

A Thesis Submitted for the Degree of PhD at the University of Warwick

Permanent WRAP URL:

<http://wrap.warwick.ac.uk/108299>

Copyright and reuse:

This thesis is made available online and is protected by original copyright.

Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it.

Our policy information is available from the repository home page.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk

THE REGULATION OF ACTIN GENE EXPRESSION

IN XENOPUS EMBRYOS.

A thesis submitted for the degree of doctor
of philosophy.

Nicholas D. Lakin.

Animal molecular genetics group
Department of Biological Sciences
University of Warwick.

April, 1991.

THE BRITISH LIBRARY DOCUMENT SUPPLY CENTRE

BRITISH THESES N O T I C E

The quality of this reproduction is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print, especially if the original pages were poorly produced or if the university sent us an inferior copy.

Previously copyrighted materials (journal articles, published texts, etc.) are not filmed.

Reproduction of this thesis, other than as permitted under the United Kingdom Copyright Designs and Patents Act 1988, or under specific agreement with the copyright holder, is prohibited.

THIS THESIS HAS BEEN MICROFILMED EXACTLY AS RECEIVED

**THE BRITISH LIBRARY
DOCUMENT SUPPLY CENTRE
Boston Spa, Wetherby
West Yorkshire, LS23 7BQ
United Kingdom**

C O N T E N T S .

Table of contents.	ii
Table of figures.	xi
Acknowledgements.	xvii
Declaration.	xviii
Summary.	xix
Abbreviations	xx

Table of contents.

	Page
Chapter 1: INTRODUCTION.	
1.1. The early development of <u>Xenopus</u> .	1
1.2. The actin gene family.	5
1.3. Actin genes and the regulation of their expression.	8
1.3.1. Actin gene structure and expression.	8
1.3.2. DNA sequence elements involved in the expression of actin genes.	13
1.3.3. Cis-acting sequences and trans-acting factors involved in the control of cardiac actin gene expression.	15
1.3.4. Cis-acting sequences and trans-acting factors involved in the control of skeletal actin gene expression.	19
1.3.5. The role of the CArG box in the expression of the proto-oncogene <u>c-fos</u> .	25
1.4. Elements involved in the expression of other muscle-specific genes.	30
1.5. MyoD and related factors: A 'master switch' for muscle determination?	37

Chapter 2: MATERIALS.

2.1. General materials.	44
2.2. Stock solutions.	45
2.3. Bacteriological media.	45
2.4. Bacteria and plasmids.	46
2.4.1. Genotypes of <u>E. coli</u> strains.	46
2.4.2. Plasmid vectors.	46

Chapter 3: METHODS.

3.1. Oocytes eggs and embryos.	47
3.1.1. Oocytes.	47
3.1.2. Eggs and embryos.	47
3.2. Micro-injection of fertilised eggs.	48
3.2.1. Preparation of DNA for micro-injection.	48
3.2.2. Micro-injection of DNA.	48
3.3. Dissection of embryos.	49
3.4. Nucleic acid isolation from embryos.	49
3.5. Standard sub-cloning techniques.	50
3.5.1. Restriction enzyme digests.	50
3.5.2. Preparation of plasmid vectors for sub-cloning.	50
3.5.3. Ligations.	50
3.5.4. <u>E. coli</u> plasmid transformations.	51
3.5.5. Polymerase chain reaction.	52
3.6. Small scale preparation of plasmid DNA.	52
3.7. Large scale preparation of plasmid DNA and	53

purification by caesium chloride/ethidium bromide centrifugation.	
3.8. Electrophoresis of nucleic acids.	55
3.8.1. Non-denaturing agarose gels.	55
3.8.2. Isolation of restriction fragments from agarose gels.	55
3.8.3. Non-denaturing polyacrylamide gels.	56
3.8.4. Denaturing polyacrylamide gels.	56
3.9. Preparation of synthetic oligonucleotides.	57
3.9.1. Gel isolation of synthetic DNA oligonucleotides.	57
3.9.2. Annealing of complementary synthetic oligonucleotides	58
3.10. Methods of radiolabelling DNA.	59
3.10.1. Nick translation.	59
3.10.2. End-labelling DNA with γ - ³² P-ATP and T4 polynucleotide kinase.	59
3.10.3. Labelling DNA fragments by in-filling.	60
3.11. Sequencing of plasmid DNA.	61
3.12. Southern blotting.	63
3.13. Primer extension.	65
3.14. Chloramphenicol acetyltransferase assays.	66
3.14.1. Preparation of chloramphenicol acetyltransferase extracts from micro-injected embryos.	66

3.14.2. Chloramphenicol acetyltransferase assays using extracts from micro-injected embryos.	66
3.15. Preparation of protein extracts for band shift assays and DNA footprinting.	68
3.15.1. <u>Xenopus laevis</u> oocytes, eggs and embryos.	68
3.15.2. Adult <u>Xenopus laevis</u> tissues.	69
3.16. Band shift assays.	69
3.17. Antibody band shift assays.	70
3.18. Methylation interference and methylation protection footprinting.	71

RESULTS AND DISCUSSION.

Chapter 4: The preliminary localisation of cis-acting regulatory sequences in the promoter of the <u>Xenopus borealis</u> skeletal actin gene.	
Introduction.	73
4.1. Preliminary characterisation of the <u>Xenopus</u> <u>borealis</u> skeletal actin gene promoter.	76
4.2. CArG box3 is essential for the full activity of the <u>Xenopus borealis</u> skeletal actin gene.	79
4.3. Discussion.	80

Chapter 5: The promoter of the Xenopus borealis
skeletal actin gene binds trans-acting
factors in vitro.

Introduction.	84
5.1. Identification of a sequence-specific trans- acting factor(s) that interacts with the <u>Xenopus borealis</u> skeletal actin gene promoter <u>in vitro</u> .	85
5.2. Preliminary localisation of factor binding sites in the promoter of the <u>Xenopus borealis</u> skeletal actin gene promoter.	87
5.3. DNA footprinting of the -167/-42 region of the <u>Xenopus borealis</u> skeletal actin gene promoter.	90
5.4. CArG box3 and CArG box1 can compete for the binding of factors to the <u>Xenopus borealis</u> skeletal actin gene.	93
5.5. Discussion.	95

Chapter 6: CArG box3 and CArG box1 bind similar, yet
distinct proteins.

Introduction.	98
6.1. CArG box1 and CArG box3 synthetic oligonucleotides bind sequence-specific factors <u>in vitro</u> .	99
6.2. CArG box3 and CArG box1 exhibit distinct competition patterns.	102

6.3. CArG box1 binds a protein which is antigenically related to the serum response factor.	104
6.4. CArG box3 binds a protein which is antigenically distinct from SRF.	108
6.5. Discussion.	111
Chapter 7: The spatial and temporal distribution of CArG box binding factors in <u>Xenopus</u> .	
Introduction.	118
7.1. CArG box binding activities are present throughout the embryo.	119
7.2. CArG box3 binds a factor(s) that is present throughout early <u>Xenopus</u> development.	121
7.3. Discussion.	123
Chapter 8: Sequences in addition to the CArG box motifs are required for the expression of the <u>Xenopus borealis</u> skeletal actin gene.	
Introduction.	129
8.1. A 156 nucleotide region of the <u>Xenopus borealis</u> skeletal actin gene promoter can direct the tissue-specific expression of a heterologous promoter.	130

8.2. CArG box3, CArG box1, or a combination of both	134
are not sufficient to drive the tissue-specific expression of a heterologous promoter.	
8.3. A region of the <u>Xenopus borealis</u> skeletal actin gene promoter upstream of the CArG box3 motif	138
is required for the expression of the gene when sequences downstream of nucleotide -42 are absent from the gene.	
8.4. Discussion.	142
Chapter 9: The upstream regulatory element of the	
<u>Xenopus borealis</u> skeletal actin gene	
binds a trans-acting factor(s) <u>in vitro</u> .	
Introduction.	148
9.1. The <u>Xenopus borealis</u> URE binds a protein(s) <u>in vitro</u> .	149
9.2. The URE, in conjunction with CArG box1 and CArG box3, is insufficient to direct the tissue-specific expression of a heterologous promoter.	150
9.3. Discussion.	153

Chapter 10: Sequences located between the CArG box1 motif and the TATA box of the <u>Xenopus borealis</u> skeletal actin gene interact with a trans-acting factor(s) <u>in vitro</u> .	
Introduction.	157
10.1. A trans-acting factor(s) binds the -83/-42 region of the <u>Xenopus borealis</u> skeletal actin gene <u>in vitro</u> .	158
10.2. Discussion.	160
Chapter 11: General discussion.	164
APPENDIX.	
Appendix 1: The expression of a human cardiac actin gene in developing <u>Xenopus laevis</u> embryos.	
Introduction.	171
A.1.1. The spatial expression of a human cardiac actin gene in stage 30 <u>Xenopus laevis</u> embryos.	172
A.1.2. The temporal expression of a human cardiac actin gene in developing <u>Xenopus laevis</u> embryos.	175
A.1.3. The analysis of sequence elements important in the expression of a human cardiac actin gene in <u>Xenopus laevis</u> embryos.	178
A.1.4. Discussion.	183

x

Appendix 2: Diagrammatic representation of the plasmids 187

p Δ gb and pXlgb.

REFERENCES. 188

TABLE OF FIGURES.

Figure.	preceding Page
1.1. Evolutionary conserved CArG box sequences in muscle-specific genes.	17
1.2. Structural comparison of the MyoD family of myogenic regulatory factors.	39
4.0. Structure of the complete <u>Xenopus borealis</u> skeletal actin gene.	74
4.1. Analysis of transcripts from embryos injected with <u>Xenopus borealis</u> skeletal actin promoter deletion plasmids.	77
4.2. Analysis of transcripts from embryos micro- injected with deletion constructs containing an additional CArG box3 motif.	79
4.3. CArG box sequences of the cardiac and skeletal actin genes.	81
5.1. Band shift analysis of the <u>Xenopus borealis</u> skeletal actin gene promoter.	86
5.2. Band shift analysis of <u>Xenopus borealis</u> skeletal actin gene promoter fragments containing progressively less 5' flanking sequences.	88
5.3A. DMS protection footprinting of the <u>Xenopus</u> <u>borealis</u> skeletal actin gene promoter.	91

5.3B.	DMS interference footprinting of the <u>Xenopus borealis</u> skeletal actin gene promoter.	91
5.4.	Competition of binding to the <u>Xenopus borealis</u> skeletal actin gene promoter by synthetic CArG box oligonucleotides.	93
5.5.	Comparison of DNA footprints generated over various actin CArG box1 motifs.	95
6.1.	CArG box oligonucleotides bind sequence-specific trans-acting factors <u>in vitro</u> .	99
6.2.1.	CArG box oligonucleotides exhibit distinct competition patterns.	102
6.2.2.	Competition analysis of CArG box binding proteins using increasing amounts of CArG box competitors.	104
6.3.1.	Comparison of CArG box and SRE sequences.	105
6.3.2.	A protein which is antigenically similar to SRF binds the <u>Xenopus borealis</u> CArG box1 motif.	106
6.4.	A protein which is antigenically distinguishable from SRF binds the <u>Xenopus borealis</u> skeletal actin CArG box3 motif.	108
7.1.1.	Distribution of CArG box binding factors throughout the embryo.	119
7.1.2.	Tissue distribution of CArG box3 binding proteins.	120

7.2. Distribution of CArG box3 binding proteins throughout <u>Xenopus</u> development.	121
8.1.1. Diagrammatic representation of the plasmid p Δ gb Δ 5(+).	132
8.1.2. Primer extension analysis of embryos injected with pXbgb Δ 6, p Δ gb, or p Δ gb Δ 5(+).	133
8.1.3. Sequence of the deleted <u>Xenopus laevis</u> β -globin promoter illustrating the two transcription start sites used by the <u>Xenopus borealis</u> skeletal actin gene promoter fragment.	134
8.2.1. Sequences of oligonucleotides used in the construction of plasmids p Δ gbB3, p Δ gbB1 p Δ gbB3/B1(+) and p Δ gbB3/B1(-).	135
8.2.2. Diagrammatic representation of the constructs p Δ gbB3, p Δ gbB1, p Δ gbB3/B1(+) and p Δ gbB3/B1(-).	136
8.2.3. Primer extension analysis of nucleic acid isolated from embryos injected with plasmids p Δ gb, pXbgb Δ 6, p Δ gb Δ 5(+), p Δ gbB3, p Δ gbB1 p Δ gbB3/B1(+) and p Δ gbB3/B1(-).	136
8.2.4. Southern blot analysis of embryos injected with plasmids p Δ gb, pXbgb Δ 6, p Δ gb Δ 5(+), p Δ gbB3, p Δ gbB1, p Δ gbB3/B1(+) and p Δ gbB3/B1(-).	137
8.3.1. Diagrammatic representation of the constructs p Δ gb Δ 6 and p Δ gb Δ 8+.	140

8.3.2.	Primer extension analysis of nucleic acid isolated from embryos injected with plasmids p Δ gb, pXbgb Δ 6, p Δ gb Δ 5(+), p Δ gb Δ 6, p Δ gb Δ 8+, p Δ gbB3, p Δ gbB1, p Δ gbB3/B1(+) and p Δ gbB3/B1(-).	140
8.3.3.	Southern blot analysis of embryos injected with plasmids p Δ gb, pXbgb Δ 6, p Δ gb Δ 5(+), p Δ gb Δ 6, p Δ gb Δ 8+, p Δ gbB3, p Δ gbB1, p Δ gbB3/B1(+) and p Δ gbB3/B1(-).	141
8.4.	Comparison of URE sequences with sequences in the -41/+28 region of the <u>Xenopus borealis</u> skeletal actin gene.	147
9.1.	The <u>Xenopus borealis</u> skeletal actin gene URE binds a trans-acting factor(s) <u>in vitro</u> .	149
9.2.1.	Diagrammatic representation of the plasmid p Δ gbCB-URE.	151
9.2.2.	Primer extension analysis of embryos injected with the plasmids p Δ gb Δ 5(+), p Δ gb Δ 6, p Δ gb Δ 8+, p Δ gbB3/B1(+) and p Δ gbCB-URE.	152
10.1.1.	Diagrammatic representation of the oligonucleotide B1/Bal.	159
10.1.2.	The -83/-42 region of the <u>Xenopus borealis</u> skeletal actin gene binds a trans-acting factor(s) <u>in vitro</u> .	159
10.1.3.	The SAPFI binding activity is present in both axis and head plus gut protein extracts.	160

11.1.	Diagrammatic representation of the cis-acting regulatory sequences, and the trans-acting factors with which they interact, that are important for the expression of the <u>Xenopus borealis</u> skeletal actin gene.	168
A.1.1.1.	Diagrammatic representation of the plasmid pHCA485.	172
A.1.1.2.	The human cardiac actin gene promoter is capable of driving the tissue-specific expression of a CAT reporter gene in the axis region of <u>Xenopus laevis</u> embryos.	172
A.1.1.3.	Diagrammatic representation of the plasmid pRE485.	174
A.1.1.4.	A 485bp fragment of the human cardiac actin gene promoter produces correctly initiated transcripts in the axis region of <u>Xenopus laevis</u> embryos.	174
A.1.2.1.	The temporal expression of a human cardiac actin gene in <u>Xenopus laevis</u> embryos.	176
A.1.3.1.	Diagrammatic representation of the human cardiac actin promoter deletion plasmids pHCA485, pHCA177, pHCA153, pHCA131, pHCA96 and pHCA64.	179
A.1.3.2.	The promoter activities of the plasmids pHCA485, pHCA177 and pHCA153 when injected into <u>Xenopus laevis</u> embryos.	179

- A.1.3.3. Diagrammatic representation of the human cardiac actin promoter deletion plasmids pRE485, pRE177, pRE153, pRE131, pRE96 and pRE64. 181
- A.1.3.4. The promoter activities of the plasmids pRE485, pRE177, pRE153, pRE131, pRE96 and pRE64 in Xenopus laevis embryos. 182

Acknowledgements.

I acknowledge my supervisor, Hugh Woodland, and thank him for his guidance and help throughout the course of this project. Grateful thanks are also due to both Hugh Woodland and Liz Jones for the many hours they spent injecting and dissecting embryos. I would also like to thank Dr. Mark Boardman for his technical assistance, in addition to his encouragement and support when things were going badly. My thanks are also due to Darrin Smith for various materials, in addition to many useful discussions.

I acknowledge Dr. T. Mohun, Bob Old and Darrin Smith for their generous gifts of anti-SRF, anti-THR and anti-Oct-1 antibodies respectively. I also acknowledge Dr. L. Kedes for the Human cardiac actin promoter plasmids. I acknowledge the financial support of the Medical Research Council.

I would also like to thank my friends at Warwick for showing me a good time. Not least, I thank Mary for putting up with me and providing a much needed escape from a claustrophobic environment. Finally, I owe the greatest thanks to my parents for their unfailing support and encouragement throughout the years.

Declaration.

All the results presented in this thesis were obtained by the author, apart from those specifically indicated in the text. All injections and almost all the dissections of embryos were performed by Hugh Woodland. In addition, the results presented in Chapter 4 were obtained by Dr. Mark Boardman in 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 acknowledged by means of reference. None of the work contained in this thesis has been used for any previous application for a degree.

SUMMARY

The main aim of this work was to identify the important *cis*-acting regulatory sequences, and the *trans*-acting factors with which they interact, which are required for the tissue-specific expression of the Xenopus borealis skeletal actin gene.

All sequences necessary and sufficient for the correct spatial and temporal expression of the Xenopus borealis skeletal actin gene are located in a 156 bp fragment of the gene that spans from nucleotide residues -197 to -42 in its promoter. This region of the skeletal actin promoter contains three imperfect repeats of a CArG sequence motif that has been demonstrated to be important in the expression of other sarcomeric actin genes. Deletion analysis of the promoter of the Xenopus borealis skeletal actin gene, using Xenopus micro-injection techniques as a transient assay system for promoter activity, have identified that CArG box3 is essential for skeletal actin gene expression.

By using band shift assays I have demonstrated that, under my assay conditions, CArG box2 is unable to bind any proteins *in vitro*. Conversely, the CArG box1 sequence exhibits two binding activities on band shift analysis. One of these is antigenically related to the transcription factor SRF, whilst the second appears to be distinct from this protein. CArG box3 also interacts with a protein *in vitro*. Although this sequence exhibits a similar shift to that of the CArG box1/SRF complex on band shift analysis, my experiments suggest that this protein is distinct from SRF.

A combination of the CArG box1 and CArG box3 motifs is unable to confer muscle-specific gene expression on a heterologous promoter. Furthermore, I have identified an upstream regulatory element (URE) in the Xenopus borealis skeletal actin gene promoter that spans from nucleotides -197 to -167 that is required for the expression of the gene, at least when sequences between nucleotide -42 and +28 are absent.

The URE of the Xenopus borealis skeletal actin gene is capable of interacting with a *trans*-acting factor(s) *in vitro*. In addition to this a further region of the gene which spans from nucleotide residues -83 to -42 is also capable of interacting with a factor(s) *in vitro*.

The mechanisms by which these multiple regulatory elements control the tissue-specific expression of the Xenopus borealis skeletal actin gene will be discussed.

Abbreviations.

ATP, dATP, } ddATP }	adenosine triphosphate, deoxyadenosine triphosphate, dideoxyadenosine triphosphate.
bp	base pair.
BSA	bovine serum albumin.
cAMP	cyclic adenosine mono-phosphate.
cdNA	complementary DNA.
Ci	Curie.
CIAP	calf intestinal alkaline phosphatase.
CTP, dCTP, } ddCTP }	cytidine triphosphate, deoxycytidine triphosphate, dideoxycytidine triphosphate.
DNA	deoxyribonucleic acid.
DTT	dithiothreitol.
dTTP, } ddTTP }	thymidine triphosphate, deoxythymidine triphosphate.
EDTA	ethylenediaminetetra-acetate.
FSH	follicle stimulating hormone.
GTP, dGTP, } ddGTP }	guanosine triphosphate, deoxyguanosine triphosphate, dideoxyguanosine triphosphate.
hCG	human chorionic gonadotrophin.
Kb	kilobase pairs.
l	litre.
MBT	mid-blastula transition.
MCK	muscle creatine kinase.
ml	millilitre.
DEPC	diethyl pyrocarbonate.

Mops	3-(N-morpholino) propanesulfonic acid.
ng	nanogram.
nt	nucleotide.
pg	picogram.
RNA	ribonucleic acid.
rpm	revolutions per minute.
rRNA	ribosomal RNA.
SDS	sodium dodecyl sulfate.
SRE	serum response element.
SRF	serum response factor.
SV40	simian virus 40.
THR	thyroid hormone receptor.
cTn I	cardiac troponin I.
sTn I	skeletal troponin I
cTNT	cardiac troponin T.
sTNT	skeletal troponin T.
Tris	Tris(hydroxymethyl)aminomethane.
tRNA	transfer RNA.
ug	microgram.
ul	microlitre.
5'/3' UTR	5'/3' untranslated region.

v/v	volume/volume.
w/v	weight/volume.

ction

CHAPTER 1

Introduction.

1.1 The early development of Xenopus.

The early stages of Xenopus embryogenesis share many general characteristics with other vertebrates. The study of this organism has enjoyed particular attention because the embryos are large and easily manipulated, as well as being readily available throughout the year. In addition to this, the embryos develop rapidly and independently, with the body plan being established and tissue-specific gene activation occurring within 24 hours of fertilisation. The following section briefly summarises the embryological changes most relevant to an understanding of the work described in this thesis.

The unfertilised Xenopus egg is a large asymmetric cell. Many yolk platelets are concentrated at its vegetal pole, whilst at the animal pole these platelets are much less dense and the region is darkly pigmented by melanin granules located near the cell surface. Whereas the animal-vegetal polarity is established during oogenesis, dorsal-ventral polarity is established by a series of events directly after fertilisation of the egg. The successful sperm always enters in the animal hemisphere of the egg. This triggers a rotation in the eggs cytoplasm, resulting in the appearance of a lightly pigmented band (the grey crescent) on the opposite side of the egg to sperm entry. The grey crescent marks the dorsal side of the

Introduction

embryo and is centred on the point where cells will later start invaginating 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 cell division) takes about 35 minutes on average (Newport and Kirschner, 1982). Only a little new transcription is observed over this period. However at stage 8, the mid-blastula transition (MBT), transcription of many genes is activated, or accelerates, and the cell divisions become asynchronous.

In terms of developmental specification, the early blastula embryo consists of two cell types, those at the animal pole and those at the vegetal pole. If animal pole cells are cultured in isolation they develop to form epidermis, whereas vegetal cells develop into predominantly endodermal tissues when cultured in isolation. Indeed, the equatorial region of a 64 cell stage embryo, which forms the mesodermal tissues in a normally developing embryo, develops into ectodermal tissue when cultured in isolation. However, if these cells are isolated later than the 64 cell stage they develop into substantial amounts of mesodermal tissue, with the addition of some ectoderm (Nakamura et al., 1970).

One interpretation of this is that mesoderm induction depends on an interaction between the animal and vegetal regions of the blastula. This theory was confirmed by the experiments of Ogi (1967, 1969) and Nieuwkoop (1969) showing that, as previously known, animal and vegetal poles of embryos cultured separately

Introduction

develop into epidermis and endoderm respectively. However, by combining animal and vegetal fragments a variety of mesodermal tissues were formed. Nieuwkoop later demonstrated that the mesoderm was formed entirely from the ectodermal component of the embryo, as a result of induction by the prospective endoderm (Sudarwati and Nieuwkoop, 1971; Nieuwkoop and Ubbels, 1972).

This relatively simple model of mesoderm induction was complicated by the findings that ventral vegetal blastomeres induced little or no muscle from animal pole cells, despite the fact that muscle of the embryo is formed from blastomeres of the ventral part of the embryo. This inconsistency of cell fates has led to the 'three signal' model of mesoderm induction as proposed by Slack and co-workers (Smith and Slack, 1983; Slack et al., 1984; Smith et al., 1985; Dale and Slack, 1987). The first two signals in this model originate from the vegetal hemisphere of the embryo. One on the dorsal side of the embryo induces predominantly notochord, whilst on the ventral side a second signal induces ventral mesoderm such as blood, mesenchyme and mesothelium. The third signal of this model originates from the newly formed dorsal mesoderm. This signal acts within the mesoderm germ line to 'dorsalize' adjacent ventral mesoderm.

In recent years work has suggested that the factors responsible for cell fates during mesoderm induction (mesoderm inducing factors, or MIFs) are related to peptide growth factors (PGFs) (for review see Smith, 1989).

Introduction

Following the onset of mesodermal induction gastrulation occurs when cells invaginate through the blastopore, starting at a position in the equatorial region opposite the sperm entry point, known as the dorsal lip of the blastopore. Endodermal and mesodermal cells gradually move inside the embryo to locations where they will subsequently form the majority of the internal organs. It is at the end of this process (stage 12½) that some mesodermal cells, which later form myotomes, begin to express the cardiac and skeletal actin genes (Gurdon et al., 1985, Wilson et al., 1986).

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, after the initial detection of the α -actin proteins (Sturges et al., 1980). Indeed, the activation of the α -actin genes during early Xenopus development occurs about 8 hours before morphological differentiation of muscle tissue takes place. This makes the study of sarcomeric actin gene expression of particular interest in development for two reasons. Firstly it acts as a early marker for muscle commitment and second, it may be regulated by gene activators which operate at the very earliest stages of muscle differentiation.

Although progress is being made it is still unclear as to how the factors responsible for mesoderm induction dictate the fate of presumptive muscle cells and activate the myogenic programme. To this end much work has been performed on studying the mechanisms by which muscle-specific genes are activated,

Introduction

with the aim of working backwards to the point when animal cells first receive a signal from vegetal cells. The remainder of this chapter is concerned with reviewing this line of research in both amphibia and other vertebrates.

1.2 The actin gene family.

Actin is a structural protein found in all eukaryotic cells. It exists in a small number of closely related isoforms which have diverse functions in various different cell types. All these isoforms can be maintained as monomers (G-actin), or they can polymerise to form filaments (F-actin).

In sarcomeric muscle (skeletal and heart muscle) F-actin exists as thin filaments which interdigitate with myosin. Myosin is attached to the actin via a globular domain on the myosin molecule which is hinged to the rest of the complex. It is believed that a cyclic process of binding, rotation around the hinge and dissociation drives a sliding action of myosin filaments over actin filaments, thus facilitating muscle contraction. This process requires the hydrolysis of ATP, and is inhibited by decreasing intracellular calcium concentrations.

In addition to the role of actin in muscle contraction other actin isoforms form a vital component of the cytoskeleton in both muscle and non-muscle cells. They are involved in different aspects of cell motility, including cell movement, cytokinesis, cytoplasmic transport, secretion and phagocytosis (for review see Clarke and Spudich, 1977).

Introduction

The multiple isoforms of actin described in the preceding paragraphs were first identified by the variation in mobility of actin proteins on iso-electric focussing gels (Garrels and Gibson, 1976; Whalen *et al.*, 1976). Using this technique at least three actin isoforms, the α , β , and δ actins, were identified. However it was not until Vandekerckhove and Weber (1978a, 1978b, 1978c and 1979) sequenced different isoforms from a number of different tissues that at least six different actin types were identified. Two of these, the β and δ actin isoforms, are co-expressed in all mammalian non-muscle cells so far studied (Vandekerckhove and Weber 1979) and are the actin types that are the components of the cytoskeleton. The α actin gene family classically consists of an α -skeletal actin which is observed solely in the skeletal muscle, and an α -cardiac actin, which predominates in the heart muscle. Smooth muscle contains two actin isoforms, an α -type which predominates in the aorta for example, and a γ -type that is at higher relative levels in the stomach.

The six classical actin isoforms are members of a highly conserved family of proteins. For example only 4 out of 375 amino acids differ between the cardiac and skeletal α -actins. Indeed, in the most widely divergent example only 24 and 25 amino acids differ between the cytoskeletal, and cardiac and skeletal actins respectively (Vandekerckhove and Weber, 1979). Interestingly, the N-terminus is the most divergent region of the actin molecule. Indeed, the first three or four amino acids are characteristic of each actin isoform. They are always

Introduction

acidic residues (either glutamate or aspartate), but their exact sequence differs between isoforms. Vandekerckhove and Weber (1981a) exploited this fact and developed a simple assay system to identify other actin isoforms in other organisms by a protein-chemical analysis of the N-terminal peptides of actin proteins.

However, the classical model, developed using mammals, of there being only six strictly conserved actin isoforms, with the skeletal, cardiac and smooth muscle actins being expressed tissue-specifically, represents a oversimplification of the situation in other vertebrates. For example, the study of cytoskeletal actins in amphibia by Vandekerckhove and Weber (1981b) led to the proposal that any order of acidic amino acids at the N-terminus of these proteins might be functional. Eight different arrangements are therefore possible (types 1-8), and at least six of these have been identified in vivo (Vandekerckhove and Weber, 1981b; Bergama et al., 1985). Despite the advances previously described, the accurate study of the distribution of various different actin types was limited by the inability of the assay system used to distinguish between closely related isoforms, such as skeletal and cardiac α -actins. It was not until the actins were studied at the nucleic acid level that a more detailed account of actin isotype expression in different cells was accomplished.

Introduction

1.3 Actin genes and the regulation of their expression.

1.3.1 Actin gene structure and expression.

The high amino acid conservation between actin isoforms means that the coding sequences of their genes are very similar. This has enabled workers to use one DNA probe to screen for many actin gene isoforms. Actin cDNAs from organisms as distantly related as Dictyostelium and Drosophila can act as suitable probes for the vertebrate genome (see Cross, 1984; Engel et al., 1981). Consequently, by hybridising a cDNA probe to a genomic Southern blot and washing at low stringency an estimate of the actin gene number can be obtained by counting the number of hybridising bands. This, and other experimental approaches have been used to estimate the number of actin related sequences in both the human and mouse genomes (Engel et al., 1981; Humphries et al., 1981; Minty et al., 1983). In contrast to initial suspicion that there were six actin related sequences in the genome (i.e. one for each actin isoform), there turn out to be approximately 20-30 actin related sequences in the human genome (Engel et al., 1981; Humphries et al., 1981) and a minimum of 20 in the mouse (Minty et al., 1983). In other organisms actin gene number varies from one in yeast (Ng and Abelson, 1980; Gallwitz and Seidel, 1980), to six in Drosophila (Tobin et al., 1980; Fyrberg et al., 1981), seven to eleven in chickens (Cleveland et al., 1980; Schwartz and Rothblum, 1980) and 11 to 20 in sea urchins (Durica et al., 1980; Schellar et al., 1981). However the fact that the hamster

Introduction

has only 5 actin like sequences in its genome (Dodemont et al., 1982) suggests that the large number of actin sequences in the human and mouse may not all be necessary. In man, at least, many copies appear to be dispersed, processed α -actin pseudogenes (Ng et al., 1985).

Actin-like sequences present in the genome appear to be unclustered and in situ hybridisation has revealed that many of these sequences are scattered on different chromosomes. Indeed, it has been shown that actin genes as closely related as the skeletal and cardiac α -actins are located on different chromosomes (Czosnek et al., 1983). This would argue against various actin isoforms being derived from the same transcriptional unit by differential processing of the RNA, as has been demonstrated with other muscle specific genes such as the myosin light chains 1 and 3 gene locus (e.g. Nabeshima et al., 1984).

Cloned actin gene sequences reveal that the coding regions of different actin isoforms show considerable similarities. This would be expected from the high degree of amino acid conservation between the different actin proteins. Generally no conservation within intron sequences has been observed in vertebrate actin genes, although Ng et al. (1985) and Nakajima-Iijima et al. (1985) have discovered sequence homologies in the introns of the human and rat cytoskeletal β -actin genes. The positions of introns are entirely conserved between species in genes encoding a single vertebrate isoform (see Buckingham and Minty, 1983), but the pattern is altered between different

Introduction

isoforms. The presence of conserved enhancer sequences within the actin gene introns, such as those found in the myosin gene, and their role in the co-ordinated expression of these genes remains doubtful (see section 1.3.3 and 1.3.4).

Examination of the remainder of the transcriptional unit reveals sequence homologies in the 5' and 3' untranslated regions (5' and 3' UTR) of actin genes. Although the 5' UTR reveals some conserved sequences, its length is normally too short and variable to make any diagnostic comparisons. Conversely, the 3' UTR is considerably longer than the 5' UTR. Sequence comparisons of the 3' UTR have shown that the region is highly conserved between actin isotypes (Yaffe *et al.* 1985). For example alignment of sequences of the rat and human α -actin genes reveals long stretches with 90-100% identity. However, this high degree of sequence homology of the 3' UTR is not shared between different actin isoforms. This has been exploited to differentiate specific actin isoforms at the mRNA level and to study the spatial and temporal expression of specific actin genes.

The classical model of actin gene expression, as determined by protein sequencing, is that the β and γ -actin isoforms are expressed in all cell types so far examined, whilst two smooth muscle actin isotypes are expressed in smooth muscle cell types. In addition to these there also exist the skeletal and cardiac α -actins which are expressed in the skeletal and cardiac muscle respectively.

However, this model of actin gene expression was complicated

Introduction

by the findings of Minty et al. (1982), who characterised a mouse foetal skeletal muscle cDNA and showed it to be homologous to a cardiac actin mRNA expressed in both foetal and adult heart tissue. Furthermore, Gunning et al. (1983) were able to isolate cardiac actin cDNAs from a human skeletal muscle cDNA library. These observations suggested that the cardiac actin gene is not limited exclusively to cardiac tissue. Indeed, in mouse foetal skeletal muscle cardiac actin transcripts are a major component of the RNA (30-40% of the total skeletal 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).

It would appear that the paired expression of the cardiac and skeletal actin genes also occurs in cardiac muscle, judging by the fact that cDNAs complementary to skeletal actin message are present in cDNA libraries prepared from cardiac actin template (Mayer et al., 1984). In addition to this it has been reported that skeletal muscle actin mRNA sequences in the heart are about 10% of their abundance in the leg muscle of newborn rats. However after 80 days of development their amount in the heart decreases by a factor of 12. Furthermore, levels of skeletal actin mRNA in rat and mouse hearts has been shown to be 2% of that of the cardiac actin mRNA (Shani et al., 1981; Minty et al., 1982).

As a general rule it would appear that both the skeletal and cardiac actin isoforms are expressed in both skeletal and heart

Introduction

muscle tissues. However, as development progresses their expression becomes increasingly differentially localised, with the skeletal muscle actin being the major actin message in skeletal muscle, and the cardiac actin message predominating in the heart. This model is supported by observations that the skeletal muscle cell line C2C12 expresses mostly cardiac actin on fusion of myoblasts to form myotomes. However this level of cardiac actin transcripts later drops and gives way to predominantly skeletal actin transcripts (Bains et al., 1984).

In Xenopus much the same pattern of expression exists. Both cardiac and skeletal actin transcripts are first observed at stage 12½-13 of development and are restricted to the region of the embryo which will develop into muscle tissue (Mohun et al., 1984; Wilson et al., 1986). At later stages of development it can be seen that both these actin transcripts are localised exclusively to the muscular tissue of the embryo somites. These two isoforms may co-exist in embryonic muscle until relatively late in development (stage 42). However in adult skeletal muscle it is predominantly the skeletal actin message that is detected, with the cardiac actin transcript being present at very low levels. Likewise cardiac actin message is the predominant α -actin message detectable in adult heart tissue (Mohun et al., 1984).

This precise and intricate pattern of expression of the cardiac and skeletal actin genes raises challenging questions as to the molecular mechanisms of actin gene expression. Furthermore, the study of the tissue-specific expression of

Introduction

actin genes provides a useful model to attempt to gain some understanding of the general mechanisms of cell-type specific gene activation. To attempt to answer some of these question, and to attempt to gain some knowledge of the activation of the myogenic program as a whole, the molecular activation of both actin and other muscle-specific genes has been the subject of much research over recent years. The following sections outline some of the progress made in understanding this complex sequence of events.

1.3.2 DNA sequence elements involved in the expression of actin genes.

The emergence of α -actin transcripts during myoblast fusion in cell culture, or prior to the formation of muscle in Xenopus embryos, suggests that at least one mechanism of controlling actin gene expression is at the transcriptional level. Furthermore, heterokaryon studies demonstrated that fusion of muscle cells with non-muscle tissues in vitro induced the expression of muscle-specific genes in an environment in which they would otherwise be silent (Blau et al., 1985). This would suggest that the emergence of muscle-specific mRNA species on the fusion of myoblasts to form multi-nucleated myotubes is not caused by an alteration in the stability of the message in the cytoplasm, or indeed due to the packaging of the RNA in an untranslatable form. Equally as important however, these studies demonstrate that the activation of muscle-specific

Introduction

genes is mediated by diffusible trans-acting molecules which are transported to the nuclei through the cytoplasm. Direct evidence that the regulation of actin gene expression is at the level of transcription has come from probing the transcriptional status of these genes in various tissues by DNaseI sensitivity techniques (Carmon et al., 1982). These studies demonstrated that the skeletal muscle α -actin gene is DNaseI sensitive, and therefore transcriptionally active, only in nuclei isolated from differentiated muscle cells. In nuclei isolated from mono-nucleated myoblasts and brain tissues, this sequence is relatively protected from DNase digestion, suggesting that the change in higher order chromatin structure which enables transcription to take place, only occurs immediately before or during cell fusion. Furthermore, it was also demonstrated that this sensitivity to DNaseI does not extend beyond 0.7Kb 5' to the transcribed region of the α -actin gene.

Significant advances in characterising the cis-acting regulatory sequences of the actin genes, and the identification of the trans-acting factors with which they interact, has been achieved in recent years. The demonstration that cloned actin genes introduced into myogenic cell lines are expressed in a similar manner to that of their endogenous counterparts has provided one assay system for the delineation of important cis-acting sequences involved in actin gene expression (for example Grichnick et al., 1986; Minty and Kedes, 1986). However, this assay system works with a varying degree of efficiency,

Introduction

depending on the cell lines used and the gene being studied (see Minty et al., 1986).

In Xenopus, Wilson et al. (1986) followed the expression of a micro-injected Xenopus borealis cardiac actin gene and demonstrated that this gene was expressed in both a temporal and spatial manner along with the endogenous actin gene. Micro-injection of cloned actin genes into developing Xenopus embryos thus provides an alternative assay system for the study of their expression to the introduction of genes into myogenic cell lines. Indeed, this assay system has been exploited to identify sequences involved in the expression of the Xenopus laevis cardiac actin gene (Mohun et al., 1986; Mohun et al., 1989a; Taylor et al., 1989).

In addition to these two assay systems transgenic animals have also been exploited to study the expression of actin genes. The remainder of this section outlines the progress made by the use of these various techniques in identifying cis-acting sequences and trans-acting factors important in the expression of the α -actin genes.

1.3.3. Cis-acting sequences and trans-acting factors involved in the control of cardiac actin gene expression.

Preliminary experiments by Minty and Kedes (1986) demonstrated that the 5' flanking region, the first exon and 28 nucleotides of the first intron of the human cardiac actin gene is sufficient to drive the tissue-specific expression of a chloramphenicol acetyl transferase (CAT) reporter gene when

Introduction

transfected into myogenic cell lines. Moreover, in amphibians micro-injection of cardiac actin genes, which have their equivalent 3' regions replaced by a reporter gene, also exhibit correct spatial and temporal expression throughout early Xenopus development (Wilson et al., 1986; Mohun et al., 1986). Indeed, the 3' region of the mouse cardiac actin gene could also be lost without affecting the expression of the gene in adult tissues, as demonstrated by engineering transgenic mice that contained the mouse cardiac actin gene (Shani, 1986). These experiments demonstrate that the 3' untranslated region of the cardiac actin genes, which are highly conserved between species, are not required for their correct expression. The fact that as little as 24 nucleotides downstream of the transcriptional start site of the Xenopus laevis cardiac actin gene is present in constructs used in the experiments of Mohun et al. (1986), would also suggest that the introns and vast majority of the 5' untranslated region of this gene is also redundant in the control of its expression.

Analysis of sequences contained in the 5'-flanking region of the human cardiac actin gene revealed two separate regions of the promoter involved in the expression of the gene. Deletion of sequences through a distal region (-443 to -395), and subsequently a proximal region (-177 to -118) resulted in significant reductions in the activity of the gene, suggesting that these sequences interact with positive trans-acting factors contained in muscle cells (Minty and Kedes, 1986). Similarly, in studies involving the Xenopus laevis cardiac

Introduction

actin gene, deletion of the promoter revealed a region of the gene spanning from nucleotides -416 to -217 which is essential for its tissue-specific expression (Mohun et al., 1986). In neither case was an increase in promoter activity observed in non-muscular tissues, suggesting the lack of importance of negative regulatory elements in the expression of the cardiac actin genes.

Examination of the promoters of both cardiac and other actin genes reveals the presence of one or more copies of a conserved CC(A/T_{rich})₆GG sequence motif, or CArG box (see figure 1.1). In the case of cardiac actin, four CArG motifs are situated in the promoters of these genes. The most proximal of these four motifs has been termed CArG box1, with the subsequent motifs being numbered through to CArG box4.

Further studies of the sequence requirements of the human cardiac actin gene by linker scanning mutation and fine deletions of the promoter, revealed that CArG box2 was required for the full activity of the gene, whilst the elimination of CArG box1 totally extinguished any promoter activity (Miwa and Kedes, 1987). The Xenopus laevis cardiac actin gene appears to be regulated in much the same manner, with only the most proximal of the four CArG motifs being required for the expression of the gene (Mohun et al., 1989a). Furthermore, it was demonstrated that CArG box2, CArG box3 and CArG box4 could replace the CArG box1 motif without significantly reducing the expression of the cardiac actin gene (Miwa and Kedes, 1987; Mohun et al., 1989a). This finding led to the speculation that

Cardiac actin.

- 1) -89 CGGCCAAATAAGAGAA Chicken.
-112 GGACCAAATAAGCAAGG Human.
-83 TACCAAATAAGGGCA Xenopus laevis.
- 2) -121 TGGCCATTTCATGGCC Chicken.
-152 GCTCCATGAATGGCC Human.
-132 CTCCATTAAATGGCT Xenopus laevis.
- 3) -153 CTGCCCTTAGATGGC Chicken.
-203 CTTCCTTACATGGT Human.
-174 TTCCATACATGGGCT Xenopus laevis.
- 4) -197 GCTCCCTATTGGCCA Chicken.
-240 GCTCCCTATTGGCCA Human.
-220 ATCCCTATTGGCCA Xenopus laevis.

Skeletal actin.

- 1) -96 TGTCCAAATATCGAGT Xenopus borealis.
-96 TGTCCAAATATCGAGT Xenopus laevis.
-95 CACCCAAATATGGC Rat.
-86 CACCCAAATATGGC Chicken.
-101 CACCCAAATATGGCTC Human.
- 2) -127 GGACCCTCAAAGGCCA Xenopus borealis.
-168 CCTTCCTTTGG Rat.
-128 CCTTCCTTTGG Chicken.
-162 GCTCCTTCCTTTGGTCA Human.
- 3) -162 CCACTATATTTGGTCA Xenopus borealis.
-221 CTCCATATACGGAAA Rat.
-177 CTCCCTATACGGAAA Chicken.
-229 ACTCCATATACGGCCC Human.

Introduction

β -actin.

1) -65	TTGCCTTTTATGCC	Human.
-64	TTGCCTTTTATGCC	Rat.
-66	TTTCCTTTTATGCC	Chicken
2) +758	TTGCCTTTTATGGTAAT	Human.
+738	TTGCCTTTTATGGTAAT	Rat.
+683	TTGCCTTTTATGGTAAT	Chicken.

Other muscle-specific genes.

-69	CCAAAATTAGGC	Rat cardiac MHC.
-70	CCAAAAGTGG	Chicken cardiac MLC2.
-122	CCAAAGAAAGG	Rat skeletal MLC2.
-120	CCAAAATAGC	Chicken cardiac Troponin T.
-1233	CCATGTAAGG	Mouse MCK.
-178	CCATACAAGG	Mouse MCK.

FIGURE 1.1. Evolutionary conserved CArG box sequences in muscle-specific genes.

CArG box1 [1]], CArG box2 [2]], CArG box3 [3]] or CArG box4 [4]] motifs are illustrated from the cardiac, skeletal and β -actin genes. CArG sequences from other muscle-specific genes including the rat cardiac myosin heavy chain, chicken cardiac myosin light chain 2, rat skeletal myosin light chain 2, chicken cardiac troponin T and mouse muscle creatine kinase genes are also illustrated. The second set of numbers illustrates the number of nucleotides upstream of the transcriptional start set where the motif is situated.

Introduction

each CArG motif interacts with the same, or functionally interchangeable trans-acting factors.

In vivo competition studies using fragments of the human cardiac actin gene support the suggestion that the CArG motif is the site of interaction for positive trans-acting factors (Miwa et al., 1987). In brief, it was shown that sequences containing either one or both of the CArG box1 and CArG box2 motifs could compete for the binding of positive trans-acting factors in vivo, thus lowering the transcriptional activity of a co-transfected human cardiac actin fusion gene that contained the CArG box1 motif. These experiments also lend support to the suggestion that the CArG box1 and CArG box2 motifs bind similar trans-acting factors.

