. 10: 3861-3876.

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.

Gunning, P., Mohun, T., Ng, S.Y., Ponte, P. and Kedes, L. (1984). Evolution of human sarcomeric-actin genes: evidence for units of selection within the 3' untranslated regions of the mRNAs. J. Mol. Evol. 20: 202-214.

Hamada, H., Perrine, M.G. and Kakunaga, T. (1982). Molecular structure and evolutionary origin of human cardiac muscle actin gene. Proc. Natl. Acad. Sci. USA 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. Nucleic Acids Res. 11, 3503-3516.

Kost, T.A., Theodorakis, N. and Hughes, S.H. (1983). The nucleotide sequence of the chick cytoplasmic alpha-actin gene. Nucleic Acids Res. 11: 8287-8301.

Luse, D.S., Haynes, J.R., VanLeeuwen, 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. Nucleic Acids Res. 9: 4339-4354.

Marotta, C.A., Strocchi, P. and Gilbert, J.M. (1978). Microheterogeneity of brain cytoplasmic and synaptoplasmic actins. J. Neurochem. 30, 1441-1451.

Maxam, A.M. and Gilbert, W. (1978). Sequencing end-labelled DNA with base-specific chemical cleavages. Methods Enzymol. 65: 499-559.

Mayer, Y., Czosnek, H., Zeelon, P.E., Yaffe, D. and Nudel, U. (1984). Expression of the genes coding for the skeletal muscle and cardiac actins in the heart. Nucleic Acids Res. 12: 1087-1100.

Miller, A.D., Curran, T. and Verma, M. (1984). c-fos protein can induce cellular transformation: a novel mechanism of activation of a cellular oncogene. Cell 36: 51-60.

Nudel, U., Zakur, R., Shani, M., Neuman, S., Levy, Z. and Yaffe, D. (1983). The nucleotide sequence of the rat cytoplasmic beta-actin gene Nucleic Acids Res. 11: 1759-1771.

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: 349-349.

Ponte, P., Ng, S.Y., Engel, J., Gunning, P. and Kedes, L. (1984). Evolutionary conservation in the untranslated regions of actin mRNAs: DNA sequence of a human beta-actin cDNA. Nucleic Acids Res. 12: 1687-1696.

Pudney, M., Varma, M.G.R. and Leake, C.J. (1973). Establishment of a cell line XTC-2 from the South African Clawed Toad, Xenopus laevis. Experimentia 29: 466-467.

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. 162: 533-551.

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.

Schuler, M.A., McOsker, P. and Keller, E.B. (1983). DNA sequence of two linked actin genes of the sea urchin S. purpuratus. Mol. Cell. Biol. 3: 448-456.

Shah, D.M., Hightower, R.C. and Meagher, R.B. (1982). Complete nucleotide sequence of a soybean actin gene. Proc. Natl. Acad. Sci. USA 79: 1022-1026.

Smith, H.O. and Birnstiel, M.L. (1976). A simple method for DNA restriction site mapping. Nucleic Acids Res. 3, 2387-2398.

Southern, E.M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98: 503-517.

Theibaud, C.H. and Fischberg, M. (1977). DNA content in the genus Xenopus. Chromosoma 59: 253-257.

Tymowska, J. and Fischberg, M. (1973). Chromosome complements of the genus Xenopus. Chromosoma 44, 335-342.

Ueyama, H., Hamada, H., Battula, N. and Kakunaga, T. (1984). Structure of a human smooth muscle actin gene (aortic type) with a unique intron site. Mol. Cell. Biol. 4: 1073-1078.

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. Proc. Natl. Acad. Sci. USA 75: 1106-1110.

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. (1981). Actin typing on total cellular extracts: A highly sensitive protein-chemical procedure able to distinguish different actins. Eur. J. Biochem. 118: 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.

Vandekerckhove, J. and Weber, K. (1984). Chordate muscle actins differ distinctly from invertebrate muscle actins. J. Mol. Biol. 179: 391-413.

Wilson, A.C., Carlson, S.S. and White, T.J. (1977). Biochemical evolution. Ann. Rev. Biochem. 46: 573-639.

Woodland, H.R., Warmington, J.R., Ballantine, J.E.M. and Turner, P.C. (1984). Are there major developmentally regulated H4 gene classes in Xenopus? Nucleic Acids Res. 12: 4939-4958.

Yaffe, D., Nudel, U., Mayer, Y. and Neuman, S. (1985). Highly conserved sequences in the 3' untranslated region of mRNAs coding for homologous proteins in distantly related species. Nucleic Acids Res. 13: 3723-3737.

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.

ACKNOWLEDGEMENTS

We thank Elizabeth Ballantine for her technical assistance, Phil Turner and Bill Bains for advice and materials, Susan Easton for her preparation of the manuscript, and the Medical Research Council for financial support. Harry Erba is a predoctoral fellow of the Medical Scientist Training Program.

APPENDIX II

The Locus Possessing Homology to an Intron Probe from pXB-C1 Also Contains an Actin Gene

In Section 4.2.1 of this thesis, it was demonstrated that two loci share strong homology to the *Hind*III intron probe in pXB-C1. One of these contains the cardiac actin gene, a complete copy of which has been cloned and injected into *Xenopus* embryos (see Chapters 4-7). However, while screening genomic libraries with the *Hind*III probe, two other clones (λXB-Cr1 and λXB-Cr2) were isolated which did not have the same restriction map as λXB-C1.

DNA from these clones was digested with various restriction enzymes, electrophoresed on an agarose gel and blotted on to nitrocellulose. The Southern blot presented in Figure I has been probed with both the *Hind*III fragment and a 1.2 kb *Eco*RI fragment from λ3H1 (see Appendix I) which contains the final two exons and some intervening sequence from the *X. borealis* β-type cytoskeletal actin gene. The *Hind*III fragment appears to be uncontaminated with actin-coding sequence from pXB-C1, as it did not cross-react at all with λ3H1 or another unidentified actin clone (see Figure Ia). Thus, λXB-Cr1 and λXB-Cr2, which are both derived from the same gene locus (e.g. see Figure Ib), have not been isolated on the basis of any homology to actin exons.

However, when the filter was reprobed with the actin-specific probe, the two clones also produced a positive signal, like all the other actin

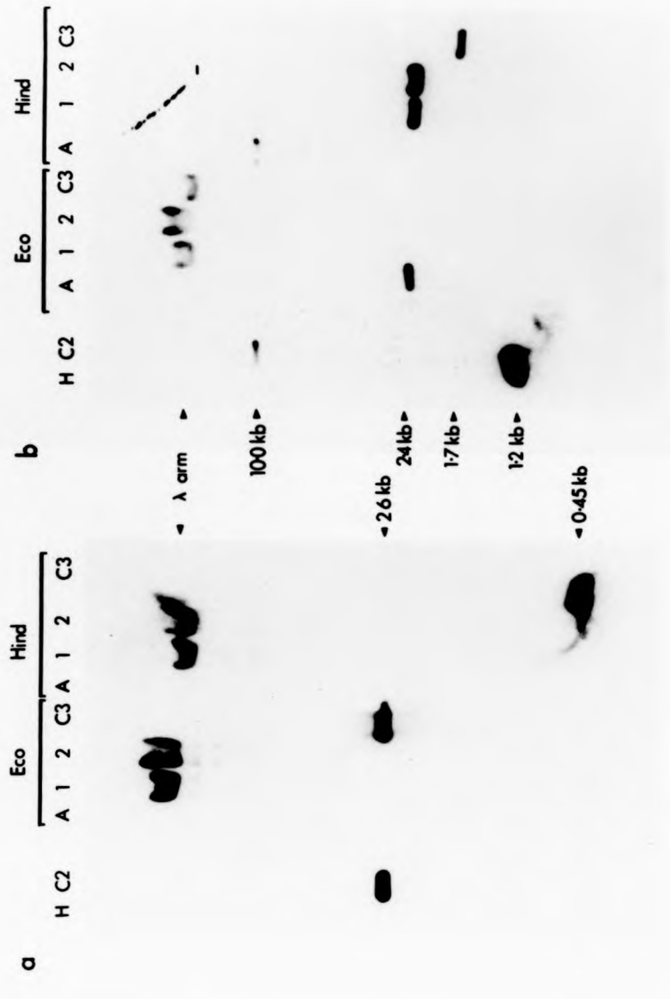


Figure I Demonstration that λXB-Cr1 and λXB-Cr2 Contain Actin Genes

Various lambda clones were digested with *Eco*RI and *Hind*III, and the resulting DnA fragments were electrophoresed on a 1.0% agarose gel and transferred to nitrocellulose. The filter was probed with the intron-specific *Hind*III fragment of pXB-C1 (a), then washed and reprobed with the actin-specific *Eco*RI fragment at the downstream end of λ3H1. Normal hybridisation and washing conditions of the filter were used for the latter probe, while, for the former, the filter was washed at 50°C in 3 x SSC.

Lambda DNAs: H = λ3H1; C2 = λXB-C2; A = an unidentified actin clone; 1 = λXB-Cr1; 2 = λXB-Cr2; C3 = λXB-C3. λ3H1 and λXB-C2 were only digested with *Eco*RI.

DNAs on the filter. It is therefore concluded that they contain actin genes too.

Only preliminary restriction mapping of λXB-Cr1 and λXB-Cr2 has yet been undertaken. λXB-Cr1 contains a slightly larger insert than λXB-Cr2. The results suggest that the region which hybridises to the *Hind*III fragment is near the end of the insert in both clones, and this has made it difficult to find internal hybridising fragments which could then be compared in size with the bands on a genomic Southern. However, a 300 bp positive *Pvu*II fragment appears to be shared by both clones and the genomic locus (data not shown).

In summary, all the data so far suggest that λXB-Cr1 and λXB-Cr2 contain the second genomic locus with strong homology to the *Hind*III intron probe. These clones include actin-coding sequences; it will be interesting to discover if the sequences encode another cardiac actin gene (see Section 4.3) and if the gene is expressed.

REFERENCES

Adelstein, R. S. (1982). Calmodulin and the regulation of the actin-myosin interaction in smooth muscle and non muscle cells. Cell 30, 349-350.

Akam, M. (1984). *Notch* mapping function obscure. Nature **317**, 113-114.

Akam, M. (1986). Mediators of cell communication? Nature **319**, 447-448.

Alonso, S., Minty, A., Bourlet, Y. and Buckingham, M. (1986). Comparison of three actin-coding sequences in the mouse; evolutionary relationships between the actin genes of warm-blooed vertebrates. J. Mol. Evol. **23**, 11-22.

Anderson, K. V., Jürgens, G. and Nüsslein-Volhard, C. (1984a). Establishment of dorsal-ventral polarity in the *Drosophila* embryo: genetic studies on the role of the *Toll* gene product. Cell 42, 779-789.

Anderson, K. V., Bokla, L. and Nüsslein-Volhard, C. (1984b). Establishment of dorsal-ventrla polarity in the *Drosophila* embryo: the induction of polarity by the *Toll* gene product. Cell 42, 791-798.

Andres, A. C., Muellener, D. B. and Ryffel, G. U. (1974). Persistence, methylation and expression of vitellogenin gene derivatives after injection into fertilised eggs of *Xenopus laevis*. Nucl. Acids Res. **12**, 2283-2302.

Atchison, M. L. and Perry, R. P. (1986). Tandem kappa immunoglobulin promoters are equally active in the presence of the kappa enhancer: implications for models of enhancer function. Cell 46, 253-262.

Awgulewitsch, A., Utset, M. F., Hart, C. P., McGinnis, W. and Ruddle, F. H. (1986). Spatial restriction in expression of a mouse homoeo box locus within the central nervous system. Nature **320**, 328-335.

Basi, G. S., Boardman, M. and Storti, R. V. (1984). Alternative splicing of a *Drosophila* tropomyosin gene generates muscle

tropomyosin isoforms with different carboxy-terminals ends. Mol. Cell. Biol. 4, 2828-2836.

Beachy, P. A. (1985). *Drosophila* proteins pave the way to gene mutations. Nature 317, 18-19.

Beachy, P. A., Helfand, S. L. and Hogness, D. S. (1985). Segmental distribution of bithorax complex proteins during *Drosophila* development. Nature 313, 545-551.

Bendig, M. M. and Williams, J. G. (1983). Replication and expression of *Xenopus laevis* globin genes injected into fertilised *Xenopus* eggs. Proc. Natl. Acad. Sci. U.S.A. 80, 6197-6201.

Bendig, M. M. and Williams, J. G. (1984). Differential expression of the *Xenopus laevis* tadpole and adult β-globin genes when injected into fertilised *Xenopus laevis* eggs. Mol. Cell. Biol. 4, 567-570.

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 λgt recombinant clones by hybridisation to single plaques *in situ*. Science 196, 180-182.

Berg, J. M. (1986). More metal-binding fingers. Nature 319, 264-265.

Bergman, Y., Rice, D., Grosschedl, R. and Baltimore, D. (1984). Two regulatory elements for immunoglobulin κ light chain gene expression. Proc. Natl. Acad. Sci. U.S.A. 81, 7041-7045.

Bergsma, D. J., Chang, K. S. and Schwartz, R. J. (1985). Novel chicken actin gene: third cytoplasmic isoform. Mol. Cell. Biol. 5, 1151-1162.

Bergsma, D. J., Grichnik, J. M., Gossett, L. M. A. and Schwartz, R. J. (1986). Delimitation and characterization of *cis*-acting DNA sequences required for the regulated expression and transcriptional control of the chicken skeletal α-actin gene. Mol. Cell. Biol. 6, 2462-2475.

Berrios, M., Osheroff, N. and Fisher, P. (1985). *In situ* localisation of DNA topoisomerase II, a major polypeptide of the *Drosophila* nuclear matrix. Proc. Natl. Acad. Sci., U.S.A. 82, 4142-4146.

Bird, A. P. (1984). DNA methylation - how important in gene control? Nature 307, 503-504.

Bird, A. P. (1986). CpG-rich islands and the function of DNA methylation. Nature 321, 209-213.

Bird, A. P., Taggart, M., Frommer, M., Miller, O. J. and Macleod, D. (1985). A fraction of the mcuse genome that is derived from islands of non methylated, CpG-rich DNA. Cell 40, 91-99.

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.

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

Bogenhagen, D. F., Wormington, W. M. and Brown, D. D. (1982). Stable transcription complexes of *Xenopus* 5S RNA genes: a means to maintain the differentiated state. Cell 28, 413-421.

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

Bray, D. (1985). Computer modelling of the behaviour of actin gels. Nature 313, 738.

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

Brent, R. and Ptashne, M. (1985). A eukaryotic transcriptional activator bearing the DNA specificty of a prokaryotic repressor. Cell 43, 729-736.

Bretscher, A. (1986). Thin filament regulatory proteins of smooth- and non-muscle cells. Nature 321, 726-727.

Briggs, M. R., Kadonaga, J. T., Bell, S. P. and Tjian, R. (1986). Purification and biochemical characterization of the promoter-specific transcription factor, Sp1. Science 224, 47-52.

Brown, D. D. (1984). The role of stable complexes that repress and activate eukaryotic genes. Cell 37, 359-365.

Brown, D. D. and Schlissel, M. S. (1985). A positive transcription factor controls the differential expression of the two 5S RNA genes. Cell 42, 759-767.

Buckingham, M. E. and Minty, A. J. (1983). Contractile protein genes, pp 365-395. In: Eukaryotic Genes: their Structure, Activity and Regulation (eds. Maclean, N., Gregory, S. P. and Flavell, R. A.), Butterworths, London.

Burdon, R. H. and Adams, R. L. P. (1980). Eukaryotic DNA methylation. TIBS 5, 294-297.

Busby, S. J. and Reeder, R. H. (1983). Spacer sequences regulate the transcription of ribosomal gene plasmids injected into *Xenopus* embryos. Cell 34, 989-996.

Busslinger, M., Hurst, J. and Flavell, R. A. (1983). DNA methylation and the regulation of globin gene expression. Cell 34, 197-206.

Butler, P. J. G. (1984). The folding of chromatin. CRC Crit. Rev. Biochem. 15, 57-91.

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

Chada, K., Magram, J., Raphael, K., Radice, G., Lacy, E. and Constantini, F. (1985). Specific expression of a foreign β-globin gene in erythroid cells of transgenic mice. Nature 314, 377-380.

Chang, K. S., Rothblum, K. N. and Schwartz, R. J. (1985). The complete sequence of the chicken α-cardiac actin gene: a highly conserved vertebrate gene. Nucl. Acids Res. 13, 1223-1237.

Chapelinsky, A. B., King, C. R., Zelenka, P. S. and Piatigorsky, J. (1985). Lens-specific expression of the chloramphenicol acetyltransferase gene promoted by 5' flanking sequences of the murine αA-crystallin gene in explanted chicken lens epithelia. Proc. Natl. Acad. Sci. U.S.A. 82, 2334-2338.

Chiu, C.-P. and Blau, H. M. (1984). Reprogramming cell differentiation in the absence of DNA synthesis. Cell 37, 879-887.

Choi, O.-R. and Engel, J. D. (1986). A 3' enhancer is required for temporal and tissue-specific transcriptional activation of the chicken adult β-globin gene. Nature 323, 731-734.

Church, G. and Gilbert, W. (1984). Genomic sequencing. Proc. Natl. Acad. Sci. U.S.A. 81, 1991-1995.

Church, G. M., Ephrussi, A., Gilbert, W. and Tonegawa, S. (1985). Cell-type-specific contacts to immunoglobulin enhancers in nuclei. Nature 313, 798-801.

Ciejek, E. M., Tsai, M.-J. and O'Malley, B. W. (1983). Actively transcribed genes are associated with the nuclear matrix. Nature 306, 607-609.

Clarke, L. and Carbon, J. (1976). A colony bank containing synthetic ColE1 hybrid plasmids representative of the entire *E. coli* genome. Cell 9, 91-99.

Clarke, M. (1984). Two states of attachment. Nature 310, 275.

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.

Cockerill, P. N. and Gerrard, W. T. (1986). Chromosomal loop anchorage of the kappa immunoglobulin gene occurs next to the enhancer in a region containing topoisomerase II sites. Cell 44, 273-282.

Collins, J. and Hohn, B. (1978). Cosmids: a type of plasmid gene-cloning vector that is packageable *in vitro* in bacteriophage λ heads. Proc. Natl. Acad. Sci. U.S.A. **75**, 4242-4246.

Cooke, J. and Webber, J. A. (1985). Dynamics of the control of body pattern in the development of *Xenopus laevis*. J. Embryol. exp. Morph. **88**, 85-112.

Coulter, D. and Wieschaus, E. (1986). Segmentation genes and the distributions of transcripts. Nature 321, 472-473.

Craig, S. W. and Pollard, T. D. (1982). Actin-binding proteins. TIBS 7, 88-92.

Cross, G. S. (1984). The structure and expression of *Xenopus borealis* actin genes. Ph.D. thesis. University of Warwick.

Cross, G. S., Wilson, C., Erba, H. P. and Woodland, H. R. (1986). Cytoskeletal actin gene families of *Xenopus borealis* and *Xenopus laevis* (submitted).

Daubas, P., Robert, B., Garner I. and Buckingham, M. (1985). A comparison between mammalian and avian fast skeletal muscle alkali light chain genes: regulatory implications. Nucl. Acids Res. 13, 4623-4643.