Band shift and DNA footprinting analysis of human cardiac actin promoter fragments demonstrated the binding of a nuclear factor(s) to a region of the promoter which spans the CArG box1 motif (Gustafson et al., 1988). Furthermore, the authors went on to demonstrate that linker scanning mutations which eliminate the activity of the gene by disruption of the CArG box1 sequence, also eliminate the binding of the trans-acting factor(s) to this sequence in vitro. Similar binding activities were observed using oligonucleotides complementary to the Xenopus laevis cardiac actin CArG box motifs (Mohun et al., 1989a). In addition to this Mohun et al. (1989a) also demonstrated that the ability of each CArG box to compete for the binding of the factor that interacts with the CArG box1 motif was in the order of CArG box1 > CArG box3 > CArG box4 >>

Introduction

CAR_G box2.

However, a recent report demonstrated that CAR_G box binding factor (CBF) was not the only factor capable of binding the human cardiac actin promoter. Gustafson and Kedes (1989) demonstrated that at least seven distinct nuclear proteins were capable of interacting with putative regulatory elements contained within the promoter of the human cardiac actin gene. However it was only recently that the function of these sites was demonstrated in vivo, and this will be discussed in later chapters.

1.3.4. Cis-acting sequences and trans-acting factors involved in the control of skeletal actin gene expression.

As in the case of the cardiac actin gene, so much progress has been made in identifying the regulatory elements which contribute to the expression of the skeletal actin gene, due to the demonstration that these genes are expressed correctly when transfected into myogenic cell lines (for example see Nudel et al., 1985). In the case of the rat skeletal actin gene it was demonstrated that 750 nucleotides of 5' flank, in addition to 2/3 of the structural gene, was sufficient to drive the tissue-specific expression of a reporter gene in myogenic cell lines (Melloul et al., 1984). Indeed, in the case of both the human and chicken skeletal actin genes it was demonstrated that only the 5' flanking region of these genes was needed to induce the expression of a reporter gene on the fusion of myoblasts to form myotubes (Grichnick et al., 1986; Muscat and Kedes, 1987).

Introduction

Furthermore, a region of the chicken skeletal actin gene promoter spanning from nucleotides -2200 to +27 fused to a reporter gene is able to direct the correct tissue-specific expression of this gene in transgenic mice (Petroopoulos *et al.*, 1989). Therefore, it would appear that, as with the cardiac actin genes, all the sequences necessary for the expression of the skeletal actin genes reside in the 5' flanking region.

Preliminary experiments using the chicken skeletal actin gene demonstrated that all sequences required for the expression of this gene are contained within a region of the promoter that spans from nucleotide residue -411 to 12 nucleotides downstream of the proposed TATA box (Grichnick *et al.*, 1986). Further experiments revealed that it was not until sequences downstream of nucleotide -200 in the promoter of this gene were deleted that a loss in transcriptional activity was observed (Bergama *et al.*, 1986). Furthermore, Muscat and Kedes (1987) described a proximal regulatory element which consisted of 153 nucleotides upstream of the transcriptional start site of the human skeletal actin gene that was sufficient to induce the tissue-specific expression of a reporter gene, although at slightly lower levels than that of a construct that contained 2Kb of 5' flank.

As expected, sequence comparison of the promoter regions of the skeletal actin genes reveals considerable evolutionary conservation between species. Furthermore, in parallel with the cardiac actin genes, imperfect repeats of the CArG box motif are also present in the promoters of the skeletal actin genes.

Introduction

However, in contrast to the cardiac actin genes, which contain four CArG boxes in their promoters, the equivalent region of the skeletal actin genes contain only three CArG box motifs.

For both the human (Muscat and Kedes, 1987) and chicken (Bergsma *et al.*, 1986) skeletal actin genes, deletions which remove the most proximal CArG box motif (CArG box1), reduce the activity of the promoter to basal levels. In addition to this, mutation of CArG box1 almost totally eliminates any promoter activity demonstrating that, as with the cardiac actin genes, this sequence element is essential for the full activity of the promoter (Walsh and Schimmel, 1988; Chow and Schwartz, 1990). Indeed, Chow and Schwartz (1990) also demonstrated that mutation of the more distal CArG box motifs also significantly reduced the activity of the promoter, demonstrating that these sequences are also important in the expression of the gene.

It is apparent from several lines of investigation that the CArG box motifs are not the only sequences required for the expression of the skeletal actin genes. For example, it has been proposed that the human skeletal actin gene has three regulatory regions consisting of distal, proximal and basal regulatory elements (Muscat and Kedes, 1987). Although both the distal and proximal regulatory regions can stimulate the expression of a heterologous basal promoter, it is not until both are allowed to act synergistically that high levels of stimulation are observed.

It is likely that the loss of important positive regulatory elements, such as CArG box1, from the skeletal actin promoter

Introduction

masks the contribution of downstream regulatory sequences on the expression of the gene. Indeed, deletion insertion mutations of a region of the chicken skeletal actin gene promoter that spans between the CArG box1 motif and the TATA box significantly reduces the activity of the promoter, suggesting that sequences within this region of the gene are important in the control of its expression (Bergsma et al., 1986). Extensive linker scanning mutation of the chicken skeletal actin gene promoter revealed four positive cis-acting regulatory elements (Chow and Schwartz, 1990). These elements include a ATAAA motif (TATA box), paired CCAAT-box associated repeats (CBAR's or CArG boxes) and an upstream T and A rich regulatory element (UTA regulatory element). Negative cis-acting regulatory elements that surround the most proximal CBAR were also proposed. Indeed, the authors went on to demonstrate that mutation of these putative regulatory elements led to the precocious induction of promoter activity in pre-fusion myoblasts, suggesting the importance of these sequences in subduing the expression of this gene in these cells.

In vitro binding studies using a fragment of the chicken skeletal actin promoter spanning from nucleotides -12 to -148 identified predominantly two separable binding complexes, designated Muscle Actin Promoter Factor 1 (MAPF1) and Muscle Actin Promoter Factor 2 (MAPF2) (Walsh and Schimmel, 1987). Whilst MAPF1 is the predominant binding activity present in nuclear extracts prepared from non-muscular cell lines, MAPF2 is the major binding activity present in nuclear extracts

Introduction

prepared from myogenic cell lines. By band shift analysis and DNA footprinting techniques it was demonstrated that both MAPF1 and MAPF2 bind the same region of DNA, namely the most proximal CArG motif. The elimination of factor binding to this sequence motif by mutation of the site results in a loss of promoter activity when introduced into myogenic cell lines, suggesting the importance of the MAPF factors in vivo (Walsh and Schimmel, 1988).

However, conflicting data have arisen from the studies of the human skeletal actin gene. In vivo competition studies demonstrated that cis-acting sequences present in the distal regulatory element of the human skeletal actin gene were distinct from control elements present in the human cardiac actin gene. However, both in vivo and in vitro competition studies investigating the proximal regulatory region of the human skeletal actin gene suggested that these sequences share common regulatory elements with the human cardiac actin gene (Muscat et al., 1988). Band shift and DNA footprinting analysis of the proximal regulatory region of the human skeletal actin gene localised the binding of a factor(s) to the CArG box1 motif situated in these sequences. In vitro binding studies also demonstrated that a single binding activity existed in extracts prepared from both myogenic and non-myogenic cell lines (Boxer et al., 1989a). The binding activity from both sources was identical in its footprint over the CArG motif, although it was situated more centrally over the CArG box than those observed in experiments performed with the chicken

Introduction

skeletal actin gene promoter. One explanation offered for this discrepancy in data was that the lower molecular weight shift observed in the experiments using the chicken skeletal actin gene promoter was a result of the degradation of the higher molecular weight MAPF1 protein on extract preparation.

In summary, it appears that within the 5' flanking region of the skeletal actin gene the CARG box1 motif plays an important role in the expression of this gene. However, there still remains some controversy as to the number and identity of the factors which interact with this sequence. What is clear is that the skeletal actin gene promoter is more complex than originally conceived. A number of other positive and negative regulatory elements which are important for the expression of the gene appear to exist. The mode of interaction of these regulatory domains with each other and the transcriptional complex itself remains unclear.

It is apparent that the CARG motif plays an essential role in the expression of both the skeletal and cardiac actin genes. Indeed, this motif is found in the upstream regions of other actin genes, such as the chicken, human and rat β -actin genes. In addition to this, CARG motifs are found in the promoters of other muscle-specific genes such as the rat cardiac myosin heavy chain, chicken cardiac myosin light chain, rat skeletal myosin light chain, chicken cardiac troponin T and muscle creatine kinase genes (see Minty and Kedes, 1986; see also figure 1.1). The CARG motif is also present in the regulatory

Introduction

region of the proto-oncogene c-fos. Indeed, by drawing comparisons with proteins that interact with the CArG motif of the c-fos gene promoter, clues as to the identity of the CArG box binding protein that interacts with the actin gene promoters have been provided. The following section therefore briefly reviews the elements involved in the expression of the c-fos gene.

1.3.5. The role of the CArG box in the expression of the proto-oncogene c-fos.

The proto-oncogene c-fos is the cellular cognate of v-fos, the transforming gene of the FBJ-murine osteosarcoma virus (Curran et al., 1983). The c-fos gene encodes a nuclear protein of 380 amino acid residues which has been implicated in the process of cell proliferation and determination. More recently, however, the c-fos gene product has been identified as a transcription factor which, in association with the product of the proto-oncogene c-jun, interacts with AP-1 protein binding-sites situated in the promoters of several genes (for review see Curran and Franza, 1988).

Studies of the expression of c-fos in various cell lines has demonstrated that transcriptional activation of the gene can be induced by addition of serum to quiescent cells (Greenberg and Ziff, 1984). This induction has also been shown to occur on the addition of a variety of growth factors, including platelet derived growth factor (PDGF) (Greenberg and Ziff, 1984; Curran, 1984), fibroblast growth factor (FGF) (Greenberg and Ziff,

Introduction

1984; Curran, 1984), nerve growth factor (NGF) (Greenberg et al., 1985; Kruijer et al., 1985) and epidermal growth factor (EGF) (Curran, 1984). Furthermore, agents that increase the intra-cellular levels of cAMP have also been shown to activate the c-fos gene (Moore et al., 1986; Morgan and Curran, 1986).

Activation of c-fos occurs rapidly, with transcripts appearing as early as 5 minutes after addition of the inducing agent (Kruijer et al., 1985) and reaching a maximum level within 20-30 minutes (Greenberg and Ziff, 1984; Kruijer et al., 1984); Mitchell et al., 1985). This increase in message is reflected in a rapid increase in the c-fos protein approximately 30 minutes after the addition of inducing agents (Kruijer et al., 1984; Curran, 1984). However, expression of the gene is transient, with levels of RNA returning to pre-induction levels 120 minutes after the addition of serum (Greenberg et al., 1985).

The study of the cis-acting elements within the c-fos gene which are responsible for this complex expression pattern has enjoyed much attention over recent years. DNaseI hypersensitive studies revealed sites at -1700, -290, +10, +240 and +700 which are released from a higher order chromatin structure to allow digestion by DNaseI (Renz et al., 1985; Deschamps et al., 1985). Indeed, sequences centred at position -290 in the c-fos promoter are highly conserved between the mouse and human, and these sequences have been shown to possess an enhancer-like structure (Renz et al., 1985).

Deletions through the human c-fos promoter revealed that

Introduction

sequences between nucleotides -332 and -276 are essential for the activation of the gene by serum and growth factors (Treisman, 1985; Gilman et al., 1986). Furthermore, this region of the gene is capable of conferring serum inducibility on a heterologous promoter. In vitro binding studies using this region of the promoter revealed that it was the binding site for a nuclear protein, termed the Serum Response Factor (SRF) (Treisman, 1986; Gilman et al., 1986; Prywes and Roeder, 1986). DNA footprinting techniques have located the binding site for this protein to be contained in a region of dyad symmetry, termed the Serum Response Element (SRE), which spans from nucleotide -320 to nucleotide -229 in the human c-fos gene promoter (Treisman, 1986). Indeed, this sequence appears to possess a CArG motif, which has been identified in the promoters of the actin genes.

A synthetic SRE has been shown to restore the serum inducibility of an otherwise silent c-fos gene, which contains only 261 nucleotides upstream of the transcriptional start site of the gene (Treisman, 1986). Furthermore, mutations within the SRE that eliminate the formation of the DNA-protein complex in vitro, also eliminate the ability of the gene to be induced by serum in vivo (Greenberg et al., 1987). However, the role of SRF in the expression of the c-fos gene is likely to be quite complex. Indeed, SRF binding to the SRE is recoverable from many cell lines and tissues in the apparent absence of growth factor stimulation (Treisman, 1986; Gilman et al., 1986).

SRF has been purified to apparent homogeneity (Treisman, 1987;

Introduction

Prywes and Roeder, 1987; Schroter *et al.*, 1987). The apparent molecular weight of SRF is 67Kd and the purified protein exhibits the same footprint over the SRE as that achieved with crude extract. However, purified SRF exhibits a slightly faster mobility on band shift analysis with SRE sequences than when crude extracts are used (Treisman, 1987).

Characterisation of a cloned SRF demonstrated that the protein binds as a homodimer to the SRE (Norman *et al.*, 1988). The DNA binding/dimerisation region of the protein does not appear to be obviously related to previously identified DNA binding domains such as the 'helix-turn-helix' and 'zinc finger' motifs (Pabo and Saver, 1984; Klug and Rhodes, 1987). However it does bear a striking homology to the yeast proteins MCM1 and ARG80 in the region of the protein known to be involved in dimerisation and DNA binding (Norman *et al.*, 1988).

More recent work has shown the protein complex over the SRE to be more complex than originally conceived, with the discovery that another protein, termed p⁶² Ternary Complex Factor (p⁶²TCF), binds the SRE-SRF complex. Furthermore, the resulting ternary complex is a requirement for efficient gene induction by serum (Shaw *et al.*, 1989). More recently it has been shown that p⁶²TCF alone cannot bind the SRE, but first requires the binding of SRF to the SRE in a dimeric form. Indeed, a region of the SRF protein of approximately 13Kd in size retains the ability to dimerise and bind DNA, in addition to forming a complex with p⁶²TCF (Schroter *et al.*, 1990).

There exist at least two cellular pathways leading to the

Introduction

induction of the c-fos gene. One involves growth factors increasing the cellular degradation of phosphoinositols, resulting in the production of diacylglycerol (DAG) and the mobilisation of calcium, which then activates protein kinase C. This pathway is thought to act through the SRE in the c-fos gene promoter. Indeed, work demonstrating that SRF is phosphorylated, and that this phosphorylation is required for DNA binding, has prompted the suggestion that this may be a target of control by protein kinase C (Prywes *et al.*, 1988; Manak *et al.*, 1990). Furthermore, by the use of inhibitors of diacylglycerol formation, which acts as a secondary messenger in the induction of protein kinase C, evidence has arisen that activation of c-fos transcription depends on SRF phosphorylation (Schalasta and Doppler, 1990).

The c-fos gene is also inducible by agents that elevate the intra-cellular levels of cAMP, such as dibutyryl-cAMP, forskolin and cholera toxin (Greenberg *et al.*, 1985; Krutjter *et al.*, 1985; Bravo *et al.*, 1987). Data from several laboratories has demonstrated that whilst deletion or mutation of the SRE blocks the induction of c-fos by growth factors, the ability of agents that elevate the intra-cellular levels of cAMP to induce the gene remains unchanged, suggesting two independent pathways of action (Gilman, 1988; Sassone-Corsi *et al.*, 1988). Deletion of the c-fos gene promoter revealed that a region spanning between nucleotide residues -65 and -57 is essential for the induction of the gene by cAMP. Indeed, this region of the promoter contains a putative cAMP response element (CRE)

Introduction

(Shang et al., 1988; Sassone-Corsi, 1988). Furthermore, the c-fos CRE is able to confer cAMP responsiveness to a heterologous promoter, and both in vivo and in vitro binding studies suggest that the element binds a bona fide CRE nuclear binding protein (CREB) (Sassone-Corsi et al., 1988).

1.4. Elements involved in the expression of other muscle-specific genes.

With the aim of identifying the essential cis-acting regulatory sequences important in muscle-specific gene expression, much work has focussed on determining the sequences involved in the expression of a variety of muscle-specific genes other than actin.

One such example of another contractile protein which has been extensively characterised is the cardiac troponin T (cTNT) gene. Mar et al. (1988) demonstrated that the efficient induction of this gene in embryonic skeletal muscle cells when myoblasts differentiate into myotubes, requires only 129 nucleotides upstream of the gene's transcriptional start site. By testing the ability of various fragments of the cTNT promoter to induce the tissue-specific expression of a heterologous promoter, sequences responsible for the restricted expression of the gene in skeletal muscle cells were localised to a region of the promoter spanning from nucleotides -50 to -129. This region of the gene contains various putative regulatory elements, including a CArG box, as identified in the actin gene promoters (see section 1.3.3), and a MyoD binding

Introduction

site (see section 1.5), in addition to a SP-1 binding site. However, two copies of a conserved motif (CATTCT), termed the M-CAT motif, were also identified (Mar and Ordahl, 1988). Surprisingly deletion of either the MyoD site or CARG box had little effect on the expression of the gene (Mar and Ordahl, 1990). Conversely, disruption of one or both of the M-CAT motifs significantly reduced the activity of the cTNT promoter, demonstrating their importance in the expression of the cTNT gene (Mar and Ordahl, 1988; Mar and Ordahl, 1990). Indeed, mutations of the M-CAT motif which facilitate a drop in the activity of the promoter, also eliminate the binding of a M-CAT binding factor (MCBF) to this sequence in vitro (Mar and Ordahl, 1990). It appears, therefore, that the regulated expression of the cTNT gene is controlled not by the CARG box, or indeed the MyoD site present in the promoter of the gene, but by two copies of a previously unidentified M-CAT motif. However, the mechanisms by which the M-CAT motif and the MCBF contribute to the tissue-specific expression of the cTNT gene remains unclear, because MCBF appears to be present in tissues of both myogenic and non-myogenic origin (Mar and Ordahl, 1990).

A further member of the troponin isoform gene family is the troponin I gene (TnI). Both the chicken and quail fast skeletal muscle TnI genes have been cloned (Baldwin et al., 1985; Nikovits et al., 1986), and subsequent analysis of sequences involved in the expression of these genes has revealed that a different mechanism for their control has evolved from that of

Introduction

its TNT counterpart.

Two distinct regulatory regions of the quail fast skeletal TnI gene exist. One element contained within the 5'-flanking region of this gene is required for its maximum expression, whilst a second larger regulatory element (1.5Kb) within the first intron is required for differentiation-specific transcription (Konieczny and Emerson, 1987). In the case of the chicken skeletal TnI gene, a region of the gene spanning from position -160 to position +60 is sufficient to drive the tissue-specific expression of a reporter gene. As in the quail skeletal TnI gene two distinct regulatory elements exist in the -160/+60 region of the gene. One region between nucleotides -160 and -40 is required for the full activity of the promoter, whilst a second region contained within the first exon of the gene is also required for optimal expression (Nikovits et al., 1990). Furthermore, the authors also demonstrated that a DNA sequence present in the first exon of the gene, which shares considerable identity to the binding site for the myogenic regulatory protein MyoD (see section 1.5), binds a trans-acting factor in vitro. More recently the internal regulatory element situated in the first intron of the quail skeletal TnI gene has been demonstrated to be capable of binding the muscle regulatory factors MyoD, Myogenin and Myf-5. However, two other regions of the internal regulatory element of the gene, which appear to bind ubiquitous proteins, are also necessary for the correct expression of the gene (Lin et al., 1991).

Myosin is another major constituent of the muscle contractile

Introduction

apparatus. Sequences which contribute to the activation of the various myosin isoform genes during myogenesis have also attracted much attention over recent years.

Work investigating the expression of the myosin heavy chain (MHC) gene has demonstrated that 1.4Kb of 5'-flanking sequence is sufficient to direct the tissue-specific expression of the gene (Bouvagnet et al., 1987). Furthermore, Bouvagnet et al. (1987) demonstrated that nucleotide residues -173 to -142 in the promoter of the gene determine tissue-specificity by inhibiting its expression in non-muscle cells, whilst requiring elements further upstream for full activity. Similarly, negative regulatory elements have been implicated in the expression of the myosin light chain 2 (MLC2) gene. The 5'-flanking region of this gene is sufficient to drive the tissue-specific expression of a reporter gene in muscle cells (Arnold et al., 1988). Furthermore, as little as 64 nucleotides upstream of the transcriptional start site of this gene is sufficient for this function (Braun et al., 1989a). However, further deletion of this promoter results in the chimeric fusion gene being expressed in both non-muscle and muscle cells. Indeed, the promoter of this gene was shown to bind nuclear factors which are present in non-muscle cells, but are absent in muscle cells (Braun et al., 1989a).

A single locus exists for the closely related myosin light chain 1 and myosin light chain 3 isoforms (MLC 1/3), with the mature RNAs being produced by differential transcription from two promoters followed by alternate splicing (Nabeshima et al.,

Introduction

1984; Periasamy et al., 1984; Strehler et al., 1985). The control of the tissue-specific expression of the MLC 1/3 gene appears to be different from other myosin isoform genes. Whilst sequences proximal to the two MLC promoters do not appear to contain tissue-specific regulatory elements, a 0.9Kb segment located >24Kb downstream of the MLC1 promoter appears to exhibit a muscle-specific enhancer activity (Donoghue et al., 1988). Indeed, this region of the gene appears to bind MyoD in vitro (Buskin and Hauschka, 1989). However, in another report, sequences important in the expression of the MLC 1 gene have also been identified upstream of its transcriptional start site (Shirakata et al., 1988).

Initial characterisation of the gene encoding the δ -subunit of the acetylcholine receptor gene, which is expressed exclusively in skeletal muscle, demonstrated that nucleotides spanning from position -148 to position +24 in the gene are sufficient to drive its cell-type specific expression (Baldwin and Burden, 1988). Indeed sequences spanning from nucleotide residues -148 to -95 have been demonstrated to contain all sequences necessary for the correct expression of the gene (Baldwin and Burden, 1989). The authors also went on to demonstrate that this nucleotide region exhibited different band shift patterns when myotube and myoblast extracts were used in these assays, and described a TGCCTGG motif present in this region of the gene which is present in the regulatory regions of a number of other muscle-specific genes.

Similarly 850bp of the 5'-flanking region of the acetyl

Introduction

choline receptor α -subunit gene has been shown to confer cell-type and developmental-type expression on a CAT reporter gene (Klarsfeld et al., 1987). All the sequences necessary for this expression have been demonstrated to be contained within a fragment of the gene's promoter spanning from nucleotide residues -110 to -45 (Piette et al., 1989). Interestingly, this region of the gene contains a TGCCTGG element previously identified in the regulatory region of the acetyl choline receptor δ -subunit gene. Furthermore, other putative regulatory elements are present in this region of the gene including a SP-1 and a MyoD binding site.

The combination of research into the sequences which regulate the expression of various muscle-specific genes has been relatively unsuccessful in identifying a general myogenic regulatory element. However, possibly the most significant steps to characterising such a sequence came from the studies of the muscle creatine kinase (MCK) gene. All sequences required for the transcriptional activation of this gene during myogenesis are contained within 3,300 nucleotides upstream of the transcriptional start site of the gene (Jaynes et al., 1986). More detailed analysis of the mouse MCK gene revealed the presence of an essential cis-acting element contained within the promoter of the gene. Indeed, this region of the gene appears to exhibit enhancer-like properties and to confer muscle-specific expression on a heterologous promoter (Sternberg et al., 1988; Jaynes et al., 1988). A similar muscle-specific enhancer was identified in a 159 nucleotide

Introduction

region of the rat MCK gene which spans from 1031 to 1190 nucleotides upstream of the transcriptional start site of the gene. Furthermore, this region of the gene was demonstrated to interact with 3 possible nuclear proteins in vitro (Horlick and Benfield, 1989).

Buskin and Hauschka (1989), by the use of in vitro binding assays, also demonstrated the binding of a factor to the MCK promoter. Furthermore, this factor was present exclusively in differentiated muscle cells, and was named MEF-1 (myocyte-specific enhancer-binding nuclear factor 1). The binding site for MEF-1 was determined by DNA footprinting techniques and disruption of the site, which eliminated MEF-1 binding, reduced the activity of the transcriptional enhancer.

In addition to this a further myocyte-specific factor has been shown to bind the MCK enhancer at a different site to MEF-1. This factor appears to be distinct from MEF-1 and has been named MEF-2. Another factor (MBF-1) was demonstrated to bind the MEF-2 site. This factor exhibits a reciprocal expression pattern to MEF-2, being present in myoblasts extracts, but being down regulated on the differentiation of myoblasts to form myotubes (Gossett et al., 1989).

The MEF-1 binding site, as described by Buskin and Hauschka (1989), has subsequently been identified in many other muscle-specific genes. More importantly, however, this factor has been demonstrated to be related to the myogenic regulatory protein MyoD.

Introduction

1.5. MyoD and related factors: A 'master-switch' for muscle determination?

Although the study of the promoters of muscle-specific genes has supplied useful information about the factors involved in their expression, the first evidence of a myogenic regulatory set of genes arose from the manipulation of myogenic cell lines.

Early evidence for the existence of myogenic regulatory genes came from experiments in which myoblasts were shown to activate muscle genes in a wide variety of non-myogenic cell lines when heterokaryons were performed between these two cell types (Blau et al., 1983; Wright, 1984). Further evidence for a gene which regulates the myogenic programme arose from studies of the mesodermal cell line C3H10T $\frac{1}{2}$ (10T $\frac{1}{2}$). These cells were shown to be converted to myoblasts by a brief exposure to the demethylating agent 5-azacytidine (Taylor and Jones, 1979). It was hypothesised that 5-azacytidine caused the limited hypomethylation of DNA in the genome, and thus activated certain loci which proceeded to determine the muscle phenotype. Indeed, the frequency of conversion of 10T $\frac{1}{2}$ cells to muscle (up to 50%) suggested that one locus, or a few closely linked loci, were activated following hypo-methylation and were responsible for establishing the myogenic lineage (Konieczny and Emerson, 1984). This hypothesis was supported by genomic transfection experiments, in which DNA from myoblasts, but not from 10T $\frac{1}{2}$ cells, was shown to convert 10T $\frac{1}{2}$ cells to myoblasts with a frequency consistent with the involvement of a single

Introduction

regulatory gene (Lassar et al., 1986).

The first muscle determination and differentiation gene to be identified was isolated by subtractive cDNA hybridisation, and was named myoblast determination gene number 1, or MyoD (Davis et al., 1987). MyoD is expressed in 10T $\frac{1}{2}$ myoblast cell lines after treatment with 5-azacytidine, in addition to other myoblast cell lines and skeletal muscle. It is not expressed in non-muscle tissue, nor in cardiac or smooth muscle. The Xenopus homologue of MyoD has subsequently been cloned. Analysis of Xenopus MyoD message demonstrates that the gene is regulated in both a spatial and temporal manner, appearing at stage 10 of development exclusively in tissues of the embryo that go on to form the somites (Hopwood et al., 1989; Scales et al., 1990).

MyoD, when transfected into several different fibroblast or adipoblast cell lines under the control of a viral promoter, converts these cells into stably proliferating myoblasts, which can subsequently differentiate under the appropriate conditions (Davis et al., 1987). Similarly, introduction of Xenopus MyoD RNA into developing embryos causes the activation of muscle-specific genes in a region of the embryo which does not normally express these genes (the isolated animal caps) (Hopwood and Gurdon, 1990).

By transfecting 10T $\frac{1}{2}$ cells with de-methylated DNA linked to a neomycin resistance gene, Pinney et al., (1988) obtained evidence for a further myogenic regulatory gene, Myd. Southern analysis reveals that Myd does not correspond to MyoD. Subsequent to this other myogenic regulatory genes distinct

Introduction

from MyoD have been isolated. Wright *et al.*, (1989) identified a further MyoD-related gene by subtractive hybridisation techniques, and named this gene myogenin. Following this, cDNAs for mouse myogenin (Edmonson and Olson, 1989), and the related factors Myf-5 (Braun *et al.*, 1989b) and MRF-4, also called Herculin and Myf-6 (Rhodes and Konieczny, 1989; Braun *et al.*, 1990; Miner and Wold, 1990) were isolated independently. All these factors are expressed exclusively in skeletal muscle, and have the ability to activate myogenesis in transfected 10T $\frac{1}{2}$ cells.

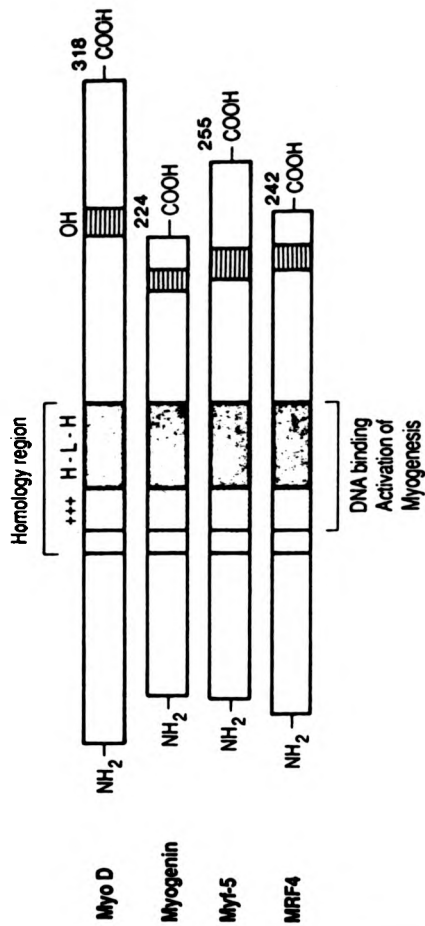
Although all the myogenic regulators identified thus far are expressed exclusively in skeletal muscle tissue, the different patterns of expression of these genes suggests they may be used at different stages of development. For example, myogenin transcripts are detected at high levels in the somite myotome of mice two days prior to the appearance of MyoD or other muscle-specific genes, suggesting that the determination of the muscle phenotype may be independent of MyoD (Sassoon *et al.*, 1989; Wright *et al.*, 1989). In addition to this, MRF-4 is the predominant myogenic regulatory gene in adult skeletal muscle and does not appear to be expressed until after birth (Rhodes and Konieczny, 1989; see also Braun *et al.*, 1990; Miner and Wold, 1990).

The predicted amino acid sequences of myogenin, Myf-5 and MRF-4 share 80% identity with MyoD in a 70 amino acid segment of the protein which contains a basic domain and a region of homology to the *myc* family of proteins (see figure 1.2).

FIGURE 1.2. Structural comparison of the MyoD family
of myogenic regulatory factors.

A linear representation of each myogenic regulatory factor is shown. The region of homology that encompasses the basic (+++) and helix-loop-helix (H-L-H) domains are shown. The corresponding region of MyoD that is sufficient for myogenesis is indicated at the bottom of the figure. The serine/threonine-rich region of homology (OH) is illustrated by stripes. The number of amino acids in each polypeptide is shown at the end of each box.

(Figure from Olson, 1990).



Introduction

Additional homology among these factors is found in a cysteine/histidine-rich region immediately amino terminal to the basic domain, and a serine/threonine-rich region that resembles a site of phosphorylation near the carboxyl terminal.

The basic/myc region has been found in all the myogenic regulatory factors including those from the mouse (Davis et al., 1987; Edmonson and Olson 1989; Miner and Wold, 1990), human (Braun et al., 1989b, 1990), rat (Rhodes and Konieczny, 1989; Wright et al., 1989), Xenopus (Hopwood et al., 1989; Scales et al., 1990; Hopwood et al., 1991), chicken (Lin et al., 1989), quail (de la Brousse and Emerson, 1990) and C. elegans (Benezra et al., 1990). Indeed, the basic/myc region of mouse MyoD has been demonstrated to be necessary and sufficient to activate myogenesis in stably transfected 10T $\frac{1}{2}$ cells (Tapscott et al., 1988).

The myc similarity domain of the myogenic regulatory factors has been postulated to adopt a helix-loop-helix conformation, in which two amphipathic α -helices are separated by an intervening loop of variable length. Indeed this helix-loop-helix (HLH) motif has been identified in a variety of regulatory gene products, including the Drosophila genes achaete-scute (Villares and Cabrera, 1987), daughterless (Caudy et al., 1988; Cronmiller et al., 1988), hairy (Rushlow et al., 1989), extramacrochaetae (Garrell and Modolell, 1990; Ellis et al., 1990), enhancer of split (Klamt et al., 1989) and twist (Thiisse et al., 1988).

Initial clues as to the mechanisms by which the myogenic

Introduction

regulators control the expression of muscle-specific genes arose from the identification of the proteins E12 and E47, which also contain the HLH motif. The observation that these two proteins were capable of binding the κ E2 DNA sequence located in the immunoglobulin kappa chain enhancer, raised the possibility that the HLH myogenic proteins were also capable of binding DNA (Murre et al., 1989a). Subsequent studies revealed that MyoD was indeed a DNA binding protein, which is capable of interacting with the κ E2 site in the immunoglobulin kappa chain enhancer when in a heterodimeric complex with either E12 or E47 (Murre et al., 1989b). Furthermore, the upstream enhancer required for the muscle-specific expression of the muscle creatine kinase gene (see section 1.4) contains two regions which bare a striking resemblance to the κ E2 site, and both MyoD and myogenin have been shown to bind these sequences in vitro when complexed with E12 (Murre et al., 1989b; Brennan and Olson, 1990). Indeed, MyoD shares antigenicity and DNA binding specificity with MEF-1, a myocyte-specific nuclear protein which has been demonstrated to bind the MCK enhancer (Lasser et al., 1989; see also section 1.5). Furthermore, Lasser et al. (1989) proceeded to demonstrate that the basic/HLH motif, which is necessary and sufficient to convert 10T $\frac{1}{2}$ fibroblasts into muscle cells (Tapscott et al., 1988), is also sufficient for specific DNA interaction.

It is apparent that the HLH motif allows the dimerisation of MyoD with E12. However this motif does not confer muscle-specific activation to the MyoD protein, as shown by the

Introduction

finding that the MyoD HLH motif can be substituted by the analogous sequence of the Drosophila T₄ achaete-scute protein with no effect on either DNA binding or muscle-specific gene activation (Davis et al., 1990). However, replacing the basic region of MyoD with the analogous sequence of other HLH proteins (the immunoglobulin enhancer binding protein E12, or the T₄ achaete-scute protein) allows DNA binding in vitro, yet abolishes muscle-specific gene activation (Davis et al., 1990). These findings suggest that the recognition code that determines muscle-specific gene activation lies within the MyoD basic region.

Although MyoD and related myogenic factors are clearly important in the activation of muscle-specific genes, the question arises as to how muscle-specific genes remain unactivated in myoblasts, despite the expression of both MyoD and E12. The recent isolation of a further HLH protein, Id, has made some progress towards answering this question (Benezra et al., 1990). Id contains a HLH dimerisation domain, yet lacks a basic region. This protein was shown to complex with both E12 and MyoD and to repress the ability of MyoD to activate the MCK promoter, presumably because Id forms heterodimers with MyoD and E12 that cannot bind DNA due to the lack of a basic domain. Furthermore, Id message is expressed in myoblasts and falls in concentration as they differentiate into myotubes. This fall in Id has been proposed to release MyoD and E12 from inactive heterodimers and allow them to complex and bind DNA, thus activating the MCK promoter.

Introduction

The identification of MEF-1 as MyoD, and the identification of the MEF-1 binding site (see section 1.4) led to the finding of binding sites for MyoD in many other muscle-specific genes, including the myosin light chain 1/3 (Donoghue et al., 1988), acetyl choline receptor α -subunit (Wang et al., 1990), troponin I (Yutzey et al., 1989; Nikovits et al., 1990; Lin et al., 1991) and Duchenne Muscular Dystrophy (Klamut et al., 1990) genes. Indeed a MyoD/MEF-1 site has also been identified in the human cardiac actin gene (Sartorelli et al., 1990) in addition to the Xenopus laevis cardiac actin gene (T. Mohun, abstract; 3rd international Xenopus meeting, 1990).

The emerging hypothesis that MyoD and related factors are the 'master switch' genes for the activation of muscle-specific gene expression is complicated by the finding that a subset of muscle-specific genes do not contain MyoD/MEF-1 binding sites. In this regard the cardiac troponin T gene contains a putative MyoD/MEF-1 binding site, however this motif does not appear to be required for the tissue-specific expression of the gene (see section 1.4). It is therefore possible that other, as yet unidentified regulatory elements, are required for the expression of certain muscle-specific genes. This, along with the possible role of MyoD in the expression of the Xenopus borealis skeletal actin gene will be discussed in subsequent chapters.

Materials and Methods

CHAPTER 2

MATERIALS.

2.1 General materials.

Restriction enzymes were from Amersham International (U.K.), Northumbria Biologicals Limited (NBL) and Bethesda Research Laboratories, Maryland U.S.A. (BRL). DNA polymerase I was obtained from Amersham International, sequencing grade polymerase from NBL and AMV reverse transcriptase from Life Sciences Inc. (U.S.A.).

All radioisotopes were supplied by Amersham International at the following specific activities: α -³²P-dGTP and α -³²P-dCTP, 3000 Ci/mmol; γ -³²P-ATP, 5000 Ci/mmol; α -³⁵S-dATP, 1000 Ci/mmol; ¹⁴C-chloramphenicol, 57 mCi/mol.

Nitrocellulose sheets (Hybond-C) and nylon filters (Hybond-N) were obtained from Amersham. Type II agarose (medium EEO) was supplied by Sigma chemical company.

Materials for bacteriological media were from Difco laboratories (Michigan, U.S.A.) and Oxoid limited (England).

E.coli transfer RNA (type XX) and salmon testis DNA were supplied by Sigma

Acrylamide was supplied by Fisons, and bisacrylamide by Kodak. X-ray film was from Fuji photo company limited (Japan).

Oligonucleotides were synthesised in the Biological Sciences department by Gill Scott, using an Applied Biosystems automated synthesiser.

Materials and Methods

All other chemicals and reagents were from BDH ('Analar' grade) or from Sigma chemical company unless otherwise stated.

2.2 Stock solutions.

Deionised formamide- formamide was stirred with Amberlite monobed resin ME-3 or MB-4 (5g per 100ml of formamide) until the pH was 7, filtered through Whatman No. 1 filter paper and then stored at -20°C.

Stock acrylamide solutions were deionised as described for formamide, except for 30 minutes, and then stored in the dark at 4°C.

TE - 10mM Tris.HCl (pH8.0), 1mM EDTA.

10x TBE - 0.9M Tris.borate pH8.3, 20mM EDTA.

20x SSC - 3M NaCl, 0.3M Na citrate, pH7.0.

50x Denhardt's medium - 1% (w/v) each of Ficoll, polyvinylpyrrolidone, bovine serum albumin.

1x Barth-X - 88mM NaCl, 1mM KCl, 2.5mM NaHCO₃, 15mM Tris.HCl pH7.6, 0.3mM Ca(NO₃)₂, 0.41mM CaCl₂, 0.82mM MgSO₄.

NAE - 0.3M Na acetate pH6.5, 1mM EDTA.

2.3 Bacteriological media.

LB broth - 5g yeast extract, 10g bactotryptone, 10g NaCl per litre. For LB plates this was supplemented with 15g/litre of bacto agar.

2x TY - 8g NaCl, 16g bactotryptone, 10g yeast extract per litre of water.

Terrific broth (TB) was used to grow bacteria for large scale

Materials and Methods

plasmid isolation. It was prepared by dissolving 12g of bactotryptone and 24g of yeast extract in 800ml of distilled water. 3ml of glycerol were added before autoclaving. Immediately prior to inoculation of this with bacteria, 100ml of 0.17M KH_2PO_4 and 100ml of K_2HPO_4 (both previously sterilised by autoclaving) were added.

All media were sterilised by autoclaving before use.

Ampicillin was used in plates and media at a final concentration of 100ug/ml.

2.4 Bacteria and plasmids.

2.4.1 Genotypes of E.coli strains.

JM101 - supE, thi, (lac-proAB), [F', traD36, pro AB, LACiqZDM15] 17-18 ditto, not tra D36.

MC1061 - F⁻, ara D139, (ara, leu)7696, lac Y74, gal U⁻, gal K⁻, hsr⁻, hsm⁺, strA.

TG2 - supE, (lac-proAB)hsd, (src-recA) 306:Tn10(tet^r) [F' traD36 proAB⁺ lacI^q lacZ M15 r_k⁻ r_m⁻ Rec⁻].

2.4.2 Plasmid vectors.

pBR322 - General purpose cloning vector (Bolivar, 1978).

pAT153 - High copy variant of pBR322 (Twigg and Sherratt, 1980).

Materials and Methods

CHAPTER 3

METHODS.

3.1 Oocytes eggs and embryos.

3.1.1 Oocytes

Oocytes were obtained by anaesthetising a female Xenopus laevis with MS222 and carrying out a partial ovariectomy. The oocytes were manually stripped from the ovary, washed and maintained in full strength Barth-X.

3.1.2 Eggs and embryos.

Females were stimulated to ovulate by injection of 100 i.u. of FSH (Folligon), followed by an injection of 500 i.u. of hCG (Chorulon) on the afternoon of the next day. Ice was added to the water to slow down the hormone effect and the frogs were left overnight. The females were allowed to lay eggs in full strength Barth-X.

A male Xenopus laevis was killed and the testis removed, placed in full strength Barth-X, and stored on ice. Typically 50-100 eggs were transferred to a petri dish and the Barth-X solution removed. A teased testis was briefly agitated over the eggs and after 30 seconds four volumes of distilled water was added to activate the sperm. After a further 5 minutes the petri dish was filled with water. Fertilised eggs rotated

Materials and Methods

within approximately 15 minutes. Eggs were then de-jellied in 2% (w/v) cysteine (pH8.0) and washed in full strength Barth-X.

Prior to gastrulation embryos were transferred to one tenth Barth-X to avoid exo-gastrulation.

3.2 Micro-injection of fertilised eggs.

3.2.1 Preparation of DNA for micro-injection.

DNA was linearised at a unique restriction site contained within the vector sequence of the construct to be injected (normally Pst I). After digestion of the DNA the reaction was extracted with phenol/chloroform and then chloroform. DNA was precipitated by the addition of sodium acetate (pH 6.5) to 0.3M and two volumes of ethanol. After centrifugation the resulting pellet was resuspended in 100ul of TE and precipitated a further three times. DNA was then resuspended at a concentration of 30-40 ug/ml.

3.2.2 Micro-injection of DNA.

DNA was injected into the animal pole of eggs fertilised in vitro whilst at the two cell stage of development. The embryos were first de-jellied before injection under Barth-X medium containing 5% ficoll. Typically 20 nl of a solution containing DNA at a concentration of 30 or 40 ug/ml was injected into each fertilised egg.

The embryos were transferred to one tenth Barth-X at the blastula stage to permit gastrulation, and stored at 14°C until

Materials and Methods

dissections were performed (see section 3.3). Dead or abnormal embryos were removed at regular intervals.

Embryos were staged according to Nieuwkoop and Faber (1956).

3.3 Dissection of embryos.

Dissections were performed at either stage 9, stage 26 or stage 30 of development. For the dissection of embryos at stage 9 the vitelline membranes were removed under Barth-X containing 5% Ficoll. Dissections were performed using forceps and a narrow gauge needle containing a mounted nose hair.

3.4 Nucleic acid isolation from embryos.

Eight to fifteen whole embryos, or embryo fragments were transferred to a glass homogeniser and homogenised in 0.27 ml of Kressmans/SDS buffer (10 mM Tris.HCl pH7.5, 1.5mM $MgCl_2$, 10mM NaCl and 1% SDS) with 0.03 ml of a 10 mg/ml stock of Proteinase K (Boehringer). The homogenate was incubated at 37°C for 30 minutes before extraction with phenol/chloroform and then chloroform. Nucleic acids were precipitated by the addition of 2.5 volumes of ethanol in the presence of 0.3M sodium acetate (pH6.5). After centrifugation for 15 minutes the pellet was washed in 70% ethanol, dried under a vacuum and brought up in 2.5ul of DEPC treated water per embryo, or embryo fragment.

Materials and Methods

3.5 Standard sub-cloning techniques.

3.5.1 Restriction enzyme digests.

Basically, these were carried out according to the manufacturer's instructions. Plasmid DNA's were generally digested for one hour using approximately 10 units of enzyme per μg of DNA. As a rule, no more than 1 μl of enzyme was added per 10 μl of digestion mix to avoid star activity of certain enzymes.

3.5.2 Preparation of plasmid vectors for sub-cloning.

Vectors were cut with the appropriate restriction enzyme(s) as described in section 3.5.1. On completion of the digestion the reaction was made to 0.1% w/v SDS in a total volume of 50 μl of 0.05M Tris.HCl (pH8.0), containing approximately 2.5 units of calf intestinal alkaline phosphatase (CIAP). The mixture was incubated at 37°C for 40 minutes before being extracted once with phenol/chloroform and then chloroform. For enzymes that generated 5' overhangs the reaction was incubated at 60°C for a further 30 minutes before extraction with phenol/chloroform. DNA was precipitated with ethanol and sodium acetate (pH6.5) and redissolved in TE (pH8.0) at an appropriate concentration.

3.5.3 Ligations.

Ligations were carried out in a 10 μl reaction containing 1x C buffer (66mM Tris.HCl pH7.6, 6.6mM MgCl_2 , 10mM DTT), 1mM ATP, vector DNA and 1 unit of T4 DNA ligase. Normally 20ng of vector

Materials and Methods

was used with a range of target DNA concentrations ranging from a ratio of 3:1 to 20:1 (target:vector). A control reaction containing vector alone was always included. Reactions were incubated at 14°C overnight.

3.5.4 E. coli plasmid transformations.

A 10ml culture of 2x TY was inoculated with a colony of E.coli picked from a minimal agar plate and incubated overnight with vigorous shaking. An appropriate volume of fresh 2x TY (normally 10ml) was inoculated with a 1/100th volume of this overnight culture and incubated at 37°C with vigorous shaking until the A₅₅₀ reached 0.5. The culture was cooled on ice for ten minutes and the cells pelleted by centrifugation at 2000 rpm for 5 minutes at 4°C. The supernatant was removed and the cells resuspended gently in 10ml of cold 0.1M MgCl₂. The suspension was incubated on ice for 10 minutes and the cells pelleted as previously described. The cells were then gently resuspended in 2ml of cold 0.1M CaCl₂. After incubation on ice for a further 30 minutes, 0.2ml aliquots of the cells were transferred to Eppendorf tubes on ice (cells could be stored at this point by adding glycerol to 15% (v/v), freezing in liquid nitrogen and storing at -70°C).

To each aliquot of cells, half the ligation mix (5ul) was added and after mixing the tubes were incubated on ice for 30 minutes. The cells were then heat shocked at 42°C for 2 minutes, quenched on ice for 5 minutes and spread onto L-agar

Materials and Methods

plates containing the appropriate antibiotic for selection of transformants. The plates were incubated at 37°C overnight.

3.5.5 Polymerase chain reaction.

Polymerase chain reactions were performed in a 100ul reaction containing 1X AB buffer (50mM KCl, 10mM Tris.HCl pH8.3, 4mM MgCl₂), 0.1mM dNTPs, 200ng of each oligonucleotide primer, approximately 100pg of small scale preparation DNA (see section 3.6), and 2.5 units of Taq DNA polymerase (Perkin Elmer Cetus). Reaction mixtures were overlayed with liquid paraffin and subjected to 30 cycles of 94°C for 1½ minutes, 50°C for 1 minute and 72°C for 1½ minutes. 10ul of the reaction mix was taken and subjected to agarose gel electrophoresis as described in section 3.8.1.

3.6 Small scale preparation of plasmid DNA.

A suitable colony was picked from a plate and used to inoculate 1.7ml of 2x TY and incubated at 37°C for 5½ hours or overnight with vigorous shaking. 0.2ml of the culture was removed, glycerol added to 15% (v/v) and frozen on dry ice. Stocks of cultures were stored at -70°C. The remaining culture was transferred to a 1.6ml Eppendorf tube and centrifuged for 3 minutes. The resulting pellet was resuspended in 0.8ml of STET (8% w/v sucrose, 5% v/v triton X-100, 50mM EDTA, 50mM Tris.HCl, pH8.0) and 25ul of freshly made lysosyme (20mg/ml stock) was added. The mixture was incubated at 96°C for 5 minutes and allowed to cool before the addition of 10ul of a 10 mg/ml stock

Materials and Methods

of RNase A. The mixture was centrifuged for 15 minutes and the pellet of cell debris removed with a micro-pipette tip. The supernatant was extracted with an equal volume of phenol/chloroform and then chloroform. Potassium acetate was added to 0.5M, and nucleic acid precipitated by the addition of $\frac{1}{2}$ volume of iso-propanol. After incubation at room temperature for 10 minutes the solution was centrifuged for 15 minutes and the resulting pellet washed with 70% ethanol. After drying, the pellet was redissolved in 50ul of TE (pH8.0) and re-precipitated by the addition of 2 volumes of ethanol in the presence of 0.3M sodium acetate (pH6.5). After centrifugation for 15 minutes the pellet was washed with 70% ethanol, dried, and resuspended in 20ul of TE (pH8.0).

3.7 Large scale preparation of plasmid DNA and purification by caesium chloride/ethidium bromide centrifugation.

A single colony was inoculated into 10ml of sterile TB containing the appropriate antibiotic, grown at 37°C with shaking until the O.D₅₅₀ was 0.2, and then used to inoculate 250ml of sterile TB containing the relevant antibiotic (0.5ml in 250ml). This culture was grown for 4-5 hours at 37°C with shaking (200 rpm) before being left overnight in the same conditions. The next day cells were incubated on ice for 20 minutes before being pelleted by centrifugation at 5000 rpm for 10 minutes at 4°C. The following quantities used are those for a single 250ml culture. The pellets were resuspended in a total of 15ml of cold STE (50mM NaCl, 5mM EDTA, 50mM Tris.HCl pH8.0),

Materials and Methods

transferred to Oakridge tubes and pelleted as above. The pellets were resuspended in 10ml of 25% sucrose buffer (25% (w/v) sucrose, 100mM Tris.HCl pH8.0). Fresh lysozyme was added (1ml of a 20mg/ml stock) and the resulting solution mixed gently. After incubation on ice for 10 minutes, 4ml of T₅₀E₂₀₀ (200mM EDTA, 50mM Tris.HCl pH8.0) was added. After a further incubation on ice for 4 minutes 15ml of cold lysis buffer (50mM Tris.HCl pH8.0, 50mM EDTA, 2% (v/v) Triton X-100) was added, mixed thoroughly but gently, and incubated on ice for 20 minutes with occasional mixing. Cell debris was then pelleted by centrifugation at 25,000 rpm for 30 minutes at 4°C. The supernatant was recovered and 1g of caesium chloride added per ml of supernatant. After the caesium chloride had dissolved 1.5ml of ethidium bromide (10mg/ml stock) was added. Using a syringe, this solution was transferred to a Beckman Vti50 heat-sealable centrifuge tube. The tubes were topped up with liquid paraffin, balanced to within 10mg and heat sealed. They were then centrifuged at 45,000 rpm in a vertical rotor for 18 hours at 20°C.

The tubes were viewed under U.V. light and the lower band (which is supercoiled plasmid DNA) was removed from the gradient using a syringe and expelled into a Beckman Vti65 heat-sealable tube. Tubes were balanced and sealed as described previously and centrifuged at 60,000 rpm for 6 hours. The tubes were viewed under U.V. light and the lower band removed as previously described. The solution was extracted at least three times with water saturated butan-1-ol to remove the ethidium

Materials and Methods

bromide, and then dialysed against 2 litres of TE overnight at 4°C. The DNA was precipitated in Corex tubes at -20°C by the addition of sodium acetate to a final concentration of 0.3M and 2 volumes of ethanol. The DNA was recovered by centrifugation at 10,000 rpm for 30 minutes at 4°C, rinsed in 70% ethanol, dried under a vacuum and resuspended in 0.5ml of TE. The solution was then extracted with neutral phenol and reprecipitated with ethanol. The DNA was recovered by centrifugation for 10 minutes at room temperature, washed and dried as above, and dissolved in 0.5ml of TE. The DNA concentration was determined by measuring the A₂₆₀.

3.8 Electrophoresis of nucleic acids.

3.8.1 Non-denaturing agarose gels.

DNA samples, to which 0.2 volumes of loading buffer (50% glycerol, 5x TBE, 0.1% bromophenol blue) had been added, were separated in 0.7 to 1.5% (w/v) agarose gels made in 1x TBE buffer containing 0.2 ug/ml ethidium bromide. Gels were run in 1x TBE buffer containing 0.5ug/ml ethidium bromide and examined on a ultraviolet light box.

3.8.2 Isolation of restriction fragments from agarose gels.

Agarose gels were run as described in section 3.8.1 and bands detected on an ultra-violet light transilluminator. A small slot of approximately 2mm in length was cut directly in front of the desired band and the gel replaced in the electrophoresis

Materials and Methods

tank. The tank was filled with 1x TBE until the level of the buffer was just below the top of the gel. The well in front of the relevant band was then filled with 1x TBE and electrophoresed for 30 seconds at 100 volts. The buffer from the well was removed and placed in a 1.6ml Eppendorf tube. The position of the desired band was ascertained by viewing under ultra-violet light. The elution of the band was repeated until none of the DNA fragment remained in the gel (normally approximately five cycles of the above procedure). The solution containing the eluted band was extracted once with phenol/chloroform and the DNA precipitated with 2 volumes of ethanol in the presence of 0.3M sodium acetate (pH6.5). After centrifugation the DNA was washed in 70% ethanol, dried under vacuum and resuspended in TE. The concentration of DNA was estimated by ethidium bromide staining.

3.8.3 Non-denaturing polyacrylamide gels.

8% polyacrylamide (19:1 bis) gels in 1x TBE were poured between 20x 20cm gel plates with 1.5mm spacers. 0.2 volumes of gel loading buffer (50% glycerol, 5x TBE, 0.1% bromophenol blue) was added to samples before being run on gels at 200 volts in 1x TBE buffer. Gels were stained in 1x TBE buffer containing ethidium bromide (0.5 ug/ml) before being viewed and photographed on a ultraviolet light box.

3.8.4 Denaturing polyacrylamide gels.

8% or 10% polyacrylamide (19:1 bis) gels containing 42% (w/v)

Materials and Methods

urea in 1x TBE were poured between 20x 40cm gel plates with 0.4mm spacers. Nucleic acid samples in denaturing gel loading buffer (90% deionised formamide, 10mM EDTA pH8.0, 0.01% bromophenol blue, 0.01% xylene cyanol) were heated at 100°C for 5 minutes, loaded onto the gel and electrophoresed at 38 watts in 1x TBE.

For oligonucleotide gel isolation (see Section 3.9.1) 20x 20cm gel plates with 1.5mm spacers were used and run at 300 volts under the same conditions as described above.

Gels were fixed in 10% (v/v) ethanol, 10% (v/v) acetic acid for 15 minutes, transferred to a sheet of blotting paper and dried at 80°C on a vacuum drier. Gels containing ³⁵S (e.g. sequencing reactions) were exposed to X-ray film at room temperature. However, gels containing ³²P were generally exposed to X-ray film with an intensifying screen at -70°C.

3.9 Preparation of synthetic oligonucleotides.

3.9.1 Gel isolation of synthetic DNA oligonucleotides.

Lyophilised oligonucleotides (usually approximately 1mg) were resuspended in 0.5ml of TE. 25ul of oligonucleotide was added to 25ul of denaturing gel loading buffer (90% deionised formamide, 10mM EDTA pH8.0, 0.01% bromophenol blue, 0.01% xylene cyanol) and the sample heated at 100°C for 5 minutes. Samples were put on ice and loaded onto a 20x 20cm denaturing acrylamide gel and electrophoresed as described in section 3.8.4.

Materials and Methods

Oligonucleotides were visualised by shadowing with ultraviolet light and the band excised from the gel using a razor blade. After cutting the gel slice into fine pieces the acrylamide was transferred to a 1.6ml Eppendorf tube, 0.3ml of elution buffer (0.5M Sodium acetate pH6.5, 1mM EDTA, 0.2% SDS) added and the mixture incubated at 37°C overnight with gentle shaking. The following day the elution buffer was removed and filtered through glass wool. A further 0.2ml of elution buffer was incubated with the acrylamide at 65°C for 5 minutes before it was filtered through glass wool and combined with the previous sample. The resulting solution was extracted with neutral phenol and the DNA precipitated on dry ice/methanol by the addition of 2 volumes of ethanol. DNA was recovered by microfugation, washed with 70% ethanol, dried under a vacuum and resuspended in TE. The DNA concentration was determined by measuring the A_{260} .

3.9.2 Annealing of complementary synthetic oligonucleotides.

Equi-molar amounts of complementary synthetic oligonucleotides were mixed in 1x medium salt restriction digest buffer (10mM Tris.HCl pH7.4, 10mM $MgSO_4$, 50mM NaCl, 1mM DTT), heated at 100°C for 10 minutes, and then left to cool to room temperature in an Eppendorf tube floating in a 1 litre beaker of water at 85°C. Annealed oligonucleotides were then precipitated at -20°C by the addition of sodium acetate to 0.3M and 2 volumes of ethanol. DNA was recovered by microfugation, washed in 70%

Materials and Methods

ethanol, dried under a vacuum and resuspended in TE at the appropriate concentration.

3.10 Methods of radiolabelling DNA.

3.10.1 Nick translation.

Linearised vectors, or DNA fragments isolated from agarose gels were labelled by nick translation for the use as probes in Southern blotting. A 20ul reaction was assembled in 1x NTB (50mM Tris.HCl pH7.2, 10mM MgSO₄, 0.1mM DTT, 50ug/ml BSA) containing 100ng of DNA, 1mM dNTPs (dATP, dCTP, dTTP), 40uCi of α -³²P-dGTP, 1ng of deoxyribonuclease (Sigma) and 10 units of DNA polymerase I (Amersham). The reaction was allowed to proceed at 14°C for 3 to 4 hours after which the mixture was run through a Sephadex G-50 gel filtration column in TE buffer to separate labelled DNA from unincorporated nucleotides. Labelled DNA (first peak of radioactivity) was collected from the column, heated at 100°C for 5 minutes, cooled on ice, and then added to Southern blot hybridisation buffer.

3.10.2 End-labelling DNA with γ -³²P-ATP and T4 polynucleotide kinase.

Single-stranded oligonucleotides (for use in primer extension assays), duplex oligonucleotides (for use in band shift assays) and DNA fragments (for use in band shift assays and DNA footprinting) were labelled using this method. Synthetic oligonucleotides are supplied unphosphorylated at the 5' end,

Materials and Methods

and can therefore be labelled by kinasing with γ - ^{32}P -ATP directly. However DNA fragments had first to be dephosphorylated using calf intestinal alkaline phosphatase (see section 3.5.2) before being added to end labelling reactions.

A 20 μ l reaction was assembled in 1x kinase buffer (50mM Tris.HCl pH7.6, 10mM MgCl_2 , 5mM DTT, 0.1 mM spermidine) containing 4 to 10ng of oligonucleotide or restriction fragment, 60uCi of γ - ^{32}P -ATP and 20 units of T4 polynucleotide kinase. The reaction was incubated at 37°C for 1 hour after which the mixture was extracted with neutral phenol. Labelled DNA was separated from unincorporated nucleotides by passing the mixture down a Sephadex G-50 column. Fractions containing labelled DNA were collected and DNA precipitated by the addition of sodium acetate to a final concentration of 0.3M and two volumes of ethanol. DNA was recovered by microfugation for 10 minutes and resuspended in TE buffer.

3.10.3 Labelling DNA fragments by in-filling.

Up to 1 μ g of restriction enzyme digested DNA was labelled by filling in the 5' overhang with a nucleotide mix containing one radio-labelled nucleotide. A reaction was assembled in 1x TM buffer (10mM Tris.HCl pH7.5, 5mM MgCl_2) containing DNA, nucleotide mix and ten units of the Klenow fragment of DNA polymerase. The nucleotide mix contained 40uCi of an appropriate α - ^{32}P -dNTP, and the other 3 dNTPs (unlabelled) to give a final concentration of 1mM. The reaction was incubated

Materials and Methods

at room temperature for 30 minutes. Labelled DNA was separated from unincorporated nucleotides by Sephadex G-50 gel filtration, and recovered by ethanol precipitation.

This method was routinely used to prepare radiolabelled DNA markers from lambda DNA digested with Eco RI and Hind III, and from pBR322 DNA digested with Hpa II.

3.11 Sequencing of plasmid DNA.

A set of deoxynucleotide and dideoxynucleotide mixes were assembled as shown below:

(A)	A ^o	C ^o	G ^o	T ^o
0.5mM dCTP	20ul	1ul	20ul	20ul
0.5mM dGTP	20ul	20ul	1ul	20ul
0.5mM dTTP	20ul	20ul	20ul	1ul
TE buffer	20ul	20ul	20ul	20ul

(B) Dideoxy NTP working solutions: 0.03mM ddATP, 0.1mM ddCTP, 0.075mM ddGTP, 0.5mM ddTTP. These concentrations were altered as necessary and reduced (generally by half) to allow reading of sequence at greater distance from the primer.

(C) Chase mix: 0.5mM of all 4 dNTPs.

Plasmids from either a small scale isolation of DNA (section 3.6), or a large scale isolation of DNA (section 3.7), were used as template.

Materials and Methods

DNA was denatured by assembling a denaturation mix with a final volume of 20ul which contained approximately 2ug of DNA in 0.2M sodium hydroxide; 0.2mM EDTA. The solution was left at room temperature for 5 minutes, and the DNA precipitated by adding 2ul of ammonium acetate (pH4.5) and 100ul of ethanol and leaving at -20°C overnight. DNA was recovered by microfugation for 15 minutes, the pellet washed in 70% ethanol and dried under a vacuum. DNA was re-suspended in 10ul of a solution made by mixing 1.5ul of reaction buffer (9ul 10x core buffer [100mM Tris.HCl pH8.0, 50mM MgCl₂] and 1ul of 700mM B-mercaptoethanol) with 6.5ul of water and 2ul of the relevant sequencing primer (2ng/ul). Primer was annealed to template by incubation at 37°C for 15 minutes. For each template, four 0.5ml Eppendorf tubes were labelled A, C, G, and T. 1ul of the A^o nucleotide mix and 1ul of the ddATP nucleotide solution was placed in the "A" tube. This was repeated for the "C", "G", and "T" tubes using the appropriate nucleotide mixes. The annealed template/primer mix was centrifuged briefly to bring down any condensation, and then 1ul of α -³⁵S-dATP and 1ul of DNA polymerase Klenow fragment (5 units/ul) added. 2.5ul of this mix was added to each of the tubes containing the nucleotide mixes. The reaction was allowed to proceed at 30°C for 20 minutes after which 1ul of chase mix was added and the tubes incubated for a further 10 minutes at 30°C. The reaction was terminated by the addition of 5ul of denaturing gel loading buffer.

One third of each reaction was analysed on a 6% denaturing polyacrylamide gel (see section 3.8.4). The gel was generally

Materials and Methods

run until the bromophenol blue was at the bottom of the gel and then a second 1/3 of the reaction loaded. Electrophoresis was continued until the bromophenol blue from this loading was at the bottom of the gel. Gels were fixed, dried and autoradiographed as described in section 3.8.4.

3.12 Southern blotting.

DNA samples were separated in non-denaturing agarose gels as described in section 3.8.1 and transferred to nitrocellulose using the method of Southern (1979). To increase the efficiency of transfer of high molecular weight DNA the gel was first treated with 0.25M HCl for 15 minutes at room temperature. This partially hydrolyses the DNA, generating smaller fragments which transfer more efficiently. After rinsing in distilled water the gel was incubated in a large volume of denaturing solution (0.8M NaCl, 0.4M NaOH) with gentle shaking for 30 minutes. This was repeated with fresh denaturing solution. The gel was rinsed with distilled water and then neutralised by incubating with gentle shaking in two changes (for 30 minutes each) of a large volume of neutralising solution (1.5M NaCl, 0.5M Tris.HCl pH7.4). Two pieces of filter paper (Whatman 3MM) moistened with 20x SSC were placed on top of a platform in a tray of 20x SSC so that the ends of the filter paper were submerged in the 20x SSC. The gel was placed on top of the filter paper and the exposed filter paper covered with cling film. A piece of nitrocellulose was cut to the same size as the gel, moistened with 2x SSC and lowered onto it's surface, being

Materials and Methods

careful to ensure that no air bubbles formed. Two pieces of filter paper cut to the same size as the gel were soaked in 2x SSC and lowered onto the surface of the nitrocellulose, again making sure that no air bubbles formed. Tissue paper was stacked on top of this and weighed down with a brick. The gel was left to blot overnight. DNA was fixed to the nitrocellulose filters by baking at 80°C under a vacuum for 3 hours.

Prehybridisation and hybridisation was carried out in heat sealed bags weighed down in a water bath set to the appropriate temperature. Alternatively, when hybridisations were performed at 37°C the heat sealed bags were placed between two glass plates and placed in a 37°C incubator. Filters were prehybridised in 20 to 50ml (depending on the size of the filter) of 5x Denhardt's, 6x SSC, 0.1% SDS, 100ug/ml E. coli tRNA, 50% deionised formamide at either 42 or 37°C for 3 hours to overnight. The prehybridisation solution was discarded and replaced with 10 to 20ml of the same buffer containing the nucleic acid probe. The bag was re-sealed and hybridised at either 42 or 37°C overnight.

Filters were first washed twice (5 minutes each) in 2x SSC, 0.1% SDS at room temperature, and then washed again under the desired conditions (which depend upon the hybridisation stringency required). Finally, the filters were blotted on filter paper and put into heat sealable bags while still damp. Filters were exposed to X-ray film with an intensifying screen at -70°C.

Materials and Methods

3.13 Primer extension.

Single-stranded synthetic oligonucleotide probe, which is antisense to the mRNA to be detected, was end labelled by the polynucleotide kinase method outlined in section 3.10.2. Hybridisation mixes with a total volume of 10ul were assembled containing 100pg of radiolabelled probe and 5ul of nucleic acid extract (two embryos, or embryo equivalents) in 1x hybridisation buffer (0.4M NaCl, 50mM PIPES pH6.0). The hybridisation mixes were drawn and sealed into heat sterilised tubes. Tubes were then incubated at 80 to 85°C for 10 minutes before being transferred to the relevant hybridisation temperature for 3 hours, or overnight. A series of hybridisation temperatures were tested and it was found, for the oligonucleotides used in the majority of cases, that 65°C seemed to give the best results.

The hybridisation mixes were transferred to Eppendorf tubes and 80ul of primer extension mix (0.5mM dATP, dTTP, dGTP and dCTP, 100mM Tris, HCl pH8.0, 10mM DTT, 12mM MgCl₂, 25ug/ml Actinomycin D) containing 10 units of AMV reverse transcriptase was added. The reactions were allowed to proceed for 1 hour at 42°C before being terminated by extraction with phenol/chloroform. DNA was precipitated on dry ice/methanol by the addition of sodium acetate (pH6.5) to a final concentration of 0.3M and 2 volumes of ethanol. DNA was recovered by microfugation for 15 minutes, washed in 70% ethanol and dried under a vacuum. The resulting pellets were resuspended in 4ul of denaturing gel loading dye (90% deionised formamide, 10mM

Materials and Methods

EDTA pH8.0, 0.01% bromophenol blue, 0.01% xylene cyanol), heated at 100°C for 5 minutes and loaded onto a 8% denaturing polyacrylamide gel (see section 3.8.4). The gel was run at 38 watts until the bromophenol blue was approximately 1 inch from the bottom of the plate. The plates were separated, the gel lifted onto blotting paper and fixed and dried as described in section 3.8.4.

3.14 Chloramphenicol acetyltransferase assays.

3.14.1 Preparation of chloramphenicol acetyltransferase extracts from micro-injected embryos.

Groups of 5 to 15 embryos, or dissected embryo fragments, were transferred to sterile homogenisers and the remaining media carefully removed. They were then homogenised in 50ul of 0.25M Tris.HCl (pH7.5). The homogenate was kept on ice when possible and spun in a microfuge at 4°C for 10 minutes. This produced a pellet of cell debris, and a clear supernatant with a white layer of lipid floating on the surface. The clear supernatant was removed taking care to avoid removing any of the lipid layer. Extracts were frozen at -70°C and stored at this temperature until required.

3.14.2 Chloramphenicol acetyltransferase assays using extracts from micro-injected embryos.

Chloramphenicol acetyltransferase (CAT) activity in extracts prepared as described in section 3.14.1 was determined

Materials and Methods

essentially as described by Gorman et al (1982). Reactions contained (in a final volume of 100ul) 0.125M Tris.HCl (pH7.5), 50ul of embryo extract, and 5ul of [14 C] chloramphenicol. Controls contained no extract, or 1ul of chloramphenicol acetyltransferase (10 u/ul, Pharmacia). A 1mg aliquot of solid acetyl coenzyme A (Sigma), which had been stored at -20°C for not more than 2 months, was dissolved in 1.5ml of sterile water to give a 0.66mg/ml solution (this stock solution could be frozen and stored at -70°C for up to a week). The acetyl CoA solution and reaction mixes were then incubated separately at 37°C for 5 minutes, after which the reactions were started by the addition of 5ul of acetyl CoA to each reaction mix. The reaction was allowed to proceed at 37°C for 1 hour and then terminated by the addition of 90ul of 0.25M Tris.HCl (pH7.5) and 1ml of cold ethyl acetate. The solutions were mixed extensively by vortexing (this extracts the chloramphenicol into the ethyl acetate). The tubes were microfuged for 5 minutes and the upper phase (ethyl acetate) transferred to a fresh tube. These were dried in a vacuum desiccator until all the solvent had evaporated. The chloramphenicol was then dissolved in 20ul of cold ethyl acetate. The samples were then spotted onto a thin layer chromatography plate (aluminium backed silica gel 60, 0.2mm thick, E. Merck, FRG) using drawn out glass capillaries. Approximately 1 to 2ul were applied at a time allowing spots to dry between applications. The TLC plate was then placed in a chromatography tank containing 200ml of 95% chloroform, 5% methanol. The TLC plate was left in place

Materials and Methods

until the solvent front was approximately 2 inches from the top of the plate, removed, and allowed to air dry. The plate was then exposed to X-ray film at room temperature.

3.15 Preparation of protein extracts for band shift assays and DNA footprinting.

3.15.1 *Xenopus laevis* oocytes, eggs and embryos.

Protein extracts from *Xenopus* oocytes, eggs and embryos were prepared essentially as described by Mohun et al., (1989).

Between 30 and 50 embryos, embryo fragments, eggs or oocytes were washed briefly in approximately 0.5ml of wash buffer (50mM Tris.HCl, 50mM KCl, 0.1mM EDTA, 2mM DTT, 2ug/ml Leupeptin and 2ug/ml Apoprotinin) before being transferred to a sterile pre-cooled glass homogeniser. All subsequent steps were carried out either on ice or at 4°C. Samples were homogenised in 5ul of extract buffer (wash buffer containing 25% (v/v) glycerol) per embryo, or embryo equivalent. However, when protein extracts were prepared from the somite, gut or ectoderm of stage 16 embryos, samples were homogenised in 5ul of extract buffer for every two embryo fractions.

Homogenates were spun in a microfuge at 4°C for 10 minutes. This results in a pellet of cell debris, and a clear supernatant with a lipid layer floating on the top. The clear supernatant was removed (avoiding taking the lipid layer) and re-centrifuged as before. The supernatant was removed (again

Materials and Methods

avoiding the removal of any lipid), divided into 25ul aliquots, and snap frozen in liquid nitrogen.

3.15.2 Adult Xenopus laevis tissues.

The protocol for this is essentially the same as for oocytes, eggs and embryos (section 3.15.1) except that minced tissues were homogenised in 2 to 3 volumes of extract buffer.

3.16 Band shift assays.

Band shift assays using protein extracts (as prepared in sections 3.15.1 and 3.15.2) were performed essentially as described by Taylor et al., (1989). 25ul binding reactions were carried out on ice for 30 minutes in 1x binding buffer (45mM KCl, 15mM HEPES pH7.9, 5mM spermidine, 1mM MgCl₂, 1mM DTT, 0.5mM PMSF, 0.1mM EDTA, 7% glycerol) containing 1.5ug of salmon sperm DNA, 10 to 20ng of non-specific duplex oligonucleotide (5'-ACAGACCGAAGCTTAGCT-3'), 0.5ng end labelled duplex oligonucleotide probe, and 5ul of protein extract. Band shift reactions containing DNA restriction fragments as a probe contained no non-specific oligonucleotide sequences. In addition to this only 1ug of salmon sperm DNA was added to reactions, with the additional non-specific DNA sequences being made up by the addition of 500ng of pBR322.

For competition analysis, the non-specific duplex oligonucleotide and the pBR322 sequences were replaced with 10 to 20ng of specific duplex oligonucleotide, or 500ng of promoter deletion plasmids respectively.

Materials and Methods

The reactions were electrophoresed on a 5% polyacrylamide gel (29:1 bis) in 0.25x TBE buffer at 200 volts for 1½ to 2 hours at 4°C. The gel was fixed in 10% (v/v) acetic acid, dried onto 3MM Whatman paper and autoradiographed with an intensifying screen at -70°C.

In some instances (i.e. band shift reactions described in chapter 9) a different set of conditions were employed in band shift assays which were essentially as described by Buskin and Hauschka (1989). On these occasions reaction conditions were identical to those previously described, with the exception that the binding buffer was adjusted so that the final concentration in reactions was: 25mM HEPES, pH7.9; 0.5mM EDTA; 0.5mM DTT; 0.5mM PMSF; 50mM KCl; 10% glycerol. In addition to this reactions were incubated on ice for 20 minutes and then at 22°C for a further 5 minutes prior to being subjected to electrophoresis as previously described.

3.17 Antibody band shift assays.

Band shift assays using antibody were essentially as described in section 3.16, except that 1mM EDTA replaced the 1mM MgCl₂ in the binding buffer, and the incubation was for 15 minutes on ice (1ul of antibody was added to the reaction after 5 minutes).

Materials and Methods

3.18 Methylation interference and methylation protection

footprinting of DNA.

DNA was labelled on one strand of the appropriate restriction fragment by polynucleotide kinase (section 3.10.2). For methylation protection assays a normal band shift reaction, which had been scaled up by a factor of five, was assembled using the labelled fragment (total volume of 125ul). After incubation on ice for 30 minutes the reaction mix was incubated at room temperature for 5 minutes before 1ul of dimethyl sulphate [DMS] (Fluka) was added. The mixture was incubated at room temperature for a further 1 minute before being loaded onto a 5% polyacrylamide gel (19:1 bis) and electrophoresed as described in section 3.16.

For methylation interference assays end labelled DNA was first partially methylated by DMS (2 minutes at room temperature) as described by Maxam and Gilbert (1980). This DNA was then used in standard band shift reactions (again scaled up by a factor of five) and electrophoresed as described in section 3.16.

Gels for both protection and interference assays were exposed to X-ray film at 4°C overnight and the retarded and free DNA bands excised from the gel. After cutting into fine pieces the gel slices were incubated in 300ul of elution buffer (0.5M sodium acetate pH6.5, 1mM EDTA, 0.2% SDS) at 37°C with gentle shaking overnight. The following day the elution buffer was removed and filtered through glass wool. Gel slices were then rinsed with a further 0.2ml of elution buffer at 65°C for 5 minutes, this was then filtered through glass wool and combined

Materials and Methods

with the previous sample. DNA was precipitated at -20°C by the addition of 1ml of ethanol. DNA was recovered by centrifugation for 15 minutes, washed in 70% ethanol and dried.

Samples were resuspended in 100ul of 1M piperidine (Fluka) and incubated at 90°C for 30 minutes (this cleaves the DNA selectively at methylated G residues). 1.2ml of butan-1-ol was added and the mixture vortexed until only one phase remained. The mixture was then centrifuged for 5 minutes and the supernatant discarded. DNA was dissolved in 150ul of 1% SDS and re-extracted with butan-1-ol as previously described. DNA was washed in 70% ethanol and dried under vacuum before being dissolved in 150ul of NaE (0.3M sodium acetate pH6.5, 0.1mM EDTA). Samples were then extracted with neutral phenol and the DNA precipitated at -20°C by the addition of 300ul of ethanol. DNA was recovered by centrifugation for 15 minutes, washed in 70% ethanol and dried. Samples were dissolved in 4ul of denaturing loading dye (90% deionised formamide, 10mM EDTA pH8.0, 0.01% bromophenol blue, 0.01% xylene cyanol).

Samples were counted by their Cerenkov emission and normalised so the same number of counts in each sample was loaded onto a 8% denaturing polyacrylamide gel. Electrophoresis was performed as described in section 3.8.4. The gels were fixed and dried prior to being autoradiographed at -70°C with an intensifying screen.

Results and discussion

CHAPTER 4

The preliminary localisation of cis-acting regulatory sequences in the promoter of the *Xenopus borealis* skeletal actin gene.

Introduction.

The introduction of cloned genes into embryos or cell lines is a powerful tool for the study of mechanisms that regulate gene expression. This avenue of research has a great advantage over the study of gene regulation in vitro in that the gene being studied is present in a relatively normal cell environment, and thus is subject to mechanisms of control that more truly reflect those that act on the endogenous gene. Micro-injection of cloned genes into developing Xenopus embryos provides a particularly attractive assay system to study gene expression for several reasons. Firstly, Xenopus embryos can be obtained in large numbers relatively easily and the large size of the egg (approximately 50 times larger than a mouse egg) enables the relatively easy injection of DNA immediately after fertilisation. Furthermore, the early development of Xenopus is particularly rapid with the main body tissues being laid down within a day of fertilisation. This, in addition to the fact that embryos can be readily dissected at early stages of development, allows for the rapid screening of injected clones for both their temporal and spatial expression.

In previous studies Busby and Reeder (1983) have successfully shown the correct temporal activation of injected Xenopus rRNA

Results and discussion

genes. Indeed, micro-injection into Xenopus embryos has also been exploited to study the expression of polymerase II genes. Kreig and Melton (1985) found that the Xenopus gene GS17 showed correct temporal regulation when injected into Xenopus laevis embryos. Furthermore, Wilson et al. (1986) demonstrated the correct spatial and temporal expression of a Xenopus borealis cardiac actin gene in Xenopus laevis embryos.

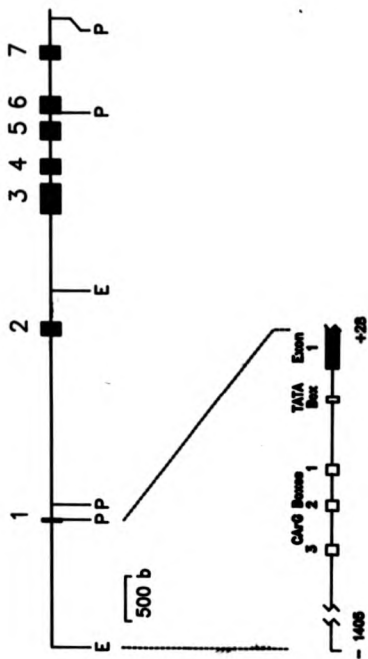
Micro-injection of the Xenopus laevis cardiac actin gene into Xenopus embryos has been used extensively to map sequences which are important for the correct expression of the gene (Mohun et al., 1986; Mohun et al., 1989a; Taylor et al., 1989). Indeed, the use of Xenopus as a transient assay system for the study of tissue-specific gene expression has been proposed to expose the exogenous gene to a more accurate reconstruction of regulatory events in vivo than other assay systems, such as the transfection of cloned genes into specific cell lines. DNA is injected into cells at the two cell stage of development. Thus, the micro-injected gene present in these two cells is passed onto daughter cells which are subjected to many diverse patterns of regulation and differentiation.

With the aim of exploiting Xenopus embryos as a transient assay system to study the expression of a skeletal actin gene, a genomic clone of a Xenopus borealis skeletal actin gene was isolated from a partial Sau 3AI library in λ 47.1 (Boardman et al., in preparation). The complete structure of this gene, including 1405 nucleotides of 5' flanking region, is illustrated in figure 4.0. This skeletal actin gene shares

FIGURE 4.0. Structure of the complete *Xenopus borealis* skeletal actin gene.

Diagrammatic representation of the *Xenopus borealis* skeletal actin gene. Exons are illustrated by filled black boxes, whilst introns and sequences of the 5' flanking region of the gene are illustrated by the intervening lines. The promoter region of the gene is expanded showing the TATA box and putative regulatory elements (CArG boxes) which are discussed later in this chapter.

Restriction sites: Eco RI (E), Pst I (P).



Results and discussion

features that are common to other actin genes. For example, the presence of an intron within the region encoding the 5' untranslated portion of the mRNA is a feature which has been observed in other sarcomeric actin genes (Fornwald et al., 1982; Hamada et al., 1982; Chang et al., 1985). In addition to this the gene is comprised of seven exons and six introns, a feature that is also shared by other skeletal actin genes (Buckingham and Minty, 1983).

The study of the expression of this skeletal actin gene reveals that like other Xenopus actin genes (see Mohun et al., 1984; Wilson et al., 1986) it is activated at stage 12½ of development. Furthermore, the transcripts from this gene are localised exclusively in tissues of the embryo that proceed to form the somites of the embryo (Boardman et al., in preparation).

To begin to localise sequences within this skeletal actin gene which are important for its spatial and temporal expression, a region of the gene containing 1405 nucleotides of 5' flanking region, in addition to 28 nucleotides of its 5' leader, was fused to a Xenopus laevis adult β -globin reporter gene. On injection of this fusion gene into developing Xenopus laevis embryos the correct spatial and temporal distribution of accurately initiated transcripts was observed (Boardman et al., in preparation). Sequences located between nucleotides -1405 and +28 are therefore sufficient to drive the correct expression of a reporter gene in micro-injected embryos, suggesting sequences downstream of nucleotide +28 are

Results and discussion

irrelevant or redundant in the transcriptional expression of the Xenopus borealis skeletal actin gene. This is in agreement with all other sarcomeric actin genes studied thus far (see chapter 1.3 and references therein).

Data outlined in this chapter describe preliminary experiments performed by Dr. M. Boardman which investigate the cis-acting regulatory elements required for the expression of the Xenopus borealis skeletal actin gene.

4.1 Preliminary characterisation of the Xenopus borealis skeletal actin gene promoter.

To further characterise the cis-acting DNA elements within the Xenopus borealis skeletal actin gene which are important for its correct expression, a series of 5' deletions through its promoter were created using exonuclease III. These deletion fragments, which extend to nucleotide +28 at their 3' border, were subsequently cloned onto the Xenopus laevis adult β -globin gene at the same position as the previously described construct containing 1405 nucleotides upstream of the actin mRNA transcriptional start site. The subsequent chimeric fusion genes were therefore identical to the construct containing 1405 nucleotides of upstream sequence, with the exception that they lacked progressively more of their 5' sequence. A diagrammatic representation of the amount of promoter contained within each of these constructs is illustrated in figure 4.1A.

Plasmids were linearised with Pst I prior to injection into the animal poles of Xenopus laevis embryos at the two cell

Results and discussion

stage of development (see chapter 3.2). At stage 9 embryos were dissected into animal, vegetal and equatorial explants. When cultured in isolation the animal, equatorial and vegetal explants of the embryo develop into ectoderm, mesoderm and endoderm respectively. Thus, it is only in the equatorial explants of the embryo that the endogenous skeletal actin gene is expressed.

Embryo explants were allowed to develop in isolation until control embryos reached stage 26 of development. Total nucleic acid was then extracted from embryo explants (see chapter 3.4) and transcripts of the micro-injected gene detected by primer extension analysis, using a primer complementary to sequences in the first exon of the Xenopus globin gene.

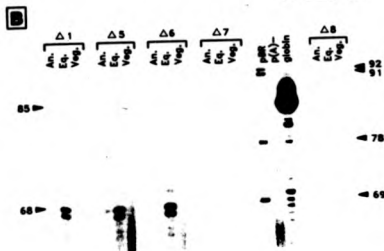
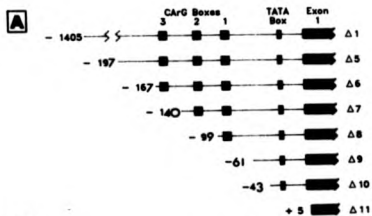
Figure 4.1B shows typical results achieved on the injection of various deletion constructs into Xenopus laevis embryos (Boardman et al., in preparation). As stated previously, 1405 of upstream region of the Xenopus borealis skeletal actin gene (construct pXbgb Δ 1) gives rise to a predicted 68 nucleotide extended product, demonstrating the correct initiation of transcripts from the Xenopus borealis skeletal actin gene promoter. Furthermore, deletion of sequences to nucleotide -167 (construct pXbgb Δ 6) has no significant effect on the expression of the injected fusion gene. This would suggest that all sequences necessary for the correct expression of the gene are situated within a fragment of the gene spanning from nucleotides -167 to +28. However, on deletion of a further 28 nucleotides (construct pXbgb Δ 7), an almost total loss of

FIGURE 4.1. Analysis of transcripts from embryos injected with
Xenopus borealis skeletal actin promoter deletion
plasmids.

A. Diagrammatic representation of the amount of 5' flanking sequences of the Xenopus borealis skeletal actin gene remaining in the constructs pXbgb Δ 1 (Δ 1) to pXbgb Δ 11 (Δ 11). CARG box and TATA box sequence motifs are represented by grey boxes, whilst the first exon of the skeletal actin gene is represented by a filled black box. All promoter fragments were fused to a Xenopus laevis adult β -globin reporter gene at nucleotide +28 in the skeletal actin gene.

B. Detection of transcripts originating from the promoter deletion constructs pXbgb Δ 1, pXbgb Δ 5, pXbgb Δ 6, pXbgb Δ 7 and pXbgb Δ 8 when injected into Xenopus embryos.

Constructs were injected at the two cell stage of development and allowed to develop to stage 9 prior to dissection into animal (An.), vegetal (Veg.) and equatorial (Eq.) regions. These embryo explants were then allowed to develop in isolation until control embryos reached stage 26 of development. Total nucleic acid was extracted and the transcripts of the micro-injected gene detected by primer extension using a primer complementary to sequences situated in the 2nd exon of the globin reporter gene. Control primer extension reactions containing either globin RNA (globin), or poly(A)⁻ RNA extracted from Xenopus ovaries (p(A)-) were also performed. DNA pBR322/Hpa II markers are present in the pBR track.



Results and discussion

promoter activity is observed when this plasmid is injected into Xenopus laevis embryos. Examination of sequences contained within this 28 nucleotide fragment reveals the presence of the most distal of three imperfectly repeated CArG sequence motifs (CArG box3), which have been identified to be present in other actin gene promoters (see chapter 1.3.2, 1.3.3 and Minty and Kedes, 1986). The limited activity of the pXbgb Δ 7 plasmid is subsequently lost on deletion of sequences to directly 5' of the most proximal of these three CArG box motifs (CArG box1) (see construct pXbgb Δ 8).

These data demonstrate that CArG box3, or sequences in its immediate vicinity, are essential for the full activity of the Xenopus borealis skeletal actin gene promoter. This is in agreement with the sequence requirements of other sarcomeric actin gene promoters such as the human cardiac actin (Miwa and Kedes, 1987) and chicken skeletal actin genes (Bergsma et al., 1986). Our data also illustrate that sequences downstream, and including CArG box1, are insufficient to drive the expression of a reporter gene, as detectable by the assay system employed in this study. This is in contrast to the sequence requirements of either the Xenopus laevis or human cardiac actin genes, which remain active on deletion of sequences to directly 5' of their CArG box1 motifs. However, this activity is much lower than that achieved with the intact promoter (Miwa and Kedes, 1987; Mohun et al., 1989a).

Results and discussion

4.2 CARg box3 is essential for the full activity of the *Xenopus borealis* skeletal actin gene.

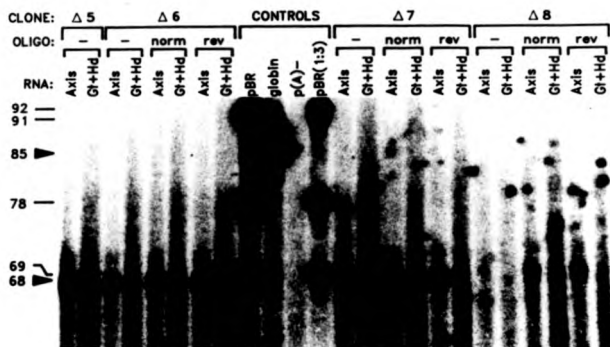
To further investigate the importance of the CARg box3 motif in the expression of the *Xenopus borealis* skeletal actin gene, it was decided to fuse a synthetic CARg box3 motif to the 5' end of each of the previously described promoter deletions and examine the activity of the resulting promoters in *Xenopus laevis* embryos. Thus, seven plasmids were constructed which contained a synthetic CARg box3 motif fused to the 5' end of the skeletal actin promoters present in the plasmids pXbgb Δ 6 to pXbgb Δ 11. The CARg motif was positioned at the 5' end of the promoters in both normal and reverse orientations. When a CARg box3 oligonucleotide was inserted into plasmids in the normal orientation the resulting plasmids were named pXbgb Δ 6+ to pXbgb Δ 11+. However, when the CARg box3 oligonucleotide was inserted into plasmids in the reverse orientation the resulting plasmids were named pXbgb Δ 6- to pXbgb Δ 11-.

Figure 4.2 shows typical results of the activity of these promoters after their injection into developing *Xenopus laevis* embryos. It is clear that whilst the addition of CARg box3 to a promoter that already contains all three CARg boxes (for example see figure 4.2, Δ 4, norm or rev) does not significantly enhance the activity of the promoter, addition of the oligonucleotide to promoters that lack CARg box3 restores the activity of the promoter to levels achieved with the intact gene (see figure 4.2, Δ 7, norm and rev). More interestingly,

FIGURE 4.2. Analysis of transcripts from embryos micro-injected with deletion plasmids containing an additional CARG box3 motif.