Davidson, E. H. (1976). Gene Activity in Early Development, 2nd edn. Academic Press, New York.

Davidson, E. H., Flytzanis, C. N., Lee, J. J., Robinson, J. J., Rose, S. J. III and Sucov, H. M. (1985). Lineage-specific gene expression in the sea urchin embryo. Cold Spring Harbor Symp. Quant. Biol. 50, 321-328.

Davidson, I., Fromental, C., Augereau, P., Wildeman, A., Zenke, M. and Chambon, P. (1986). Cell type-specific protein binding to the enhancer of simian virus 40 in nuclear extracts. Nature **323**, 544-548.

Dente, L., Cesareni, G. and Cortese, R. (1983). pEMBL: a new family of single-stranded plasmids. Nucl. Acids Res. **11**, 1645-1655.

Desplan, C., Theis, J. and O'Farrell, P. H. (1985). The *Drosophila* developmental gene, *engrailed*, a sequence-specific DNA binding activity. Nature **318**, 630-635.

Dodemont, H. J., Soriano, P., Quax, W. J., Ramaekers, T., 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. EMBO J. **1**, 167-171.

Doyle, H. J., Harding, K., Hoey, T. and Levine, M. (1986). Transcripts encoded by a homoeo box gene are restricted to dorsal tissues of *Drosophila* embryos. Nature **323**, 76-79.

Duboule, D., Baron, A., Mähl, P. and Galliot, B. (1986). A new homeo-box is present in overlapping cosmid clones which define the mouse *Hox-1* locus. EMBO J. **5**, 1973-1980.

Dworkin, M. B., Shrutkowski, A., Baumgarten, M. and Dworkin-Rastle, E. (1984). The accumulation of prominent tadpole mRNAs occurs at the beginning of neurulation in *Xenopus laevis* embryos. Dev. Biol. **106**, 289-295.

Dynan, W. S. (1985). A single protein that binds to enhancers, promoters and replcation origins? TIG **1**, 269-270.

Dynan, W. S. and Tjian, R. (1983). The promoter-specific transcription factor Spl binds to upstream sequences in the SV40 early promoter. Cell **35**, 79-87.

Dynan, W. S. and Tjian, R. (1985). Control of eukaryotic messenger RNA synthesis by sequence-specific DNA-binding proteins. Nature 316, 774-778.

Dynan, W. S., Saffer, J. D., Lee, W. S. and Tjian, R. (1985). Transcription factor Spl recognises promoter sequences from the monkey genome that are similar to the simian virus 40 promoter. Proc. Natl. Acad. Sci. U.S.A. 82, 4915-4919.

Dynan, W. S., Sazer, S., Tjian, R. and Schimke, R. T. (1986). The transcription factor Spl reccgnises a DNA sequence in mouse dihydrofolate reductase promoter. Nature 319, 246-248.

Earnshaw, W. C., and Heck, M. M. S. (1985). Localization of topoisomerase II in mitotic chromosomes. J. Cell Biol. 100, 1716-1725.

Earnshaw, W. C., Halligan, B., Cooke, C. A., Heck, M. M. S. and Lui, L. F. (1985). Topoisomerase II is a structural component of mitotic chromosome scaffolds. J. Cell Biol. 100, 1706-1715.

Egly, J. M., Miyamoto, N., Moncollin, V. and Chambon, P. (1984). Is actin a transcription initiation factor for RNA polymerase B? EMBO J. 3, 2363-2371.

Eldrige, J., Zehner, Z. and Paterson, B. M. (1985). Nucleotide sequence of the chicken cardiac α-actin gene: absence of strong homologies in the promoter and 3'-untranslated regions with the skeletal α-actin sequence. Gene 36, 55-63

Elgin, S. C. R. (1984). Anatomy of hypersensitive sites. Nature 309, 213-214.

Emerson, B. M. and Felsenfeld, G. (1984). Specific factor conferring nuclease hypersensitivity at the 5' end of the chicken adult β-globin gene. Proc. Natl. Acad. Sci. U.S.A. 81, 95-99.

Emerson, B. M., Lewis, C. D. and Felsenfeld, G. (1985). Interaction of specific nuclear factors with the nuclease-hypersensitive region of the chicken adult β-globin gene: nature of the binding domain. Cell 41, 21-30.

Engel, J. N., Gunning, P. W. and Kedes, L. M. (1981). Isolation and characterisation of human actin genes. Proc. Natl. Acad. Sci. U.S.A. 78, 4674-4678.

Engels, W. R. (1983). The P family of transposable elements in *Drosophila*. Ann. Rev. Genet. 17, 315-344.

Engels, W. R. (1984). A *trans*-acting product needed for P factor transposition in *Drosophila*. Science 226, 1194-1196.

Ephrussi, A., Church, G. M., Tonegawa, S. and Gilbert, W. (1985). B lineage-specific interactions of an immunoglobulin enhancer with cellular factors *in vivo* Science 227, 134-140.

Erba, H. P., Gunning, P. and Kedes, L. (1986). Nucleotide sequence of the human γ cytoskeletal actin mRNA: anomalous evolution of vertebrate non-muscle actin genes. Nucl. Acids Res. 14, 5275-5294.

Etkin, L. D., Pearman, B., Roberts M. and Bektesh, S. L. (1984). Replication, integration and expression of exogenous DNA injected into fertilised eggs of *Xenopus laevis*. Differentiation 26, 194-202.

Everett, R. D., Baty, D. and Chambon, P. (1983). The repeated GC-rich motifs upstream from the TATA box are important elements of the SV40 early promoter. Nucl. Acids Res. 11, 2447-2464.

Ferguson, B., Jones, N., Richter, J. and Rosenberg, M. (1984). Adenovirus Ela gene product expressed at high levels in *Escherichia coli* is functional. Science 224, 1343-1346.

Flytzanis, C. N., McMahon, A. P., Hough-Evans, B. R., Katula, K. S., Britten, R. J. and Davidson, E. H. (1985). Persistence and integration of cloned DNA in postembryonic sea urchins. Dev. Biol.

108, 431-442.

Forbes, D. J., Kirschner, M. W. and Newport, J. W. (1983). Spontaneous formation of nucleus-like structures around bacteriophage DNA microinjected into *Xenopus* eggs. Cell 34, 13-23.

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

Friedman, R. L. and Stark, G. R (1985). α-interferon-induced transcription of HLA and metallothionein genes containing homologous upstream sequences. Nature 314, 637-639.

Frischauf, A. M., Lehrach, H., Poustka, A. and Murray, N. E. (1983). Lambda replacement vectors carrying polylinker sequences. J. Mol. Biol. 170, 827-842.

Fyrberg, E. A., Bond, B. J., Hershey, N. D., Mixter, K. S. and Davidson, N. (1981). The actin genes of *Drosophila*: protein coding sequences 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.

Gallwitz, D. and Seidel, R. (1980). Molecular cloning of the actin gene from the yeast *Saccharomyces cerevisisiae*. Nucl. Acids 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*. Proc. Natl. Acad. Sci. U.S.A. 77, 2546-2550.

Garel, A., Zolan, M. and Axel, R. (1977). Genes transcribed at diverse rates have a similar conformation in chromatin. Proc. Natl. Acad. Sci. U.S.A. 74, 4867-4871.

Garner, M. M. and Revzin, A. (1986). The use of gel electrophoresis to detect and study nucleic acid-protein interactions. TIBS 11, 395-396.

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

Gasser, S. M. and Laemmli, U. K. (1986a). The organisation of chromatin loops: characterisation of a scaffold attachment site. EMBO J. 5, 511-517.

Gasser, S. M. and Laemmli, U. K (1986b). Cohabitation of scaffold binding regions with upstream/enhancer elements of three developmentally regulated genes of *D. melanogaster*. Cell 46, 521-530.

Gehring, W. J. (1985). The molecular basis of development. Sci. Am. 253, 136-146.

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 estrogen-inducible vitellogenin genes are expressed when totally methylated. Cell 33, 43-51.

Gerhart, J. G. (1980). Mechanisms regulating pattern formation in the amphibian egg and early embryo, pp 135-315. In: Biological Regulation and Development, II, (eds. Goldberger, R. F.), Plenum Press, New York.

Gilbert, D. M. (1986). Temporal order of replication of *Xenopus laevis* 5S ribosomal RNA genes in somatic cells. Proc. Natl. Acad. Sci. U.S.A. 83, 2924-2928.

Ginsberg, A. M., King, B. O. and Roeder, R. G. (1984). *Xenopus* 5S gene transcription factor, TFIIIA: characterization of a cDNA clone and measurement of RNA levels throughout development. Cell 39, 479-489.

Goldberg, D. A., Posakony, J. and Maniatis, T. (1983). Correct developmental expression of a cloned alcohol dehydrogenase gene transduced into the *Drosophila* germ line. Cell 34, 59-73.

Goldman, R. D., Schloss, J. A. and Starger, J. M. (1976) Organisational changes of actin-like microfilaments during animal cell movement, pp 217-240. In: Cell Motility, Book A, Cold Spring Harbor Conferences on Cell Proliferation, Vol. 3, (eds. Goldman, R., Pollard, T. and Rosenbaum, J.), Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Goodbourn, S., Burstein, H. and Maniatis, T. (1986). The human β-interferon gene enhancer is under negative control. Cell 45, 601-610.

Gough, N. (1981). The rearrangements of the immunoglobulin genes. TIBS 6, 203-205.

Greenwald, I. (1985). *lin12*, a nematode homeotic gene, is homologous to a set of mammalian proteins that includes epidermal growth factor. Cell 43, 583-590.

Grichnik, J. M., Bergsma, D. J. and Schwartz, R. J. (1986). Tissue restricted and stage specific transcription is maintained within 411 nucleotides flanking the 5' end of the chicken α-skeletal actin gene. Nucl. Acids Res. 14, 1683-1701.