Primer extension analysis of total nucleic acid extracted from embryos injected with plasmids pXbgb Δ 6+ (Δ 6, norm), pXbgb Δ 6- (Δ 6, rev), pXbgb Δ 7+ (Δ 7, norm), pXbgb Δ 7- (Δ 7, rev), pXbgb Δ 8+ (Δ 8, norm) and pXbgb Δ 8- (Δ 8, rev). Plasmids that contain no additional CARG box3 motif (pXbgb Δ 5, pXbgb Δ 6, pXbgb Δ 7 and pXbgb Δ 8) were also assayed (- track).

Plasmids were injected into embryos at the 2 cell stage of development and allowed to develop to stage 26 before being dissected into axis (axis) and head plus gut (Gt + Hd) regions. Total nucleic acid was subsequently isolated from these embryo fractions and the transcripts of the micro-injected gene detected by primer extension using a primer complementary to sequences situated in the second exon of the globin reporter gene. Control primer extension reactions containing either globin RNA (globin), or poly(A)⁻ RNA extracted from Xenopus ovaries (p(A)⁻) were also performed. DNA pBR322/Hpa II markers are present in the pBR track.



Results and discussion

however, the addition of a synthetic CArG box3 in a position directly 5' of CArG box1 (see figure 4.2, $\Delta 8$, norm and rev) restores the activity of an otherwise silent promoter. Thus, CArG box3 in conjunction with sequences between -99 and +28 of the skeletal actin gene, is capable of adjusting the level of expression of a reporter gene to the same level as seen in a plasmid which contains sequences from -165 to +28 of the actin gene. It would therefore appear that sequences in and adjacent to the centre CArG box (CArG box2) are redundant for the expression of the gene, at least in stage 26 embryos.

4.3 Discussion.

Deletion analysis of the Xenopus borealis skeletal actin gene promoter has identified the 5' border for essential cis-acting regulatory sequences as 167 nucleotides upstream of the transcriptional start site of the gene. Thus, sequences present between nucleotide -167 and the point of fusion to the reporter gene (nucleotide +28) contain all sequences necessary for the correct expression of the gene. Other muscle-specific genes such as the muscle creatine kinase gene (Jaynes et al., 1988; Horlick and Benfield, 1989) and myosin 1/3 light chain genes (Donoghue et al., 1988), possess essential cis-acting regulatory sequences which are positioned at large distances upstream or downstream of the transcriptional start site of the gene. However, data presented in this chapter would argue against such sequences being important in the expression of the

Results and discussion

Xenopus borealis skeletal actin gene.

The relatively small amount of 5' flanking region capable of directing the correct expression of the Xenopus borealis skeletal actin gene is in keeping with data obtained from the study of other skeletal actin genes. For example, 200 nucleotides upstream of the transcriptional start site of the chicken skeletal actin gene have been demonstrated to drive the expression of a CAT fusion gene as effectively as promoters containing 2.0Kb of upstream sequence (Bergsma et al., 1986). The same would also appear to be true for the human cardiac actin gene, although more 5' flank (485 nucleotides) has to be present for the full activity of the promoter (Minty and Kedes, 1986).

Sequence comparison of the promoter of the Xenopus borealis skeletal actin gene with other sarcomeric actin gene promoters reveals the presence of three conserved imperfect repeats (see figure 4.3). The sequences of these repeats is in general CC(A/T rich)₆GG, and thus would appear to be the CArG motifs which were first identified in the promoter of the human cardiac actin gene (Minty and Kedes, 1986). Indeed these CArG motifs have been observed in the promoters of all sarcomeric actin genes studied thus far (Minty and Kedes, 1986; Mohun et al., 1986; Walsh and Schimmel, 1987; Chow and Schwartz, 1990), in addition to the promoter of the Xenopus cytoskeletal actin gene (Mohun and Garrett, 1987).

Deletion of sequences in, and adjacent to the CArG box3 motif of the Xenopus borealis skeletal actin gene promoter results in

A. Cardiac actin.

CArg box1.

-89	CGGCCAAATAAGAGAA	Chicken.
-112	GGACCAATAAGGCAAGG	Human.
-93	TACCAATAAGGGCA	<u>Xenopus laevis.</u>

CArg box2.

-121	TGGCCATTTCATGGCC	Chicken.
-152	GCTCCATGAATGGCC	Human.
-132	CTCCATTAATGGCT	<u>Xenopus laevis.</u>

CArg box3.

-153	CTGCCCTTAGATGCC	Chicken.
-203	CTTCCTTACATGGT	Human.
-174	TTCCATACATGGGCT	<u>Xenopus laevis.</u>

CArg box4.

-197	GCTCCCTATTGGCCA	Chicken.
-240	GCTCCCTATTGGCCA	Human.
-220	ATCCCTATTGGCCA	<u>Xenopus laevis.</u>

Results and discussion

B. Skeletal actin.

CARG box1.

-96	TGTCCTAAATATGGAGT	<u>Xenopus borealis.</u>
-96	TGTCCTAAATATGGAGT	<u>Xenopus laevis.</u>
-95	CACCTAAATATGGC	Rat.
-86	CACCTAAATATGGC	Chicken.
-101	CACCTAAATATGGCTC	Human.

CARG box2.

-137	GGACCTCAAGGCCA	<u>Xenopus borealis.</u>
-168	CCTCTTTGC	Rat.
-128	CCTCTTTGC	Chicken.
-162	GCTCCTCTTTGGTCA	Human.

CARG box3.

-165	CCACTATATTGGTCA	<u>Xenopus borealis.</u>
-221	CTCCATATACGGAAA	Rat.
-177	CTCCTTATACGGAAA	Chicken.
-229	ACTCCATATACGGCCC	Human.

FIGURE 4.3.

A. Sequence comparison of CARG box motifs of the chicken (Eldridge et al., 1985), Human (Minty and Kedes, 1986) and Xenopus laevis (Mohun et al., 1989a) cardiac actin genes.

B. Sequence comparison of Xenopus borealis skeletal actin CARG box motifs with the equivalent motifs from the rat (Ordahl and Cooper, 1983), chicken (Nudel et al., 1985) human (Boxer et al., 1989a) and Xenopus laevis (Stutz and Sphor, 1986) skeletal actin genes.

Results and discussion

an almost total loss in its transcriptional activity. Furthermore, the addition of a synthetic CArG box3 oligonucleotide to the 5' end of a promoter lacking sequences upstream of the CArG box1 motif restores activity of an otherwise silent fusion gene. Thus, it is the loss of CArG box3, and not sequences that are also deleted from the skeletal actin promoter on the preparation of the plasmid pXbgb Δ 7, which is responsible for the observed loss of activity of the promoter. Furthermore, this experiment also demonstrates that the decrease in fusion gene activity on the deletion of promoter sequences is due to the loss of important regulatory elements, and not to the positioning of inhibitory sequence elements present in the vector closer to the skeletal actin gene promoter. More interestingly however, this experiment would also suggest the redundancy of CArG box2 and adjacent sequences in the expression of the Xenopus borealis skeletal actin gene.

Work investigating the sequence requirements for the effective expression of cardiac actin genes suggests that CArG box1 is of primary importance for their expression (Miwa and Kedes, 1987; Mohun et al., 1989a). However, studies of the chicken skeletal actin gene demonstrate that deletion of a putative CArG box3 from the promoter region of the gene results in a substantial loss in transcriptional activity, despite the presence of the two more proximal CArG motifs (Bergama et al., 1986). In addition to this, deletion of the promoter to directly 5' of CArG box1 results in an almost total loss of expression of the

Results and discussion

gene. It would therefore appear that although CArG box1 may be the essential CArG sequence involved in the expression of the cardiac actin genes, it may be that other upstream sequences are important in the expression of the skeletal actin genes.

It is apparent from experiments described in this chapter that the CArG box3 motif is essential for the expression of the Xenopus borealis skeletal actin gene. Experiments also demonstrate that sequences in and adjacent to the CArG box2 motif are not required for the effective expression of the skeletal actin gene in stage 26 Xenopus embryos. Although experiments have not been performed which directly address the importance of the CArG box1 motif in the expression of the Xenopus borealis skeletal actin gene, work investigating the expression of other actin genes would support the suggestion that this sequence is important in the expression of the gene.

Results and Discussion

CHAPTER 5

The promoter of the *Xenopus borealis* skeletal actin gene binds trans-acting factors in vitro.

Introduction.

Experiments described in chapter 4 demonstrate that as little as 167 nucleotides of upstream sequence of the *Xenopus borealis* skeletal actin gene promoter is sufficient for the competent expression of the gene when injected into *Xenopus laevis* embryos. Furthermore, deletion of the CArG box3 motif from the promoter of the gene results in an almost total loss of promoter activity in vivo. It is therefore possible to speculate that the CArG box3 motif, and possibly other sequence elements present in the promoter of the skeletal actin gene, are capable of interacting with positive trans-acting factors. No increase in promoter activity was observed when any of the deletion plasmids were injected into *Xenopus* embryos, suggesting that no negative regulatory elements are present within the promoter of the *Xenopus borealis* skeletal actin gene.

Experiments outlined in this chapter investigate the binding of potential trans-acting factors to the promoter of the *Xenopus borealis* skeletal actin gene. Furthermore, experiments are also described which identify the specific sequences in the promoter of the gene with which these factors interact.

Results and Discussion

5.1 Identification of a sequence-specific trans-acting factor(s) that interacts with the *Xenopus borealis* skeletal actin gene promoter in vitro.

In order to characterise factors which interact with the promoter of the *Xenopus borealis* skeletal actin gene, a fragment of the 5' flanking region of the gene was subjected to band shift analysis using protein extracts prepared from the axis region of stage 30 *Xenopus laevis* embryos. In preliminary experiments an Eco RI/Bal I restriction fragment from the deletion construct pXbg Δ 6 was used as a probe in band shift assays. This restriction fragment contains sequences of the *Xenopus borealis* skeletal actin promoter downstream of nucleotide -167, to a Bal I site positioned immediately 5' to the putative TATA box of the gene (nucleotide -42) (see figure 5.1A). CArG box3, which has been demonstrated to be essential for the expression of the gene (see chapter 4), is therefore present within this fragment. Furthermore, CArG box1 and CArG box2 are also contained within these sequences; the former of these having been implicated in the expression of other sarcomeric actin genes (Miwa and Kedes, 1987; Mohun *et al.*, 1989a; Walsh and Schimmel, 1988).

Restriction fragments of the *Xenopus borealis* skeletal actin gene promoter were end-labelled with $\gamma^{32}\text{P}$ -ATP after treatment with calf intestinal alkaline phosphatase (see chapter 3.5.2). Fragments were then incubated in binding reactions containing increasing amounts of non-specific competitor DNA in the form of sheared salmon sperm DNA (ssDNA) or poly dI.dC. Absence of a

Results and Discussion

non-specific DNA competitor in band shift reactions results in a smear of the probe when subjected to gel electrophoresis, due to the binding of a variety of non-specific proteins to the DNA probe. However, the addition of either ssDNA or poly dI.dC to binding reactions results in the emergence of two predominant shifted bands (Figure 5.1B). The faster migrating of these two shifted bands was not consistently observed in subsequent experiments using other extract preparations, questioning the authenticity of this binding activity. Conversely, the second shifted band was constantly observed in band shift analysis using this promoter fragment. Indeed, this band can be seen to persist even at high concentrations of non-specific DNA competitors (for example 5ug of ssDNA and 4ug of poly dI.dC). In the light of this experiment it was decided to use 1ug of ssDNA, in addition to 0.5ug of plasmid DNA, in all subsequent binding reactions as this produced the cleanest result, without significantly decreasing the intensity of these shifted bands.

To investigate whether the interaction of factors with the skeletal actin promoter is sequence-specific, competition of this binding activity was performed using promoter sequences contained within the construct pXbgb Δ 6. To achieve this pBR322 DNA present in normal binding reactions was replaced with an equivalent amount of pXbgb Δ 6 in competition binding reactions.

As illustrated in figure 5.1C, the previously described binding activity is apparent in normal binding reactions containing no competitor (track N, figure 5.1C). However, competition with a 12.5 molar excess of skeletal actin promoter

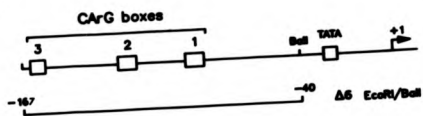
FIGURE 5.1. Band shift analysis of the *Xenopus borealis* skeletal actin gene promoter.

A. Diagrammatic representation of the Eco RI/Bal I restriction fragment of the *Xenopus borealis* skeletal actin gene promoter used in band shift analysis. The complete promoter of the gene is illustrated above the restriction fragment, showing the three conserved CArG motifs and the TATA box.

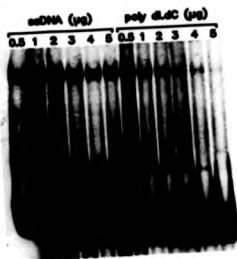
B. Band shift reactions using the Eco RI/Bal I restriction fragment of the *Xenopus borealis* skeletal actin gene as a probe, in conjunction with protein extracts prepared from the axis region of stage 30 embryos. Increasing amounts of either salmon sperm DNA (ssDNA) or poly dI.dC non-specific competitor was added to binding reactions. The amount of non-specific DNA added to reactions is signified above the relevant tracks.

C. Analysis of the capability of the pXbgb Δ 6 promoter to compete for the binding of factors to the Eco RI/Bal I fragment of the *Xenopus borealis* skeletal actin gene. Track N represents no competition and contains 500ng of pBR322, whilst track C represents competition with 500ng of the plasmid pXbgb Δ 6 replacing the pBR322 (12.5 molar excess of promoter).

A.



B.



C.



Results and Discussion

sequences results in a significant decrease in the intensity of the shifted band (track C, figure 5.1C). This decrease in the binding activity is unlikely to be caused by the addition of plasmid sequences to binding reactions, as deletion constructs which are added to competition reactions replace pBR322 DNA that is present in normal binding reactions.

A fragment of the Xenopus borealis skeletal actin gene promoter which spans from nucleotides -167 to -42 is therefore capable of interacting with a sequence-specific trans-acting factor(s) in vitro.

5.2 Preliminary localisation of the binding sites for factors in the promoter of the Xenopus borealis skeletal actin gene.

The existence of factors which are capable of binding to the promoter of the Xenopus borealis skeletal actin gene raises several interesting questions. Firstly, is the promoter capable of interacting with one or several trans-acting factors, and secondly, to which sequences in the promoter do these factors bind?

In order to localise the region(s) of the Xenopus borealis skeletal actin gene promoter which are capable of binding factors a series of 5' promoter deletion fragments were used in band shift assays. Restriction fragments which contain sequences of the skeletal actin gene spanning from the Bal I site (nucleotide -42) to the Eco RI site at the 5' end of the promoter were isolated from the deletion constructs pXbgb Δ 6, pXbgb Δ 7, pXbgb Δ 8 and pXbgb Δ 9. The 5' end of each of these

Results and Discussion

restriction fragments corresponds to nucleotide positions -167, -140, -99 and -61 respectively in the promoter of the gene (see figure 5.2A). Each fragment was end labelled with $\gamma^{32}\text{P}$ -ATP and used in band shift assays in conjunction with protein extracts prepared from the axis region of stage 30 Xenopus laevis embryos.

Results illustrated in figure 5.2B show the previously described shifted band in binding reactions using the -167/-42 restriction fragment of the Xenopus borealis skeletal actin gene as a probe. The deletion of 27 nucleotides of 5' sequence from this fragment (pXbgb Δ 7, track 2) appears to result in an increase in the mobility of the retarded band, when compared with the retarded band present in pXbgb Δ 6 binding reactions. It is an attractive idea that this increase in mobility may be due to the elimination of a factor(s) which binds the promoter fragment, thus resulting in a smaller protein-DNA complex which can migrate more rapidly through the gel. Although this may be the case, the effect is more likely to result from the reduced size of the DNA fragment contributing to this relatively rapid mobility, rather than the number of proteins complexed with it.

It is apparent that the loss of 27 nucleotides from the 5' end of the -167/-42 skeletal actin promoter fragment results in only a slight decrease in the intensity of the shifted band. Whilst this slight decrease in the binding activity may represent the elimination of sequences capable of binding a trans-acting factor(s), the fact that CARG box3 is essential for the full expression of the gene makes it surprising that no

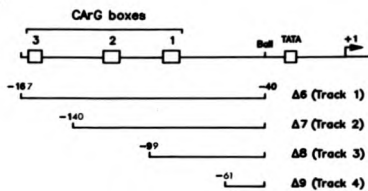
FIGURE 5.2 Band shift analysis of *Xenopus borealis* skeletal actin gene promoter fragments containing progressively less 5' flanking sequences.

A. Diagrammatic representation of the regions of the *Xenopus borealis* skeletal actin promoter contained within the Eco RI/Bal I restriction fragments isolated from plasmids pXbgb Δ 6, pXbgb Δ 7, pXbgb Δ 8 and pXbgb Δ 9.

B. Band shift analysis of promoter deletion fragments using protein extracts prepared from the axis region of stage 30 *Xenopus laevis* embryos. Binding reactions contained promoter fragments isolated from the plasmids pXbgb Δ 6 (track 1), pXbgb Δ 7 (track 2), pXbgb Δ 8 (track 3) or pXbgb Δ 9 (track 4). Track C represents the competition for binding of factors to the -167/-42 fragment of the skeletal actin promoter by a 12.5 molar excess of the Xbgb Δ 6 promoter (500ng of plasmid pXbgb Δ 6).

se
ti
b
p
w
t
p
c
f
5
1
P
a
o
s
b
s
f
H
s
R
R

A.



B.



Results and Discussion

significant binding to these sequences is apparent. One explanation for this is that proteins which bind to CArG box3 also have the ability to bind the related CArG box1 motif. Thus, on deletion of CArG box3 from the promoter of the gene, proteins that usually bind the CArG box3 motif bind to CArG box1. Alternatively, the CArG box3 motif present in the -167/-42 restriction fragment may not be capable of binding a trans-acting factor when used in these band shift assays, due to the motif being situated directly adjacent to the 5' end of the promoter. The protein may not be able to properly 'grip' the DNA strand resulting in a decreased binding efficiency. The fact that DNA footprinting data (see chapter 5.3) does not detect any protection over CArG box3, despite the fact that a synthetic CArG box3 oligonucleotide can bind a protein(s) in vitro (see chapter 6) lends support to this hypothesis.

Deletion of the promoter to nucleotide -99 (construct pXbgb Δ 8, track 3) results in a further increase in the mobility of the shifted band. More interestingly however is the marked decrease in the intensity of this band, indicating a marked decrease in the ability of the promoter to bind factors. This observation is slightly confusing, when taken in conjunction with micro-injection data which demonstrate that a promoter deleted to nucleotide -99 (plasmid pXbgb Δ 8), containing an additional CArG box3 motif, is transcriptionally active in Xenopus embryos (see chapter 4.2). One explanation for this inconsistency in data is that the removal of sequences which span between nucleotides -140 and -99 may indeed result

Results and Discussion

in the loss of factors which are capable of binding this region of the promoter in vitro. However these factors may be redundant in the control of skeletal actin gene expression in vivo.

Deletion of sequences to nucleotide -61 results in an almost total loss of factor binding to the resulting DNA fragment (figure 5.2B, track 4). This final deletion removes the most proximal CARG box motif which has been implicated in the expression of other sarcomeric actin genes (Miwa and Kedes, 1987; Mohun et al., 1989a; Walsh and Schimmel, 1988). It is interesting to speculate that CARG box1 is the DNA binding site for factors which interact with the pXgb Δ 8 promoter fragment. However, more accurate methods of identifying the specific sequences of the promoter which interact with these factors are needed to corroborate this hypothesis.

5.3. DNA footprinting of the -167/-42 region of the *Xenopus borealis* skeletal actin gene promoter.

In an attempt to more accurately define sequences within the *Xenopus borealis* skeletal actin gene promoter which interact with regulatory factors, it was decided to localise factor binding sites in the -167/-42 region of the skeletal actin gene by DNA footprinting techniques.

The Eco RI/Bal I restriction fragment of the *Xenopus borealis* skeletal actin gene (nucleotides -167 to -42) was isolated from the plasmid pXgb Δ 6. This restriction fragment was then end-labelled on either the sense or anti-sense strand of DNA and

Results and Discussion

subjected to either dimethyl-sulphate (DMS) protection, or DMS interference footprinting assays using protein extracts prepared from the axis region of stage 30 Xenopus laevis embryos (see chapter 3.18).

Results of a DMS protection footprint are illustrated in figure 5.3A. It is apparent that the methylation of guanine residues by DMS is protected at nucleotide residues -84 and -85 on the sense strand of the promoter, and at nucleotide residues -92 and -93 on the anti-sense strand. Both these pairs of nucleotides correspond to the 5' and 3' borders of the CArG box1 motif. However, no significant protection of nucleotides is observed immediately upstream or downstream of this sequence. Similar results were obtained using DMS interference footprinting assays studying the anti-sense strand of the -167/-42 promoter fragment (figure 5.3B). These data demonstrate that the observed abolition of any detectable binding activity to the skeletal actin gene promoter on deletion of nucleotides -99 to -61 was indeed due to the loss of the CArG box1 sequence motif.

Interestingly, no protection of sequences was observed over the CArG box2 motif. Indeed, this lack of protection extended over the whole region of the promoter spanning from nucleotides -140 to -99. Although these data are in agreement with micro-injection studies that demonstrate the lack of importance of the CArG box2 sequence in the expression of the gene, they are somewhat surprising in light of band shift assays which demonstrated a significant loss in binding activity on the

FIGURE 5.3A. DMS protection footprinting of the *Xenopus borealis* skeletal actin gene promoter.

DMS protection footprinting of the *Xenopus borealis* skeletal actin gene promoter spanning from nucleotides -167 to -42.

The Eco RI/Bal I restriction fragment of the plasmid pXbgb Δ 6 was end labelled on either the sense or anti-sense strand. These promoter fragments were then subjected to DMS protection footprinting using protein extracts prepared from the axis region of stage 30 *Xenopus laevis* embryos (see chapter 3.18).

Footprinting of both the sense and anti-sense strand of DNA is illustrated and the positions of each CArG box motif on the ladder of DNA fragments is shown. In addition to bound (track B) and free (track F) DNA being analysed, a sample consisting of a partial chemical cleavage of the probe at G residues was also loaded onto the gel (track G).

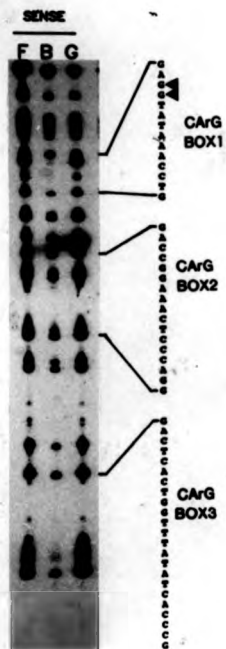
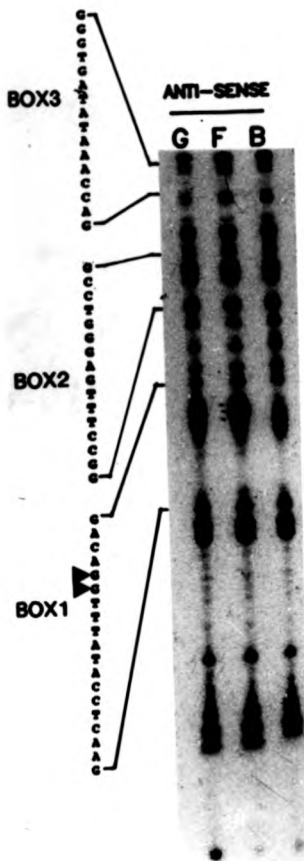


FIGURE 5.3B. DMS interference footprinting of the *Xenopus borealis* skeletal actin gene promoter.

DMS interference footprinting of the anti-sense strand of the *Xenopus borealis* skeletal actin gene promoter spanning from nucleotide -167 to nucleotide -42.

The Eco RI/Bal I restriction fragment of the plasmid pXbgt Δ 6 was end labelled on the anti-sense strand. This promoter fragment was then subjected to DMS interference footprinting using protein extracts prepared from the axis region of stage 30 *Xenopus laevis* embryos (see chapter 3.18).

Free (track F) and bound (track B) fractions of the binding reaction are illustrated in addition to a partial chemical cleavage of the probe at G residues (track G). Sequences of each of the CArG box motifs at their relevant positions on the DNA strand are illustrated.



Results and Discussion

deletion of these sequences. One possible explanation for this apparent loss in binding activity is that deletion of sequences spanning from nucleotides -140 to -99 affects the binding of proteins to the CArG box1 motif. This could be facilitated by the positioning of CArG box1 at the end of the DNA fragment, thus resulting in a decreased efficiency of factors binding this sequence.

Protection of sequences at the 5' end of the anti-sense strand of the promoter are more difficult to identify, because this region of DNA is too far up the gel to read accurately. However, on the sense strand only a little, if any protection can be observed over the CArG box3 sequence. These data are in agreement with those obtained from using deletion fragments in band shift assays, which showed only a slight decrease in the activity of the shifted band on the deletion of the CArG box3 motif. However, micro-injection data that showed the critical importance of the CArG box3 motif in the expression of the Xenopus borealis skeletal actin gene would suggest that a positive trans-acting factor does interact with this sequence in vivo. The possible reasons for this apparent lack of binding to CArG box3 have been discussed previously. What is evident from these experiments however is that a trans-acting factor can bind to the CArG box1 sequence motif, suggesting that this element, in addition to CArG box3, may also be involved in the expression of the Xenopus borealis skeletal actin gene.

Results and Discussion

5.4 CARG box3 and CARG box1 can compete for the binding of factors to the *Xenopus borealis* skeletal actin gene.

To investigate further whether CARG boxes are indeed the sites of interaction for positive trans-acting factors, the ability of these sequences to compete for the binding of factors to the *Xenopus borealis* skeletal actin gene was studied.

To achieve this three double-stranded oligonucleotides were synthesised which contain sequences complementary to the CARG box3, CARG box2 and CARG box1 sequence motifs. Restriction sites were designed to surround the core CARG motifs in the event of the oligonucleotides being sub-cloned at a later date. The sequences of these oligonucleotides are illustrated in figure 5.4A.

Band shift reactions containing the -167/-42 fragment of the *Xenopus borealis* skeletal actin gene promoter were constructed. The ability of each CARG box motif to compete for the binding of trans-acting factors to this promoter fragment was analysed by the addition of a 40-fold molar excess of unlabelled CARG box oligonucleotide to binding reactions.

Results illustrated in figure 5.4B demonstrate the previously described shifted band present in binding reactions containing no competitor. Furthermore, in this particular experiment an additional shifted band with a slower mobility than that of the predominant complex is also apparent. However, this additional band was not present in subsequent experiments and thus, although suggesting the possibility of additional factors binding the promoter fragment, its authenticity is

FIGURE 5.4. Competition of binding to the *Xenopus borealis* skeletal actin gene promoter by synthetic CArG box oligonucleotides.

A. Diagrammatic representation of sequences present in the synthetic oligonucleotides complementary to CArG box1, CArG box2 and CArG box3.

Restriction sites: E= Eco RI; S= Sac I; K= Kpn I.

B. Band shift assays challenging the formation of protein-DNA complexes by the addition of a 40-fold molar excess of CArG box oligonucleotides to binding reaction using the -167/-42 fragment of the *Xenopus borealis* skeletal actin gene promoter as a probe. Binding reactions contained protein extracts prepared from the axis region of stage 30 *Xenopus laevis* embryos.

Track 1; no competitor. Track 2; 12.5 molar excess of construct pXbgh Δ 6 (500 ng). Track 3; 40-fold molar excess of CArG box3. Track 4; 40-fold molar excess of CArG box2. Track 5; 40-fold molar excess of CArG box1.

A.

CarG box1.

AAATTGGAGCTCTCCAAATATGGAGTATC
 CCTCGAGAGGTTTATACCTCATGGTAA
 E S K E

CarG box2.

AAATTGGAGCTCACCTCAAAGCGGTATC
 CCTCGAGAGGGAGTTTCCGCTGGTAA
 E S K E

CarG box3.

AAATCGAATCTATATTTGGTTCG
 GCTCGAGATATAAACCTGGCTAA
 E S K E

B.

1 2 3 4 5



Results and Discussion

questionable.

On competition with a 12.5 molar excess of the pXbg Δ 6 promoter an expected decrease in the binding activity is observed. More interesting, however, is the finding that a 40-fold molar excess of both CARG box1 and CARG box3 can effectively compete for the binding of factors to the Xenopus borealis skeletal actin gene promoter (figure 5.4B, tracks 3 and 5). This is in agreement with footprinting data that previously identified the binding of a factor(s) to CARG box1. Moreover these data would suggest that the CARG box3 motif is also capable of binding a protein(s) in vitro. More important, however, is the fact that both CARG box1, and to a lesser extent CARG box3, can almost totally eliminate any detectable binding to the skeletal actin gene promoter. This would suggest that if CARG box1 is the major site for a protein-DNA complex, then CARG box3 also has the ability to compete for this factor, although at a lower efficiency than CARG box1.

In agreement with footprinting data that demonstrate the absence of any protein-DNA complexes forming over the CARG box2 sequence, so a CARG box2 oligonucleotide appears to have little ability to compete for the binding of factors to the skeletal actin gene promoter, under our assay conditions (figure 5.4B, track 4). This finding is not totally unexpected as in vivo analysis of sequences which are important for the expression of the gene suggest the redundancy of CARG box2 and adjacent sequences in the expression of the gene.

Results and Discussion

5.5 Discussion.

Experiments outlined in this chapter demonstrate that sequence-specific trans-acting factors are capable of binding the Xenopus borealis skeletal actin gene promoter. Indeed analysis of binding to 5' deletion fragments of the promoter, in addition to DMS footprinting experiments, have demonstrated that the major site for this factor binding is the CArG box1 sequence motif. This lends support to the suggestion that CArG box1 is important in the expression of the Xenopus borealis skeletal actin gene, in addition to CArG box3.

The pattern of factor binding to the Xenopus borealis skeletal actin gene promoter conflicts with data obtained from similar experiments that investigate the binding of trans-acting factors to the chicken skeletal actin gene promoter. In these experiments two sequence-specific binding activities named MAPF1 and MAPF2 were observed on band shift analysis of the promoter (see chapter 1.3.4 and references therein). In addition to this both the MAPF1 (the predominant binding activity in non-myogenic nuclear extracts) and MAPF2 (the predominant binding activity in myogenic nuclear extracts) factors bind the 3' half of CArG box1, overlapping onto sequences directly 3' to this motif (Walsh and Schimmel, 1987; also see fig 5.5). One explanation for this apparent discordance in data is that the sequences directly 3' to either the CArG box1 or CArG box3 sequence motifs of the Xenopus borealis skeletal actin gene show no great similarity to those present in the chicken skeletal actin gene promoter. Therefore,

Results and Discussion

-98	TCTGTCCAAATATGGAGTTC AGACAGGTTTATACCTCAAG	-79	<u>Xenopus borealis</u> skeletal actin.
-96	GACACCCAAATATGGCGACG CTGTGGGTTTATACCGCTGC	-77	Chicken skeletal actin.
-102	ACACCCAAATATGGCTCGAGA TGTGGGTTTATACCGAGTCT	-84	Human skeletal actin.
-113	GGGACCAAATAAGGCAAGG CCCTGGGTTTATTCGGTTCC	-95	Human cardiac actin.
-95	AGCTACCAAATAAGGGCAGG TCGATGGTTTATTCCTCC	-76	<u>Xenopus laevis</u> cardiac actin.

FIGURE 5.5. Comparison of the Xenopus borealis skeletal actin gene CArG box1 DMS footprint with the footprints of equivalent motifs from the chicken skeletal actin (Walsh and Schimmel, 1987), human skeletal actin (Boxer et al., 1989a), human cardiac actin (Boxer et al., 1989a) and Xenopus laevis cardiac actin (Taylor et al., 1989) genes.

Strong footprinting of the residues are indicated by the symbol ▲, whereas weak interaction with the residues are illustrated by the symbol △.

Results and Discussion

if the MAPF proteins are indeed sequence-specific DNA binding proteins which contact the 3' half of the CARG sequence motif, then they would be incapable of contacting the Xenopus borealis skeletal actin gene promoter.

Studies using the human skeletal actin gene produce similar data to ours, with predominantly one binding activity being present when the promoter of this gene is subjected to band shift analysis (Boxer et al., 1989a). Indeed, the DNA footprint over the CARG box1 motif of this gene is similar to the DNA footprint generated over the Xenopus borealis skeletal actin gene CARG box1 motif (see fig 5.5). In addition to this the DNA footprint of the Xenopus borealis skeletal actin gene CARG box1 sequence is also similar to those of the cardiac actin gene CARG box1 motifs. It is interesting to speculate therefore, that these motifs bind the same, or similar proteins, and this will be discussed further in chapter 6.

Band shift analysis and DNA footprinting experiments would suggest little or no binding of factors to the CARG box3 motif. However, as discussed previously the lack of detectable binding to this sequence element in our assays may be due to the positioning of the CARG motif directly adjacent to the 5' end of the DNA fragment. Indeed, competition analysis using oligonucleotides would argue that CARG box3 is capable of interacting with factors that bind to the Xenopus borealis skeletal actin gene promoter. The finding that CARG box3 almost eliminates any binding to the skeletal actin gene promoter would suggest that CARG box3 is also capable of interacting

Results and Discussion

with factors that bind the CARG box1 motif. The observation that CARG box1 is a more efficient competitor than CARG box3 would suggest that CARG box1 has a greater affinity for these factors than CARG box3. Indeed this has also been proposed to be the situation with the Xenopus cardiac actin gene (Mohun et al., 1989a).

CARG box2 appears to exhibit no significant binding of proteins in vitro. Indeed, a CARG box2 oligonucleotide lacks the ability to compete effectively for the binding of factors to the promoter region of the Xenopus borealis skeletal actin gene. This finding is in accordance with micro-injection data that demonstrates the lack of importance of CARG box2 and adjacent sequences in the expression of the gene.

In summary it would appear that CARG box3 is essential for the expression of the Xenopus borealis skeletal actin gene. Furthermore, under the assay conditions employed in this study, both CARG box1 and CARG box3 have the ability to bind sequence specific trans-acting factors in vitro, whilst CARG box2 does not. Preliminary competition analysis would also suggest that CARG box1 and CARG box3 have the ability to compete for each others binding activities, and this will be investigated further in chapter 6.

Results and Discussion

CHAPTER 6

CARG box3 and CARG box1 bind similar, yet distinct proteins in vitro.

Introduction.

Band shift and DNA footprinting analysis of the Xenopus borealis skeletal actin gene promoter, using relatively large DNA fragments, demonstrates that both CARG box1 and CARG box3 are capable of binding proteins in vitro (see chapter 5). Furthermore, competition for the binding of factors to the Xenopus borealis skeletal actin gene promoter with CARG box oligonucleotides would suggest that CARG box1 and CARG box3 are capable of interacting with each other's proteins. This finding raises the question as to whether CARG box3 and CARG box1 do indeed bind the same protein(s), or whether these factors are similar, yet distinct.

Experiments investigating the expression of other actin genes have also identified a trans-acting factor(s) which binds the CARG box1 motif present in the promoters of these genes (Mohun et al., 1989a; Miwa and Kedes, 1987; Walsh and Schimmel, 1987; Boxer et al., 1989a). Furthermore Mohun et al. (1989a) went on to propose that the factor which binds the CARG box1 motif of the Xenopus laevis cardiac actin gene is also capable of binding the other three CARG motifs present in the promoter of this gene. However, the affinity of the protein for each CARG motif differs so that its affinity for CARG box1 > CARG box3 > CARG

Results and Discussion

box4 >> CArG box2.

A preliminary requirement in understanding how the CArG box contributes to the expression of the Xenopus borealis skeletal actin gene is to identify the transcription factors with which this sequence motif interacts. It was therefore decided that experiments should be designed to identify the protein(s) which are capable of binding the CArG motifs present in the promoter of the Xenopus borealis skeletal actin gene. Data presented in this chapter illustrate that only CArG box1 and CArG box3 oligonucleotides demonstrate binding of trans-acting factors in vitro, at least under our assay conditions. Furthermore, the affinities of CArG box sequences for the binding of these factors is investigated in greater depth, and the protein which binds the CArG box1 motif identified.

6.1 CArG box1 and CArG box3 synthetic oligonucleotides bind sequence specific factors in vitro.

In order to more extensively characterise the factors which bind to CArG box1 and CArG box3 it was decided to study the capability of CArG box oligonucleotides to bind proteins in vitro.

To achieve this double-stranded oligonucleotides containing each of the CArG motifs (see figure 5.4A) were end labelled with γ ³²P-ATP and subjected to band shift analysis using protein extracts prepared from the axis regions of stage 30 Xenopus laevis embryos (see chapter 3.16).

Results from this experiment are illustrated in figure 6.1. No

FIGURE 6.1. CArG box oligonucleotides bind sequence-specific
trans-acting factors in vitro.

Oligonucleotides containing the CArG box1, CArG box2 and CArG box3 motifs were used as probes in binding reactions containing protein extracts prepared from stage 30 Xenopus laevis embryos. Both normal binding reactions (track N) and competition reactions containing a 60-fold molar excess of the homologous unlabelled CArG box oligonucleotide (track C) are represented. The 60-fold molar excess of CArG box oligonucleotide competitor replaced an equivalent amount of non-specific oligonucleotide present in normal binding reactions.

CArG
box3
N C



CArG
box2
N C



CArG
box1
N C



◀CBA1

◀CBA2

Results and Discussion

apparent binding of factors is observed in band shift reactions containing CArG box2 oligonucleotides, as judged by the altered migration of the probe through polyacrylamide gels. This is in agreement with data presented in chapter 5, which demonstrates that no proteins appear to bind the CArG box2 motif situated in the promoter of the Xenopus borealis skeletal actin gene. In contrast however, the CArG box3 oligonucleotide exhibits a single retarded band when subjected to band shift analysis. Furthermore, this binding activity appears to be facilitated by a sequence-specific interaction of factors with oligonucleotide sequences, by the criterion that the binding activity is extinguished on competition with a 60-fold molar excess of unlabelled CArG box3 oligonucleotide. It would appear, therefore, that CArG box3 is capable of interacting with a sequence-specific trans-acting factor(s) in vitro.

CArG box1 oligonucleotides exhibit an apparently identical shifted band to that present in CArG box3 binding reactions, as judged by its mobility through a polyacrylamide gel. From this point onwards this binding activity present in both CArG box1 and CArG box3 binding reactions will be referred to as CArG Binding Activity 1 (CBA1). The intensity of the CArG box1 CBA1 can be seen to be greater than that of its CArG box3 counterpart. On the assumption that CArG box1 and CArG box3 do indeed bind the same protein to form this CBA1 DNA-protein complex, it is possible to speculate that CArG box1 has a greater affinity for this protein than CArG box3. Indeed, experiments which investigate the ability of CArG box1 and CArG

Results and Discussion

box3 to compete for the binding of factors to the Xenopus borealis skeletal actin gene promoter would support this hypothesis (see chapter 5.4).

In addition to the CBA1 shifted band a faster migrating binding activity is also present exclusively in CArG box1 binding reactions (CBA2). This binding activity is also extinguished on competition with a 60-fold molar excess of CArG box1, and thus would appear to be caused by the binding of a second sequence-specific factor. The CBA2 shifted band was consistently observed on band shift analysis of CArG box1 sequences, despite the presence of protease inhibitors. This, in addition to the finding that an increase in the CBA2 shifted band is accompanied by a corresponding increase, and not a decrease in the CBA1 shifted band, would argue against the possibility of the CBA2 shifted band being produced by the degradation of the protein that forms the CBA1 complex. Furthermore, data outlined in chapter 5.4 suggest that CArG box3 is capable of interacting with proteins that bind the CArG box1 motif. Thus, if the CBA2 binding activity is a product of partial degradation of the CBA1 protein, then the CBA2 shifted band would be expected to be present in CArG box3 binding reactions. The fact that this has never been observed, even on prolonged exposure of gels, would also argue against the CBA2 activity being the result of a partial proteolysis of the CBA1 protein.

In addition to the CBA1 and CBA2 shifted bands in CArG box1 binding reactions, various other retarded bands are also apparent. However, these binding activities are not extinguished

Results and Discussion

on competition with a molar excess of CArG box1 sequences. Thus, these 'non-specific' bands are likely to be facilitated by the binding of general DNA binding proteins to the CArG box1 probe, rather than the binding of a sequence-specific trans-acting factor(s).

6.2 CArG box3 and CArG box1 exhibit distinct competition patterns.

With the aim of further investigating the ability of CArG box1 and CArG box3 binding proteins to interact with each others sequences it was decided to study the ability of each CArG motif to compete for the binding of factors to the other CArG boxes. To achieve this, binding reactions containing either CArG box3, CArG box2 or CArG box1 were constructed using protein extracts prepared from the axis region of stage 30 Xenopus laevis embryos. The ability of each CArG box to compete for the binding activities present in these reactions was tested by the addition of a 40-fold molar excess of either unlabelled CArG box1, CArG box2 or CArG box3.

Results illustrated in figure 6.2.1 demonstrate the expected absence of any competable shifted bands in CArG box2 binding reactions.

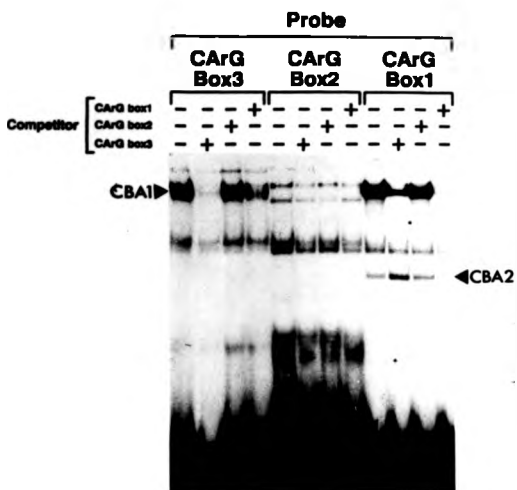
In contrast however, CArG box1 binding reactions reveal the characteristic CBA1 and CBA2 shifted bands which are lost on competition with a 40-fold molar excess of unlabelled CArG box1. As might be expected, the CBA2 shifted band is only abolished on competition with CArG box1, further demonstrating that this

FIGURE 6.2.1. CArG box oligonucleotides exhibit distinct competition patterns.

Band shift assay reactions were constructed using either CArG box1, CArG box2 or CArG box3 as probes, in conjunction with protein extracts prepared from the axis region of stage 30 Xenopus laevis embryos.

Competition was performed by the addition of a 40-fold molar excess of either CArG box1, CArG box2 or CArG box3 oligonucleotides to binding reactions. The CArG box competitor added to binding reactions substituted an equivalent amount of non-specific oligonucleotide present in normal binding reactions.

The specific oligonucleotide used as a probe in each binding reaction is illustrated above the relevant tracks. The presence (+), or absence (-) of oligonucleotide competition in binding reactions is also illustrated.



Results and Discussion

protein-DNA complex is capable of forming exclusively over the CArG box1 motif. In agreement with experiments described in chapter 5.4, which study the ability of CArG boxes to compete for the binding of factors to the promoter of the Xenopus borealis skeletal actin gene, is the finding that the CBA1 shifted band is reduced on competition with both the CArG box1 and CArG box3 sequences. It is therefore possible to speculate that CArG box1 and CArG box3 share the ability to bind at least one common factor in vitro. However, as previously suggested it would appear that CArG box1 has a greater affinity for this factor than CArG box3, as judged by the increased ability of CArG box1 over CArG box3 to compete for the binding of factors to the CArG box1 sequence.

However, the results of experiments studying the ability of CArG motifs to compete for the binding activity present in CArG box3 band shift reactions shed a different light on this assumption. As described previously, CArG box3 binding reactions exhibit predominantly one retarded band when subjected to band shift analysis. However, unlike the seemingly identical shift in CArG box1 binding reactions, this binding activity is most effectively competed for by the addition of a 40-fold molar excess of CArG box3, whereas CArG box1 is a less effective competitor. Thus, CArG box1 CBA1 is most effectively competed for by CArG box1, whereas CArG box3 CBA1 is most effectively competed for by CArG box3.

This observation is more convincing in experiments which study the capability of various concentrations of CArG box1 and CArG

Results and Discussion

box3 to compete for each others binding activity. In these experiments it can be seen that CArG box1 extinguishes the CBA1 binding activity present in CArG box1 binding reactions more efficiently than CArG box3 (see figure 6.2.2). However, the opposite appears to be the case in CArG box3 binding reactions. These data suggest that the CBA1 shifted band present in CArG box1 binding reactions, and the CBA1 shifted band present in CArG box3 binding reactions, are caused by two distinct factors. The first factor would appear to bind preferentially to CArG box1, whereas the second would appear to bind preferentially to CArG box3. However, the finding that CArG box1 and CArG box3 have the ability to compete for each others binding activities would suggest that each factor is capable of binding the other's sequence, although at a lower affinity than that of the homologous binding reactions.

Whether these apparently different factors are completely dissimilar, or whether they are closely related is as yet unknown. In an attempt to answer this question, it is of primary importance to identify at least one of the factors which interacts with these CArG motifs.

6.3 CArG box1 binds a protein that is antigenically related to the serum response factor.

The transcriptional activation of the proto-oncogene c-fos can be stimulated by a number of extracellular stimuli, including serum, growth factors and agents that elevate the intracellular levels of cAMP (see chapter 1.3.5 and references therein).

FIGURE 6.2.2. Competition analysis of CARG box binding
proteins using increasing amounts of CARG
box competitors.

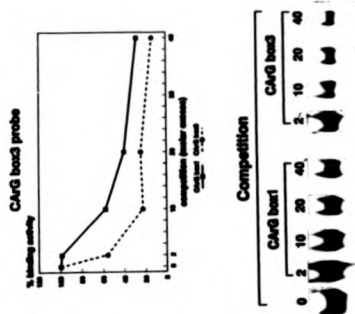
Binding reactions were constructed which contained either CARG box1 or CARG box3 oligonucleotides as probes. Either no competitor or a 2, 10, 20 or 40-fold molar excess of CARG box1 or CARG box3 was added to these reactions prior to the addition of protein extract prepared from stage 30 Xenopus laevis embryos. Band shift reactions were run on polyacrylamide gels as described in chapter 3.16 and autoradiographed at -70°C with an intensifying screen to enable the localisation of retarded bands. Each CBAl shifted band was then excised from the gel and counted by their Cerenkov emission.

The % binding activity represents the activity of the CBAl shifted bands in competition binding reactions expressed as a percentage of the activity of the CBAl band in reactions containing no competitor.

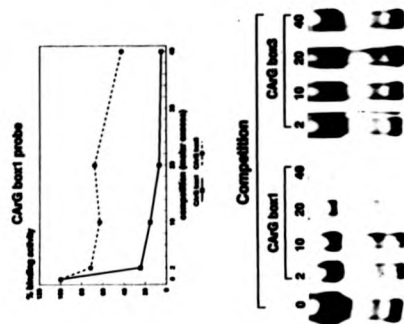
A. The effect of a 2, 10, 20 or 40-fold molar excess of unlabelled CARG box1 or CARG box3 on the formation of the CBAl complex in reactions containing CARG box3 as a probe.

B. The effect of a 2, 10, 20 or 40-fold molar excess of unlabelled CARG box1 or CARG box3 on the formation of the CBAl complex in reactions containing CARG box1 as a probe.

A.



B.



Results and Discussion

Whilst the activation of this gene by increased levels of cAMP is thought to be mediated via a cAMP response element present in the promoter of the gene, a cis-acting sequence responsible for the activation of the gene by serum and growth factors is contained within a element of dyad symmetry (DSE) present in the 5' flanking region of the gene (Treisman, 1985). Situated at the centre of the DSE is the serum response element (SRE), which has been shown to be the binding site of the serum response factor (SRF) (Treisman, 1986; Gilman *et al.*, 1986; Prywes and Roeder, 1986). Inspection of the SRE DNA sequence reveals the presence of a CArG box motif. This CArG sequence resembles the CArG box1 motif present in a number of actin gene promoters, including that of the Xenopus borealis skeletal actin gene (see figure 6.3.1A). Furthermore, the methylation footprint created by the binding of SRF to the CArG motif of the SRE exhibits a striking resemblance to the footprint created by the binding of a factor to the CArG box1 motif of the Xenopus borealis skeletal actin gene (see figure 6.3.1B).

The similarity between these two regulatory elements raises the question as to whether they interact with the same, or related transcription factors. To address this question we obtained a poly-clonal antibody raised against the C-terminal half of Xenopus SRF (a generous gift from Dr. T. Mohun). This antibody was exploited in antibody band shift assays to determine whether it could recognise one or more of the proteins which complex with the Xenopus borealis skeletal actin CArG box1 motif, and thus identify them as being antigenically similar to SRF.

Results and Discussion

A.

-95	CACCCAAATATGGC	Rat.
-86	CACCCAAATATGGC	Chicken.
-101	CACCCAAATATGGCTC	Human.
-96	TGTCCAAATATGGAGT	<u>Xenopus laevis</u> .
-95	TGTCCAAATATGGAGT	<u>Xenopus borealis</u> .
	CCTTATAAGG	<u>Xenopus c-fos</u> SRE.
	CCATATTAGG	Human <u>c-fos</u> SRE.

B.

TGTCCAAATATGGAGT ACAGCTTTATACCTCA ▲▲	<u>Xenopus borealis</u> skeletal actin CArG box1.
TGTCCATATTAGGACA ACAGGTATAATCCTGA ▲▲	Human <u>c-fos</u> SRE

FIGURE 6.3.1. Comparison of CArG box and SRE sequences.

A. Sequence comparison of skeletal actin CArG box1 motifs with the CArG elements present in the SRE of the human (Treisman, 1986) and Xenopus (Mohun et al., 1989b) c-fos genes.

B. Comparison of the methylation footprints created over the Xenopus borealis skeletal actin CArG box1 motif and the CArG motif present in the human c-fos SRE (Treisman, 1986).

Results and Discussion

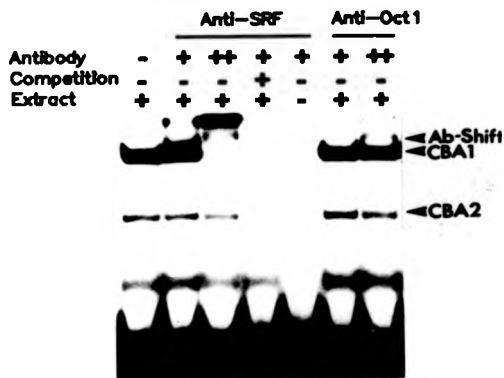
CARG box1 was end labelled with ^{32}P -ATP and used in antibody band shift assays with protein extracts prepared from stage 30 Xenopus laevis embryos. Either a 1/10 dilution or concentrated form of anti-SRF antibody was used in these assays.

Results shown in figure 6.3.2 illustrate the effect of adding anti-SRF to CARG box1 binding reactions. With no antibody present the previously described CBA1 and CBA2 shifted bands are observed. Furthermore, control reactions containing anti-serum raised against the transcription factor Oct-1 (a gift of Dr. D. Smith) exhibit identical binding patterns to reactions containing no anti-serum. However, on addition of a 1/10 dilution of anti-SRF antibody to binding reactions an additional shifted band with a slightly slower mobility than the CBA1 complex is observed. This is caused by the binding of anti-SRF to either the CBA1 or CBA2 DNA-protein complexes, slowing the migration of the resulting ternary complex through a polyacrylamide gel. Indeed, this additional shift appears to be caused by an interaction of the anti-serum to a sequence-specific factor that binds the CARG box1 motif, as demonstrated by the band being extinguished on competition with a 40-fold molar excess of unlabelled CARG box1.

Addition of excess anti-SRF antibody to binding reactions results in a total loss of the CBA1 shifted band, due to the binding of sufficient antibody to the complex to arrest its migration into the gel. The finding that anti-SRF antibody recognises the protein responsible for the CBA1 protein-DNA complex demonstrates that a protein which is antigenically

FIGURE 6.3.2. A protein which is antigenically similar to SRF binds the *Xenopus borealis* skeletal actin CArG box1 motif.

Antibody band shift assays using CArG box1 as a probe, and containing protein extracts prepared from the axis region of stage 30 *Xenopus laevis* embryos, were performed as described in chapter 3.17. The presence in binding reactions of either a 1/10 dilution (+) or concentrated form (++) of anti-SRF or anti-Oct1 anti-serum (Smith and Old, 1991) is indicated above the relevant tracks. The presence (+) or absence (-) of a 40-fold molar excess of unlabelled CArG box1 competitor, or of protein extracts is also illustrated.



Results and Discussion

related to SRF is capable of binding the CARG box1 motif. Of equal interest however is the observation that the CBA2 shifted band appears to be unaffected by the addition of anti-SRF to binding reactions, even at high concentrations. This supports the suggestion that the protein that is responsible for CBA2 is distinct from the SRF-like protein which causes the CBA1 complex. The possibility does still remain, however, that the CBA2 shifted band results from the cleavage of the SRF protein, removing the epitopes recognised by the anti-SRF antibody but leaving its ability to bind DNA intact. However, the portion of SRF to which the poly clonal anti-serum was raised does contain some of the DNA binding domain (T. Mohun, personal communication). Therefore, at high concentrations of anti-SRF it would be expected that some antibodies are present which are capable of recognising epitopes within the DNA binding domain, and thus inducing an additional shift of the CBA2 complex. The absence of this effect, even at high concentrations of anti-SRF antibody, would argue against the CBA2 shifted band being the result of proteolysis of the SRF-like protein.

It is apparent therefore from experiments that exploit the specificity of an anti-SRF antibody, that CARG box1 does indeed bind SRF, or at least a protein which is antigenically related to this protein. In addition to this it would also appear that CARG box1 has the ability to bind a second protein which is distinct from the SRF.

Results and Discussion

6.4 CArG box3 binds a protein which is antigenically distinct from SRF.

Competition analysis of the proteins which bind the CArG motifs of the Xenopus borealis skeletal actin gene suggest that CArG box1 and CArG box3 bind similar, yet distinguishable proteins. This, in addition to the discovery that the CBA1 complex observed in CArG box1 binding reactions is created by the binding of an SRF-like protein, raises the question as to whether CArG box3 also binds an SRF-like protein. Indeed, comparison of the nucleotide sequences of CArG box1 and CArG box3 reveals a match of 7 out of 10 of the nucleotides of the CArG motif. Furthermore, CArG box3 appears to be an almost perfect inverted repeat of the CArG box1 motif, showing a match of 9 out of the 10 nucleotides when compared in the opposite orientation to CArG box1 (see figure 6.4A). It is therefore a possibility that although the CArG box1 and CArG box3 motifs appear to bind dissimilar proteins, they may both bind SRF-related proteins. To investigate this possibility it was decided to repeat CArG box1 antibody band shift experiments in parallel with antibody band shift experiments using CArG box3 as a probe.

Results illustrated in figure 6.4B show the characteristic binding patterns achieved with both CArG box1 and CArG box3 when no anti-serum is added. Control binding reactions containing antibody raised against the Xenopus laevis thyroid hormone receptor (anti-THR) (a gift from Dr. R. W. Old) show no significant alteration in the binding activities present in

FIGURE 6.4. A protein which is antigenically distinguishable from SRF binds the *Xenopus borealis* skeletal actin CArG box3 motif.

A. Sequence comparison of the *Xenopus borealis* skeletal actin gene CArG box3 and CArG box1 motifs. The CArG box1 motif is shown in either the correct (+) or reverse (-) orientations. Conserved sequences are indicated by the sign * above the relevant nucleotides.

B. Antibody band shift assays using either CArG box1 or CArG box3 as probes were performed in combination with protein extracts prepared from the axis region of stage 30 *Xenopus laevis* embryos.

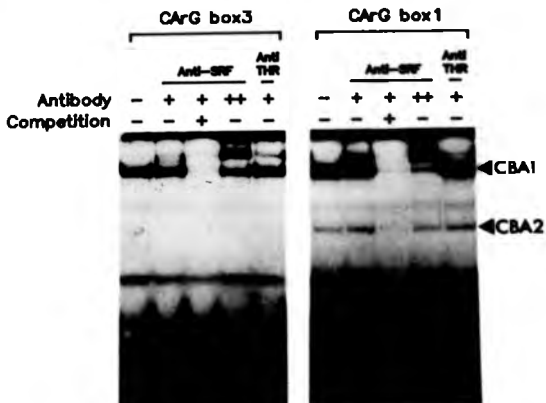
The presence in binding reactions of either a 1/10 dilution (+) or concentrated form (++) of anti-SRF or anti-THR anti-serum is indicated above the relevant tracks. The presence (+) or absence (-) of a 40-fold molar excess of unlabelled CArG box1 competitor is also illustrated

Results and Discussion

A.

-162	CTATATTGG	-153	CARG box3.
	* * * *		
-93	CCAAATATGG	-84	CARG box1 (+).
	* * * * *		
-84	CCATATTGG	-93	CARG box1 (-).

B.



Results and Discussion

either CARG box1 or CARG box3 binding reactions. However, small amounts of an additional shifted band are induced by the addition of anti-THR antibody, although further investigation of this binding activity revealed it to be non-specific. Indeed this band could even be observed on the addition of anti-THR to binding reactions containing no protein extracts (data not shown; D. Smith, personal communication).

In agreement with data produced in the preceding section, the addition of a 1/10 dilution of anti-SRF antibody to CARG box1 binding reactions induces an additional shifted band, with the CBA1 shifted band being totally extinguished on the addition of a more concentrated form of the antibody. Furthermore, an additional shifted band, which has not previously been observed, is also present on the addition of excess anti-SRF antibody. However, this band appears to migrate at a slightly slower rate than the additional band produced on the addition of a 1/10 dilution of the anti-SRF antibody to binding reactions. Indeed, it appears to migrate at approximately the same rate as the non-specific band observed in binding reactions containing anti-THR antibody. Thus, although this band may be caused by the additional binding of anti-SRF to the CARG box1 Protein-DNA complex, it is more likely to be caused by a non-specific interaction of the anti-serum with the CARG box probe.

Addition of the anti-SRF antibody to CARG box3 binding reactions yields strikingly different results from those obtained on addition of the antibody to CARG box1 binding reactions. Anti-SRF, when added at a 1/10 dilution to CARG box3

Results and Discussion

binding reactions induces only a small additional shift of the CBA1 complex, when compared to the equivalent track of CArG box1 binding reactions. More important, however, is the finding that addition of excess anti-SRF antibody to CArG box3 binding reactions has only a slight effect on the CBA1 shifted band, with the majority of it remaining. At these high concentrations of anti-SRF an additional shift is observed. This additional binding activity may represent the binding of some SRF-like proteins to the CArG box3 motif, and may contain some of the protein-DNA complex lost from the CBA1 shifted band in these reactions. However, the combination of the additional shifted band and the remaining CBA1 shift appears to have a greater intensity than that of the CBA1 band present in control reactions. This would suggest that some of the additional shift is due to a non-specific interaction of the anti-SRF antibody with the CArG box3 probe. Furthermore, the mobility of this shifted band appears to be slower than that of the band present in CArG box1 reactions containing a 1/10 dilution of the anti-SRF antibody. Indeed, it exhibits a similar mobility to that of the non-specific band observed in control anti-serum reactions, supporting the suggestion that this band is non-specific.

What is clear is that the majority of the CBA1 shifted band in CArG box3 binding reactions remains intact on addition of quantities of anti-SRF which are capable of extinguishing the CBA1 shifted band in CArG box1 binding reactions. This demonstrates that the majority of proteins complexed with CArG

Results and Discussion

box3 are antigenically distinct from those which bind to CARG box1.

6.5 Discussion.

Data presented in this chapter demonstrates the binding of sequence-specific trans-acting factors to the CARG box1 and CARG box3 sequences situated in the promoter of the Xenopus borealis skeletal actin gene. CARG box1 exhibits two distinct DNA-protein complexes when assayed by band shift analysis. One of these complexes has been shown to be formed by the binding of a protein which is antigenically related to the transcription factor SRF which binds the promoter of the c-fos gene. The second complex formed with the CARG box1 sequence appears to be facilitated by the binding of a factor which is antigenically distinct from the SRF.

CARG box3 forms a complex with identical mobility to that of the CARG box1-SRF complex, when subjected to band shift analysis. However both competition and antibody band shift assays demonstrate that although some of this complex may be attributed to the binding of an SRF-like protein to CARG box3, the majority is formed by the binding of a protein(s) that is antigenically distinguishable from the SRF protein that interacts with CARG box1.

The finding that CARG box1 does indeed bind an SRF-like protein is not unexpected, as judged by the similarity of the CARG box1 sequence to that of the c-fos SRE. Indeed, SRF has been shown to contact the G residues situated at either end of the CARG motif

Results and Discussion

present in the c-fos SRE (Treisman, 1986; Frywes and Roeder, 1986; Schroter et al., 1987; see figure 5.4B). These nucleotides appear to be essential for the binding of SRF to DNA (Leung and Miyamoto, 1989), and are conserved between the c-fos SRE and the Xenopus borealis skeletal actin gene CArG box1 motif. However, the A/T core of the CArG box appears to differ between the c-fos SRE and the Xenopus borealis skeletal actin CArG box1 motifs without affecting the binding of SRF. Studies using the human SRE have shown that alteration of this core sequence does not significantly effect the binding of SRF to the CArG motif, providing this core remains predominantly A/T rich (Leung and Miyamoto, 1990). Furthermore, by using PCR techniques to isolate a variety of SRF binding sites, Pollock and Treisman (1990) identified a CArG motif identical to that of the Xenopus borealis skeletal actin gene CArG box1 motif which was capable of binding human SRF. It is possible to speculate, therefore, that the CArG box1 sequence is indeed capable of binding a bona fide SRF protein.

The CArG box1 motifs present in other actin gene promoters have also been shown to bind SRF. Preliminary experiments studying the human skeletal and human cardiac actin genes identified their respective CArG box1 motifs to bind a protein of the same molecular weight as SRF (Boxer et al., 1989a). Further investigation of this CArG box binding factor (CBF) demonstrated it to be indistinguishable from SRF by chromatographic profiles, glycerol gradient sedimentation, temperature stability and DNA binding properties, in addition to being recognised by an anti-

Results and Discussion

SRF antibody (Boxer et al., 1989b). In addition to this, the Xenopus laevis cardiac actin CARG box1 motif also appears to bind SRF. Furthermore, SRF is also capable of binding this motif in vivo, by virtue of the fact that the Xenopus laevis cardiac actin CARG box1 and the Xenopus c-fos SRE are functionally interchangeable (Taylor et al., 1989).

With the exception of the chicken skeletal actin gene, all studies mentioned thus far suggest that SRF is the only protein which is capable of binding the CARG motif present in actin gene promoters. No other binding activities such as the CBA2 shifted band described in this report have been identified. However, by altering the conditions of band shift assays, four other binding activities have been identified which are capable of interacting with the CARG box1 motif of the Xenopus laevis cardiac actin gene (Taylor and Gurdon: Poster; Third International Xenopus Meeting, 1990).

The identity of the protein(s) which facilitate the CBA2 shift are as yet unknown, but it is interesting to speculate that this binding activity may be due to the binding of a second protein, p62TCF, which has been demonstrated to bind the c-fos SRE in addition to SRF (see chapter 1.3.5 and references therein). However, a requirement for p62TCF binding is the formation of the SRE-SRF complex (Shaw et al., 1989). Therefore, if the CBA2 complex is indeed caused by the binding of p62TCF to CARG box1, then this ternary complex would be expected to exhibit a slower mobility than the CARG box1-SRF complex on band shift analysis. The finding that CBA2 migrates more rapidly through a

Results and Discussion

polyacrylamide gel than the SRF-CaRG box1 complex, and the discovery that anti-SRF does not recognise the CBA2 binding activity, would argue against the binding of p62TCF to this motif. In addition to this, p62TCF contacts the c-fos SRE immediately 5' to the CaRG box motif present in the c-fos SRE (Shaw et al., 1989). The absence of any flanking sequences in the CaRG box1 oligonucleotide would also argue against the binding of p62TCF to CaRG box1.

As stated previously, band shift analysis of the chicken skeletal actin CaRG box1 motif revealed the presence of two shifted bands (Walsh and Schimmel, 1987). It is therefore possible that the CBA2 shifted band observed on analysis of the Xenopus borealis skeletal actin gene CaRG box1 motif is caused by the binding of proteins similar to those which facilitate the MAPF2 binding activity present in chicken skeletal actin gene studies. However, unlike the MAPF2 protein, the protein which forms the CBA2 complex appears to be distributed throughout the embryo (see chapter 7), arguing against this being caused by the binding of MAPF2.

Competition of SRF-CaRG box1 complexes of the Xenopus laevis cardiac actin gene suggest that the other 3 CaRG motifs present in the promoter of this gene also have the ability to bind SRF. Indeed, the affinities of SRF for each of these CaRG motifs has been proposed to be in the order of CaRG box1 > CaRG box3 > CaRG box4 >> CaRG box2 (Mohun et al., 1989a). This reflects data outlined in this report which use the Xenopus borealis skeletal actin gene CaRG motifs in band shift assays. These data show the

Results and Discussion

relative affinity for each CARG box to compete for SRF binding to CARG box1 to be in the order of CARG box1 > CARG box3 >>> CARG box2. However, in the studies of Mohun et al., (1989a) reciprocal competition experiments were not performed. Our more extensive competition experiments suggest that two similar factors exist which are capable of binding the CARG motifs. The first of these factors has a high affinity for the CARG box1 motif, whilst the second has a high affinity for the CARG box3 motif. Indeed, antibody band shift experiments appear to corroborate this suggestion, demonstrating that the factor which binds CARG box3 is antigenically distinct from the SRF protein which binds CARG box1.

As previously stated, the core CARG box3 sequence is a perfect inverted repeat of the core CARG box1 sequence, with the exception of one nucleotide. This nucleotide mismatch, which occurs at position -161 in the promoter, is one of the four G residues which have been identified to be the contact points of the SRF protein with the CARG motif. Taking this into account it is not unexpected that this single nucleotide change can facilitate the binding of a different protein to the CARG motif.

Subtle changes in the binding sites of other transcription factors has been shown to dramatically alter the proteins with which these sequences interact. Indeed, in some studies the same DNA sequence has been shown to bind a multitude of different, although normally related transcription factors. For example the octamer sequence, ATGCAAATNA, has been shown to bind two different transcription factors with diverse roles in gene

Results and Discussion

expression (for review see Schaffner, 1989). One of these factors, termed Oct-1, is ubiquitous and directs the expression of the histone H2B gene by binding an octamer motif situated in the promoter of this gene. Conversely the same octamer motif is present in the promoter of the immunoglobulin genes and has been shown to direct their tissue-specific expression by interacting with the lymphoid-specific transcription factors Oct-2A and Oct-2B. Several Drosophila homeodomain proteins have also been shown to bind the same DNA sequence. For example, the DNA sequence TCAATTAAAT is the binding site for the proteins encoded by the homeobox genes engrailed, even skipped, fushi tarazo, paired and Zerknult (see Biggin and Tjian, 1989). Furthermore, the helix-loop-helix family of transcription factors bind a common consensus sequence termed the E box, which was first identified in the immunoglobulin enhancers (Church et al., 1985).

It is therefore not inconceivable that a change of one nucleotide in a transcription factor binding site, as in the case of the CARG box1 and CARG box3 sequence motifs, can change the protein with which the sequence interacts. Indeed, this has been demonstrated to occur with other DNA binding domains. For example, the insertion of a C residue into the binding site for the transcription factor AP-1, converts it into a putative cAMP response element which is capable of binding the transcription factor CREB (see Ziff, 1990). Studies of the DNA binding sites of the POU-domain family of transcription factors have demonstrated that the transcription factor Pit-1, which contributes towards the tissue-specific expression of the

Results and Discussion

prolactin and growth hormone genes, and the aforementioned Oct-2 transcription factor, recognise regulatory DNA elements which differ by only two nucleotides. Furthermore, a double point mutation of the Pit-1 binding site present in the prolactin gene promoter allows this DNA sequence to bind Oct-2 instead of Pit-1, and facilitates the expression of the gene exclusively in B-cells (Elsholtz et al., 1990).

Experiments discussed in this chapter would suggest that the interactions of the Xenopus borealis skeletal actin gene promoter with regulatory proteins is likely to be complex. Indeed, the question arises as to whether other as yet undetected transcription factors might bind to the Xenopus borealis skeletal actin gene promoter. The identification of a protein which binds the CArG box1 sequence, and the discovery of a distinct protein which binds the CArG box3 sequence, prompts the question as to how the SRF and CArG box3 binding protein direct the tissue-specific expression of the Xenopus borealis skeletal actin gene, and this will be discussed in chapter 7.

Results and Discussion

CHAPTER 7

The spatial and temporal distribution of CARG box binding factors in *Xenopus*.

Introduction.

Previous experiments have demonstrated the importance of CARG box3 in the spatial expression of the *Xenopus borealis* skeletal actin gene. This motif, in addition to a more proximal CARG box1 motif, is capable of interacting with trans-acting factors in vitro. Two proteins appear to interact with the CARG box1 motif, one of which is antigenically related to the transcription factor SRF. In addition to this a third protein, which is antigenically distinguishable from SRF, appears to interact with the CARG box3 sequence in vitro.

The CARG box motif has been identified in the regulatory regions of other, but not all muscle-specific genes studied to date (see figure 1.1, chapter 1). Furthermore, this sequence motif is also present in the regulatory regions of the β -actin genes (see figure 1.1, chapter 1), in addition to the c-fos (see chapter 1.3.5) and interleukin-2 receptor (Phan-Dinh-Tuy et al., 1988) genes.

In order to understand how the CARG box motifs contribute to the spatial and temporal expression of the *Xenopus borealis* skeletal actin gene, it was decided to examine the temporal and spatial distribution of the previously described CARG box binding proteins. Experiments described in this chapter

Results and Discussion

investigate the ability of the CArG motif to interact with factors present in protein extracts prepared from either a region of the embryo in which the skeletal actin gene is expressed, or a region of the embryo where the actin gene is inactive. In addition to this experiments are also described which investigate the ability of factors present in extracts prepared from embryos at different stages of development to interact with the CArG box3 motif.

7.1 CArG box binding activities are present throughout the embryo.

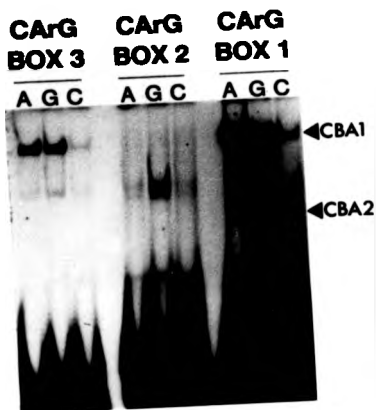
All band shift experiments described thus far have used protein extracts prepared from a region of the embryo which contains tissue that is actively expressing the skeletal actin gene (i.e the axis region). In order to determine whether the CArG box binding activities are present exclusively in a region of the embryo that expresses the skeletal actin gene, protein extracts were prepared from a region of the embryo which expresses little, if any skeletal actin (head plus gut), in addition to a region of the embryo that expresses the actin gene (axis).

Oligonucleotides containing either CArG box1, CArG box2 or CArG box3 were end labelled with γ - ^{32}P -ATP. These sequences were then used in band shift assays containing extracts prepared from either the axis or head plus gut regions of stage 30 Xenopus laevis embryos.

Results shown in figure 7.1.1 illustrate the CBA1 shift

FIGURE 7.1.1. Distribution of CArG box binding factors
throughout the embryo.

Protein extracts were prepared from the axis (A) or head plus gut (G) regions of stage 30 embryos. These extracts were then used in band shift assays containing either CArG box1, CArG box2 or CArG box3 as a probe. The CArG box used as a probe is illustrated above the relevant tracks on the figure. Track C represents competition with a 40-fold molar excess of unlabelled CArG box in binding reactions containing axis protein extracts. CArG box oligonucleotide used as competitor replaced an equivalent amount of non-specific oligonucleotide in normal binding reactions.



Results and Discussion

present in CarG box3 binding reactions, in addition to the SRF-like and CBA2 shifts present in CarG box1 binding reactions, when extracts prepared from the axis region of stage 30 embryos are used in band shift assays. More importantly however, when extracts prepared from the head plus gut regions of embryos are used in band shift assays the same pattern of retarded bands is achieved (Compare track A with track G). Indeed, no great variation in the intensity of these bands is observed between extracts. In agreement with previous data no binding of factors to CarG box2 sequences is apparent in extracts prepared from either of the embryo fractions.

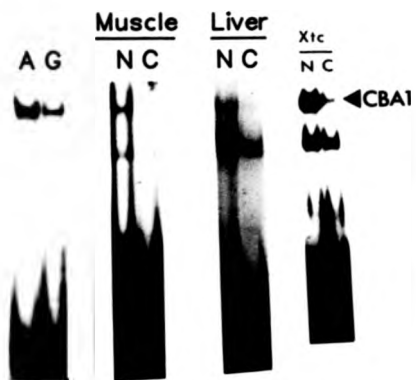
It would appear therefore that the previously described CarG box binding activities are present throughout the embryo. To investigate the tissue distribution of the CarG box3 binding activity in further detail, protein extracts were prepared from adult skeletal muscle and adult liver tissues. In addition to this protein extracts were also prepared from the Xenopus cell line XTC, which is derived from mesodermal tissues that do not express skeletal actin (Pudney et al., 1973). These tissues offer a purer source of protein extract and thus reduce the possibility of contaminating tissues which may occur with the relatively crude dissection of embryos into axis and head plus gut regions.

Muscle, liver and XTC extracts were used in band shift assays in conjunction with oligonucleotides containing the CarG box3 motif. Results of this experiment illustrated in figure 7.1.2 demonstrate that the CBA1 binding activity is present in

FIGURE 7.1.2. Tissue distribution of CArG box3 binding protein(s).

Protein extracts were prepared from either liver or muscle tissues obtained from a adult Xenopus laevis frog. Protein extracts were also prepared from the cell line XTC. These extracts were then used in band shift assays containing CArG box3 as a probe. Track N represents normal binding reactions containing no competitor. Track C represents competition reactions containing a 40-fold molar excess of unlabelled CArG box3. CArG box3 oligonucleotide used as a competitor replaced an equivalent amount of non-specific oligonucleotide in normal binding reactions.

Binding reactions containing protein extracts prepared from either the axis (A) or head plus gut (G) regions of stage 30 Xenopus laevis embryos are also illustrated.



Results and Discussion

protein extracts prepared from tissues of both muscular and non-muscular origin. The CBA1 binding activity is therefore present in tissues that do not express the skeletal actin gene in addition to those that do. Furthermore, this experiment also demonstrates that the CArG box3 CBA1 binding activity apparent in extracts prepared from the head plus gut regions of embryos is not due to contamination of these embryo fragments with muscular tissue.

7.2 CArG box3 binds a factor(s) that is present throughout early *Xenopus* development.

In addition to being expressed in a spatially restricted manner, the *Xenopus borealis* skeletal actin gene is also expressed in a stage-specific manner, being activated at stage 12½ of development (see chapter 1.3.1 and references therein). Results illustrated in the preceding section demonstrate that the CArG box binding activities identified in this study are present in tissues of the embryo where the skeletal actin gene is inactive, in addition to those where it is expressed.

To address the question as to whether the factor(s) that interact with the CArG box3 sequence are expressed at a particular stage of *Xenopus* development, protein extracts were prepared from embryos at different stages of development. These extracts were then analysed for the presence of factors which are capable of binding the CArG box3 motif by band shift analysis.

Results illustrated in figure 7.2 reveal that the CBA1 binding

FIGURE 7.2. Distribution of CARG box3 binding proteins
throughout Xenopus development.

Protein extracts were prepared from either the eggs of a Xenopus laevis frog, or from embryos at stage 6 (st 6), 10 (st 10), and 15 (st 15) of development. Extracts were also prepared from either the axis (A) or head plus gut (G) regions of stage 30/31 (st 30/31) Xenopus laevis embryos. Extracts were used in band shift assay reactions containing CARG box3 as a probe. Binding reactions containing no competitor (tracks N) are illustrated along with reactions that contain a 40-fold molar excess of unlabelled CARG box3 as a competitor (tracks C). CARG box3 oligonucleotide sequences contained in competition reactions replaced an equivalent amount of non-specific oligonucleotide present in normal binding reactions.

Due to the long period of time that the gel was allowed to run no free DNA is apparent in this figure.

[illegible]

Results and Discussion

activity is present not only in extracts prepared from both the axis and head plus gut regions of stage 30/31 embryos, but is also present in all other extracts examined. Thus, the proteins which interact with CARG box3 are not only expressed, but are capable of binding this sequence at stages of embryo development when the skeletal actin gene is inactive.

Interestingly, a second retarded band with a faster mobility than CBA1 is apparent in extracts prepared from embryos at early stages of development. This binding activity is transient, being present in extracts prepared from eggs, in addition to embryos at stage 6 of development, but decreasing at stage 10 until very little remains at stage 15. This profile of binding to CARG box3 is interesting in that the factor(s) responsible for this protein-DNA complex appears to be capable of binding CARG box3 at stages of embryo development when the skeletal actin gene is inactive. However, immediately prior to and during the stages of development when the skeletal actin gene is expressed, this binding activity is no longer apparent. It is interesting to speculate, therefore, that this binding activity corresponds to a negative trans-acting factor that is capable of binding CARG box3. However this binding activity was not consistently observed in subsequent experiments using other protein extracts. This, in addition to the absence of the binding activity in the head plus gut region of stage 30/31 embryos, would argue against the authenticity of this observation.

What is obvious from this experiment is that the protein(s)

Results and Discussion

responsible for the CARG box3 CBA1 are present, and capable of binding the CARG box3 sequence throughout early Xenopus development.

7.3 Discussion.

Experiments described in this chapter demonstrate no detectable difference in the proteins that interact with CARG box3 and CARG box1 between tissues. Indeed, the motifs exhibit identical binding activities in both regions of the embryo that express the skeletal actin gene and regions that do not. It is apparent, therefore, that the CBA1 and CBA2 binding activities present in CARG box1 binding reactions do not correspond to the tissue-specific MAPF1 and MAPF2 factors identified to interact with the CARG box1 motif of the chicken skeletal actin gene (Walsh and Schimmel, 1987; see also chapter 1.3.4). Conversely, our data are in agreement with those obtained from the study of the human skeletal, human cardiac and Xenopus cardiac actin genes, which demonstrate the binding of apparently ubiquitous factors to the CARG box1 sequence (Boxer et al., 1989a; Taylor et al., 1989).

The CBA1 binding activity present in CARG box1 binding reactions has previously been identified to be antigenically related to SRF (see chapter 6). SRF binds the SRE present in the promoter of the c-fos gene, and this binding activity has been demonstrated to be recoverable from many cell lines and tissues (Treisman, 1986; Gilman et al., 1986). It is therefore not unexpected that the SRF-like protein which binds the CARG

Results and Discussion

box1 motif is present throughout the embryo. However, these experiments also reveal that the other CARG box binding activities identified are also present in both the axis and head plus gut regions of the embryo. Indeed, the binding activity observed in CARG box3 binding reactions was also recoverable from an adult tissue (i.e the liver), in addition to a cell line (XTC), that does not express the skeletal actin gene.

From experiments described in chapter 4 it is apparent that the CARG box3 motif, and probably the CARG box1 motif, is essential for the tissue-specific expression of the Xenopus borealis skeletal actin gene. Positive cis-acting regulatory elements that contribute towards the expression of other tissue-specific genes have been shown to interact with positive trans-acting factors exclusively in the tissues where these genes are expressed. For example, tissue-specific factors have been demonstrated to bind regulatory elements present in the muscle creatine kinase (Buskin and Hauschka, 1989; Gossett et al., 1989), growth hormone (for review see Karin et al., 1990) and immunoglobulin (for review see Schaffner, 1989) genes. It is apparent, therefore, that if the CARG box is indeed the element which confers tissue-specific expression onto the Xenopus borealis skeletal actin gene, then the expression of this gene is subjected to more subtle mechanisms of control than originally conceived.

How then is it possible for these apparently ubiquitous CARG box binding factors to govern the tissue-specific expression of

Results and Discussion

the Xenopus borealis skeletal actin gene? Assuming that no other sequences exist which are responsible for the tissue-specific expression of the gene, then several mechanisms are possible. One of these is the post-translational modification of one or more of the CArG box binding proteins exclusively in skeletal muscle. This modification of the protein(s) would then cause the transcriptional activation of the skeletal actin gene exclusively in developing muscle, despite its occupying the sequence in a number of other tissues. In this regard phosphorylation has previously been implicated in the control of transcription factor function. For example, the heat shock element (HSE) in the promoter of these genes is recognised by a DNA binding protein called HSF, or HSTF. Increased transcription of yeast heat shock genes correlates with increased phosphorylation of HSTF that is already bound to the heat shock gene promoter (see Mitchell and Tjian, 1989). More recently it has been demonstrated that phosphorylation of the transcription factor CREB increases the activation of this protein on a reporter gene containing a CREB binding site (see Berk and Schmidt, 1990). Indeed, SRF is a phospho-protein which is phosphorylated on serine residues in vivo (Prywes et al., 1988).

A possible level of regulation of the Xenopus borealis skeletal actin gene therefore exists by the phosphorylation of the SRF-like protein which binds the CArG box1 motif. However phosphorylation of SRF/CBF is a requirement of DNA binding (Prywes et al., 1988; Boxer et al., 1989b). Therefore, if

Results and Discussion

phosphorylation of SRF is a mechanism of control of the skeletal actin gene, then a tissue-specific binding of the SRF-like protein to the CArG box1 motif would be expected. The absence of this tissue-specific binding would argue against the phosphorylation of SRF as a mechanism of controlling skeletal actin gene expression.

The possibility still remains that like the HSTF protein binding the HSE in yeast, the CBA2 and/or CArG box3 CBAL proteins may be activated by phosphorylation, despite their occupying the CArG motif in both a phosphorylated and dephosphorylated state. However, the phosphorylation of a protein which binds a short sequence of DNA would be expected to increase the mobility of the resulting DNA-protein complex through a gel, due to its increased negative charge. Therefore, although this mechanism of control is a possibility with the CArG box binding factors, the lack of difference in the mobility of the CArG box binding activities between tissues would argue against it.

Another explanation as to how the apparently ubiquitous CArG box binding activities direct the tissue-specific expression of the Xenopus borealis skeletal actin gene is that further proteins may interact with these proteins after they have bound the CArG motif. This second set of proteins could be expressed exclusively in developing skeletal muscle and although they do not bind the CArG box motif directly, they could activate the expression of the gene by binding the ubiquitous factors that interact with the CArG motif. Experiments to test this

Results and Discussion

hypothesis have been unsuccessful. However, this mechanism of control appears to be unlikely as the additional binding of proteins to the CARG motif exclusively in muscle tissue would induce a further tissue-specific shift of the apparently ubiquitous binding activities. The lack of this observation would argue against this as a mechanism of control.

The mechanism by which the ubiquitously expressed dorsal gene product achieves the ventral activation of the Drosophila genes twist and snail, and the related NF- κ B protein the tissue-specific expression of the immunoglobulin κ chain gene, is thought to occur by tissue-specific translocation of these proteins to the nucleus (see Hunt, 1989). This control mechanism of cell type-specific gene expression cannot be discounted as being important in the expression of the Xenopus borealis skeletal actin gene, since our experiments used whole cell extracts rather than nuclear extracts of embryos. Thus, any restricted nuclear localisation of the ubiquitous CARG box binding factors would not be detected in our experiments. However, experiments from other laboratories which study the tissue distribution of CARG box binding factors (namely SRF) have used nuclear extracts and observed similar results to ours (e.g. Boxer et al., 1989a). If we assume that similar mechanisms of regulation exist between the human and Xenopus skeletal actin genes, then it is unlikely that this mechanism of gene regulation is important in the case of the Xenopus borealis skeletal actin gene.

It is clear that the control of the tissue-specific expression

Results and Discussion

of the Xenopus borealis skeletal actin gene is complex. If the CArG box motif is indeed the element that confers tissue-specific expression to the skeletal actin gene, then a variety of mechanisms exist which would enable the apparently ubiquitous CArG box binding factors to direct this expression. However, it is also possible that other, as yet unidentified sequences present in the regulatory region of the gene are important for its expression. Indeed, one or more of a combination of these elements could be responsible for the strict spatial and temporal expression of the gene. This possibility is investigated in more detail in further chapters.

Results and Discussion

CHAPTER 8

Sequences in addition to the CArG box motifs are required for the expression of the *Xenopus borealis* skeletal actin gene.

Introduction.

Data presented in previous chapters has demonstrated the importance of the CArG box3 sequence motif in the expression of the *Xenopus borealis* skeletal actin gene. Furthermore, trans-acting factors are able capable of interacting with both the CArG box1 and CArG box3 sequences in vitro (see chapter 6). However, it is also apparent that the proteins which interact with these sequences are distributed throughout the embryo and exhibit no striking tissue-specific distribution (see chapter 7).

It is possible that the CArG box motifs alone are instrumental in determining the tissue-specific expression of the *Xenopus borealis* skeletal actin gene, and the possible mechanisms by which this may occur have been discussed previously (see chapter 7). However, although the CArG motif is obviously important in the expression of the *Xenopus borealis* skeletal actin gene, no evidence exists to demonstrate that these sequences, in conjunction with basal promoter elements such as the TATA box, are sufficient for the tissue-specific expression of the gene. In this regard it is becoming increasingly clear that the cell-type-specific expression of other muscle genes, such as the human cardiac actin (Sartorelli et al., 1990),

Results and Discussion

Xenopus cardiac actin (T. Mohun, Abstract; 3rd International Xenopus meeting, 1990) and quail fast skeletal troponin I (Lin et al., 1991) genes, is governed by a combination of both tissue-specific and ubiquitous transcription factors.

It was therefore decided to investigate whether the CArG box motif alone is sufficient for the tissue-specific expression of the Xenopus borealis skeletal actin gene. To achieve this a series of experiments were designed to test whether CArG box1, CArG box3, or a combination of both are sufficient to drive the tissue-specific expression of a heterologous promoter exclusively in the region of the embryo that expresses the endogenous skeletal actin gene. In addition to these experiments data is also presented in this chapter which demonstrate that a further region of the Xenopus borealis skeletal actin gene promoter, in addition to the CArG motifs, is required for the correct expression of the gene.

8.1 A 156 nucleotide region of the Xenopus borealis skeletal actin gene promoter can direct the tissue-specific expression of a heterologous promoter.

In order to establish a region of the Xenopus borealis skeletal actin gene which is sufficient for the tissue-specific expression of the gene a plasmid was obtained from Dr. M. Boardman which contained a region of the skeletal actin promoter (nucleotides -197 to -42) fused to a heterologous TATA box (plasmid p Δ gb Δ 5(+)).

In order to make this construct the plasmid p Δ gb was

Results and Discussion

exploited. This construct is derived from the plasmid pXlgb, which contains the complete Xenopus laevis β -globin gene, in addition to 477 nucleotides of 5' flank, inserted into the vector pAT153 (See Appendix 2). To make p Δ gb an internal deletion was made in the 5' promoter of the Xenopus laevis globin gene which spans from nucleotides -455 to -66. Thus, the plasmid p Δ gb contains the TATA box of the Xenopus laevis β -globin gene (nucleotides -29 to -25), yet retains only 90 nucleotides of the 5' flanking region of the gene (see appendix 2).

In order to construct p Δ gb Δ 5(+) the plasmid pXbgb Δ 5 (see chapter 4) was digested with Xho II, and the end blunted by in-filling prior to digestion with Bal I. This excises a 156 bp fragment from the plasmid which extends from 11 nucleotides upstream of the TATA box of the Xenopus borealis skeletal actin gene (nucleotide -42) to 35 nucleotides upstream of the CARG box3 motif (nucleotide -197). This fragment was subsequently purified after separation from vector sequences by agarose gel electrophoresis.

The plasmid p Δ gb was digested at a unique Eco RI site situated at the 5' end of the deleted Xenopus laevis β -globin promoter. This site was then blunted by in-filling and the Xenopus borealis skeletal actin gene promoter fragment inserted into this blunt Eco RI site. Recombinants were screened for promoter inserts in the correct orientation by restriction enzyme analysis.

The resulting plasmid (p Δ gb Δ 5(+)) contains a region of the

Results and Discussion

Xenopus borealis skeletal actin gene promoter spanning from nucleotides -197 to -42 fused precisely 59 nucleotides upstream of the Xenopus laevis β -globin gene TATA box (see figure 8.1.1).

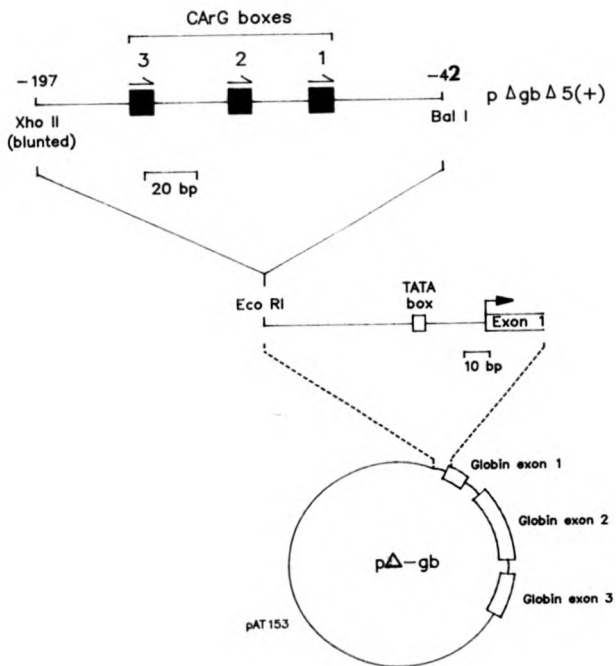
To test whether this 156 bp region of the Xenopus borealis skeletal actin gene promoter is sufficient to drive the expression of an otherwise silent $p\Delta gb$ construct exclusively in the axis region of embryos, the $p\Delta gb\Delta 5(+)$ plasmid was linearised with Pst I and injected into Xenopus laevis embryos at the two cell stage of development. In addition to this, embryos were also injected with either $p\Delta gb$ or $pXbgb\Delta 6$, the latter of which contains the actin promoter, plus 28 nucleotides of the first exon of the gene, fused to a globin reporter gene in its first exon (see chapter 4.1).