Grosschedl, R. and Birnstiel, M. L. (1980). Identification of regulatory sequences in the prelude sequences of an H2A histone gene by the study of specific deletion mutants *in vivo*. Proc. Natl. Acad. Sci. U.S.A. 77, 1432-1436.

Grosschedl, R., Weaver, D., Baltimore, D. and Constantini, F. (1984). Introduction of a μ immunoglobulin gene into the mouse germ line: specific expression in lymphoid cells and synthesis of functional antibody. Cell 38, 647-658.

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. Acids Res. 10, 4951-4971.

Guinta, D. R., Tso, J. Y., Narayanswami, S., Hamkalo, B. A. and Korn, L. J. (1986). Early replication and expression of oocyte-type 5S RNA genes in *Xenopus* somatic cell line carrying a translocation. Proc. Natl. Acad. Sci. U.S.A. 83, 5150-5154.

Gunning, P., Ponte, P., Blau, H. and Kedes, L. (1983). α-skeletal and α-cardiac actin genes are coexpressed in adult human skeletal muscle and heart. Mol. Cell. Biol. 3, 1985-1995.

Gunning, P., Mohun, T., Ng, S.-Y., Ponte, P. and Kedes, L. (1984a). Evolution of human sarcomeric-actin genes: evidence for units of selection within the 3′ untranslated regions of the mRNAs. J. Mol. Biol. 20, 202-214.

Gunning, P., Ponte, P., Kedes, L., Hickey, R. J. and Skoultchi, A. I. (1984b). Expression of human cardiac actin in mouse L cells: a sarcomeric actin associates with a nonmuscle cytoskeleton. Cell 36, 709-715.

Gurdon, J. B. (1974). The Control of Gene Expression in Animal Development. Clarendon Press, Oxford.

Gurdon, J. B. and Melton, D. A. (1981). Gene transfer in amphibian eggs and oocytes. Ann. Rev. Genet. 15, 189-218.

Gurdon, J. B., Brennan, S., Fairman, S. and Mohun, T. J. (1984). Transcription of muscle-specific actin genes in early *Xenopus* development: nuclear transplantation and cell dissociation. Cell 38, 691-700.

Gurdon, J. B., Mohun, T. J., Fairman, S. and Brennan, S. (1985a). All components required for the eventual activation of muscle-specific actin genes are localised in the subequatorial region of an uncleaved egg. Proc. Natl. Acad. Sci. U.S.A. **82**, 139-143.

Gurdon, J. B., Fairman, S., Mohun, T. J. and Brennan, S. (1985b). Activation of muscle-specific actin genes in *Xenopus* development by an induction between animal and vegetal cells of a blastula. Cell **41**, 913-922.

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., Petrino, M. G. and Kakunaga, T. (1982). Molecular structure and evolutionary origin of the human cardiac muscle actin gene. Proc. Natl. Acad. Sci. U.S.A. **79**, 5901-5905.

Hammer, R. E., Pursel, V. G., Rexroad, Jr., C. E., Wall, R. J., Bolt, D. J., Ebert, K. M., Palmiter, R. D. and Brinster, R. L. (1985). Production of transgenic rabbits, sheep and pigs by microinjection. Nature **315**, 680-683.

Hanauer, A., Levin, M., Heilig, R., Daegelen, D., Kahn, A. and Mandel, J. L. (1983). Isolation and characterisation of cDNA clones for human skeletal muscle α-actin. Nucl. Acids Res. **11**, 3503-3516.

Harbers, K., Schnieke, H., Stuhlmann, H., Jähner, D. and Jaenisch, R. (1981). DNA methylation and gene expression: endogenous retroviral genome becomes infectious after molecular cloning. Proc. Natl. Acad. Sci. U.S.A. **78**, 7609-7613.

Harland, R. M., Weintraub, H. and McKnight, S. L. (1983). Transcription of DNA injected into *Xenopus* oocytes is influenced by template topology. Nature **302**, 38-43.

Hartshorne, T. A., Blumberg, H. and Young, E. T. (1986). Sequence homology of the yeast regulatory protein ADR1 with *Xenopus* transcription factor TFIIIA. Nature 320, 283-287.

Harvey, R. P., Tabin, C. J. and Melton, D. A. (1986). Embryonic expression and nuclear localization of *Xenopus* homeobox (*Xhox*) gene products. EMBO J. 5, 1237-1244.

Hayward, L. J. and Schwartz, R. J. (1986). Sequential expression of chicken actin genes during myogenesis. J. Cell Biol. 102, 1485-1493.

Hiromi, Y. and Hotta, Y. (1985). Actin gene mutations in *Drosophila*: heat shock activation in the indirect flight muscles. EMBO J. 4, 1681-1687.

Hiromi, Y., Kuroiwra, A. and Gehring, W. (1985). Control elements of the *Drosophila* segmentation gene *fushi tarazu*. Cell 43, 603-613.

Hiromi, Y., Okamoto, H., Gehring, W. J. and Hotta, Y. (1986). Germline transformation with *Drosophila* mutant actin genes induces constitutive expression of heat shock genes. Cell 44, 293-301.

Hohn, B. and Murray, K. (1977). Packaging recombinant DNA molecules into bacteriophage particles *in vitro*. Proc. Natl. Acad. Sci. U.S.A. 74, 3259-3263.

Howard, K. and Ingham, P. (1986). Regulatory interactions between the segmentation genes *fushi tarazu*, *hairy*, and *engrailed* in the *Drosophila* blastoderm. Cell 44, 949-957.

Hu, M. C.-T., Sharp, S. B. and Davidson, N. (1986). The complete sequence of the mouse skeletal α-actin gene reveals several conserved and inverted repeat sequences outside of the protein-coding region. Mol. Cell Biol. 6, 15-25.

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

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

Izant, J. G. and Weintraub, H. (1984). Inhibition of thymidine kinase gene expression by anti-sense RNA: a molecular approach to genetic analysis. Cell **36**, 1007-1015.

Jackson, D. A. (1986). Organization beyond the gene. TIBS **11**, 249-252.

Jackson, D. A. and Cook, P. R. (1985a). A general method for preparing chromatin containing intact DNA. EMBO J. **4**, 913-918.

Jackson, D. A. and Cook, P. R. (1985b). Transcription occurs at a nucleoskeleton. EMBO J. **4**, 919-925.

Jackson, I. J., Schofield, P. and Hogan, B. (1985). A mouse homoeo box gene is expressed during embryogenesis and in adult kidney. Nature **317**, 745-748.

Jakobovits, E. B., Bratosin, S. and Aloni, Y. (1980). A nucleosome-free region in SV40 minichromosomes. Nature **285**, 263-265.

Jaynes, J. B., Chamberlain, J. S., Buski, J. N., Johnson, J. E. and Hauschka, S. D. (1986). Transcriptional regulation of the muscle creatine kinase gene and regulated expression in transfected mouse myoblasts. Mol. Cell Biol. **6**, 2855-2864.

Jonas, E., Sargent, T. D. and Dawid, I. B. (1985). Epidermal keratin gene expressed in embryonic *Xenopus laevis*. Proc. Natl. Acad. Sci. U.S.A. **82**, 5413-5417.

Jones, E. A. and Woodland, H. R. (1986). Development of the ectoderm in *Xenopus*: tissue specification and the role of cell association and division. Cell **44**, 345-355.

Jones, K. A. and Tjian, R. (1985). Sp1 binds to promoter sequences and activates HSV "immediate-early" gene transcription *in vitro*. Nature **317**, 179-182.

Jones, K. A., Yamamoto, K. R. and Tjian, R. (1985). Two distinct transcription factors bind to the HSV thymidine kinase promoter *in vitro*. Cell **42**, 559-572.

Jones, O. W. and Berg, P. (1966). Studies on the binding of RNA polymerase to polynucleotides. J. Mol. Biol. **22**, 199-209.

Jones, P. A. (1985). Altering gene expression with 5-azacytidine. Cell **40**, 485-486. Cell **40**, 485-486.

Jongstra, J., Reudelhuber, T. L. Oudet, P., Benoist, C., Chae, C.-B., Jeltsch, J.-M., Mathis, D. J. and Chambon, P. (1984). Induction of altered chromatin structures by simian virus 40 enhancer and promoter elements. Nature **307**, 708-714.

Jost, J. P. and Seldran, M. (1984). Association of transcriptionally active vitellogenin II gene with the nuclear matrix of liver. EMBO J **3**, 2005-2008.

Joyner, A. L., Lebo, R. V., Kan, Y. W., Tjian, R., Cox, D. R. and Martin, G. R. (1985). Comparative chromosome mapping of a conserved homeo box region in mouse and human. Nature **314**, 173-175.

Jürgens, G. (1985). A group of genes controlling the spatial expression of the bithorax complex in *Drosophila*. Nature **316**, 153-155.

Karin, M., Haslinger, A., Holtgreve, A., Richards, R. I., Krauter, P., Westphal, H. M. and Beato, M. (1984). Characterization of DNA sequences through which cadmium and glucocortioid hormones induce human metallothionein - II_A gene. Nature **308**, 513-519.

Karn, J., Brenner, S., Barnett, L. and Cesareni, G. (1980). Novel bacteriophage λ vector. Proc. Natl. Acad. Sci. U.S.A. **77**, 5172-5176.

Keegan, L., Gill, G. and Ptashne, M. (1986). Separation of DNA binding from the transcription-activating function of a eukaryotic regulatory protein. Science **231**, 699-704.

Keshet, I., Yisraeli, J. and Cedar, H. (1985). Effect of regional DNA methylation on gene expression. Proc. Natl. Acad. Sci. U.S.A. 82, 2560-2564.

Keshet, I., Lieman-Hurwitz, J. and Cedar, H. (1986). DNA methylation affects the formation of active chromatin. Cell 44, 535-543.