Micro-injected embryos were allowed to proceed to stage 26 of development before being dissected into axis and head plus gut regions. Total nucleic acid was extracted as described in chapter 3.4 and the transcripts of the micro-injected genes detected by primer extension. The oligonucleotide used in primer extension assays in these experiments is anti-sense to a sequence located in the first exon of the Xenopus laevis β -globin gene. Therefore, correctly initiated transcripts originating from the plasmid $p\Delta gb\Delta 5(+)$ result in an extended product of 85 nucleotides. However, due to the different nature of the plasmid $pXbgb\Delta 6$, correctly initiated transcripts arising from this plasmid result in a primer extension product of 68 nucleotides.

FIGURE 8.1.1. Diagrammatic representation of the plasmid
p Δ gb Δ 5(+).

The -197/-42 fragment of the Xenopus borealis skeletal actin gene promoter is represented in the upper part of the figure. The CArG box motifs are represented by red boxes, with their orientation in the resulting fusion construct being represented by arrows above the box.

The -197/-42 fragment of the Xenopus borealis skeletal actin gene promoter was inserted into the construct p Δ gb at an Eco RI site situated at the 5' end of the Xenopus laevis β -globin gene promoter in the sense orientation. Xenopus laevis β -globin sequences present in the p Δ gb plasmid are represented in black, whilst pAT153 sequences are shown in blue.



Results and Discussion

Figure 8.1.2 illustrates the results of a primer extension assay performed on total nucleic acid extracts prepared from the axis or head plus gut regions of embryos injected with either p Δ gb, pXbgb Δ 6, or p Δ gb Δ 5(+). As expected, no transcripts originating from the micro-injected gene are apparent in either the axis or head plus gut regions of embryos injected with p Δ gb. On the contrary, injection of pXbgb Δ 6 results in correctly initiated transcripts arising from the micro-injected gene appearing exclusively in the axis region of embryos. Some larger extension products are observed in the gut region of embryos injected with pXbgb Δ 6. However, these products were not consistently observed in subsequent experiments.

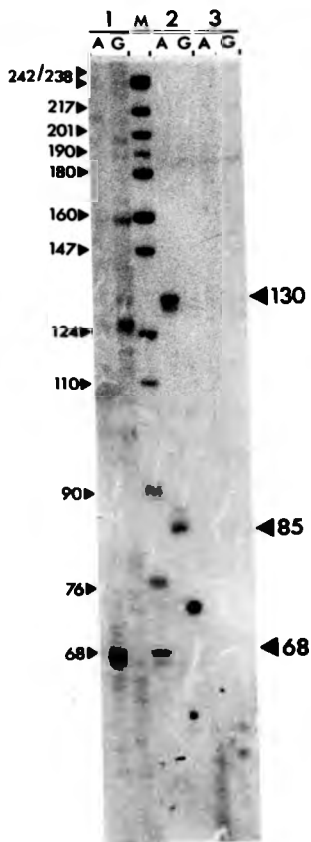
More importantly however, the injection of p Δ gb Δ 5(+) into embryos results in a primer-extended product of approximately 85 nucleotides exclusively in the axis region of embryos. Thus, the fragment of the Xenopus borealis skeletal actin gene promoter which spans from nucleotides -197 to -42 is capable of directing correctly initiated transcription from a heterologous promoter. Furthermore, this transcription appears to be tissue-specific.

A second primer extension product also exists exclusively in the axis region of embryos injected with the plasmid p Δ gb Δ 5(+). The predicted size of this second extension product is approximately 130 nucleotides. Assuming that the size of this fragment is 130 nucleotides, then the site of initiation of this transcript is at position -45 in the

FIGURE 8.1.2. Primer extension analysis of embryos injected with pXbgb Δ 6, p Δ gb, or p Δ gb Δ 5(+).

Embryos were injected with either pXbgb Δ 6 (tracks 1A and 1G), p Δ gb Δ 5(+) (tracks 2A and 2G), or p Δ gb (tracks 3A and 3G) at the two cell stage of development. When embryos reached stage 26 of development they were dissected into axis (A) and head plus gut (G) regions. Total nucleic acid was extracted and the transcripts of the micro-injected gene detected by primer extension.

The size of DNA markers (track M) are illustrated on the left hand side of the figure, whilst the size of the extension products are represented on the right hand side.



Results and Discussion

promoter of the deleted Xenopus laevis β -globin gene. One explanation as to the initiation of transcription from this site is that a putative TATA motif is present in the deleted β -globin promoter 28 nucleotides upstream of this second transcription initiation site (see figure 8.1.3). It would appear, therefore, that the Xenopus borealis skeletal actin gene promoter fragment is capable of driving the tissue-specific expression of the heterologous promoter from two TATA motifs.

All sequences necessary and sufficient for the tissue-specific expression of the Xenopus borealis skeletal actin gene are therefore contained within a 156 bp fragment of the promoter spanning from nucleotides -197 to -42. This region of the promoter contains all three CArG box motifs, of which CArG box1 and CArG box3 have previously been implicated in the expression of the gene (see chapter 4). However, other sequences are present in this promoter fragment and it is possible that these may also contribute towards the expression of the gene.

8.2 CArG box3, CArG box1, or a combination of both are not sufficient to drive the tissue-specific expression of a heterologous promoter.

It is apparent from data presented in chapter 4 that CArG box3 is essential for the expression of the Xenopus borealis skeletal actin gene. However, both CArG box3 and CArG box1 interact with factors distributed throughout the embryo (see chapter 7). To determine whether one or more of the CArG motifs

FIGURE 8.1.3. Sequence of the deleted *Xenopus laevis* β -globin promoter illustrating the two transcription start sites used by the *Xenopus borealis* skeletal actin gene promoter fragment.

The deleted *Xenopus laevis* β -globin gene promoter present in the construct p Δ gb is shown. Two TATA boxes are utilised by the -197/-42 nucleotide fragment of the *Xenopus borealis* skeletal actin gene promoter and these are positioned at nucleotides -465/-461 (red underline) or -29/-25 (black underline) in the promoter of the β -globin gene. These two TATA boxes initiate transcription at positions -45 (red arrow) or +1 (black arrow) in the β -globin gene respectively. The first exon of the *Xenopus laevis* β -globin gene is illustrated by blue coloured nucleotides.

-477
|
GAATTCAGGCATATAAAAAACGGCAGAGTGGAAGG
TATA box 2
-465/-461
-464/-65
deletion

-51
|
GGCAGTGCATCCTTACAGCTACATAAAGTCTGATGGATG
Transcription initiation 2
-45
TATA box 1
-29/-25

-12
|
GAGAATTAGAGCACTTGTTCTTTTGCAGAAGCTCAGAA
Transcription initiation 1
+1

Results and Discussion

is sufficient to drive the expression of the Xenopus borealis skeletal actin gene, or whether additional sequence elements are required, it was decided to test the ability of either CArG box1, CArG box3, or a combination of both to direct the tissue-specific expression of the deleted β -globin promoter.

To address this question a series of double-stranded oligonucleotides were synthesised which contain sequences complementary to CArG box1, CArG box3, or both CArG box1 and CArG box3 (oligonucleotide CArG B3/B1) (see figure 8.2.1). The core sequences of the CArG B3/B1 oligonucleotide were designed such that the sequences between the two CArG boxes, in addition to the CArG motifs themselves, are identical to the sequences of the closely arranged CArG box1 and CArG box3 motifs situated in the promoter of the plasmid pXbgb Δ B+ (see chapter 4.2). This oligonucleotide was therefore virtually identical to a short region of a promoter which has previously been shown to be capable of directing the tissue-specific expression of a reporter gene in Xenopus laevis embryos (see chapter 4.2).

The plasmid p Δ gb was digested with Eco RI and each double-stranded oligonucleotide, which possess Eco RI sticky ends, inserted into this site. Oligonucleotides lacked terminal phosphates to avoid multiple inserts of these sequences. Transformed bacteria were screened for inserts by restriction enzyme analysis. Positives were screened further for the insertion of oligonucleotides in the desired orientation by double-stranded DNA sequencing.

Plasmids were constructed which consisted of a single copy of

Results and Discussion

CArG box3.

AATTCGAGCTCT**ATATTTT**GGTACCG

GCACGAGATATAAAACCATGGCTTAA
E S K E

CArG box1.

AATTCGAGCTCT**CCAAATAT**CGAGGTACC

GCTCGAGAGGTTTATACCTCCATGGTTAA
E S K E

CArG B3/B1.

AATTCGAGCTCT**ATATTTT**GGTACCCAATTGGTTCTGT**CCAAATAT**GGAGTC

GCTCGAGATATAAAACCATGGTTAAGCCAAGACAGGTTTATACCTCAGTTAA
E S K E

FIGURE 8.2.1. Sequences of oligonucleotides used in the construction of plasmids $p\Delta_{gbB3}$, $p\Delta_{gbB1}$, $p\Delta_{gbB3/B1(+)}$ and $p\Delta_{gbB3/B1(-)}$.

CArG box3, CArG box1 or CArG B3/B1 oligonucleotide sequences are shown. CArG box motifs are represented by bold type. Restriction enzyme sites are marked below the sequence (E= Eco RI; S= Sac I; K= Kpn I).

Results and Discussion

the CARG B3/B1 oligonucleotide inserted into the Eco RI site of p Δ gb in either the sense (p Δ gbB3/B1(+)) or anti-sense (p Δ gbB3/B1(-)) orientations. Additional plasmids were also constructed which consisted of a single copy of either the CARG box1 (p Δ gbB1) or CARG box3 (p Δ gbB3) oligonucleotides inserted into the Eco RI site in the sense orientation (see figure 8.2.2).

The constructs pXbgb Δ 6, p Δ gb, p Δ gb Δ 5(+), p Δ gbB3/B1(+), p Δ gbB3/B1(-), p Δ gbB1 or p Δ gbB3 were linearised with Pst I and injected into Xenopus laevis embryos at the two cell stage of development. Embryos were allowed to develop until stage 26 of development before being dissected into axis and head plus gut regions. Total nucleic acid was prepared from embryo dissections, and the transcripts of the micro-injected gene detected by primer extension as described in chapter 8.1.

Figure 8.2.3 illustrates the previously described extension products present exclusively in the axis region of embryos injected with the plasmid p Δ gb Δ 5(+) (track 3A), whilst no transcripts appear to be present in either the axis or head plus gut regions of embryos injected with p Δ gb (tracks 1A and 1G). However, it is also apparent that no transcripts originating from the micro-injected fusion gene are present in embryos containing the constructs p Δ gbB1, p Δ gbB3, p Δ gbB3/B1(-) or p Δ gbB3/B1(+) (tracks 4A to 7G).

Southern blot analysis of nucleic acid extracts, using vector DNA as a probe, reveals that approximately equivalent amounts of DNA are present in embryos injected with each of the

FIGURE 8.2.2. Diagrammatic representation of the constructs

p Δ gbB3, p Δ gbB1, p Δ gbB3/B1(+) and
p Δ gbB3/B1(-).

Oligonucleotides cloned into the Eco RI site of the construct p Δ gb to create the plasmids p Δ gbB3, p Δ gbB1, p Δ gbB3/B1(+) and p Δ gbB3/B1(-) are illustrated at the top of the figure. CARG box motifs are illustrated by red boxes.

The construct p Δ gb is shown in the lower half of the figure, with the deleted promoter of the Xenopus laevis β -globin gene being expanded to illustrate the site of insertion of oligonucleotides. Xenopus laevis sequences are illustrated in black, whilst pAT153 sequences are shown in blue.

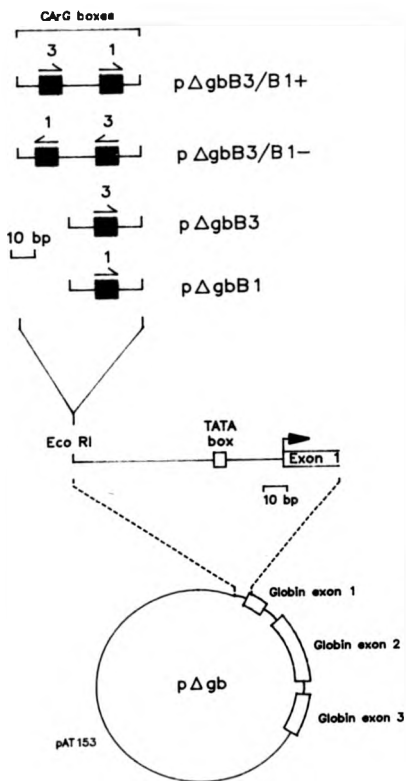
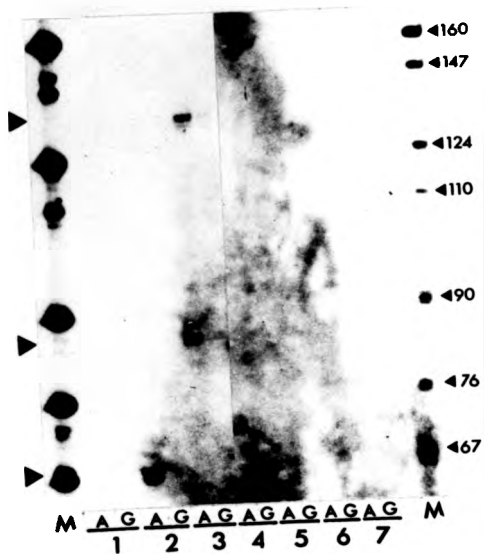


FIGURE 8.2.3. Primer extension analysis of nucleic acid isolated from embryos injected with plasmids p Δ gb, pXbgb Δ 6, p Δ gb Δ 5(+), p Δ gbB3, p Δ gbB1, p Δ gbB3/B1(+) and p Δ gbB3/B1(-).

Embryos were injected at the two cell stage of development with either p Δ gb (tracks 1A and 1G), pXbgb Δ 6 (tracks 2A and 2G), p Δ gb Δ 5(+) (tracks 3A and 3G), p Δ gbB3 (tracks 4A and 4G), p Δ gbB1 (tracks 5A and 5G), p Δ gbB3/B1(+) (tracks 6A and 6G) or p Δ gbB3/B1(-) (tracks 7A and 7G). Embryos were allowed to proceed to stage 26 of development before being dissected into axis (A) and head plus gut (G) regions. Total nucleic acid was extracted and the transcripts from the micro-injected gene detected by primer extension.

The size of DNA markers (track M) is indicated on the right hand side of the figure.



Results and Discussion

plasmids (see figure 8.2.4.). Indeed, it transpires that embryo fractions which are actively expressing the micro-injected gene (i.e. $p\Delta gb\Delta 5(+)$) contain slightly less DNA than those that are not. In addition to this, RNA present in total nucleic acid extracts appears to be relatively undegraded, as judged by the intact nature of rRNA when subjected to agarose gel electrophoresis (data not shown). It would therefore appear that the lack of transcripts originating from the micro-injected constructs $p\Delta gbB1$, $p\Delta gbB3$, $p\Delta gbB3/B1(-)$ and $p\Delta gbB3/B1(+)$ is due to a lack of transcriptional activity of these fusion genes, and not to a lack of exogenous DNA in certain embryos, or the degradation of mRNA on nucleic acid preparation.

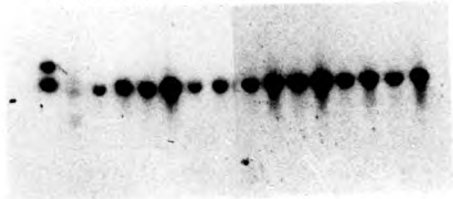
These data demonstrate that CARG box1, CARG box3, or a combination of both are insufficient to drive the tissue-specific expression of a heterologous promoter. Thus, although the CARG box motif is essential for the expression of the Xenopus borealis skeletal actin gene, it would appear that it has to operate in conjunction with other as yet unidentified sequences elements. The finding that the -197/-42 promoter fragment is capable of directing the expression of a heterologous promoter makes it possible to speculate that these additional sequences are contained within this fragment of the gene.

FIGURE 8.2.4. Southern blot analysis of embryos infected with plasmids p Δ gb, pXbgb Δ 6, p Δ gb Δ 5(+), p Δ gbB3, p Δ gbB1, p Δ gbB3/B1(+) and p Δ gbB3/B1(-).

Nucleic acid extract equivalent to one axis or head plus gut region of the embryo was digested with Hind III and electrophoresed through a 0.7% agarose gel. DNA was transferred to nitrocellulose as described in chapter 3.12 and micro-injected DNA detected by probing filters with the vector pAT153. After hybridisation of the probe filters were washed twice in 3X SSC for 10 minutes at 50°C prior to a final wash in 3X SSC for 10 minutes at room temperature. Filters were blotted dry and autoradiographed at -70°C.

Nucleic acid extract prepared from the axis (A) and head plus gut (G) fractions of embryos injected with the constructs p Δ gb (tracks 1), pXbgb Δ 6 (tracks 2), p Δ gb Δ 5(+) (tracks 3), p Δ gbB3 (tracks 4), p Δ gbB1 (tracks 5), p Δ gbB3/B1(+) (tracks 6) and p Δ gbB3/B1(-) were analysed. Track C represents 600pg of the plasmid p Δ gb digested with Hind III loaded onto the gel as a positive control.

1 2 3 4 5 6 7
C AG AG AG AG AG AG



Results and Discussion

8.3 A region of the *Xenopus borealis* skeletal actin gene promoter upstream of the CArG box3 motif is required for the expression of the gene when sequences downstream of nucleotide -42 are absent from the gene.

The discovery that CArG box1, CArG box3, or a combination of both CArG box1 and CArG box3 is unable to drive the expression of a heterologous promoter suggests that other sequences in addition to the CArG motifs are required for the tissue-specific expression of the *Xenopus borealis* skeletal actin gene. Furthermore, the finding that a fragment of the promoter spanning from nucleotides -197 to -42 contains all sequences that are necessary for the expression of the gene, makes it possible to speculate that these additional regulatory sequence elements are contained within this 156 nucleotide fragment of the promoter.

In order to locate additional sequences which are important in the expression of the *Xenopus borealis* skeletal actin gene a further two constructs were made. These plasmids contained progressively less of the -197/-42 skeletal actin promoter fragment fused to the heterologous promoter contained within the construct p Δ gh (see figure 8.3.1). The first of these plasmids contains a 126 bp fragment of the skeletal actin promoter that spans from nucleotides -42 to -167. However, the second of these plasmids contains a region of the skeletal actin promoter that spans from nucleotides -42 to -99, in addition to a synthetic CArG box3 oligonucleotide fused to nucleotide -99.

Results and Discussion

In order to make these plasmids either pXbgb Δ 8+, or pXbgb Δ 6 were digested with Pst I and Bal I. This results in the excision from these plasmids of a fragment which extends from nucleotide -42 in the respective promoters of the Xenopus borealis skeletal actin gene, to a Pst I site contained in the ampicillin resistance gene of the vector. These restriction fragments were subsequently isolated after separation from vector sequences by agarose gel electrophoresis.

The plasmid p Δ gb was digested with Eco RI and the ends of the DNA blunted by in-filling with the Klenow fragment of DNA polymerase, prior to being digested with Pst I. The terminal phosphates of the resulting fragments were removed by calf intestinal alkaline phosphatase, to avoid the reconstitution of p Δ gb in ligation reactions.

The Pst I/Bal I fragment isolated from the plasmids pXbgb Δ 6 and pXbgb Δ 8+ was inserted into the Pst I/blunt fragment of the vector. Because p Δ gb, pXbgb Δ 6 and pXbgb Δ 8+ are all pAT153 based plasmids, the insertion of these fragments into p Δ gb reconstitutes the vector sequences, which includes the ampicillin resistance gene.

Transformed E.coli colonies were isolated and screened for inserts by restriction enzyme analysis. Finally, positives were screened for the correct insertion of DNA by double-stranded DNA sequencing.

The resulting plasmids were named p Δ gb Δ 6 and p Δ gb Δ 8+, depending on the origin of the inserted promoter fragment. The p Δ gb Δ 6 plasmid contains a fragment of the Xenopus borealis

Results and Discussion

skeletal actin gene promoter, which extends 11 nucleotides upstream of the TATA box (nucleotide -42) to 5 nucleotides upstream of the CArG box3 motif (nucleotide -167), fused onto the end of the deleted β -globin promoter contained in the plasmid p Δ gb. However, the plasmid p Δ gb Δ 8+ contains a fragment of the skeletal actin gene promoter that extends from nucleotide -42 to a CArG box3 motif placed directly adjacent to nucleotide -99, fused to the heterologous promoter (see figure 8.3.1).

The plasmids pXbgb Δ 6, p Δ gb, p Δ gb Δ 5(+), p Δ gb Δ 6, p Δ gb Δ 8+, p Δ gbB3, p Δ gbB1, p Δ gbB3/B1(+) and p Δ gbB3/B1(-) were linearised with Pst I and injected into embryos at the two cell stage of development. Micro-injected embryos were allowed to proceed to stage 26 of development, before being dissected into axis and head plus gut regions. Total nucleic acid was extracted and the transcripts of the micro-injected gene detected by primer extension as previously described (see section 8.1).

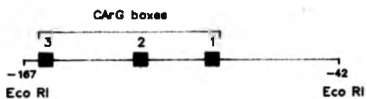
Data illustrated in figure 8.3.2 reveals the previously described extension products present in reactions containing nucleic acid isolated from the axis region of embryos injected with p Δ gb Δ 5(+). In addition to this embryos injected with either p Δ gb, p Δ gbB3, p Δ gbB1, p Δ gbB3/B1(+) or p Δ gbB3/B1(-) show no extended products in either the axis or head plus gut regions of the embryo. Furthermore, no extension products are observed in either region of embryos injected with p Δ gb Δ 6 or p Δ gb Δ 8+.

FIGURE 8.3.1 Diagrammatic representation of the constructs

p Δ gb Δ 6 and p Δ gb Δ 8+.

Xenopus borealis skeletal actin gene promoter fragments contained within the constructs p Δ gb Δ 6 and p Δ gb Δ 8+ are represented at the top of the figure. CARG box motifs are illustrated in red.

The construct p Δ gb is shown in the lower half of the figure, with the deleted promoter of the Xenopus laevis β -globin gene being expanded to illustrate the site of insertion of the Xenopus borealis skeletal actin gene promoter fragments. Xenopus laevis sequences are illustrated in black, whilst pAT153 sequences are shown in blue.



$p\Delta gb\Delta 6$



$p\Delta gb\Delta 8+$

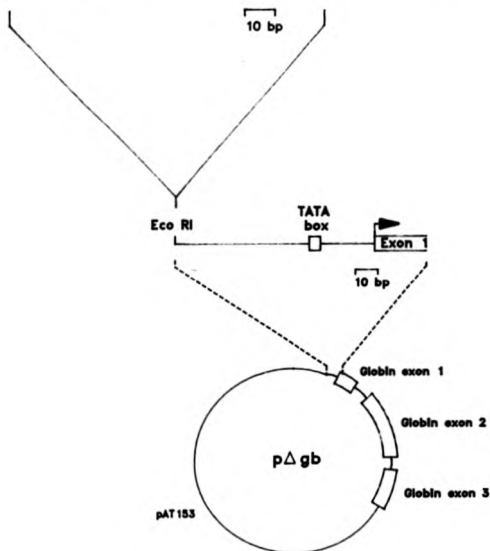


FIGURE 8.3.2. Primer extension analysis of nucleic acid isolated from embryos injected with plasmids
p Δ gb, pXbgb Δ 6, p Δ gb Δ 5(+), p Δ gb Δ 6,
p Δ gb Δ 8+, p Δ gbB3, p Δ gbB1, p Δ gbB3/B1(+) and
p Δ gbB3/B1(-).

Embryos were injected at the two cell stage of development with either p Δ gb (tracks 1A and 1G), pXbgb Δ 6 (tracks 2A and 2G), p Δ gb Δ 5(+) (tracks 3A and 3G), p Δ gb Δ 6 (tracks 4A and 4B), p Δ gb Δ 8+ (tracks 5A and 5G), p Δ gbB3 (tracks 6A and 6G), p Δ gbB1 (tracks 7A and 7G), p Δ gbB3/B1(+) (tracks 8A and 8G) or p Δ gbB3/B1(-) (tracks 9A and 9G). Embryos were allowed to proceed to stage 26 of development before being dissected into axis (A) and head plus gut (G) regions. Total nucleic acid was extracted and the transcripts from the micro-injected gene detected by primer extension.

The size of DNA markers (track M) is indicated on the right hand side of the figure.

1	2	3	4	5	6	7	8	9
A	G	A	G	A	G	A	G	A

M

147

124

110

90

76

67

Results and Discussion

The apparent lack of transcripts originating from certain fusion genes may occur by less micro-injected DNA persisting in these samples. To establish whether this was the case, the relative amounts of exogenous DNA present in each sample were compared by southern analysis using vector sequences to detect the micro-injected gene. Figure 8.3.3 demonstrates that none of the embryo fractions contain substantially less DNA than is apparent in the axis region of embryos injected with the plasmid p Δ gb Δ 5(+). In addition to this, the analysis of rRNA by agarose gel electrophoresis reveals RNA contained in nucleic acid samples to be intact (data not shown). Thus, the lack of mRNA in embryos injected with certain fusion genes would appear to be due to a lack of transcriptional activity of these genes, and not due to the absence of micro-injected DNA, or the degradation of mRNA on nucleic acid extract preparation.

From these data it is possible to conclude that an upstream regulatory element (URE) present in a region of the Xenopus borealis skeletal actin gene that extends from nucleotides -197 to -168 is essential for the expression of the gene in the context of these plasmids. The lack of activity of the p Δ gb Δ 6 construct is somewhat surprising in light of experiments which demonstrate that the URE region of the skeletal actin promoter is not essential when sequences between nucleotide -41 and +28 of the gene are present (see construct pXbgb Δ 6, chapter 4). However, the loss of sequences downstream of nucleotide -42 make the URE an absolute requirement for the expression of the Xenopus borealis skeletal actin gene. It is therefore possible

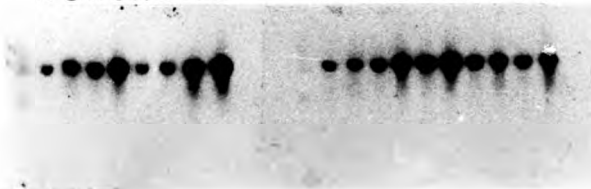
FIGURE 8.3.3. Southern blot analysis of embryos injected with plasmids p Δ gb, pXbgb Δ 6, p Δ gb Δ 5(+), p Δ gb Δ 6, p Δ gb Δ 8+, p Δ gbB3, p Δ gbB1, p Δ gbB3/B1(+) and p Δ gbB3/B1(-).

Nucleic acid extract equivalent to one axis or head plus gut region of the embryo was digested with Hind III and electrophoresed through a 0.7% agarose gel. DNA was transferred to nitrocellulose as described in chapter 3.12 and micro-injected DNA detected by probing filters with the vector pAT153. After hybridisation of the probe, filters were washed twice in 3X SSC for 10 minutes at 50°C prior to a final wash in 3X SSC for 10 minutes at room temperature. Filters were blotted dry and autoradiographed at -70°C.

Nucleic acid extract prepared from the axis (A) and head plus gut (G) fractions of embryos injected with the constructs p Δ gb (tracks 1), pXbgb Δ 6 (tracks 2), p Δ gb Δ 5(+) (tracks 3), p Δ gb Δ 6 (tracks 4), p Δ gb Δ 8+ (tracks 5), p Δ gbB3 (tracks 6), p Δ gbB1 (tracks 7), p Δ gbB3/B1(+) (tracks 8) and p Δ gbB3/B1(-) (tracks 9) were analysed.

1 2 3 4
A G A G A G A G

5 6 7 8 9
A G A G A G A G A G



Results and Discussion

to speculate that there are two alternative redundant positive control elements in the wild-type gene, although the URE is essential for the expression of the gene in the absence of sequences downstream of nucleotide -42.

8.4 Discussion.

Data presented in this chapter demonstrate that a 156 nucleotide fragment of the Xenopus borealis skeletal actin gene promoter is sufficient to drive the tissue-specific expression of a heterologous promoter exclusively in the axis region of embryos. Thus, a region of the promoter which spans from nucleotides -197 to -42 contains all sequences necessary for the expression of the gene. Indeed, the CArG box3 and CArG box1 motifs are contained within this fragment of the promoter and these sequence elements have been previously implicated in the expression of both the Xenopus borealis skeletal actin and other sarcomeric actin genes (see chapters 1.4.2, 4.1 and references therein).

Experiments presented in this chapter demonstrate that CArG box1, CArG box3, or a combination of both are insufficient to drive the tissue-specific expression of a heterologous promoter. Furthermore, constructs containing the CArG box motifs alone demonstrate no expression in the head plus gut regions of embryos. It would appear unlikely, therefore, that the CArG box is a promoter element that drives the transcription of the Xenopus borealis skeletal actin gene ubiquitously, with expression of this gene being restricted to

Results and Discussion

muscle tissue by the repression of the gene in non-muscle cells by negative regulatory elements.

One possible explanation for the lack of transcriptional activity of chimeric fusion promoters containing the CARG box motifs alone is that this element is situated too near the TATA box to function effectively. However, the distances between the most proximal CARG motif and the heterologous TATA box in the constructs p Δ gbB3/B1(+), p Δ gbB3/B1(-), p Δ gbB3 and p Δ gbB1 are 63, 65, 64 and 66 nucleotides respectively. The distance between the CARG box1 motif and the TATA box in the wild-type Xenopus borealis skeletal actin gene promoter is 52 nucleotides. Thus, the most proximal CARG motif present in the promoters of the chimeric fusion constructs is approximately the same distance from the heterologous TATA box as the CARG box1 motif is from the TATA box in the wild-type gene. It would appear unlikely, therefore, that the distance between the CARG motif and the heterologous TATA element in the constructs p Δ gbB3/B1(+), p Δ gbB3/B1(-), p Δ gbB3 and p Δ gbB1 is responsible for their lack of transcriptional activity when injected into Xenopus embryos. Equally, the positioning of the CARG motifs on the DNA helix relative to each other and the TATA box does not appear to be essential for the expression of the human cardiac actin gene (Miwa and Kedes, 1987). Although the positioning of the CARG motifs relative to the TATA box has not been directly addressed in this study, the findings of Miwa and Kedes (1987) would argue against this being responsible for the lack of transcriptional activity of certain chimeric fusion

Results and Discussion

promoter constructs. In this regard, the construct p Δ gb Δ 5(+) contains a fragment of the Xenopus borealis skeletal actin gene promoter which is capable of driving the expression of a Xenopus laevis β -globin gene by utilising two TATA boxes. Indeed, these two TATA boxes are on virtually opposite sides of the DNA helix to each other with respect to the CArG box1 motif present in this plasmid.

It is apparent from this study that the CArG boxes alone are insufficient to drive the tissue-specific expression of a heterologous promoter. Therefore, other sequences are required in addition to the CArG box to direct the tissue-specific expression of the Xenopus borealis skeletal actin gene. This discovery is not totally unexpected since the CArG motif has been identified in the regulatory regions of a number of genes with diverse patterns of expression such as the c-fos (see chapter 1.4.3. and references therein), β -actin (see figure 1.1, chapter 1) and the interleukin-2-receptor (Phan-Dinh-Tuy et al., 1988) genes. Thus, it is possible to speculate that although the CArG box motif is essential for the expression of a variety of genes, the tissue-specificity of the Xenopus borealis skeletal actin gene is determined by other, as yet unidentified regulatory sequences.

It is becoming increasingly clear that the cell-type-specific expression of a variety of genes is complex and requires multiple regulatory sequence elements. For example, only 218 nucleotides of 5' flanking region of the rat insulin II gene is required to drive the expression of a reporter gene (Crowe and

Results and Discussion

Tsai, 1989). However, linker scanning mutation and deletion studies of this promoter fragment revealed the presence of multiple DNA regulatory elements required for the full expression of the gene. Similarly, the growth hormone gene also contains multiple regulatory elements, including a cAMP response element and a binding site for the pituitary-specific transcription factor Pit-1 (see Karin et al., 1990). Furthermore, it has been proposed that ubiquitous stimulatory factors drive the expression of the immunoglobulin heavy chain gene, but that tissue-specificity is conferred to this gene by negative regulatory elements repressing this gene in non-lymphoid cells (Imler et al., 1987).

The observation that the CArG box is unable to confer muscle-specificity to a heterologous promoter has also been observed in studies involving the Xenopus laevis cardiac actin gene (T. Mohun; personal communication). Indeed, this gene has been shown to require the CArG box1 motif, a MyoD binding site and a further, as yet unidentified regulatory element for its correct expression (T. Mohun, abstract; 3rd international Xenopus meeting, 1990). Likewise, the human cardiac actin gene requires a combination of a CArG box1 motif, a MyoD site and a SP-1 binding site for its correct expression (Sartorelli et al., 1990). However, a region of the chicken skeletal actin gene promoter (nucleotides -73 to -100), which contains the CArG box1 motif of this gene, has been demonstrated to drive the muscle specific expression of a c-fos gene promoter truncated to nucleotide -56 (Walsh, 1989).

Results and Discussion

The construct $p\Delta gb\Delta 5(+)$ and $p\Delta gb\Delta 6$ are identical, with the exception that $p\Delta gb\Delta 6$ lacks a region of the Xenopus borealis skeletal actin gene promoter (nucleotides -197 to -168) which is present in the $p\Delta gb\Delta 5(+)$ construct. The construct $p\Delta gb\Delta 5(+)$ is transcriptionally active in the axis regions of embryos injected with this plasmid, however the $p\Delta gb\Delta 6$ construct is not. This demonstrates the importance of the -197/-168 region (URE) of the Xenopus borealis skeletal actin promoter in the expression of the gene. Paradoxically, the construct $pXgb\Delta 6$, which lacks the URE, is expressed in a correct tissue-specific manner when injected into developing Xenopus laevis embryos. However this construct contains sequences of the Xenopus borealis skeletal actin gene promoter (-41 to +28) which are absent from both the $p\Delta gb\Delta 5(+)$ and $p\Delta gb\Delta 6$ constructs. One possible explanation for this discrepancy in data is that the -41/+28 region of the gene contains sequences which are capable of substituting for regulatory sequences contained in the URE. If this is the case, then it is not known whether regulatory elements contained within the -41/+28 region of the gene are identical to those contained in the URE. However, it is interesting to note that a region of the chicken skeletal actin gene promoter spanning from nucleotides -202 to -12 has been observed to contain partial dyad symmetry and is capable of directing the transcription of a reporter gene in a bi-directional manner (Grichnik et al., 1988). Indeed, sequences present in the URE of the Xenopus borealis skeletal actin gene are present in a

Results and Discussion

inverted orientation in the -41/+28 region of the gene (see figure 8.4).

It is apparent from data presented in this chapter that sequences in addition to the CArG box motifs are required for the expression of the Xenopus borealis skeletal actin gene. By fusing promoter deletions to a heterologous promoter it was demonstrated that an upstream regulatory element (URE) exists in a region of the skeletal actin promoter that spans from nucleotides -197 to -168. Whether the URE contains a tissue-specific promoter element, or whether it contains a ubiquitous transcription factor binding site, like the SP-1 site in the human cardiac actin gene promoter, is unknown.

Results and Discussion

-174	AGGGAGAG	-167	URE.

+28	AGGGAGAG	+21	-41/+28 region (lower strand).
-197	GATCTG	-192	URE.
	*** **		
-3	GATGTG	-8	-41/+28 region (lower strand).
-184	GTTGAAGGGGA	-174	URE.
	**** ** **		
-8	GTTGCGGTGA	-18	-41/+28 region (lower strand).

FIGURE 8.4. Comparison of URE sequences with sequences in the -41/+28 region of the *Xenopus borealis* skeletal actin gene.

Sequences were compared between the URE and the -41/+28 region of the *Xenopus borealis* skeletal actin gene. All sequences of the -41/+28 region of the gene illustrated are of the lower strand of DNA. The position of each sequence is illustrated by the nucleotide position in the gene.

Results and Discussion

CHAPTER 9

The upstream regulatory element of the *Xenopus borealis* skeletal actin gene binds a trans-acting factor(s) in vitro.

Introduction.

It is apparent from data presented in the preceding chapter that CArG box3, and probably CArG box1, are essential for the expression of the *Xenopus borealis* skeletal actin gene. However, these sequence motifs alone are unable to confer muscle-specific expression on a heterologous promoter. Unlike the chicken skeletal actin gene, whose CArG box1 motif appears to be sufficient for its expression (Walsh, 1989), data presented in this report argue that other sequences in conjunction with the CArG motif are responsible for the tissue-specific expression of the *Xenopus borealis* skeletal actin gene. Indeed, in the case of the human cardiac actin gene the CArG box1 motif, along with an SP-1 binding site, have been demonstrated to be essential for the expression of the gene. However, tissue-specificity is conferred upon this gene by a MyoD site situated downstream of the CArG box1 motif present in the promoter of this gene (Sartorelli et al., 1990).

By exploiting the capability of a 156 nucleotide fragment of the *Xenopus borealis* skeletal actin gene to drive the tissue-specific expression of a heterologous promoter, a further upstream regulatory element (URE) in the promoter of the skeletal actin gene (nucleotides -197 to -168) has been

Results and Discussion

identified (see chapter 8). Experiments outlined in this chapter demonstrate the ability of the Xenopus borealis skeletal actin gene URE to interact with a trans-acting factor(s) in vitro. Furthermore, the distribution of this binding activity throughout the embryo is examined. In addition to this, experiments are also described which investigate the capability of the URE, in conjunction with the CARG box1 and CARG box3 sequence motifs, to drive the expression of a heterologous promoter in Xenopus laevis embryos.

9.1 The Xenopus borealis URE binds a proteins(s) in vitro.

In order to investigate whether the Xenopus borealis skeletal actin gene URE is capable of interacting with trans-acting factors in vitro, a double-stranded oligonucleotide was synthesised which contained sequences present in this region of the skeletal actin gene promoter (nucleotides -197 to -168) (see figure 9.1a). This oligonucleotide was end labelled with γ -³²P-ATP (see chapter 3.10.2) and used in band shift assays in conjunction with protein extracts prepared from stage 16 Xenopus laevis embryos. To test the distribution throughout the embryo of any trans-acting factors with which the URE interacts, embryos were dissected into somite, presumptive gut and ectoderm regions. Protein extracts prepared from these embryo fractions were then used in band shift assays using the URE sequence as a probe.

Results illustrated in figure 9.1b demonstrate the presence of two binding activities in band shift reactions containing

FIGURE 9.1. The *Xenopus borealis* skeletal actin gene URE binds a trans-acting factor(s) in vitro.

A. Sequence of the oligonucleotide URE. Restriction sites are represented under the sequence of the oligonucleotide. *Xenopus* sequences contained in the URE oligonucleotide are indicated by a dashed line above the sequence.

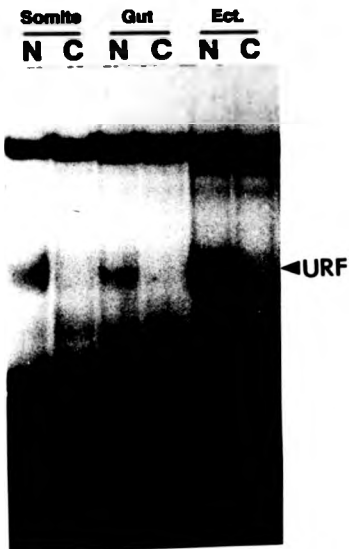
B. Protein extracts were isolated from either the somite (SOMITE), gut (GUT) or ectoderm (ECT.) fractions of stage 19 *Xenopus laevis* embryos. These extracts were then employed in band shift assays using the URE oligonucleotide as a probe. Conditions for band shift assays in this experiment are different from those normally used and are essentially as described by Buskin and Hauschka (1989) (see chapter 3.16).

Binding reactions contained either no competitor (N tracks), or competitor in the form of a 40-fold molar excess of unlabelled URE oligonucleotide. URE oligonucleotides used in competition reactions replaced an equivalent amount of non-specific competitor in normal binding reactions.

A.

 GGATCTGTATACTGGTTGAACGGGAGCGGAGAAGCT
TCGACCTAGACATATGACCAACTTCCCCTCCCTCT
 Sac I Xho II Sac I

B.



Results and Discussion

protein extracts prepared from the somites of stage 16 Xenopus laevis embryos. However, on competition with a 40-fold molar excess of URE only the lower of these two bands is abolished (figure 9.1b, compare track N with track C). It would therefore appear that only the lower of these two retarded bands is representative of a sequence-specific trans-acting factor(s) that binds the URE sequence. This binding activity has been termed the Upstream Regulatory Factor, or URF (see figure 9.1b).

The URF appears to be present in all fractions of the embryo examined (figure 9.1b, compare N tracks in somite, gut and ect. lanes). Indeed, the pattern of shifted bands appears to be identical in all protein extracts examined. Thus, the URF is present and capable of binding the URE in both fractions of the embryo that express the skeletal actin gene (i.e. somites), in addition to fractions of the embryo that express little, if any skeletal actin (i.e. gut and ectoderm). It is also apparent that no negative trans-acting factors, which are present exclusively in fractions of the embryo which do not express the skeletal actin gene, are capable of interacting with the URE, as detectable by the assay system employed in this study.

9.2 The URE, in conjunction with CARG box1 and CARG box3, is insufficient to direct the tissue-specific expression of a heterologous promoter.

The discovery of a further regulatory element in the promoter of the Xenopus borealis skeletal actin gene, and the

Results and Discussion

demonstration that this sequence is capable of interacting with a trans-acting factor(s) in vitro, raises the question as to whether this sequence, in conjunction with CArG box3 and CArG box1, is sufficient for the correct expression of the skeletal actin gene. To address this question it was decided to test whether the URE, in conjunction with the CArG motifs, is sufficient to drive the tissue-specific expression of a heterologous promoter. To achieve this the plasmid p Δ gbCB-URE was constructed. This plasmid contains the URE oligonucleotide inserted directly upstream of the two CArG box motifs present in the plasmid p Δ gbB3/B1(+) (for p Δ gbB3/B1(+) plasmid description see chapter 8.2).

The construction of p Δ gbCB-URE was achieved by digesting p Δ gbB3/B1(+) at a unique Sac I site positioned immediately upstream of the CArG box3 motif contained within this plasmid. The URE oligonucleotide, which possesses Sac I sticky ends, was inserted into the Sac I site of p Δ gbB3/B1(+) in the sense orientation. Oligonucleotides added to ligation reactions lacked terminal phosphate groups to avoid multiple copies of these sequences being inserted into the vector. Recombinants were screened for the correct insertion of oligonucleotides by the polymerase chain reaction (see chapter 3.5.5), and positives checked further by double strand DNA sequencing (see chapter 3.11) after large scale preparation of plasmid DNA.

Thus, a construct was created which contains CArG box3 and CArG box1, with sequences complementary to the URE inserted directly upstream of the CArG box3 motif (see figure 9.2.1).

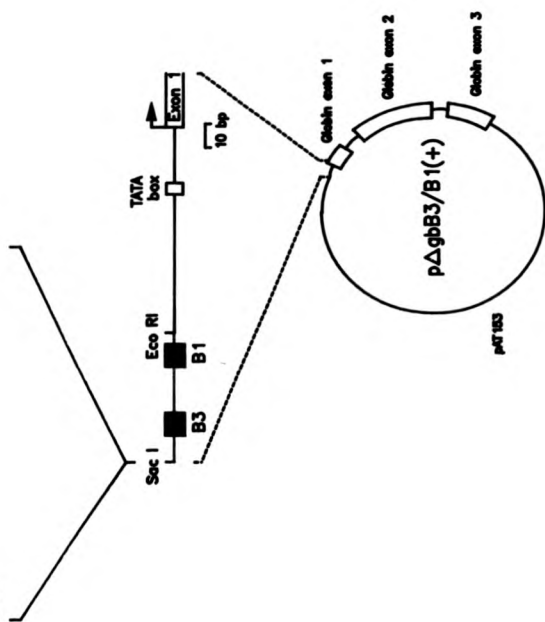
FIGURE 9.2.1. Diagrammatic representation of the plasmid
p Δ gbCB-URE.

The oligonucleotide cloned into the Sac I site of the plasmid p Δ gbB3/B1(+) to create the construct p Δ gbCB-URE is illustrated at the top of the figure.

The construct p Δ gbB3/B1(+) is shown in the lower half of the figure, with the deleted promoter of the Xenopus borealis β -globin gene containing the CArG box3 and CArG box1 motifs (red boxes) expanded to illustrate the site of insertion of the oligonucleotide. Xenopus laevis and Xenopus borealis sequences are shown in black and red respectively, whilst pAT153 sequences are shown in blue.

GCATCTGTATACTGTTCAAGCGGAGGAGAGCT
 TCGACCTAGACATATGACCGACTTCGCTCGCTCT

pΔgbCB-URE



Results and Discussion

Indeed, the sequences fused to the 5' end of the heterologous promoter are identical to those present in the plasmid pXhgb Δ 5 which span from the 5' end of the Xenopus borealis skeletal actin gene promoter (position -197) to the nucleotide directly 3' to the CArG box3 motif (position -152). However, a 5 nucleotide sequence which spans from between nucleotide -167 to the CArG box3 motif in the wild-type gene (position -163) is not reconstituted in this fusion promoter.

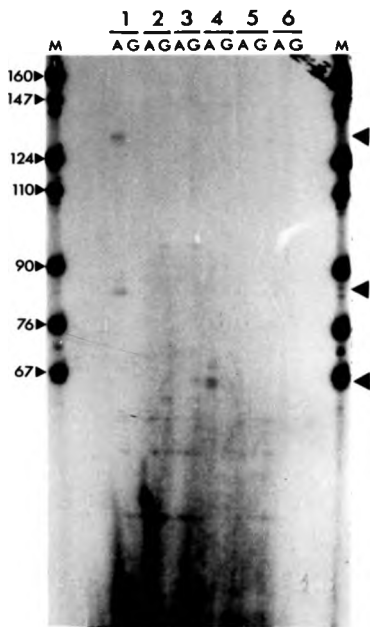
The plasmids p Δ gb Δ 5(+), p Δ gb Δ 6, p Δ gb Δ 8+, pXhgb Δ 6, p Δ gbB3/B1(+) and p Δ gbCB-URE were linearised with Pst I and injected into embryos at the two cell stage of development. Micro-injected embryos were allowed to proceed to stage 26 of development before being dissected into axis and head plus gut regions. Total nucleic acid was extracted from these embryo fractions and transcripts of the micro-injected gene detected by primer extension as previously described (see chapter 8.1).

Results illustrated in figure 9.2.2 demonstrate the previously described extension products present in primer extension reactions which contain nucleic acid extracts isolated from the axis region of embryos injected with the construct p Δ gb Δ 5(+). As previously observed, no transcription is apparent from p Δ gb Δ 6 or p Δ gbB3/B1(+) in either the axis or head plus gut regions of embryos injected with these plasmids. However, it is also apparent from these data that the plasmid p Δ gbCB-URE is also transcriptionally inactive when micro-injected into Xenopus embryos, as judged by the absence of any detectable extension products in primer extension reactions.

FIGURE 9.2.2. Primer extension analysis of embryos injected with the plasmids $p\Delta gb\Delta 5(+)$, $p\Delta gb\Delta 6$, $n\Delta zh\Delta 8+$, $pXbgb\Delta 6$, $p\Delta gbB3/B1(+)$ and $p\Delta gbCB-URE$.

Embryos were injected at the two cell stage of development with either $p\Delta gb\Delta 5(+)$ (tracks 1A and 1G), $p\Delta gb\Delta 6$ (tracks 2A and 2G), $p\Delta gb\Delta 8+$ (tracks 3A and 3G), $pXbgb\Delta 6$ (tracks 4A and 4G), $p\Delta gbB3/B1(+)$ (tracks 5A and 5G) or $p\Delta gbCB-URE$ (tracks 6A and 6G). Embryos were allowed to proceed to stage 26 of development before being dissected into axis (A), or head plus gut (G) regions. Total nucleic acid was extracted and the transcripts of the micro-injected gene detected by primer extension as described in chapter 8.1.

The size of DNA markers (M tracks) is indicated on the left hand side of the figure.



Results and Discussion

It would appear, therefore, that the addition of the URE to directly 5' of the CArG box motifs present in the plasmid p Δ gbB3/B1(+) failed to induce the tissue-specific expression of an otherwise silent fusion gene in Xenopus laevis embryos. This inability of the URE, in conjunction with the CArG box1 and CArG box3 motifs, to drive the expression of a heterologous promoter suggests that further sequences are required for the expression of the Xenopus borealis skeletal actin gene.

9.3 Discussion.

Experiments outlined in chapter 8 demonstrate that a region of the Xenopus borealis skeletal actin gene which spans from nucleotides -197 to -168 (URE) is essential for the expression of the gene. Furthermore, data presented in this chapter demonstrate that this region of the gene is capable of binding a trans-acting factor(s) in vitro. This binding activity appears to be ubiquitous, being reproducible in band shift assays using protein extracts prepared from both regions of the embryo which express the skeletal actin gene (somites), in addition to regions that express little, if any skeletal actin (gut and ectoderm). However, no direct evidence exists to demonstrate that the URF shifted band present in binding reactions containing somite protein extracts is caused by an identical protein(s) which facilitates the URF shift in reactions containing gut or ectoderm protein extracts. Indeed, the oligonucleotide used in band shift reactions described in this chapter is relatively large (35 nucleotides). Thus, the

Results and Discussion

possibility exists that two different binding sites are present in this region of the promoter. It is possible to speculate that a positive trans-acting factor present in the somite fraction of embryos is capable of interacting with one site of the URE. However, a distinct negative trans-acting factor, present in the gut and ectoderm fractions of the embryo, may interact with a different or overlapping site of the URE to the positive trans-acting factor in somite extracts. Until convincing DNA footprints of the URF binding activity are achieved using protein extracts prepared from both muscular and non-muscular origin, this hypothesis cannot be discounted. However, the identical nature of the URF shifted band when different extracts are compared would argue that the binding activity present in extracts of different origins is facilitated by the same protein(s).

When the URE is placed directly upstream of the CArG box3 and CArG box1 motifs present in the plasmid p Δ gbB3/B1(+), the combination of these sequences fails to drive the tissue-specific expression of a heterologous promoter (see section 9.2). A combination of the CArG box3, CArG box1 and URE sequence elements are therefore insufficient for the expression of the Xenopus borealis skeletal actin gene. Although fusion of the URE oligonucleotide into the Sac I site of the plasmid p Δ gbB3/B1(+) places this regulatory element in exactly the same position with respect to the CArG box3 motif as it is in the wild type promoter, five nucleotides which span between the URE and the CArG box3 motif are not reconstituted in this

Results and Discussion

plasmid. Thus, if a cis-acting element is centred around position -168 in the promoter of the Xenopus borealis skeletal actin gene, it is not reconstituted in the plasmid p Δ gbCB-URE. However, assuming that a cis-acting element is not centred around position -168 in the promoter of the Xenopus borealis skeletal actin gene, then it is apparent that other sequences in addition to CArG box3, CArG box1 and the URE are essential for the expression of the gene. This possibility will be discussed in further detail in chapter 10.

By drawing comparisons with other extensively characterised muscle-specific genes it is possible to speculate about the role of the URE in the expression of the skeletal actin gene. For example, the quail fast skeletal troponin I gene contains an internal regulatory element (IRE) in the first intron of this gene (Konieczny and Emerson, 1987; see also chapter 1.5). This IRE has been demonstrated to contain a MyoD binding site which is essential for the expression of the gene. However, in addition to this a further two sequence elements in the IRE, which interact with ubiquitous factors, are also essential for the expression of the troponin I gene (Lin et al., 1991). Similarly, the human cardiac actin gene requires a combination of three sequence elements for its effective expression. Two of these elements, namely a CArG box and SP-1 motif, bind the ubiquitous transcription factors SRF and SP-1 respectively. However a third cis-acting element, which binds the myogenic-specific transcription factor MyoD, is also required for the expression of the gene (Sartorelli et al., 1990). It would

Results and Discussion

appear therefore, that certain muscle-specific genes require a combination of both muscle-specific and ubiquitous transcription factors for their correct expression. In this regard, the ubiquitous nature of the URF, in addition to the CArG box binding factors (see chapter 7), makes it possible to speculate that the URE and CArG box, like the SP-1 and CArG box motifs in the human cardiac actin gene promoter, are sites for the interaction of ubiquitous transcription factors. If this is the case, then the other regulatory elements proposed to exist in the promoter of the Xenopus borealis skeletal actin gene may bind tissue-specific factors and confer muscle-specific expression to the skeletal actin gene.

Results and Discussion

CHAPTER 10

Sequences located between the CArG box1 motif and the TATA box of the *Xenopus borealis* skeletal actin gene interact with a trans-acting factor(s) in vitro.

Introduction.

The observation that the construct p Δ gbCB-URE is transcriptionally inactive when injected into *Xenopus laevis* embryos demonstrates that sequences in addition to the URE, CArG box1 and CArG box3 motifs are required for the expression of the *Xenopus borealis* skeletal actin gene (see chapter 9). Furthermore, the demonstration that a fragment of the skeletal actin promoter which spans from nucleotides -197 to -42 is capable of driving the tissue-specific expression of a heterologous promoter, indicates that all sequences necessary for the expression of the gene are contained within this 156 nucleotide fragment (see chapter 8). Two possible regions of the promoter are capable of housing these additional regulatory elements. The first spans from between the CArG box3 and CArG box1 motifs situated in the promoter of the gene. However, a chimeric fusion gene, which lacks sequences between the CArG box3 and CArG box1 motifs, is transcriptionally active when injected into *Xenopus laevis* embryos (see construct pXbgh Δ 8+, chapter 4.2). This discovery would argue against this region of the skeletal actin gene containing regulatory elements important for its expression. However, a second potential

Results and Discussion

regulatory region of the gene exists which spans from directly 3' of the CArG box1 motif (nucleotide -83) to 11 nucleotides upstream of the TATA box (nucleotide -42). Indeed, a region that spans between the CArG box1 motif and the TATA box of the chicken skeletal actin gene has been shown to be essential for the expression of this gene in myogenic cell lines (Bergsma et al., 1986). Furthermore, the equivalent region of the human cardiac actin gene has been shown to contain functionally important binding sites for the transcription factors SP-1 and MyoD (Sartorelli et al., 1990).

As a preliminary investigation into the possibility of additional regulatory elements being present in the -83/-42 region of the Xenopus borealis skeletal actin gene it was decided to investigate whether a trans-acting factor(s) was capable of interacting with these sequences in vitro. Experiments described in this chapter demonstrate the binding of a trans-acting factor(s) to the -83/-42 region of the Xenopus borealis skeletal actin gene in vitro. This finding, in addition to the possibility of other regulatory elements being present in this region of the gene will be discussed.

10.1 A trans-acting factor(s) binds the -83/-42 region of the Xenopus borealis skeletal actin gene in vitro.

To investigate whether a region of the Xenopus borealis skeletal actin gene which spans from nucleotides -83 to -42 is capable of interacting with trans-acting factors in vitro the oligonucleotide B1/Bal was synthesised. This 50 nucleotide

Results and Discussion

stretch of DNA contains sequences complementary to the -83/-42 region of the Xenopus borealis skeletal actin gene surrounded by Eco RI sticky ends (see figure 10.1.1).

B1/Bal was end labelled with $\gamma^{32}\text{P}$ -ATP and used in band shift assays containing protein extracts prepared from the axis region of stage 30 Xenopus laevis embryos.

Results illustrated in figure 10.1.2 demonstrate the presence of four binding activities when the oligonucleotide B1/Bal is analysed for DNA binding by band shift analysis using protein extracts prepared from the axis regions of stage 30 Xenopus laevis embryos. Competition for factor binding by a 20 or 40-fold molar excess of unlabelled B1/Bal extinguishes only one of these shifted bands. This competable binding activity has been designated as Skeletal Actin Promoter Factor 1 (SAPF1). Competition with a 20 and 40-fold molar excess of either CARG box1 or CARG box3 oligonucleotides (for sequences see chapter 6) does not significantly reduce the intensity of any of the retarded bands present in binding reactions. Thus, out of the four shifted bands present in these band shift assays only one, namely SAPF1, is generated by a sequence-specific DNA-binding protein(s). Indeed, the inability of either CARG box1 or CARG box3 oligonucleotides to compete for SAPF1 binding demonstrates that the loss of this band on competition with B1/Bal is due to sequence-specific competition, as opposed to a general effect of adding a 40-fold molar excess of oligonucleotide to binding reactions.

To investigate the distribution of the SAPF1 binding activity

FIGURE 10.1.1. Diagrammatic representation of the
oligonucleotide B1/Bal.

The Xenopus borealis skeletal actin gene promoter is illustrated showing the CArG box motifs (red boxes), TATA box (blue box) and first exon (black box). The -83/-42 region of the gene is expanded and represented as sequences contained within the oligonucleotide B1/Bal. Xenopus borealis sequences contained in B1/Bal are illustrated by a line above the oligonucleotide sequence.

X.borealis skeletal
actin promoter

~ 1405 ————

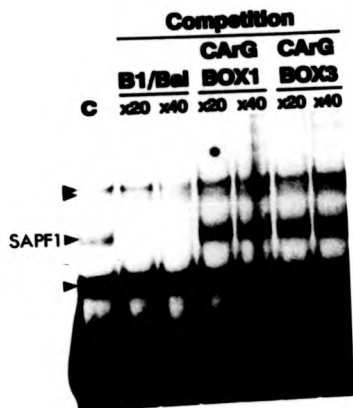


AATTGAGTTCACCGAGTGTGGGGCTTTCCCTGGGCAACACCAATGGCCA
 GTCAGTGGCTCACACCCCGAAGGGACCCGTTGTGTTTACCCGGTTTAA

Oligonucleotide B 1/Bal

FIGURE 10.1.2. The -83/-42 region of the *Xenopus borealis*
skeletal actin gene binds a trans-acting
factor(s) in vitro.

The oligonucleotide B1/Bal was used in band shift assays in conjunction with protein extracts prepared from the axis region of stage 30 *Xenopus laevis* embryos. Track C represents a binding reaction which contains no competitor. Competition for DNA binding by a 20 (X20) and 40-fold (X40) molar excess of either B1/Bal, CArG box1 or CArG box3 are illustrated in competition tracks.



Results and Discussion

throughout the embryo, band shift assays using the B1/Bal oligonucleotide were repeated using extracts prepared from either the axis or head plus gut regions of stage 30 Xenopus laevis embryos.

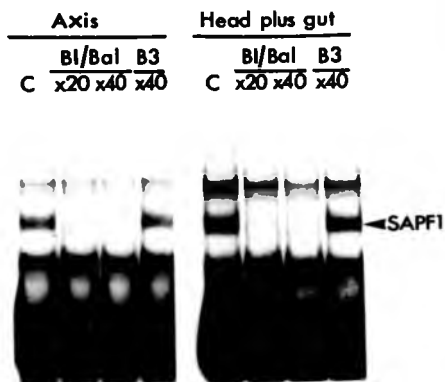
Results from this experiment (see figure 10.1.3.) demonstrate the pattern of shifted bands to be identical in binding reactions containing protein extracts prepared from either fraction of the embryo. It would appear, therefore, that the protein(s) responsible for the SAPF1 binding activity is present and capable of binding the B1/Bal sequence in extracts prepared from both regions of the embryo that express the skeletal actin gene (i.e. axis), in addition to those that do not (i.e. head plus gut).

10.2 Discussion.

Data presented in this chapter demonstrate the binding of a trans-acting factor(s) to a region of the Xenopus borealis skeletal actin gene which spans from between the CArG box motif and the TATA box of the gene (nucleotides -83 to -42). Experiments have not been performed which specifically investigate the role of these sequences in the expression of the Xenopus borealis skeletal actin gene. However, previously described micro-injection experiments make it possible to speculate that the -83/-42 region of the skeletal actin gene is required for its correct expression in Xenopus embryos (see introduction to this chapter). In this regard an equivalent region of the human cardiac actin gene has been demonstrated to

FIGURE 10.1.3. The SAPFl binding activity is present in both axis and head plus gut protein extracts.

Labelled B1/Bal was used in band shift assays in conjunction with protein extracts prepared from either the axis or head plus gut regions of stage 30 Xenopus laevis embryos. Binding reactions containing no competitor are illustrated (tracks C). Competition for DNA binding by a 20 (X20) and 40-fold (X40) molar excess of either B1/Bal or CArG box3 (B3) is illustrated in competition tracks.



Results and Discussion

possess two essential cis-acting regulatory elements. One of these elements binds the ubiquitous transcription factor SP-1, whilst the second interacts with the muscle-specific transcription factor MyoD (Sartorelli et al., 1990). More interestingly however, the chicken skeletal actin gene has also been demonstrated to possess regulatory sequences which are present between the CArG box1 and TATA box motifs of this gene (Bergsma et al., 1986). Inspection of this regulatory region of the chicken skeletal actin gene reveals the presence of a putative SP-1 binding site (5'-GGGCGG-3'). However, Bergsma et al. (1986) did not perform any experiments which investigated the role of this specific DNA sequence in the expression of the gene.

No SP-1 binding site is apparent in the -83/-42 region of the Xenopus borealis skeletal actin gene. Indeed, no significant sequence identities appear to exist when this region of the Xenopus borealis skeletal actin gene is compared with the equivalent region of the chicken skeletal actin gene.

Although the identity of the SAPF1 binding activity is unknown, the factor(s) which facilitates this DNA shift appears to be present in extracts prepared from a region of the embryo that does not express the skeletal actin gene, in addition to a region that does (see figure 10.1.3.). If the apparently ubiquitous SAPF1 binding activity is responsible for conferring tissue-specific expression onto the skeletal actin gene, then the mechanisms by which the expression of this gene is controlled are likely to be more subtle than originally

Results and Discussion

thought. The possible mechanisms by which a ubiquitous transcription factor is capable of governing the tissue-specific expression of a gene have been outlined previously (see chapter 7).

However, the possibility exists that other regulatory sequences are present in the -83/-42 region of the Xenopus borealis skeletal actin gene which have not been detected in the experiments outlined in this chapter. In this regard, a 6 nucleotide sequence (5'-CAAATG-3') in the -83/-42 region of the skeletal actin promoter (nucleotide positions -43 to -48), bears a striking identity to the proposed consensus sequence for the transcription factor MyoD (CANNTG) (see Murre et al., 1989a). If this six nucleotide sequence is indeed an authentic MyoD binding site then it is possible to speculate that this factor may act in conjunction with the ubiquitous URF, SAPFl and CArG box binding factors to drive the tissue-specific expression of the skeletal actin gene. Indeed, a combination of MyoD and ubiquitous transcription factors has been proposed to control the expression of other muscle-specific genes, such as the human cardiac (Sartorelli et al., 1990), Xenopus cardiac (T. Mohun; abstract, 3rd international Xenopus meeting, 1990) and skeletal troponin I (Nikovits et al., 1990; Lin et al., 1991) genes.

To investigate the possible role of the -43/-48 putative MyoD site in the expression of the Xenopus borealis skeletal actin gene, band shift assays were performed using an oligonucleotide containing sequences complementary to the putative skeletal

Results and Discussion

actin MyoD site. These experiments were performed using protein extracts prepared from either muscular or non-muscular regions of the embryo. No convincing tissue-specific binding activities were apparent in these experiments (data not shown). However, the validity of this observation is questionable due to the finding that a bona fide MyoD site, which is present in the MCK enhancer (see Buskin and Hauschka, 1989), also failed to exhibit any tissue-specific shift in these experiments. Thus, although it is possible to speculate as to the importance of this putative MyoD site in the expression of the Xenopus borealis skeletal actin gene, it is not yet clear whether this DNA sequence is capable of binding MyoD or related factors.

Results and Discussion

CHAPTER 11

General discussion.

Experiments outlined in this report investigate the *cis*-acting sequences, and the *trans*-acting factors with which they interact, that are required for the tissue-specific expression of the Xenopus borealis skeletal actin gene. Experiments performed in this laboratory have demonstrated that like all other sarcomeric actin genes studied to date, so the 3' untranslated region and intragenic sequences of the Xenopus borealis skeletal actin gene are not required for its correct expression (Boardman *et al.*, in preparation). Indeed, by exploiting the capability of a fragment of the Xenopus borealis skeletal actin gene promoter to drive the tissue-specific expression of a heterologous promoter, all sequences which are required for the correct expression of the gene have been demonstrated to be contained within a fragment of the skeletal actin gene promoter that spans from nucleotides -197 to -42 (see chapter 8).

Sequence comparison of the Xenopus borealis skeletal actin gene promoter with the regulatory regions of other actin genes reveals the presence of three conserved CArG sequence motifs in the -197/-42 region of the promoter. Indeed, deletion analysis of the actin gene promoter reveals that the most distal of these CArG motifs (CArG box3) is essential for the expression of the gene in stage 26 Xenopus laevis embryos. Furthermore, a region of the promoter which contains the CArG box2 sequence

Results and Discussion

motif (nucleotides -139 to -99) appears to be redundant or irrelevant in the expression of the gene in stage 26 Xenopus laevis embryos. In this regard a promoter which lacks sequences between nucleotide positions -139 and -99, but retains the CArG box3 motif and sequences that span between nucleotides -99 and +28 of the gene, is active when injected into developing Xenopus laevis embryos (see chapter 4.2).

No experiments have been performed that directly address the importance of the CArG box1 motif in the expression of the Xenopus borealis skeletal actin gene. However, the equivalent motif present in the promoters of other skeletal actin and cardiac actin genes has been demonstrated to be essential for their expression (Minty and Kedes, 1986; Miwa and Kedes, 1987; Mohun et al., 1989a; Walsh and Schimmel, 1988; Chow and Schwartz, 1990). Thus, if a similar mechanism of control of the sarcomeric actin genes exists between species, then it is possible to suggest that the CArG box1 motif of the Xenopus borealis skeletal actin gene is important in the expression of this gene.

The importance of the CArG box3 motif, and possibly the CArG box1 motif, in the expression of the Xenopus borealis skeletal actin gene is reflected in the ability of these sequences to interact with sequence-specific trans-acting factors in vitro (see chapters 5, 6 and 7). However, the CArG box2 sequence motif does not appear to bind any trans-acting factors, as detected by the assay system employed in this study.

CArG box1 is capable of forming two DNA-protein complexes in

Results and Discussion

vitro. Furthermore, by exploiting the antibody band shift technique, one of these DNA-protein complexes has been demonstrated to be facilitated by a protein that is antigenically related to the transcription factor SRF (see chapter 6). This is in agreement with work performed on the Xenopus cardiac (Taylor et al., 1989), human cardiac (Boxer et al., 1989b) and human skeletal (Boxer et al., 1989b) actin genes, which also demonstrate the binding of SRF to their respective CARG box1 motifs. Indeed, the Xenopus laevis cardiac actin CARG box1 motif has been demonstrated to be functionally interchangeable with the CARG motif present in the serum response element of the c-fos gene promoter (Taylor et al., 1989).

The identity of the second CARG box1 DNA binding protein(s) is as yet unknown, although this protein(s) appears to be antigenically distinguishable from SRF.

CARG box3 exhibits a similar shift to the SRF/CARG box1 binding activity when subjected to band shift analysis. However, competition analysis and antibody band shift analysis of this binding activity demonstrates that the protein(s) that facilitates this shift is distinct from SRF. However, whether the CARG box3 binding protein (CBA1) is completely different from SRF, or whether they are related proteins remains unclear. Indeed, it is feasible that the CBA1 and SRF proteins are differentially spliced products that arise from the same gene, although the apparent lack of any detectable epitopes between these two proteins would argue against this.

Results and Discussion

Experiments described in chapter 8 demonstrate that a combination of the CArG box1 and CArG box3 sequence motifs are unable to drive the tissue-specific expression of a heterologous promoter. Thus, other sequences in addition to the CArG boxes are required for the expression of the Xenopus borealis skeletal actin gene. In this regard it is becoming increasingly clear that a number of muscle-specific genes require multiple cis-acting regulatory elements for their effective expression (for examples see Sartorelli et al., 1990; Chow and Schwartz, 1990; Lin et al., 1991; Horlick and Benfield, 1989; Bouvagnet et al., 1987). Indeed, by investigating the capability of various fragments of the Xenopus borealis skeletal actin promoter to drive the expression of a heterologous promoter, an upstream regulatory element (URE) in the skeletal actin promoter (nucleotides -197 to -168) has been identified (see chapter 8). However, the URE only appears to be essential for the expression of the gene when sequences that span between nucleotides -41 and +28 are absent from chimeric fusion gene plasmids.

The URE is capable of interacting with a trans-acting factor(s) in vitro and this has been named the upstream regulatory factor (URF). However, a combination of CArG box1, CArG box3 and URE sequences is unable to confer muscle-specific expression onto a heterologous promoter (see chapter 9). It was therefore reasoned that other regulatory elements that are required for the expression of the skeletal actin gene exist in a region of the promoter that spans from nucleotides -83 to

Results and Discussion

-42. Indeed, when an oligonucleotide that corresponds to these sequences is subjected to band shift analysis the sequence-specific binding of a protein(s) is apparent (see chapter 10). However, the specific sequences in the -83/-42 region of the promoter with which this protein(s) interacts are unknown.

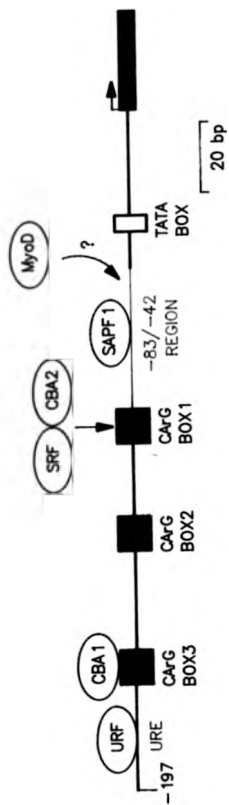
It is apparent that, as in the case of a number of other muscle-specific genes, so the Xenopus borealis skeletal actin gene is controlled by a number of positive cis-acting regulatory elements. Three regulatory regions of the Xenopus borealis skeletal actin gene have been identified, namely the CArG box1, CArG box3 and URE sequences. In addition to this a possible fourth regulatory element exists in the -83/-42 region of the promoter. A diagrammatic representation of these regulatory elements, and the proteins with which they interact, is illustrated in figure 11.1.

All the binding activities described in this report (i.e. CBA1, CBA2, SRF, URF and SAPF1) appear to be distributed throughout the embryo. The question arises then as to how these apparently ubiquitous binding activities are capable of directing the expression of the Xenopus borealis skeletal actin gene exclusively in the muscle tissue of developing embryos? If no other tissue-specific trans acting-factors bind the skeletal actin promoter then a number of mechanisms exist by which one or more apparently ubiquitous transcription factors can direct the tissue-specific expression of a gene, and these have been described in chapter 7. However, by drawing comparisons with the proposed mechanisms of control of other muscle-specific

FIGURE 11.1. Diagrammatic representation of the cis-acting regulatory sequences, and the trans-acting factors with which they interact, that are important for the expression of the *Xenopus borealis* skeletal actin gene.

197 nucleotides upstream of the transcriptional start site of the *Xenopus borealis* skeletal actin gene are illustrated. The URE (blue), CArG box (red) and -83/-42 (green) regulatory regions of the promoter are shown. The TATA box is also indicated, as is the first exon of the gene which is illustrated as a black box.

The sites of interaction of SRF, CBA2, CBA1, URF, and SAPF1 with the skeletal actin promoter are shown. In addition to this, the possible site of interaction of MyoD with the promoter is illustrated.



Results and Discussion

genes, it is possible to suggest a further mechanism of control of the Xenopus borealis skeletal actin gene.

The extensive characterisation of other muscle-specific genes has uncovered a number of cis-acting regulatory elements that are important for their tissue-specific expression (see chapter 1.4). However, it was not until recently that a common sequence motif, which interacts with the MyoD family of regulatory proteins, was identified to be important in the expression of a number of muscle-specific genes (see chapter 1.5). However, it was also apparent from these studies that for effective expression of the gene, a number of ubiquitous transcription factors had to act in parallel with MyoD on the gene promoter. (for examples see Sartorelli et al., 1990; Lin et al., 1991).

It is interesting that a putative MyoD binding site is situated in the -83/-42 region of the Xenopus borealis skeletal actin gene promoter. It is tempting to speculate, therefore, that the previously described ubiquitous proteins that bind the skeletal actin promoter are required to drive the expression of this gene and that tissue-specificity is conferred upon this gene by the putative MyoD site.

Preliminary investigation of the putative MyoD site has been unable to detect any sequence-specific binding of proteins to this sequence. However, under our assay conditions, no sequence-specific binding to a bona fide MyoD site which is present in the MCK enhancer was observed. Thus, the importance of MyoD in the expression of the Xenopus borealis skeletal actin gene remains unclear and further experiments must be

Results and Discussion

performed to ascertain the importance of this putative MyoD binding site in the expression of the gene.

APPENDIX 1

Appendix

A P P E N D I X 1

The expression of a human cardiac actin gene in developing *Xenopus laevis* embryos.

Introduction.

Data presented in this report and data from other laboratories demonstrate the potential of the micro-injection of genes into *Xenopus* embryos as an efficient assay system for their expression. Indeed, it has been argued that the introduction of genes into developing *Xenopus* embryos exposes them to a more accurate reconstruction of developmental regulatory events than other assay systems, such as the transfection of genes into specific cell lines (see chapter 4, introduction).

Data presented in this section investigate the potential of exploiting *Xenopus* micro-injection techniques to examine the regulation of the human cardiac actin gene. To achieve this, a human cardiac actin gene promoter, fused to a CAT reporter gene, was micro-injected into developing *Xenopus laevis* embryos and the spatial and temporal expression of this gene examined. In addition to this, data are presented which examine the important cis-acting regulatory sequences involved in the expression of the human cardiac actin gene in developing *Xenopus laevis* embryos.

Appendix

A.1.1. The spatial expression of a human cardiac actin gene in stage 30 *Xenopus laevis* embryos.

To examine the spatial regulation of a human cardiac actin gene in developing *Xenopus laevis* embryos the plasmid pHCA485 was obtained from Dr. L. Kedes (see Minty and Kedes, 1986). This plasmid contains 485 nucleotides of human cardiac actin upstream sequence, in addition to the first exon and 24 nucleotides of the first intron of the gene, fused to the bacterial chloramphenicol acetyltransferase (CAT) reporter gene (see figure A.1.1.1.).

The plasmid pHCA485 was linearised with Pat I and injected into *Xenopus laevis* embryos at the two cell stage of development. Embryos were allowed to proceed until stage 30 before being dissected into axis, head and gut regions. CAT extracts were prepared from both dissected and whole embryos as described in chapter 3.14.1 and the CAT activity of these extracts determined as described in chapter 3.14.2.

It is apparent from data illustrated in figure A.1.1.2 that a region of the human cardiac actin promoter that spans from nucleotides -485 to +68 in the gene is capable of driving the expression of a CAT gene in whole *Xenopus laevis* embryos (see track W). Furthermore, dissection of embryos reveals that this CAT activity is restricted almost exclusively to the axis region of embryos (compare track A with tracks H and G). Thus it appears that 485 nucleotides of human cardiac actin gene upstream sequence, in addition to 68 nucleotides downstream of the transcriptional start site of the gene, is capable of

FIGURE A.1.1.1. Diagrammatic representation of the plasmid
pHCA485.

In the lower half of the figure the plasmid pHCA485 is shown illustrating the human cardiac actin (red box), CAT (blue box) and SV40 splice and polyadenylation signal (black box) sequences.

In the top part of the figure the human cardiac actin promoter region is expanded to illustrate the point at which the gene is fused to the CAT reporter gene (nucleotide +68). The first exon of the gene is illustrated by a blue box, whilst the four CArG boxes in the promoter of the gene are illustrated by red boxes.

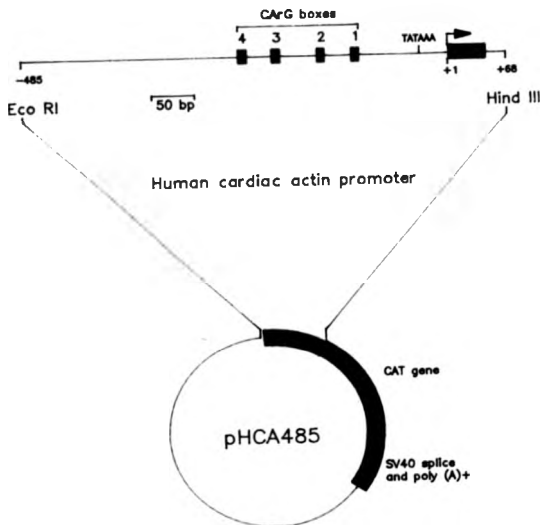


FIGURE A.1.1.2. The human cardiac actin gene promoter is capable of driving the tissue-specific expression of a CAT reporter gene in the axis region of *Xenopus laevis* embryos.

Embryos were injected with the plasmid pHCA485 and allowed to proceed to stage 30 of development prior to being dissected into axis, head and gut regions. CAT extracts were prepared from the axis (A), gut (G) and head (H) fractions of dissected embryos, in addition to whole embryos (W). CAT assays were performed as described in chapter 3.14.2.



A G H W

pHCA485

Appendix

directing the correct spatial expression of a CAT reporter gene in stage 30 Xenopus laevis embryos.

To establish whether the transcripts originating from the human cardiac actin fusion gene are correctly initiated a plasmid which contained the human cardiac actin gene promoter (nucleotides -485 to +68) fused to a mouse globin reporter gene was constructed. The construction of this plasmid was necessary because CAT mRNA transcripts were not detectable in our assay system. To make this construct the plasmid pAW103 (a generous gift from Dr. R. W. Old) was exploited. This plasmid contains a fragment of the histone 1 and 3 genes fused to the 2nd intron and 3rd exon of the mouse β -globin gene contained in the vector pAT153.

The plasmid pAW103 was digested at a Bam HI site situated at the histone/vector fusion point of the plasmid and the site infilled using the klenow fragment of DNA polymerase. The resulting DNA fragment was then cut at a Hind III site situated in the 2nd intron of the mouse β -globin gene. The terminal phosphates of the resulting fragments were removed by calf intestinal alkaline phosphatase to prevent the reconstitution of pAW103 in ligation reactions. Thus, a blunt/Hind III fragment of pAW103, which contains vector sequences in addition to the 3rd exon and part of the 2nd intron of the mouse β -globin gene, was created.

The plasmid pHCA485 was digested with Eco RI and the site infilled prior to the resulting restriction fragment being digested by Hind III. This results in the excision of a

Appendix

blunt/Hind III fragment from pHCA485 which contains 485 nucleotides of human cardiac actin gene upstream sequence in addition to 68 nucleotides downstream of the transcriptional start site of the gene (see figure A.1.1.3). This fragment was isolated after separation from vector sequences by agarose gel electrophoresis and inserted into the blunt/Hind III sites of pAW103.

Thus, the construct pRE485 was created. This plasmid contains sequences extending from nucleotide -485 to +68 of the human cardiac actin gene fused to the second intron of the mouse β -globin gene (see figure A.1.1.3).

The construct pRE485 was linearised with Pst I and injected into Xenopus laevis embryos at the two cell stage of development. Embryos were allowed to develop until stage 30 before being dissected into axis, head and gut regions. Total nucleic acid was extracted from both dissected and whole embryos and the transcripts of the micro-injected gene detected by primer extension (see chapter 3.13). The oligonucleotide employed in primer extension reactions described in this chapter (MG-1) is complementary to sequences in the 3rd exon of the mouse globin gene. Thus, correctly initiated transcripts originating from the micro-injected gene result in a 67 nucleotide extension product in primer extension assays employing the MG-1 oligonucleotide.

Data presented in figure A.1.1.4. illustrate that a 67 nucleotide extended product is apparent in primer extension reactions containing nucleic acid extracted from whole embryos

FIGURE A.1.1.3. Diagrammatic representation of the plasmid
pRE485.

In the lower part of the figure the plasmid pRE485 is shown illustrating the human cardiac actin (red) and mouse β -globin (black) sequences. Sequences of the vector (pAT153) are illustrated in green.

In the upper part of the figure the human cardiac actin promoter region is expanded to illustrate the point at which the gene is fused to the 2nd intron of the mouse β -globin gene (nucleotide +68). The first exon of the gene is illustrated by a blue box, whilst the four CArG boxes in the promoter of the gene are illustrated by red boxes.

The cloning strategy employed in constructing this plasmid results in the loss of the Eco RI site at the 5' end of the human cardiac actin promoter.

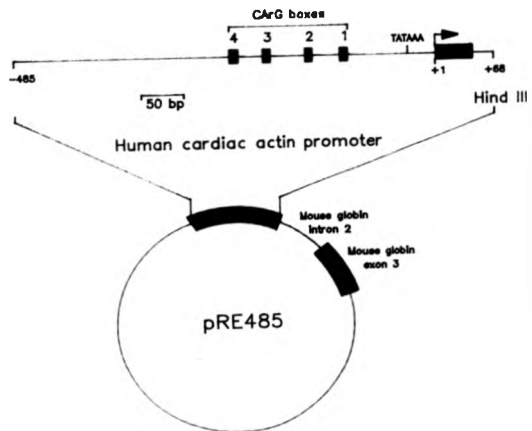
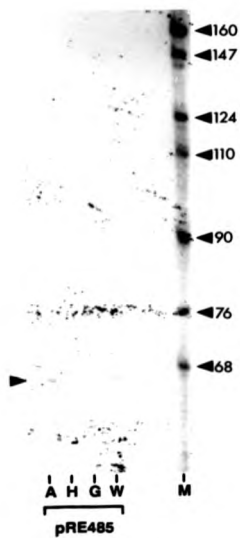


FIGURE A.1.1.4. A 485bp fragment of the human cardiac actin gene promoter produces correctly initiated transcripts in the axis region of *Xenopus laevis* embryos.

Embryos were injected with pRE485 at the two cell stage of development. Embryos were allowed to proceed until stage 30 before being dissected into axis, head and gut regions. Nucleic acid was extracted from whole embryos (W), in addition to the axis (A), head (H) and gut (G) regions of dissected embryos. Transcripts of the micro-injected gene were detected by primer extension using the oligonucleotide MG-1.

The size of DNA markers (track M) is indicated on the right hand side of the figure. The 67 nucleotide extension product which is produced on the correct initiation of transcripts originating from the human cardiac actin gene is illustrated by an arrow on the left hand side of the figure.



Appendix

injected with the plasmid pHCA485 (see track W). Indeed, analysis of nucleic acid extracted from the axis, head and gut regions of embryos injected with this construct demonstrate the extended product to be present exclusively in the region of the embryo where the Xenopus laevis cardiac actin gene is expressed (i.e. the axis). Furthermore, no other bands in addition to the 67 nucleotide extension product are apparent in these reactions, demonstrating the correct initiation of transcripts.

It is possible to conclude from these experiments that a region of the human cardiac actin gene which extends from nucleotides -485 to +68 is capable of directing the correct spatial expression of a reporter gene in stage 30 Xenopus laevis embryos.

A.1.2. The temporal expression of the human cardiac actin gene in developing Xenopus laevis embryos.

It is apparent from experiments described in section A.1.1 that 485 nucleotides of human cardiac actin gene upstream sequence, in addition to 68 nucleotides downstream of the transcriptional start site of the gene, is sufficient for the correct spatial expression of the gene in stage 30 Xenopus laevis embryos. To investigate whether the human cardiac actin gene is also capable of being expressed in a correct temporal manner when injected into developing Xenopus laevis embryos, it was decided to study the expression of the pHCA485 plasmid throughout the development of Xenopus.

To achieve this pHCA485 was linearised with Pst I and injected

Appendix

into embryos at the two cell stage of development. In addition to this embryos were also injected with the plasmid pSV2-CAT, which contains the CAT gene under the control of a SV40 promoter.

Embryos were allowed to proceed to stage 9 of development before being dissected into animal, vegetal and equatorial regions. When these embryo explants are allowed to develop in isolation the animal, vegetal and equatorial regions develop into ectoderm, endoderm and mesoderm respectively. Thus, it is only the equatorial explant of the embryo that expresses the endogenous cardiac actin gene.

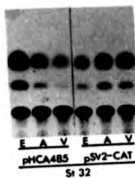
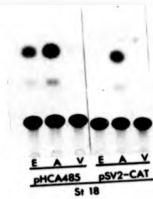
Animal, equatorial and vegetal explants of embryos were allowed to develop in isolation. CAT extracts were prepared from a number of embryo explants when control embryos reached stages 10, 13½, 18, 26 and 32 of development. The CAT activity of these extracts was determined as described in chapter 3.14.1. The ratio of vegetal, animal and equatorial extracts used in CAT assays was 3:2:1 respectively. This was necessary to account for the difference in cell density of the different embryo explants.

Data presented in figure A.1.2.1 illustrates that CAT activity is present in both the animal and equatorial explants of stage 10 embryos injected with the plasmid pSV2-CAT. However, no CAT activity is apparent in the vegetal explants of these embryos. This can be attributed to the lack of exogenous DNA in these explants, as Southern blot analysis of extracts failed to detect any exogenous DNA. However, exogenous DNA was detected

FIGURE A.1.2.1. The temporal expression of a human cardiac actin gene in *Xenopus laevis* embryos.

The plasmids pHCA485 or pSV2-CAT were injected into embryos at the two cell stage of development and embryos allowed to develop to stage 9 before being dissected into animal, vegetal and equatorial regions. CAT extracts were prepared from animal (A), vegetal (V) and equatorial (E) explants when control embryos had reached stages 10, 13½, 18, 26 and 32 of development.

CAT assays using extracts prepared from embryos at different stages of development were performed as described in chapter 3.14.2.



Appendix

at all stages of development examined in the animal and equatorial explants of embryos injected with either the pHCA485 or pSV2-CAT (data not shown).

In contrast to embryos injected with pSV2-CAT no CAT activity is apparent in either the animal, equatorial or vegetal explants of stage 10 embryos injected with pHCA485. Thus, it would appear that the human cardiac actin gene is transcriptionally inactive at stage 10 of Xenopus development. This is not unexpected as the Xenopus cardiac actin gene is not activated until stage 12½ of Xenopus development (Mohun et al., 1984; Wilson et al., 1986). Indeed, it is apparent from figure A.1.2.1 that the human cardiac actin gene is activated at approximately the same stage of Xenopus laevis development as the endogenous cardiac actin gene, as judged by CAT activity being present in embryos at stage 13½ of development when injected with pHCA485. However, this activation of the human cardiac actin gene does not appear to be tissue-specific, as CAT activities are predominant in the animal explants of embryos injected with pHCA485. It is not clear whether the pHCA485 plasmid is expressed in the vegetal explants of stage 13½ embryos, as Southern blot analysis revealed the lack of exogenous DNA in these embryo fractions (data not shown).

The lack of spatial regulation of the human cardiac actin gene in stage 13½ embryo explants is also apparent in explants which are allowed to develop until control embryos have reached stage 18 of development. However, as development proceeds the human cardiac actin gene appears to be expressed in an increasingly

Appendix

tissue-specific manner (compare the animal and equatorial regions of stage 18 embryos with the equivalent explants of stage 26 and 32 embryos). Indeed, CAT activity appears to be restricted to predominantly equatorial explants by the time control embryos have reached stage 32 of development. The observed localisation of CAT activity to the equatorial explants of stage 32 embryos does not appear to be caused by the differential replication of DNA in these samples, as Southern blot analysis revealed that approximately equivalent amounts of exogenous DNA are present in the equatorial and animal explants of embryos (data not shown).

It would appear that 485 nucleotides of human cardiac actin gene upstream sequence, in addition to 68 nucleotides downstream of the transcriptional start site of the gene, is capable of activating the expression of the human cardiac actin gene at approximately the same stage of Xenopus development as the endogenous cardiac actin gene. However, it is also apparent from this experiment that it is not until the latter stages of Xenopus development that the human cardiac actin gene is expressed in a strict spatial manner.

A.1.3. The analysis of sequence elements important in the expression of the human cardiac actin gene in Xenopus laevis embryos.

In order to characterise sequences important for the expression of the human cardiac actin gene in Xenopus laevis embryos a series of plasmids containing 5' deletions of the

Appendix

human cardiac actin gene promoter were obtained from Dr. L. Kedes (see Minty and Kedes, 1986; Miwa and Kedes, 1987). These constructs are identical to the plasmid pHCA485, with the exception that they lack progressively more of the promoter region of the human cardiac actin gene (see figure A.1.3.1).

The plasmids pHCA485, pHCA177, pHCA153, pHCA131, pHCA96, pHCA64 and pSV2-CAT were linearised with Pst I and injected into embryos at the two cell stage of development. Embryos were allowed to proceed to stage 30 of development before being dissected into axis, head and gut regions. CAT extracts were prepared from whole and dissected embryos and the CAT activity of these extracts determined as described in chapter 3.14.2.

Data illustrated in figure A.1.3.2A displays typical results on the injection of pHCA485 and pHCA177 into Xenopus laevis embryos. As previously observed, the injection of the plasmid pHCA485 into Xenopus laevis embryos results in CAT activity being localised to the axis region of embryos. However, deletion of the human cardiac actin gene promoter to nucleotide -177 (plasmid pHCA177) results in a substantial reduction in the promoter activity of the fusion gene when it is injected into Xenopus laevis embryos. This is in accordance with data described by Minty and Kedes (1986), which identified a distal regulatory element in the -485/-178 region of the human cardiac actin gene promoter that is essential for the full activity of the gene in myogenic cell lines.