Kilchherr, F., Baumgartner, S., Bopp, D., Frei, E. and Noll, M. (1986). Isolation of the *paired* gene in *Drosophila* and its spatial expression during early embryogenesis. Nature 321, 493-499.

Killary, A. M. and Fournier, R. E. K. (1984). A genetic analysis of extinction: *trans*-dominant loci regulate expression of liver-specific traits in hepatoma hybrid cells. Cell 38, 523-534.

Kimmel, C. B. and Warga, R. M. (1986). Tissue-specific cell lineages originate in the gastrula of the zebra fish. Science 231, 365-368.

Kirschner, M. and Mitchison, T. (1986). Beyond self-assembly: from microtubules to morphogenesis. Cell 45, 329-342.

Klein, S., Sablitzky, F. and Rajewsky, K. (1984). Deletion of the IgH enhancer does not reduce immunoglobulin heavy chain production of a hybridoma IgD class switch variant. EMBO J. 3, 2473-2476.

Klein, S., Gerster, T., Picard, D., Radbruch, A. and Schaffner, W. (1985). Evidence for transient requirement of the IgH enhancer. Nucl. Acids Res. 13, 8901-8912.

Kmiec, E. B. and Worcel, A. (1985). The positive transcription factor of the 5S RNA gene induces a 5S DNA-specific gyration in *Xenopus* oocyte extracts. Cell 41, 945-953.

Kolata, G. (1984). Globin gene studies create a puzzle. Science 223, 470-471.

Korn, L. J. and Gurdon, J. B. (1981). The reactivation of developmentally inert 5S genes in somatic nuclei injected into *Xenopus* oocytes. Nature 289, 461-465.

Korn, L. J., Gurdon, J. B. and Price, J. (1982). Oocyte extracts reactivate developmentally inert *Xenopus* 5S genes in somatic nuclei. Nature 300, 354-355.

Kost, T. A., Theodorakis, N. and Hughes, S. H. (1983). The nucleotide sequence of the chick cytoplasmic β-actin gene. Nucl. Acids Res. 11, 8287-8301.

Krieg, P. A. and Melton, D. A. (1984). Functional messenger RNAs are produced by SP6 *in vitro* transcription of cloned cDNAs. Nucl. Acids Res. 12, 7057-7070.

Krieg, P. A. and Melton, D. A. (1985). Developmental regulation of a gastrula-specific gene injected into fertilised *Xenopus* eggs. EMBO J. 4, 3463-3471.

Kruczek, I. and Doerfler, W. (1983). Expression of the chloramphenicol acetyltransferase gene in mammalian cells under the control of adenovirus type 12 promoters. Effect of promoter methylation on gene expression. Proc. Natl. Acad. Sci. U.S.A. 80, 7586-7590.

Krumlauf, R., Hammer, R. E., Tilghman, S. M. and Brinster, R. L. (1985). Developmental regulation of α-fetoprotein genes in transgenic mice. Mol. Cell. Biol. 5, 1639-1648.

Landolfi, N. F., Capra, J. D. and Tucker, P. A. (1986). Interaction of cell-type-specific nuclear proteins with immunoglobulin V_H promoter region sequences. Nature 323, 548-551.

Lane, C. D., Marbaix, G. and Gurdon, J. B. (1971). Rabbit haemoglobin synthesis in frog cells: the translation of reticulocyte 9S RNA in frog oocytes. J. Mol. Biol. 61, 73-91.

Larsen, A. and Weintraub, H. (1982). An altered DNA conformation detected by S1 nuclease occurs at specific regions in active chick globin chromatin. Cell 29, 609-622.

Laskey, R. A., Mills, A. D. and Morris, N. R. (1977). Assembly of SV40 chromatin in a cell-free system from *Xenopus* eggs. Cell 10, 237-243.

Laski, F. A., Rio, D. C. and Rubin, G. M. (1986). Tissue specificity of *Drosophila* P element transposition is regulated in prokaryotes and eukaryotes. Cell 44, 107-111.

Laughon, A. and Scott, M. P. (1984). Sequence of a *Drosophila* segmentation gene: protein structure homology with DNA-binding proteins. Nature 310, 25-31.

Lawrence, J. B. and Singer, R. H. (1986). Intracellular localization of messenger RNAs for cytoskeletal proteins. Cell 45, 407-415.

Lazarides, E. and Weber, K. (1974). Actin antibody: the specific visualization of actin filaments in non-muscle cells. Proc. Natl. Acad. Sci. U.S.A. 71, 2268-2272.

Leavitt, J., Leavitt, A. and Atallah, A. M. (1980). Dissimilar modes of expression of β- and γ-actin in normal and leukemic human T lymphocytes. J. Biol. Chem. 255, 4984-4987.

Lewin, R. (1984). Why is development so illogical? Science 224, 1327-1329.

Lin, F.-L., Sperle, K. and Sternberg, N. (1985). Recombination in mouse L cells between DNA introduced into cells and homologous chromosomal sequences. Proc. Natl. Acad. Sci. U.S.A. 82, 1391-1395.

Lin, S. and Riggs, A. D. (1975). The general affinity of *lac* repressor for *E. coli* DNA: implications for gene regulation in procaryotes and eukaryotes. Cell 4, 107-111.

Loenen, W. A. and Brammar, W. J. (1980). A bacteriophage lambda vector for cloning large DNA fragments made with several restriction enzymes. Gene 10, 249-259.

Luchnick, A. N., Bakayer, V. V., Yugai, A. A., Zbarsky, I. B. and Georgiev, G. P. (1985). DNase I hypersensitive minichromosomes of

SV40 possess an elastic torsional strain in DNA. Nucl. Acids Res. 13, 1135-1149.

Luse, D. S. and Roeder, R. G. (1980). Accurate transcription initiation on a purified mouse β-globin DNA fragment in a cell-free system. Cell 20, 691-699.

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. Acids Res. 9, 4339-4354.

Lusky, M. and Botchan, M. (1981). Inhibition of SV40 replication in simian cells by specific pBR322 DNA sequences. Nature 293, 79-81.

Mace, H. A. F., Pelham, H. R. B. and Travers, A. A. (1983). Association of an S1 nuclease-sensitive structure with short direct repeats 5′ of *Drosophila* heat shock genes. Nature 304, 555-557.

Macleod, D. and Bird, A. P. (1983). Transcription in oocytes of highly methylated rDNA from *Xenopus laevis* sperm. Nature 306, 200-203.

Mandel, J. and Chambon, P. (1979). DNA methylation: organ specific variations in the methylation pattern within and around ovalbumin and other chicken genes. Nucl. Acids Res. 7, 2081-2103.

Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982). Molecular Cloning - a Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Mason, J. O., Williams, G. T. and Neuberger, M. S. (1985). Transcription cell type specificity is conferred by an immunoglobulin V_H gene promoter that includes a functional consensus sequence. Cell 41, 479-487.

Mattaj, I. W., Leinherd, S., Jiricny, J. and De Robertis, E. M. (1985). An enhancer-like sequence within the *Xenopus* U2 gene promoter facilitates the formation of stable transcription complexes. Nature

316, 163-167.

Maxam, A. M. and Gilbert, W. (1978). Sequencing end-labelled DNA with base-specific chemical cleavages. Meth. in Enz. 65, 499-559.

Mayer, Y., Czosnek, H., Zeelon, P. E., Yaffe, D. and Nudel, U. (1984). Expression of the genes coding for skeletal muscle and cardiac actins in the heart. Nucl. Acid. Res. 12, 1087-1100.

McGhee, J. D. and Ginder, G D. (1979). Specific DNA methylation sites in the vicinity of the chicken β-globin genes. Nature 280, 419-420.

McGhee, J. D., Wood, W., Dolan, M., Engel, J. D. and Felsenfeld, G. (1981). A 200 base pair region at the 5' end of the chicken adult β-globin gene is accessible to nuclease digestion. Cell 27, 45-55.

McGinnis, W., Levine, M. S., Hafen, E., Kuroiwa, A. and Gehring, W. J. (1984a). A conserved DNA sequence in the homoeotic genes of the *Drosophila* Antennapedia and bithorax complexes. Nature 308, 428-433.

McGinnis, W., Garber, R. L., Wirz, J., Kuroiwa, A. and Gehring, W. J. (1984b). A homologous protein-coding sequence in *Drosophila* homeotic genes and its conservation in other metazoans. Cell 37, 403-408.

McKnight, S. L. and Gavis, E. R. (1980). Expression of the herpes thymidine kinase gene in *Xenopus laevis* oocytes: an assay for the study of deletion mutants constructed *in vitro*. Nucl. Acids Res. 8, 5931-5948.

McKnight, S. L. and Kingsbury, R. (1982). Transcriptional control signals of a eukaryotic protein-coding gene. Science 217, 316-324.

McKnight, G. S. and Palmiter, R. D. (1979). Transcriptional regultaion of the ovalbumin and conalbumin genes by steroid hormones in chick oviduct. J. Biol. Chem. 254, 9050-9058.

McKnight, S. and Tjian, R. (1986). Transcriptional selectivity of viral genes in mammalian cells. Cell 46, 795-805.

Medford, R. M., Nguyen, H. T., Destree, A. T., Summers, E. and Nadal-Ginard, B. (1984). A novel mechanism of alternative RNA splicing for the developmentally regulated generation of troponin T isoforms from a single gene. Cell **38**, 409-421.

Melloul, D., Aloni, B., Calvo, J., Yaffe, D. and Nudel, U. (1984). Developmentally regulated expression of chimeric genes containing muscle actin DNA sequences in transfected myogenic cells. EMBO J. **3**, 983-990.

Melton, D. A. (1985). Injected anti-sense RNAs specifically block messenger RNA translation *in vivo*. Proc. Natl. Acad. Sci. U.S.A. **82**, 144-148.

Melton, D. A., Krieg, P. A., Rebagliati, M. R., Maniatis, T., Zinn, K. and Green, M. R. (1984). Efficient *in vitro* synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. Nucl. Acids Res. **12**, 7035-7056.

Miller, J., McLachlan, A. D. and Klug, A. (1985). Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. EMBO J. **4**, 1609-1614.

Miller, J. H. and Reznikoff, W. S. Eds. (1978). The Operon. Cold Spring Harbor Laboratory, New York.

Minty, A. and Kedes, L. (1986). Upstream regions of the human cardiac actin gene that modulate its transcription in muscle cells: presence of an evolutionarily conserved repeated motif. Mol. Cell. Biol. **6**, 2125-2136.

Minty, A. J., Alonso, S., Caravatti, M. and Buckingham, M. E. (1982). A fetal skeletal muscle mRNA in the mouse and its identity with cardiac actin 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, 77-103.

Minty, A., Blau, H. and Kedes, L. (1986). Two-level regulation of cardiac actin gene transcription: muscle-specific modulating factors can accumulate before gene activation. Mol. Cell. Biol. 6, 2146-2156.

Mlodzik, M., Fjose, A. and Gehring, W. J. (1985). Isolation of *caudal*, a *Drosophila* homeo box-containing gene with maternal expression, whose transcripts form a concentration gradient at the pre-blastoderm stage. EMBO J. 4, 2961-2969.

Mohun, T. J., Brennan, S., Dathan, N., Fairman, S. and Gurdon, J. B. (1984). Cell type-specific activation of actin genes in the early amphibian embryo. Nature 311, 716-721.

Müller, M. M., Carrasco, A. E. and De Robertis, E. M. (1984). A homeo-box-containing gene expressed during oogenesis in *Xenopus*. Cell 39, 157-162.

Murray, N. E. (1983). Phage lambda and molecular cloning. In: The Bacteriophage Lambda II. (ed. Hershey, A. D.), Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Nabeshima, Y., Fujii-Kuriyama, Y., Muramatsu, M. and Obata, K. (1984). Alternative transcription and two modes of splicing result in two myosin light chains from one gene. Nature 308, 333-338.

Nakajima-Iijima, S., Hamada, H., Reddy, P. and Kakanuga, T. (1985). Molecular structure of the human cytoplasmic β-actin gene: interspecies homology of sequences in the introns. Proc. Natl. Acad. Sci. U.S.A. 82, 6133-6137.

Nakamura, O., Takasaki, H. and Nagata, A. (1978). Further studies of the prospective fate of blastomeres at the 32-cell stage of *Xenopus*

laevis embryos. Med. Biol. **56**, 355-366.

Nasmyth, K. A. (1982). Molecular genetics of yeast mating type. Ann. Rev. Genet. **16**, 439-500.

Newport, J. and Kirschner, M. (1982a). A major developmental transition in early *Xenopus* embryos: I. Characterisation and timing of cellular changes at the midblastula stage. Cell **30**, 675-685.

Newport, J. and Kirschner, M. (1982b). A major developmental transition in early *Xenopus* embryos: II. Control of the onset of transcription. Cell **30**, 687-696.

Ng, R. and Abelson, J. (1980). Isolation and sequence of the gene for actin in *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. U.S.A. **77**, 3912-3916.

Ng, S.-Y., Gunning, P., Eddy, R., Ponte, P., Leavitt, J., Shows, T. and Kedes, L. (1985). Evolution of the functional human β-actin gene and its multi-pseudogene family: conservation of non-coding regions and the chromosomal dispersion of pseudogenes. Mol. Cell Biol. **5**, 2720-2732.

Nickol, J. M. and Felsenfeld, G. (1983). DNA conformation at the 5' end of the chicken adult β-globin gene. Cell **35**, 467-477.

Nieuwkoop, P. D. (1969). The formation of the mesoderm in urodelean amphibians. I. Induction by the endoderm. Wilhelm Roux' Archiv. **162**, 341-373.

Nieuwkoop, P. D. and Faber, J. (1956). Normal Table of *Xenopus laevis* (Daudin), Amsterdam: North-Holland.

Noma, Y., Sideras, P., Naito, T., Bergstedt-Lindquist, S., Azuma, C., Severinson, E., Tanabe, T., Kinashi, T., Matsuda, F., Yaoita, Y. and Honjo, T. (1986). Cloning of cDNA encoding the murine IgG1 induction factor by a novel strategy using SP6 promoter. Nature **319**, 640-646.

North, G. (1986). Descartes and the fruitfly. Nature **322**, 404-405.

Nudel, U., Zakut, R., Shani, M., Neuman, S., Levy, Z. and Yaffe, D. (1983). The nucleotide sequence of the rat cytoplasmic β-actin gene. Nucl. Acids Res. 11, 1759-1771.

Nüsslein-Volhard, C. and Wieschaus, E. (1980). Mutations affecting segment number and polarity in *Drosophila*. Nature 287, 795-801.

Ōgi, K. I. (1969). Regulative capacity in the early amphibian embryo. Najoya Univ. Dept. Gen. Educ. Res. Bull. 13, 31-40.

Ordahl, C. P. and Cooper, T. A. (1983). Strong homology in promoter and 3' untranslated regions of chick and rat α-actin genes. Nature 303, 348-349.

Orr-Weaver, T. L., Szostak, J. W. and Rothstein, R. J. (1981). Yeast transformation: a model system for the study of recombination. Proc. Natl. Acad. Sci. U.S.A. 78, 6354-6358.

Otey, C. A., Kalnoski, M. H., Lessard, J. L. and Bulinski, J. C. (1986). Immunolocalisation of the gamma isoform of nonmuscle actin in cultured cells. J. Cell Biol. 102, 1726-1737.

Palmiter, R. D. and Brinster, R. L. (1985). Transgenic mice. Cell 41, 343-345.

Palmiter, R. D., Mulvihill, E., McKnight, S. and Senear, A. (1977). Regulation of gene expression in the chick oviduct by steroid hormones. Cold Spring Harb. Symp. Quant. Biol. 42, 639-647.

Pardo, J. V., Pittenger, M. F. and Craig, S. W. (1983). Subcellular sorting of isoactins: selective association of γ actin with skeletal muscle mitochondria. Cell 32, 1093-1103.

Parker, C. S. and Topol, J. (1984a). A *Drosophila* RNA polymerase II transcription factor contains a promoter-region-specific DNA-binding activity. Cell 36, 357-369.

Parker, C. S. and Topol, J. (1984b). A *Drosophila* RNA polymerase II transcription factor binds to the regulatory site of an *hsp* 70 gene.

Cell 37, 273-283.

Pasteels, J. (1939). Sur l'origine du matérial candal des Urodèles. Bull. roy. Acad. Belg. Cl. Sci. 25, 660-666.

Paterson, B. M. and Eldridge, J. D. (1984). α-cardiac actin is the major sarcomeric isoform expressed in embryonic avian skeletal muscle. Science 224, 1436-1438.

Paulson, J. R. and Laemmli, U. K. (1977). The structure of histone depleted metaphase chromosomes. Cell 12, 817-828.

Payvar, F., DeFranco, D., Firestone, G. L., Edgar, B., Wrange, O., Okret, S., Gustatsson, J.-A. and Yamamota, K. R. (1983). Sequence-specific binding of glucocorticoid receptor to MTV DNA at sites within and upstream of the transcribed region. Cell 35, 381-392.

Pelham, H. R. B., Wormington, W. M. and Brown, D. D. (1981). Similar 5S RNA transcription factors in *Xenopus* oocytes and somatic cells. Proc. Natl. Acad. Sci. U.S.A. 78, 1760-1764.

Plon, S. E. and Wang, J. C. (1986). Transcription of the human β-globin gene is stimulated by an SV40 enhancer to which it is physiclly linked but topologically uncoupled. Cell 45, 575-580.

Plumb, M. A., Nicolas, R. H., Wright, C. A. and Goodwin, G. H. (1985). Multiple sequence-specific DNA binding activities are eluted from chicken nuclei at low ionic strengths. Nucl. Acids Res. 13, 4047-4065.

Pollard, T. D. and Craig, S. W. (1982). Mechanism of actin polymerization. TIBS 7, 55-58.

Porter, S. D. and Smith, M. (1986). Homoeo-domain homology in yeast *MATα2* is essential for repressor activity. Nature 320, 766-768.

Preiss, A., Rosenberg, U. B., Kienlin, A., Seifert, E. and Jäckle, H. (1985). Molecular genetics of *Kruppel*, a gene required for segmentation of the *Drosophila* embryo. Nature 313, 27-32.

Pringle, J. R. and Hartwell, L. H. (1981). The *Saccharomyces cerevisiae* cell cycle, pp 97-142. In: The Molecular Biology of the Yeast *Saccharomyces* (eds. Strathern, J. N., Jones, E. W. and Broach, J. R), Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Probst, E., Kressman, A. and Birnstiel, M. L. (1979). Expression of sea urchin histone genes in the oocyte of *Xenopus laevis*. J. Mol. Biol. **135**, 709-732.

Ptashne, M. (1984). Repressors. TIBS **9**, 142-145.

Ptashne, M. (1986). Gene regulation by proteins acting nearby or at a distance. Nature **322**, 697-701.

Rabin, M., Hart, C. P., Ferguson-Smith, A., McGinnis, W., Levine, M. and Ruddle, F. H. (1985). Two homoeo box loci mapped in evolutionarily related mouse and human chromosomes. Nature **314**, 175-177.

Rebagliati, M. R., Weeks, D. L., Harvey, R. P. and Melton, D. A. (1985). Identification and cloning of localized maternal RNAs from *Xenopus* eggs. Cell **42**, 769-777.

Reynolds, F. H., Premkumar, E. and Pitha, P. (1975). Interferon activity produced by translation of human interferon messenger RNA in cell-free ribosomal systems and in *Xenopus* oocytes. Proc. Natl. Acad. Sci. U.S.A. **72**, 4881-4885.

Rhodes, D. and Klug, A. (1986). An underlying repeat in some transcriptional control sequences corresponding to half a double helical turn of DNA. Cell **46**, 123-132.

Rigby, P. W. J., Dieckmann, M., Rhodes, C. and Berg, P. (1977). Labeling deoxyribonucleic acid to high specific activity *in vitro* by nick translation with DNA polymerase I. J. Mol. Biol. **113**, 237-251.

Robert, B., Daubas, P., Akimenko, M.-A., Cohen, A., Garner, I., Guernet, J.-L. and Buckingham, M. (1984). A single locus in the mouse encodes both myosin light chains 1 and 3, a second locus corresponds to a

related pseudogene. Cell 39, 129-140.

Robertson, E., Bradley, A., Kuehn, M. and Evans, M. (1986). Germ-line transmission of genes introduced into cultured pluripotential cells by retroviral vector. Nature 323, 445-448.

Rosenberg, U. B., Preiss, A., Seifert, E., Jackle, H. and Knipple, D. C. (1985). Production of phenocopies by *Kruppel* antisense RNA injection into *Drosophila* embryos. Nature 313, 703-706.

Rosenberg, U. B., Schröder, C., Preiss, A., Kienlin, A., Côté, S., Riede, I. and Jäckle, H. (1986). Structural homology of the product of the *Drosophila Kruppel* gene with *Xenopus* transcription factor IIIA. Nature 319, 336-339.

Rozek, C. E. and Davidson, N. (1983). *Drosophila* has one myosin heavy chain gene with three developmentally regulated transcripts. Cell 32, 23-34.

Rubin, M. R., Toth, L. E., Patel, M. D., D'Eustachio, P. and Nguyen-Huu, M. C. (1986). A mouse homeo box gene is expressed in spermatocytes and embryos. Science 233, 663-667.

Rungger, D., Matthias, P. D. and Huber, J. P. (1981). Transcription of complex structural genes in the *Xenopus* oocyte system, pp 28-32. In: International Cell Biology, 1980-1981 (eds. Schweiger, H. G.), Springer-Verlag, Berlin.

Rusconi, S. and Schaffner, W. (1981). Transformation of frog embryos with the rabbit β-globin gene. Proc. Natl. Acad. Sci. U.S.A. 78, 5051-5055.

Ryoji, M. and Worcel, A. (1984). Chromatin assembly in *Xenopus* oocytes: *in vivo* studies. Cell 37, 21-32.

Ryoji, M. and Worcel, A. (1985). Structure of the two distinct types of minichromosomes that are assembled on DNA injected in *Xenopus* oocytes. Cell 40, 923-932.

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.

Sander, M. and Hsieh, T. (1985). *Drosophila* topoisomerase II double-strand DNA cleavage: analysis of DNA sequence homology at the cleavage site. Nucl. Acids Res. 13, 1057-1071.

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. and Dawid, I. B. (1983). Differential gene expression in the gastrula of *Xenopus laevis*. Science 222, 135-139.

Sargent, T. D., Jamrich, M. and Dawid, I. B. (1986). Cell interactions and the control of gene activity during early development of *Xenopus laevis*. Dev. Biol. 114, 238-246.

Sassone-Corsi, P., Wildeman, A. and Chambon, P. (1985). A *trans*-acting factor is responsible for the simian virus 40 enhancer activity *in vitro*. Nature 313, 458-463.

Scheer, U., Hinssen, H., Franke, W. W. and Jockusch, B. M. (1984). Microinjection of actin-binding proteins and actin antibodies demonstrates involvement of nuclear actin in transcription of lampbrush chromosomes. Cell 39, 111-122.

Scheidereit, C., Geisse, S., Westphal, H. M. and Beato, M. (1983). The glucocorticoid receptor binds to defined nucleotide sequences near the promoter of mouse mammary tumour virus. Nature 304, 749-752.

Scherer, S. and Davis, R. W. (1979). Replacement of chromosome segments with altered DNA sequences constructed *in vitro*. Proc. Natl. Acad. Sci. U.S.A. 76, 1951-1955.

Schlissel, M. S. and Brown, D. D. (1984). The transcriptional regulation of 5S RNA genes in chromatin: the roles of active stable transcription complexes and histone H1. Cell 37, 903-913.

Schnieke, A., Harbers, K. and Jaenisch, R. (1983). Embryonic lethal mutation in mice induced by retrovirus insertion into the α1(I) collagen gene. Nature 304, 315-320.

Schöler, H. R. and Gruss, P. (1985). Cell type-specific transcriptional enhancement *in vitro* requires the presence of *trans*-acting factors. EMBO J. 4, 3005-3013.

Schon, E., Evans, T., Welsh, J. and Efstratiadis, A. (1983). Conformation of promoter DNA: fine mapping of S1-hypersensitive sites. Cell 35, 837-848.

Schröder, H. C., Zahn, R. K. and Muller, W. E. G. (1982). Role of actin and tubulin in the regulation of poly(A) polymerase-endoribonuclease IV complex from calf thymus. J. Biol. Chem. 257, 2305-2309.

Schuler, M. A., McOsker, P. and Keller, E. B. (1983). DNA sequence of two linked actin genes of the sea urchin *S. purpuratus*. Mol. Cell Biol. 3, 448-456.

Segall, J., Matsui, T. and Roeder, R. G. (1980). Multiple factors are required for the accurate transcription of purified genes by RNA polymerase III. J. Biol. Chem. 255, 11986-11991.

Seiler-Tuyns, A., Eldridge, J. D. and Paterson, B. M. (1984). Expression and regulation of chicken actin genes introduced into mouse myogenic and nonmyogenic cells. Proc. Natl. Acad. Sci. U.S.A. 81, 2980-2984.

Shani, M. (1985). Tissue-specific expression of rat myosin light-chain 2 in transgenic mice. Nature 314, 283-286.

Shani, M. (1986). Tissue-specific and developmentally regulated expression of a chimeric actin-globin gene in transgenic mice. Mol.

Cell. Biol. 6, 2624-2631.

Shepherd, J. C. W., McGinnis, W., Carrasco, A. E., De Robertis, E. M. and Gehring, W. J. (1984). Fly and frog homeo domains show homologies with yeast mating type regulatory proteins. Nature 310, 70-71.

Shermoen, A. W. and Beckendorf, S. K. (1982). A complex of interacting DNAase I-hypersensitive sites near the *Drosophila* glue protein gene, *Sgs4*. Cell 29, 601-607.

Shiokawa, K., Sameshima, M., Tashiro, K., Miura, T., Nakakura, N. and Yamana, K. (1986). Formation of nucleus-like structure in the cytoplasm of λ-DNA-injected fertilized eggs and its partition into blastomeres during early embryogenesis in *Xenopus laevis*. Dev. Biol. 116, 539-542.

Shott, R. J., Lee, J. J., Britten, R. J. and Davidson, E. H. (1984). Differential expression of the actin gene family of *Strongylocentrotus purpuratus*. Dev. Biol. 101, 295-306.

Shuey, D. J. and Parker, C. S. (1986). Bending of promoter DNA on binding of heat shock transcription factor. Nature 323, 459-461.

Sinden, R. R., Carlson, J. O. and Pettijohn, D. E. (1980). Torsional tension in the DNA double helix measured with trimethyl psoralen in living *E. coli*: analagous measurements in insect and human cells. Cell 21, 773-783.

Singh, H., Sen, R., Baltimore, D. and Sharp, P. A. (1986). A nuclear factor that binds to a conserved sequence motif in transcriptional control elements of immunoglobulin genes. Nature 319, 154-158.

Slack, J. M. W. (1983). From Egg to Embryo. Determinative Events in Development, Cambridge University Press, Cambridge.

Slack, J. M. W. and Forman, D. (1980). An interaction between dorsal and ventral regions of the marginal zone in early amphibian embryos.

J. Embryol. exp. Morph. 56, 283-299.

Smithies, O., Gregg, R. G., Boggs, S. S., Koralewski, M. A. and Kucherlapati, R. S. (1985). Insertion of DNA sequences into the human chromosomal β-globin locus by homologous recombination. Nature **317**, 230-234.

Southern, E. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**, 503-517.

Spradling, A. C. (1981). The organisation and amplification of two chromosomal domains containing *Drosophila* chorion genes. Cell 27, 193-201.

Spradling, A. C. and Rubin, G. M. (1983). The effect of chromosomal position on the expression of the *Drosophila* xanthine dehydrogenase gene. Cell **34**, 47-57.

Staudt, L. M., Singh, H., Sen, R., Wirth, T., Sharp, P. A. and Baltimore, D. (1986). A lymphoid-specific protein binding to the octamer motif of immunoglobulin genes. Nature **323**, 640-643.

Sternberg, N., Tiemeier, D. and Enquist, L. (1977). *In vitro* packaging of λ *Dam* vector containing *Eco*RI DNA fragments of *Escherichia coli* and phage P1. Gene **1**, 255-280.

Stewart, T. A. and Mintz, B. (1981). Successive generations of mice produced from an established culture line of euploid teratocarcinoma cells. Proc. Natl. Acad. Sci. U.S.A. **78**, 6314-6318.

Stollar, D. (1986). Antibodies to DNA. CRC Crit. Rev. Biochem. **20**, 1-36.

Stossel, T. P. (1984). Contribution of actin to the structure of the cytoplasmic matrix. J. Cell. Biol. **99**, 15S-21S.

Streisinger, G., Walker, C., Dower, N., Knauber, D. and Singer, F. (1981). Production of clones of homozygous diploid zebra fish

(*Brachydanio rerio*). Nature **291**, 293-296.

Struhl, G. and White, R. A. H. (1985). Regulation of the *Ultrabithorax* gene of *Drosophila* by other bithorax complex genes. Cell **43**, 507-519.

Sturgess, E. A., Ballantine, J. E. M., Woodland, H. R., Mohun, P. R., Lane, C. D. and Dimitriadis, G. J. (1980). Actin synthesis during the early development of *Xenopus laevis*. J. Embryol. exp. Morph. **58**, 303-320.

Stutz, F. and Spohr, G. (1986). Isolation and characterization of sarcomeric actin genes expressed in *Xenopus laevis* embryos. J. Mol. Biol. **187**, 349-361.

Swift, G. H., Hammer, R. E., MacDonald, R. J. and Brinster, R. L. (1984). Tissue-specific expression of the rat pancreatic elastase I gene in transgenic mice. Cell **38**, 639-646.

Takahashi, K., Vigneron, M., Matthes, H., Wildeman, A., Zenke, M. and Chambon, P. (1986). Requirement of stereospecific alignments for initiation from the simian virus 40 early promoter. Nature **319**, 121-126.

Thomas, K. R., Folger, K. R. and Capecchi, M. R. (1986). High frequency targeting of genes to specific sites in the mammalian genome. Cell **44**, 419-428.

Tokunaga, K., Yoda, K. and Sakiyama, S. (1985). Structure of a processed gene of mouse cytoplasmic λ-actin transposed onto a BAM5 sequence: insertion has created 13 base-pair direct repeats. Nucl. Acids Res. **13**, 3031-3042.

Townes, T. M., Lingrel, J. B., Chen, H. Y., Brinster, R. L. and Palmiter, R. D. (1985). Erythroid-specific expression of human β-globin genes in transgenic mice. EMBO J. **4**, 1715-1723.

Turner, P. C. and Woodland, H. R. (1983). Histone gene number and organisation in *Xenopus*: *Xenopus borealis* has a homogeneous major cluster. Nucl. Acids Res. **11**, 971-986.

Ueyama, H., Hamada, H., Battula, N. and Kakunaga, T. (1984). Structure of a human smooth muscle actin gene (aortic type) with a unique intron site. Mol. Cell. Biol. **4**, 1073-1078.

Vale, R. D., Schnapp, B. J., Mitchison, T., Steuer, E., Reese, T. S. and Sheetz, M. P. (1985). Different axoplasmic proteins generate movement in opposite directions along microtubules *in vitro*. Cell **43**, 623-632.

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). 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. Proc. Natl. Acad. Sci. U.S.A. **75**, 1106-1110.

Vandekerckhove, J. and Weber, K. (1978c). The amino acid sequence of actin from chicken skeletal muscle actin and chick gizzard smooth muscle actin. FEBS Letters **102**, 219-222.

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. (1981a). Actin typing on total cellular extracts: a highly sensitive protein-chemical procedure able to distinguish different actins. Eur. J. Biochm. **113**, 595-603.

Vandekerckhove, J. and Weber, K. (1981b). Diversity of expression of non-muscle actin in amphibia. J. Mol. Biol. 152, 413-426.

Vandekerckhove, J. and Weber, K. (1984). Chordate muscle actins differ distinctly from invertebrate muscle actins. J. Mol. Biol. 179, 391-413.

Vandekerckhove, J., Bugaisky, G. and Buckingham, M. (1986). Simultaneous expression of skeletal muscle and heart actin proteins in various striated muscle tissues and cells. J. Biol. Chem. 261, 1838-1843.

Vincent, A., Colot, H. V. and Rosbash, M. (1985). Sequence and structure of the *Serendipity* locus of *Drosophila melanogaster*. J. Mol. Biol. 186, 149-166.

Voss, S. D., Schlokat, U. and Gruss, P. (1986). The role of enhancers in the regulation of cell-type-specific transcriptional control. TIBS 11, 287-289.

Wabl, M. R. and Burrows, P. D. (1984). Expression of immunoglobulin heavy chain at a high level in the absence of a proposed immunoglobulin enhancer element in *cis*. Proc. Natl. Acad. Sci. U.S.A. 81, 2452-2455.

Wahl, G. M., Stern, M. and Stark, G. R. (1979). Efficient transfer of large DNA fragments from agarose gels to diazobenzyloxymethal-paper and rapid hybridisation by using dextran sulphate. Proc. Natl. Acad. Sci. U.S.A. 76, 3683-3687.

Walker, M. D., Edlund, T., Boulet, A. M. and Rutter, W. J. (1983). Cell-specific expression controlled by the 5'-flanking region of insulin and chymotrypsin genes. Nature 306, 557-561.

Wang, Y.-L. (1984). Reorganization of actin filament bundles in living fibroblasts. J. Cell. Biol. 99, 1478-1485.

Wasylyk, C. and Wasylyk, B. (1986). The immunoglobulin heavy-chain B-lymphocyte enhancer efficiently stimulates transcription in non-lymphoid cells. EMBO J. 5, 553-560.

Weinberger, J., Baltimore, D. and Sharp, P. A. (1986). Distinct factors bind to apparently homologous sequences in the immunoglobulin heavy-chain enhancer. Nature 322, 846-848.

Weubtraub, H. (1983). A dominant role for DNA secondary structure in forming hypersensitive structures in chromatin. Cell 32, 1191-1203.

Weintraub, H. (1984). Histone-H1-dependent chromatin superstructures and the suppression of gene activity. Cell 38, 17-27.

Weintraub, H. (1985). Assembly and propagation of repressed and derepressed chromosomal states. Cell 42, 705-711.

Weisbrod, S. (1982). Active chromatin. Nature 297, 289-295.

Weisbrod, S. and Weintraub, H. (1979). Isolation of a subclass of nuclear proteins responsible for conferring a DNase I-sensitive structure on globin chromatin. Proc. Natl. Acad. Sci. U.S.A. 76, 630-634.

Whalen, R. G., Butler-Brown, G. S. and Gros, F. (1976). Protein synthesis and actin heterogeneity in calf muscle cells in culture. Proc. Natl. Acad. Sci. 73, 2018-2022.

Wharton, K. A., Johansen, K. M., Xu, T. and Artavanis-Tsakonas, S. (1985). Nucleotide sequence from the neurogenic locus *Notch* implies a gene product that shares homology with proteins containing EGF-like repeats. Cell 43, 567-581.

White, R. A. H. and Wilcox, M. (1984). Protein products of the bithorax complex in *Drosophila*. Cell 39, 163-171.

Wickens, M. P. and Gurdon, J. B. (1980). Post-transcriptional processing of simian virus 40 late transcripts in injected forg oocytes. J. Mol. Biol. **163**, 1-26.

Wickens, M. P., Woo, S., O'Malley, B. W. and Gurdon, J. B. (1980). Expression of a chicken chromosomal ovalbumin gene injected into frog oocyte nuclei. Nature **285**, 628-634.

Williams, D. A., Orkin, S. H. and Mulligan, R. C. (1986). Retrovirus-mediated transfer of human adenosine deaminase sequences into cells in culture and into murine haematopoietic cells *in vivo*. Proc. Natl. Acad. Sci. U.S.A. **83**, 2566-2570.

Wilson, C., Cross, G. S. and Woodland, H. R. (1986). Tissue-specific expression of actin genes injected into *Xenopus* embryos. Cell (in the press).

Wood, W. B. (1966). Host specificity of DNA produced by *Escherichia coli*: bacterial mutations affecting the restriction and modification of DNA. J. Mol. Biol. **16**, 118-133.

Woodland, H. R. and Jones, E. A. (1985). Interacting systems in amphibia. Nature **318**, 102-104.

Woodland, H. R. and Jones, E. A. (1986). Unscrambling egg structure. Nature **319**, 261-262.

Wormington, W. M. and Brown, D. D. (1983). Onset of the 5S RNA gene regulation during *Xenopus* embryogenesis. Dev. Biol. **99**, 248-257.

Wu, C. (1984a). Two protein-binding sites in chromatin implicated in the activation of the heat-shock genes. Nature **309**, 229-234.

Wu, C. (1984b). Activating protein factor binds *in vitro* to upstream control sequences in heat shock gene chromatin. Nature **311**, 81-84.

Xiao, H. and Lis, J. T. (1986). A consensus sequence polymer inhibits *in vivo* expression of heat shock genes. Mol. Cell. Biol. **6**, 3200-3206.

Yaffe, D., Nudel, U., Mayer, Y. and Neuman, S. (1985). Highly conserved sequences in the 3' untranslated region of mRNAs coding for homologous proteins in distantly related species. Nucl. Acids Res. **13**, 3723-3737.

Yisraeli, J., Adelstein, R. S., Melloul, D., Nudel, U., Yaffe, D. and Cedar H. (1986). Muscle-specific activation of a methylated chimeric actin gene. Cell **46**, 409-416.

Yu, S.-F., von Rüden, T., Kantoff, P. W., Garber, C., Seiberg, M., Rüther, U., Anderson, W. F., Wagner, E. F. and Gilboa, E. (1986). Self- inactivating retroviral vectors designed for transfer of whole genes into mammalian cells. Proc. Natl. Acad. Sci. U.S.A. **83**, 3194-3198.

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.

Zimmer, W. and Schwartz, R. J. (1982). Amplification of chicken actin genes during myogenesis, pp 137-146. In: Gene Amplification (ed. R. T. Schimke), Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Zinn, K. and Maniatis, T. (1986). Detection of factors that interact with the human β-interferon regulatory region *in vivo* by DNase I footprinting. Cell **45**, 611-618.

Zissler, J., Signer, E. R. and Schaefer, F. (1971). The role of recombination in growth of bacteriophage lambda, p 455. In: The Bacteriophage Lambda (ed. Hershey, A. D.), Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.