Deletion of the human cardiac actin gene promoter to nucleotide -153 (plasmid pHCA153) results in a further loss of

FIGURE A.1.3.1. Diagrammatic representation of the human cardiac actin promoter deletion plasmids pHCA485, pHCA171, pHCA153, pHCA131, pHCA96 and pHCA64.

In the lower half of the figure the plasmid pHCA485 is shown illustrating the human cardiac actin (red box), CAT (blue box) and SV40 splice and polyadenylation signal (black box) sequences. Sequences of the vector (pAT153) are illustrated in green.

In the upper half of the figure the human cardiac actin promoter region is expanded to illustrate the point at which the gene is fused to the CAT reporter gene (nucleotide +68). The first exon of the gene is illustrated by a blue box, whilst the four CArG boxes in the promoter of the gene are illustrated by red boxes.

The positions to which the promoter is deleted in the plasmids pHCA171, pHCA153, pHCA131, pHCA96 and pHCA64 are illustrated along the expanded human cardiac actin promoter. The plasmids pHCA153, pHCA131, pHCA96 and pHCA64 have a Bgl II site situated at the 5' end of the human cardiac actin gene instead of an Eco RI site which is present in the same position of the plasmids pHCA485 and pHCA177.

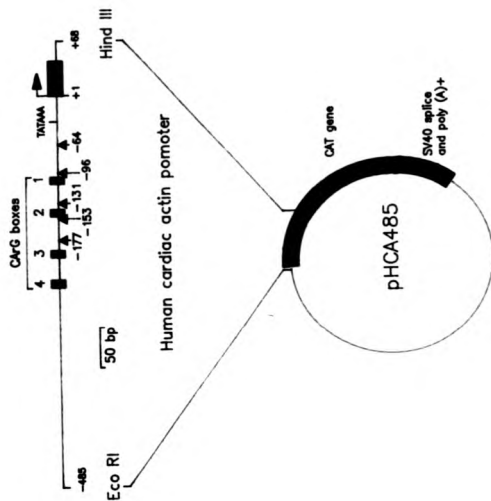
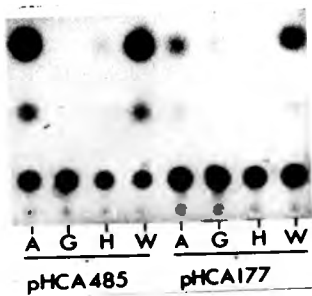


FIGURE A.1.3.2. The promoter activities of the plasmids
pHCA485, pHCA177 and pHCA153 when injected
into Xenopus laevis embryos.

A. The plasmids pHCA485 and pHCA177 were injected into embryos at the two cell stage of development. Embryos were allowed to proceed to stage 30 of development prior to being dissected into axis, head and gut regions. CAT extracts were prepared from whole embryos (W), in addition to the axis (A), head (H) and gut (G) regions of dissected embryos. CAT assays were performed as described in chapter 3.14.2.

B. The plasmids pHCA485, pHCA153 and pSV2-CAT were injected into embryos at the two cell stage of development. Embryos were allowed to proceed to stage 30 of development prior to being dissected into axis, head and gut regions. CAT extracts were prepared from the axis (A), head (H) and gut (G) regions of dissected embryos. CAT assays were performed as described in chapter 3.14.2.

A.



B.



Appendix

activity of the fusion gene when injected into Xenopus laevis embryos (see figure A.1.3.2B). Furthermore, subsequent deletion of the promoter to nucleotide -131 results in the complete loss of any detectable activity of the resulting fusion gene in Xenopus laevis embryos (data not shown). Thus, the removal of sequences spanning from between nucleotide -153 and -131 in the promoter of the human cardiac actin gene results in the loss of any detectable activity of the human cardiac actin fusion gene in stage 30 Xenopus laevis embryos. Indeed, this region of the human cardiac actin promoter contains the CARG box2 motif, which has previously been implicated in the expression of the gene (Miwa and Kedes, 1987).

To investigate whether transcripts originating from the plasmids pHCA485, pHCA177, pHCA153, pHCA131, pHCA96 and pHCA64 are correctly initiated in Xenopus laevis embryos the human cardiac actin gene promoter fragments of these plasmids were fused to the 2nd intron and 3rd exon of the mouse β -globin gene contained in the plasmid pAW103 (see section A.1.1.).

To achieve this pHCA177 was digested with Eco RI and the site blunted by in-filling, prior to digestion with Hind III. The resulting blunt/Hind III fragment of the human cardiac actin gene promoter (nucleotides -177 to +68) was isolated after separation from vector sequences by agarose gel electrophoresis.

The plasmid pAW103 was digested with Bam HI and the site in-filled prior to the fragment being digested with Hind III. The terminal phosphates of the resulting fragments were removed by

Appendix

calf intestinal alkaline phosphatase to avoid the reconstitution of vector sequences in ligation reactions. The blunt/Hind III fragment of the human cardiac actin gene promoter was then inserted into the blunt/Hind III fragment of pAW103 to create the plasmid pRE177 (see figure A.1.3.3).

The Eco RI restriction site at the 5' end of the human cardiac actin gene promoter present in the plasmids pHCA485 and pHCA177 is replaced by a Bgl II restriction site in the plasmids pHCA153, pHCA131, pHCA96 and pHCA64 (see Miwa and Kedes, 1987). Thus, a slightly different cloning strategy was employed to create plasmids containing these respective promoter fragments fused to the mouse β -globin sequences contained within the plasmid pAW103.

The plasmids pHCA153, pHCA131, pHCA96 and pHCA64 were digested with Bgl II and Hind III, resulting in the excision of the human cardiac promoter sequences from these plasmids. These promoter fragments were then isolated after separation from vector sequences by agarose gel electrophoresis. The plasmid pAW103 was digested with Bam HI and Hind III and the terminal phosphates removed from the resulting fragments by calf intestinal alkaline phosphatase.

The respective human cardiac actin gene promoter fragments isolated from the plasmids pHCA153, pHCA131, pHCA96 and pHCA64 were then inserted into the Bam HI/Hind III site of pAW103 to create the plasmids pRE153, pRE131, pRE96 and pRE64.

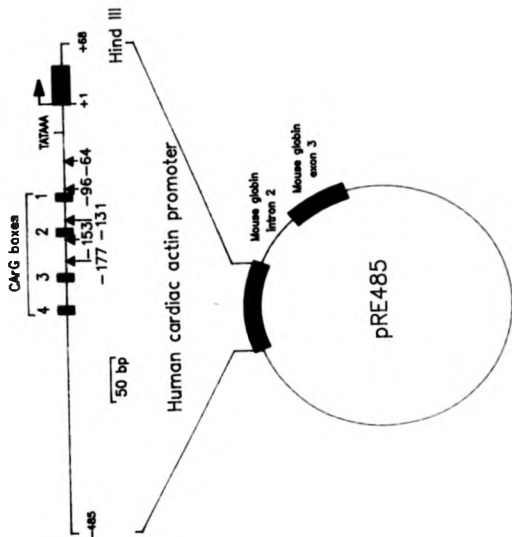
Thus the plasmids pRE485, pRE177, pRE153, pRE131, pRE96 and pRE64 were created. These constructs contain progressively less

FIGURE A.1.3.3. Diagrammatic representation of the human cardiac actin promoter deletion plasmids pRE485, pRE171, pRE153, pRE131, pRE96 and pRE64.

In the lower part of the figure the plasmid pRE485 is shown illustrating the human cardiac actin (red) and mouse β -globin (black) sequences. Sequences of the vector (pAT153) are illustrated in green.

In the upper part of the figure the human cardiac actin promoter region is expanded to illustrate the point at which the gene is fused to the mouse β -globin reporter gene (nucleotide +68). The first exon of the gene is illustrated by a blue box, whilst the four CArG boxes in the promoter of the gene are illustrated by red boxes.

The positions to which the promoter is deleted in the plasmids pRE171, pRE153, pRE131, pRE96 and pRE64 are illustrated along the expanded human cardiac actin promoter. The Bgl II site in the plasmids pRE153, pRE131, pRE96 and pRE64 was lost during the construction of these plasmids due to the ligating of a Bgl II site to a Bam HI site. In addition to this, the Eco RI site in the plasmids pRE485 and pRE177 was also lost on the creation of these plasmids due to the in-filling of the site prior to being ligated to a blunt Bam HI site.



Appendix

of the human cardiac actin gene promoter fused to the 2nd intron of the mouse globin reporter gene at nucleotide +68 in the cardiac actin gene (see figure A.1.3.3.). The number in the plasmid nomenclature is representative of the number of nucleotides upstream of the transcriptional start site of the human cardiac actin gene present in these plasmids.

The plasmids pRE485, pRE177, pRE153, pRE131, pRE96 and pRE64 were linearised with Pst I and injected into developing Xenopus laevis embryos at the two cell stage of development. Embryos were allowed to proceed to stage 30 before being dissected into axis, head and gut regions. Nucleic acid was extracted from whole and dissected embryos and transcripts originating from the micro-injected gene detected by primer extension as described in section A.1.1.

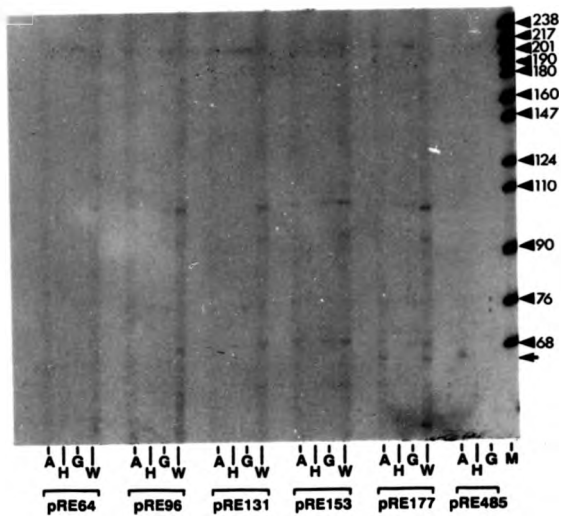
Figure A.1.3.4 illustrates the previously described 67 nucleotide extension product present in the axis region of embryos injected with pRE485. Indeed, a 67 nucleotide extension product is also present exclusively in the axis region of embryos injected with pRE177. Although the intensity of this band appears to be the same as the equivalent band in pRE485 reactions, only half as much nucleic acid extract was added to pRE485 reactions in comparison to other samples. Taking this into account, then it is apparent that deletion of the human cardiac actin promoter to nucleotide -177 results in a decrease in the transcriptional activity of the resulting fusion gene.

Further deletion of the human cardiac actin gene promoter to nucleotide -153 results in a total loss of detectable

FIGURE A.1.3.4. The promoter activities of the plasmids pRE485, pRE177, pRE153, pRE131, pRE96, and pRE64 in *Xenopus laevis* embryos.

The plasmids pRE485, pRE177, pRE153, pRE131, pRE96 and pRE64 were injected into embryos at the two cell stage of development. Embryos were allowed to proceed to stage 30 of development prior to being dissected into axis, head and gut regions. Total nucleic acid was prepared from whole embryos (W), in addition to the axis (A), head (H) and gut (G) regions of dissected embryos. Transcripts originating from the micro-injected gene were detected by primer extension analysis using the oligonucleotide MG-1.

The size of DNA markers (track M) is indicated on the right hand side of the figure. The 67 nucleotide extension product, which is produced on the correct initiation of transcripts originating from the human cardiac actin gene, is illustrated by an arrow on the right hand side of the figure.



Appendix

transcription of the fusion gene in Xenopus laevis embryos. Indeed, all subsequent promoter deletions are transcriptionally inactive as detected by the assay system employed in these experiments. This is in contrast to experiments investigating the ability of deleted promoter fragments to drive the expression of a CAT reporter gene. These data demonstrate that 153 nucleotides of human cardiac actin upstream sequence (plasmid pHCA153) is capable of directing the tissue-specific expression of a CAT reporter gene, although at lower levels than the pHCA485 plasmid. This discrepancy in data may be due to CAT assays being a more sensitive assay for promoter activity than primer extension assays.

The investigation into the effect of promoter deletions on the transcriptional activity of the human cardiac actin gene promoter yields similar results to experiments investigating the capability of promoter deletions to drive the expression of a CAT gene. However, multiple extension products are apparent in some of the primer extension reactions. Thus, although a percentage of the transcripts originating from the human cardiac actin gene in Xenopus embryos are correctly initiated, there also appears to be certain amounts of incorrectly initiated transcripts originating from the micro-injected gene.

A.1.4. Discussion.

Data presented in this section demonstrate that 485 nucleotides of human cardiac actin gene upstream sequence, in addition to 68 nucleotides downstream of the transcriptional

Appendix

start site of the gene, is sufficient to drive the tissue-specific expression of a reporter gene in Xenopus laevis embryos. Furthermore, transcripts originating from this fusion gene appear to be correctly initiated, as determined by primer extension analysis.

The sequence requirements for the effective expression of the human cardiac actin gene have been extensively characterised (Minty and Kedes, 1986; Miwa and Kedes, 1987; Sartorelli et al., 1990). These studies demonstrated that as little as 113 nucleotides of human cardiac actin upstream sequence is sufficient to direct the tissue-specific expression of the gene in myogenic cell lines (Miwa and Kedes, 1987). Indeed, although distal regulatory elements exist which are required for the full activity of the human cardiac actin gene, all sequences necessary for the tissue-specific expression of the gene are contained downstream and inclusive of the CArG box1 motif (see Sartorelli et al., 1990).

In contrast to these data no transcription from the construct pHCA131 is apparent in Xenopus laevis embryos, despite it containing all human cardiac actin sequences which have been identified to be sufficient for the expression of the gene in myogenic cell lines. One possible explanation for this discrepancy in data is that the human cardiac actin gene is expressed less efficiently in Xenopus laevis embryos than it is in myogenic cell lines. Therefore, deletions of the human cardiac actin gene promoter that significantly reduce the activity of the gene in myogenic cell lines, may totally

Appendix

abolish the activity of the gene in Xenopus laevis embryos. In this regard pHCA131 lacks the distal regulatory region and the CArG box2 motif of the human cardiac actin gene. Deletion of these sequences from the human cardiac actin gene promoter significantly reduces the activity of the gene in myogenic cell lines (Minty and Kedes, 1986; Miwa and Kedes, 1987). Thus, the deletion of these sequences from the human cardiac actin gene promoter may totally abolish the activity of the gene in the less sensitive Xenopus assay system.

In addition to 485 nucleotides of human cardiac actin gene upstream sequence being able to direct the correct tissue-specific expression of a reporter gene in stage 30 Xenopus laevis embryos, it also appears to activate the reporter gene at approximately the correct stage of Xenopus development (see section A.1.2.). However, at early stages of Xenopus development expression of the fusion gene does not appear to be tissue-specific. Indeed, it is not until the latter stages of development that the gene is expressed predominantly in the region of the embryo that expresses the endogenous cardiac actin gene. The reason for this apparent infidelity in the expression of the human cardiac actin gene in the early stages of Xenopus development is unknown. Similar problems have not been experienced when a Xenopus skeletal actin fusion gene is injected into developing embryos (data not shown). Indeed, this type of assay system has been used extensively in the study of sequence requirements for the expression of the Xenopus laevis cardiac actin gene (see Gurdon et al., 1985; Mohun et al.,

Appendix

1986).

Due to the inconsistency of expression patterns of the human cardiac actin gene in the early stages of Xenopus development, and the apparent insensitivity of Xenopus micro-injection techniques in identifying the sequences important in the expression of the human cardiac actin gene, it was decided that future work would concentrate on the study of the expression of the Xenopus borealis skeletal actin gene.

APPENDIX 2

Diagrammatic representation of the plasmids p Δ gb and pXlgb.

A. Diagrammatic representation of the plasmid p Δ gb. The lower part of the figure illustrates the complete plasmid, showing the Xenopus laevis β -globin sequences in black. Sequences of the vector (pAT153) are shown in blue.

The upper part of the figure shows the promoter of the Xenopus laevis β -globin gene illustrating the site of a 389 nucleotide internal deletion.

B. Diagrammatic representation of the plasmid pXlgb. The lower part of the figure shows the complete plasmid, illustrating sequences of the Xenopus laevis β -globin gene in black. Sequences of the vector are illustrated in blue.

References

R E F E R E N C E S

References

- Arnold, H. H., Tannich, E. and Paterson, B. (1988).
The promoter of the chicken myosin light chain 2 gene shows cell-specific expression in transfected primary cultures of muscle. *Nucleic Acids Res.* **16**: 2411-2429.
- Bains, W., Ponte, P., Blau, H. and Kedes, L. (1984).
Cardiac actin is the major actin gene product in skeletal muscle cell differentiation in vitro. *Mol. Cell. Biol.* **4**: 1449-1453.
- Baldwin, T. J. and Burden, S. (1988).
Isolation and characterisation of the mouse acetylcholine receptor delta subunit gene: identification of a 148-bp cis-acting region that confers myotube-specific expression. *J. Cell Biol.* **107**: 2271-2279.
- Baldwin, T. J. and Burden, S. J. (1989).
Muscle-specific expression controlled by a regulatory element lacking a MyoD binding site. *Nature* **341**: 716-720.
- Baldwin, A., Kittler, E. and Emerson, C. (1985).
Structure and regulation of a fast skeletal muscle Troponin I gene. *Proc. Natl. Acad. Sci. USA.* **82**: 8080-8084.
- Benezra, R., Davis, R. L., Lockshon, D., Turner, D. L. and Weintraub, H. (1990).
The protein Id: a negative regulator of helix-loop-helix DNA binding proteins. *Cell* **61**: 49-59.
- Bergsma, D. J., Chang, K. S. and Schwartz, R. J. (1985).
Novel chicken actin gene: third cytoplasmic isoform. *Mol. Cell. Biol.* **5**: 1151-1162.

References

- Bergsma, D. J., Grichnik, J. M., Gosset, L. M. A. and Schwartz, R. J. (1986).
Delimitation and characterisation of cis-acting DNA sequences required for the regulated expression and transcriptional control of the chicken α -actin gene. *Mol. Cell. Biol.* 6: 2462-2475.
- Berk, A. J. and Schmidt, M. C. (1990).
How do transcription factors work? *Genes Dev.* 4: 151-155.
- Biggin, M. D. and Tjian, R. (1989).
Transcription factors and the control of Drosophila development. *Trends In Genetics* 5: 377-383.
- Blau, H. M., Chiu, C.-P. and Webster, C. (1983).
cytoplasmic activation of human nuclear genes in skeletal heterokaryons. *Cell* 32: 1171-1180.
- Blau, H. M., Pavlath, G. K., Hardeman, E. C., Chiu, C.-P., Silberstein, L., Webster, S. G., Miller, S. C. and Webster, C. (1985).
Plasticity of the differentiated state. *Science* 230: 758-766.
- Bouvagnet, P. F., Strehler, E. E., White, G. E., Strehler-Page, M. A., Nadal-Ginard, B. and Mahdavi, V. (1987).
Multiple negative and positive 5' regulatory elements control the cell type-specific expression of the embryonic skeletal myosin heavy-chain gene. *Mol. Cell. Biol.* 7: 4377-4389.
- Boxer, L. M., Miwa, T., Gustafson, T. A. and Kedes, L. H. (1989a).
Identification and characterisation of a factor that binds to two human sarcomeric actin gene promoters. *J. Biol. Chem.* 264: 1284-1299.

References

- Boxer, L. M., Prywes, R., Roeder, R. G. and Kedes, L. H. (1989b).
The sarcomeric CArG binding factor is indistinguishable from the c-fos
serum response factor. *Mol. Cell. Biol.* **9**: 515-522.
- Braun, T., Tannich, E., Buschhausen-Denker, G. and Arnold, H. H. (1989a).
Promoter upstream elements of the chicken cardiac myosin light-chain 2-A
gene interact with trans-acting regulatory factors for muscle-specific
transcription. *Mol. Cell. Biol.* **9**: 2513-2525.
- Braun, T., Buschhausen-Denker, G., Bober, E., Tannich, E. and Arnold, H. H.
(1989b).
A novel human muscle factor related to but distinct from MyoD1 induces
myogenic conversion in 10T $\frac{1}{2}$ fibroblasts. *EMBO J.* **8**: 701-709.
- Braun, T., Bober, E., Winter, B., Rosenthal, N. and Arnold, H. H. (1990).
Myf-6, a new member of the human gene family of myogenic determination
factors: evidence for a gene cluster on chromosome 12. *EMBO J.* **9**: 821-831.
- Bravo, R., Neubergh, M., Burckhardt, J., Almendral, J., Wallich, R. and
Müller, R. (1987).
Involvement of common and cell-type-specific pathways in c-fos gene
control: Stable induction by cAMP in macrophages. *Cell* **48**: 251-260.
- Brennan, T. J. and Olson, E. N. (1990).
Myogenin resides in the nucleus and acquires high affinity for a conserved
enhancer element on heterodimerisation. *Genes Dev.* **4**: 582-595.
- Buckingham, M. E. and Minty, A. J. (1983).
In: *Eukaryotic genes: their structure, activity and regulation.* (Eds
Maclean, N., Gregory, S. P. and Flavell, R. A.). 365-395 (Butterworths,
London, 1983).

References

- Busby, S. J. and Reader, R. H. (1983).
Spacer sequences regulate the transcription of ribosomal gene plasmids injected into Xenopus embryos. *Cell* **34**: 989-996.
- Buskin, J. N. and Hauschka, S. D. (1989).
Identification of a myocyte nuclear factor that binds to the muscle-specific enhancer of the mouse muscle creatine kinase gene. *Mol. Cell. Biol.* **9**: 2627-2640.
- Carmon, Y., Czosnek, H., Nudel, U., Shani, M. and Yaffe, D. (1982).
DNase I sensitivity of genes expressed during myogenesis. *Nucleic Acids Res.* **10**: 3085-3098.
- Caudy, M., Vassin, H., Branc, M., Tuma, R., Yeh Jan, L. and Jan, Y. N. (1988).
Daughterless a Drosophila gene essential for both neurogenesis and sex determination, has sequence similarities to myc and the acheate-scute complex. *Cell* **55**: 1061-1067.
- 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. *Nucleic Acids Res.* **13**: 1223-1237.
- Chow, K.-L. and Schwartz, R. J. (1990).
A combination of closely associated positive and negative cis-acting promoter elements regulates transcription of the skeletal α -actin gene. *Mol. Cell. Biol.* **10**: 528-538.
- Church, G. M., Ephrussi, E., Gilbert, W. and Tonogawa, S. (1985).
Cell-type-specific contacts to immunoglobulin enhancers in nuclei. *Nature* **313**: 798-801.

References

- Clarke, M. and Spudich, J. A. (1977).
Non-muscle contractile proteins: the role of actin and myosin in cell mobility and shape determination. *Ann. Rev. Biochem.* **46**: 797-822.
- Cleveland, D. W., Lopata, M. A., MacDonald, R. J., Cowan, N. J., Rutter, W. J. and Kirschner, M. W. (1980).
Number and evolutionary conservation of α - and β -tubulin and β - and δ -actin genes using specific cloned cDNA probes. *Cell* **20**: 95-105.
- Cronmiller, C., Schedl, P. and Cline, T. (1988).
Molecular characterisation of daughterless a Drosophila sex determination gene with multiple roles in development. *Genes Dev.* **2**: 1666-1676.
- Cross, G. S. (1984).
The structure and expression of Xenopus borealis actin genes. Ph.D. thesis, University of Warwick.
- Crowe, D. T. and Tsai, M.-J. (1989).
Mutagenesis of the rat insulin II 5'-flanking region defines sequences important for expression in HIT cells. *Mol. Cell. Biol.* **9**: 1784-1789.
- Curran, T. (1984).
Induction of c-fos gene and protein by growth factors precedes activation of c-myc. *Nature* **313**: 716-720.
- Curran, T., MacConnell, W. P., van Straaten, F. and Verma, I. M. (1983).
Structure of FBJ murine osteosarcoma virus genome. Molecular cloning of its associated helper virus and the cellular homologue of the c-fos gene from mouse and human cells. *Mol. Cell. Biol.* **3**: 914-921.
- Curran, T. and Franza, B. R. (1988).
Fos and jun: the AP-1 connection. *Cell* **55**: 395-397.

References

- Czosnek, H., Nudel, U., Mayer, Y., Barker, P. E., Pravtcheva, D. D., Ruddle, E. H. and Yaffe, D. (1983).
The genes coding for the cardiac muscle actin, the skeletal muscle actin and the cytoplasmic β -actin are located on three different mouse chromosomes. *EMBO J.* **2**: 1977-1979.
- Dale, L. and Slack, J. M. W. (1987).
Regional specification within the mesoderm of early embryos of Xenopus laevis. *Development* **100**: 279-295.
- Davis, R. L., Weintraub, H. and Lassar, A. B. (1987).
Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell* **51**: 987-1000.
- Davis, R. L., Cheng P.-F., Lassar, A. B. and Weintraub, H. (1990).
The MyoD DNA binding domain contains a recognition code for muscle-specific gene activation. *Cell* **60**: 733-746.
- de la Brousse, C. F. and Emerson, C. P. (1990).
Localised expression of a myogenic regulatory gene, gmf-1, in the somite dermatome of avian embryos. *Genes Dev.* **4**: 567-581.
- Deschamps, J., Meylink, F. and Verma, I. M. (1985).
Identification of a transcriptional enhancer element upstream of the proto-oncogene fos. *Science* **230**: 1174-1177.
- Dodemont, H. J., Soriano, P., Quax, W. J., Ramackers, 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.

References

Donoghue, M., Ernst, H., Wentworth, B., Nadal-Ginard, B. and Rosenthal, N. (1988).

A muscle-specific enhancer is located at the 3' end of the myosin light-chain 1/3 gene locus. *Genes Dev.* **2**: 1779-1790.

Durica, D. S., Schloss, J. A. and Crain, W. R. (1980).

Organisation of actin gene sequences in the sea urchin: molecular cloning of an intron-containing DNA sequence coding for a cytoplasmic actin. *Proc. Natl. Acad. Sci. USA.* **77**: 5683-5687.

Edmondson, D. G. and Olson, E. N. (1989).

A gene with homology to the myc similarity region of MyoD1 is expressed during myogenesis and is sufficient to activate the muscle differentiation programme. *Genes Dev.* **3**: 628-640.

Eldridge, 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.

Ellis, H. M., Spann, D. R. and Posakeony, J. W. (1990).

Extramacrochaetae, a negative regulator of sensory organ development in Drosophila, defines a new class of helix-loop-helix proteins. *Cell* **61**: 27-38.

Elsholtz, H. P., Albert, V. R., Treacy, M. N. and Rosenfeld, M. G. (1990).

A two base change in a POU factor binding site switches pituitary-specific to lymphoid-specific gene expression. *Genes Dev.* **4**: 43-51.

Engel, J. N., Gunning, P. W. and Kedes, L. M. (1981).

Isolation and characterisation of human cardiac actin genes. *Proc. Natl. Acad. Sci. USA.* **78**: 4674-4678.

References

- 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. *Nucleic Acids Res.* 10: 3861-3876.
- Fyrberg, E. A. Bond, B. J., Hershey, N. D., Mixer, 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.
- Gallwitz, D. and Seidel, R. (1980).
Molecular cloning of the actin gene from the yeast Saccharomyces cerevisiae. *Nucleic Acids Res.* 8: 1043-1059.
- Garell, J. and Modolell, J. (1990).
The Drosophila extramacrochaetae locus, an antagonist of proneural genes that, like these genes, encodes a helix-loop-helix protein. *Cell* 61: 39-48.
- Garrels, J. I. and Gibson, W. (1976).
Identification and isolation of multiple forms of actin. *Cell* 9: 793-805.
- Gilman, M. Z. (1988).
The c-fos serum response element responds to protein kinase C-dependant and -independent signals but not cyclic AMP. *Genes Dev.* 2: 394-402.
- Gilman, M. Z., Wilson, R. N. and Weinberg, R. A. (1986).
Multiple protein-binding sites in the 5'-flanking region regulate c-fos expression. *Mol. Cell. Biol.* 6: 4305-4316.
- Gorman, C. M., Moffat, L. E. and Howard, B. H. (1982).
Recombinant genomes that express chloramphenicol acetyltransferase in mammalian cells. *Mol. Cell. Biol.* 2: 1044-1051.

References

- Gossett, L. A., Kelvin, D. J., Sternberg, E. A. and Olson, E. N. (1989).
A new myocyte-specific enhancer binding factor that recognises a conserved element associated with muscle-specific genes. *Mol. Cell. Biol.* **9**: 5022-5033.
- Greenberg, M. E. and Ziff, E. B. (1984).
Stimulation of 3T3 cells induces transcription of the c-fos proto-oncogene. *Nature* **311**: 433-438.
- Greenberg, M. E., Greene, L. A. and Ziff, E. B. (1985).
Nerve growth factor and epidermal growth factor induce rapid transient changes in proto-oncogene transcription in PC12 cells. *J. Biol. Chem.* **260**: 14101-14110.
- Greenberg, M. E., Siegfried, Z. and Ziff, E. B. (1987).
Mutation of the c-fos gene dyad symmetry element inhibits serum inducibility of transcription in vivo and the nuclear regulatory factor binding in vitro. *Mol. Cell. Biol.* **7**: 1217-1225.
- 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. *Nucleic Acids Res.* **14**: 1683-1701.
- Grichnik, J. M., French, B. A. and Schwartz, R. J. (1988).
The chicken skeletal α -actin gene promoter region exhibits partial dyad symmetry and a capacity to drive bi-directional transcription. *Mol. Cell. Biol.* **8**: 4587-4597.
- Gunning, P., Ponte, P., Blau, H. and Kedes, L. (1983).
 α -skeletal and α -cardiac actin genes are co-expressed in adult skeletal muscle and heart. *Mol. Cell. Biol.* **3**: 1985-1995.

References

- Gurdon, J. B., Fairman, S., Mohun, T. J. and Brennan, S. (1985).
Activation of muscle-specific actin genes in Xenopus development by an induction between animal and vegetal cells of a blastula. *Cell* **41**: 913-922.
- Gustafson, T. A. and Kedes, L. H. (1989).
Identification of multiple proteins which interact with functional regions of the human cardiac α -actin promoter. *Mol. Cell. Biol.* **9**: 3269-3283.
- Gustafson, T. A., Miwa, T., Boxer, L. and Kedes, L. (1988).
Interaction of nuclear proteins with muscle-specific regulatory sequences of the human cardiac actin promoter. *Mol. Cell. Biol.* **7**: 4100-4119.
- 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. USA.* **79**: 5901-5905.
- Hopwood, N. D., Pluck, A. and Gurdon, J. B. (1989).
MyoD expression in the forming somites is an early response to mesoderm induction in Xenopus embryos. *EMBO J.* **8**: 3409-3417.
- Hopwood, N. D. and Gurdon, J. B. (1990).
Activation of muscle genes without myogenesis by ectopic expression of MyoD in frog embryo cells. *Nature* **347**: 197-200.
- Hopwood, N. D., Pluck, A. and Gurdon, J. B. (1991).
Xenopus Myf-5 marks early muscle cells and can activate muscle genes ectopically in early embryos. *Development* **111**: 551-560.
- Horlick, R. A. and Benfield, P. A. (1989).
The upstream muscle-specific enhancer of the rat muscle creatine kinase gene is composed of multiple elements. *Mol. Cell. Biol.* **9**: 2396-2413.

References

- Humphries, S. E., Whittall, R., Minty, A., Buckingham, M. E. and Williamson, R. (1981).
There are approximately 20 actin genes in the human genome. *Nucleic Acids Res.* **9**: 4895-4908.
- Hunt, T. (1989).
Cytoplasmic anchoring proteins and the control of nuclear localisation. *Cell* **59**: 949-951.
- Imler, J.-L., Lemaire, C., Wasylyk, C. and Wasylyk, B. (1987).
Negative regulation contributes to the tissue-specificity of the immunoglobulin chain enhancer. *Mol. Cell. Biol.* **7**: 2558-2567.
- Jaynes, J. B., Chamberlain, J. S., Buskin, 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.
- Jaynes, J. B., Johnson, J. E., Buskin, J. N., Gartside, C. L. and Hauschka, S. (1988).
The muscle creatine kinase gene is regulated by multiple upstream elements, including a muscle-specific enhancer. *Mol. Cell. Biol.* **8**: 62-70.
- Karin, M., Castrillo, J.-L. and Theill, L. E. (1990).
Growth hormone gene regulation: a paradigm for cell-type-specific gene activation. *Trends In Genetics* **6**: 92-96.
- Klamt, C., Knust, E., Tietze, K. and Campos-Ortega, J. A. (1989).
Closely related transcripts encoded by the neurogenic gene complex enhancer of split of Drosophila melanogaster. *EMBO J.* **8**: 203-210.

References

- Klamut, H. J., Gangopadhyay, S. B., Worton, R. G. and Ray, P. N. (1990). Molecular and functional analysis of the muscle-specific promoter region of the Duchenne muscular dystrophy gene. *Mol. Cell. Biol.* **10**: 193-205.
- Klarsfeld, A. P., Daubas, B., Bourachot, B. and Changeux, J.-P. (1987). A 5'-flanking region of the chicken acetyl-choline receptor alpha subunit gene confers tissue-specificity and developmental control of expression in transfected cells. *Mol. Cell. Biol.* **7**: 951-955.
- Klug, A. and Rhodes, D. (1987). 'Zinc fingers': a novel motif involved for nucleic acid recognition. *Trends Biochem. Sci.* **12**: 464-469.
- Konieczny, S. F. and Emerson, C. P. (1984). 5-azacytidine induction of stable mesodermal stem cell lineages from 10T $\frac{1}{2}$ cells: evidence for regulatory genes controlling determination. *Cell* **38**: 791-800.
- Konieczny, S. F. and Emerson, C. P. (1987). Complex expression of the muscle-specific contractile protein troponin I gene. *Mol. Cell. Biol.* **7**: 3065-3075.
- Kreig, P. A. and Melton, D. A. (1985). Developmental regulation of a gastrula-specific gene injected into fertilised Xenopus eggs. *EMBO J.* **4**: 3463-3471.
- Kruijer, W., Cooper, J. A., Hunter, T. and Verma, I. M. (1984). Platelet-derived growth factor induces rapid but transient expression of the c-fos gene and protein. *Nature* **312**: 711-716.
- Kruijer, W., Schubert, D. and Verma, I. M. (1985). Induction of the proto-oncogene fos by nerve growth factor. *Proc. Natl. Acad. Sci. USA.* **82**: 7330-7334.

References

- Lassar, A. B., Paterson, B. M. and Weintraub, H. (1986).
Transfection of a DNA locus that mediates the conversion of 10T $\frac{1}{2}$ fibroblasts to myoblasts. *Cell* 47: 649-656.
- Lassar, A. B., Buskin, J. N., Lockshon, D., Davis, R. L., Apone, S., Hauschka, S. D. and Weintraub, H. (1989).
MyoD is a sequence-specific DNA binding protein requiring a region of myc homology to bind the muscle creatine kinase enhancer. *Cell* 58: 823-831.
- Leung, S. and Miyamoto, N. (1989).
Point mutational analysis of the human c-fos serum response factor binding site. *Nucleic Acids Res.* 17: 1177-1195.
- Lin, H., Yutzey, K. E. and Konieczny, S. F. (1991).
Muscle-specific expression of the troponin I gene requires interactions between helix-loop-helix muscle regulatory factors and ubiquitous transcription factors. *Mol. Cell. Biol.* 11: 267-280.
- Lin, Z.-Y., Dechesne, C. A., Eldridge, J. and Paterson, B. M. (1989).
An avian muscle factor related to MyoD1 activates muscle-specific promoters in non-muscle cells of different germ layer origin and in BrdU-treated myoblasts. *Genes Dev.* 3: 986-996.
- Manak, J. R., de Bisschop, N., Kris, R. M. and Prywes, R. (1990).
Casein kinase II enhances the DNA binding activity of serum response factor. *Genes Dev.* 4: 955-967.
- Mer, J. H. and Ordahl, C. P. (1988).
A conserved CATTCCT motif is required for the skeletal muscle-specific activity of the cardiac Troponin T gene promoter. *Proc. Natl. Acad. Sci. USA.* 85: 6404-6408.

References

- Mar, J. H. and Ordahl, C. P. (1990).
M-CAT binding factor, a novel trans-acting factor governing muscle-specific transcription. *Mol. Cell. Biol.* **10**: 4271-4283.
- Mar, J. H., Antin, F. B., Cooper, T. A. and Ordahl, C. P. (1988).
Analysis of the upstream regions governing expression of the chicken cardiac Troponin T gene in embryonic cardiac and skeletal muscle cells. *J. Cell Biol.* **107**: 573-585.
- Maxam, A. and Gilbert, W. (1980).
Sequencing end labelled DNA with base-specific chemical cleavages. *Meth. Enzymol.* **65**: 499-560.
- 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 heart. *Nucleic Acids Res.* **12**: 1087-1100.
- Melloul, D., Aloni, B., Calvo, J., Yaffe, D. and Nudel, U. (1984).
Developmentally regulated expression of chimeric genes containing actin DNA sequences in transfected myogenic cells. *EMBO J.* **3**: 983-990.
- Miner, J. H. and Wold, B. (1990).
Herculin, a fourth member of the MyoD family of myogenic regulatory genes. *Proc. Natl. Acad. Sci. USA.* **87**: 1089-1093.
- Minty, A. and Kedes, L. (1986).
Upstream regions of the human cardiac actin gene that modulate its transcription in muscle cells: presence of an evolutionary conserved 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.

References

- Minty, A. J., Alonso, S., Guenet, J. L. and Buckingham, M. E. (1983).
Number and organisation 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 modulatory factors can accumulate before gene activation. *Mol. Cell. Biol.* **6**: 2137-2148.
- Mitchell, P. J. and Tjian, R. (1989).
Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science* **245**: 371-378.
- Mitchell, R. L., Zokas, L., Schreiber, R. D. and Verma, I. M. (1985).
Rapid induction of the expression of proto-oncogene fos during human monocytic differentiation. *Cell* **40**: 209-217.
- Miwa, T. and Kedes, L. (1987).
Duplicated CArG box domains have positive and mutually dependant regulatory roles in expression of the human cardiac actin gene. *Mol. Cell. Biol.* **7**: 2803-2813.
- Miwa, T., Boxer, L. and Kedes, L. (1987).
CArG boxes in the human α -cardiac actin gene are core binding sites for positive trans-acting regulatory factors. *Proc. Natl. Acad. Sci. USA.* **84**: 6702-6706.
- Mohun, T. J. and Garrett, N. (1987).
An amphibian cytoskeletal-type actin gene is expressed exclusively in muscle tissue. *Development* **101**: 393-402.

References

- Mohun, T.J., Brennan, S., Dethan, N., Fairman, S. and Gurdon, J. B. (1984).
Cell type-specific activation of actin genes in the early amphibian embryo.
Nature **311**: 716-721.
- Mohun, T., Garrett, N. and Gurdon, J. (1986).
Upstream sequences required for tissue-specific activation of the cardiac
actin gene in Xenopus laevis embryos. EMBO J. **5**: 3185-3193.
- Mohun, T., Taylor, M., Garrett, N. and Gurdon, J. (1989a).
The CARG promoter sequence is necessary for muscle-specific transcription
of the cardiac actin gene in Xenopus embryos. EMBO J. **8**: 1153-1161.
- Mohun, T. J., Garrett, N. and Taylor, M. V. (1989b).
Temporal and tissue-specific expression of the proto-oncogene c-fos during
development in Xenopus laevis. Development **107**: 835-846.
- Moore, J. P., Todd, J. A., Hesketh, T. R. and Metcalfe, J. C. (1986).
c-fos and c-myc gene activation, ionic signals, and DNA synthesis in
thymocytes. J. Biol. Chem. **261**: 8158-8162.
- Morgan, J. I. and Curran, T. (1986).
Role of ion flux in the control of c-fos expression. Nature **322**: 716-720.
- Murre, C., McCaw, P. S. and Baltimore, D. (1989a).
A new DNA binding and dimerisation motif in immunoglobulin enhancer
binding, daughterless, MyoD and myc proteins. Cell **56**: 777-783.
- Murre, C., McCaw, P. S., Vaessin, H., Caudy, M., Jan, L. Y., Jan, Y. N.,
Cabrera, C. V., Buskin, J. N., Hauschka, S. D., Lassar, A. B., Weintraub,
H. and Baltimore, D. (1989b).
Interactions between heterologous helix-loop-helix proteins generate
complexes that bind to a common DNA sequence. Cell **58**: 537-544.

References

- Muscat, G. O. and Kedes, L. (1987).
Multiple 5'-flanking regions of the human skeletal actin gene synergistically modulate muscle-specific expression. *Mol. Cell. Biol.* **7**: 4089-4099.
- Muscat, G. O., Gustafson, T. A. and Kedes, L. (1988).
A common factor regulates skeletal and cardiac α -actin gene transcription in muscle. *Mol. Cell. Biol.* **8**: 4120-4133.
- Nabeshima, Y.-I., Fujii-Kuryama, Y., Muramatsu, M. and Ogata, K. (1984).
Alternative transcription and two modes of splicing result in two myosin light chains from one gene. *Nature* **308**: 333-338.
- Nakajima-Iijima, S., Hamada, H., Reddy, P. and Kakamiga, T. (1985).
Molecular structure of the human cytoplasmic β -actin gene: interspecies homologies of sequences in introns. *Proc. Natl. Acad. Sci. USA.* **82**: 6133-6137.
- Nakamura, O., Takasaki, H. and Ishihara, M. (1970)
Differentiation during cleavage in Xenopus laevis. I. Acquisition of self differentiation capacity of the dorsal marginal zone. *Proc. Japan. Acad.* **46**: 694-699.
- Newport, T. J. and Kirschner, M. (1982).
A major developmental transition in early Xenopus embryos: I. Characterisation and timing of cellular changes at the mid-blastula stage. *Cell* **30**: 675-685.
- Ng, R. and Abelson, J. (1980).
Isolation and sequence of the gene for actin in Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci. USA.* **77**: 3913-3916.

References

- Ng, S.-Y., Gunning, P., Eddy, R., Ponte, P., Leavitt, J., Shows, T. and Kedes, L. (1985).
Evolution of the functional human α -actin gene and its multiple pseudogene family: conservation of non-coding regions and the chromosomal dispersion of pseudogenes. *Mol. Cell. Biol.* **5**: 2720-2732.
- 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 tables of Xenopus laevis (Daudin), Amsterdam: North Holland.
- Nieuwkoop, P. D. and Ubbels, G. A. (1972).
The formation of mesoderm in urodelean amphibians, IV. Quantitative evidence for the purely "ectodermal" origin of the entire mesoderm and of the pharyngeal endoderm. *Wilhelm Roux' Archiv.* **169**: 185-199.
- Nikovits, W., Kuncio, G. and Ordahl, C. P. (1986).
The chicken fast skeletal Troponin I gene: Exon organisation and sequence. *Nucleic Acids Res.* **14**: 3377-3390.
- Nikovits, W., Mar, J. H. and Ordahl, C. P. (1990).
Muscle-specific activity of skeletal Troponin I promoter requires interaction between upstream regulatory sequences and elements contained within the first transcribed exon. *Mol. Cell. Biol.* **10**: 3468-3482.
- Norman, C., Runswick, M., Pollock, R. and Treisman, R. (1988).
Isolation and properties of cDNA clones encoding SRF, a transcription factor that binds to the c-fos serum response element. *Cell* **55**: 989-1003.
- Nudel, U., Greenberg, D., Ordahl, C. P., Saxel, O., Newman, S. and Yaffe, D. (1985).

References

- Developmentally regulated expression of a chick muscle-specific gene in stably transfected rat myogenic cells. *Proc. Natl. Acad. Sci. USA.* **82**: 3106-3109.
- Ogi, K.-I. (1967).
Determination in the development of the amphibian embryo. *Sci. Rep. Tohoku Univ. IV (Biol)* **33**: 239-247.
- Ogi, K.-I. (1969).
Regulatory capacity in the early amphibian embryo. *Res. Bull. Dept. Gen. Ed. Nagoya University* **13**: 31-40.
- Olson, E. N. (1990).
MyoD family: a paradigm for development? *Genes Dev.* **4**: 1454-1461.
- Ordahl, C. P. and Cooper, T. A. (1983).
Strong homology in promoter and 3'-untranslated regions of the chick and rat α -actin genes. *Nature* **303**: 348-349.
- Pabo, C. and Saver, R. (1984).
Protein-DNA recognition. *Annu. Rev. Biochem.* **53**: 293-321.
- 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.
- Periasamy, M., Strehler, E. E., Garfinkle, L. I., Gubits, R. M., Ruiz-Opazo, N. and Nadal-Ginard, B. (1984).
Fast skeletal muscle myosin light chains 1 and 3 are produced from a single gene by a combined process of differential RNA transcription and splicing. *J. Biol. Chem.* **259**: 13595-13604.

References

- Petropoulos, C. J., Rosenberg, M. P., Jenkins, N. A., Copeland, N. G. and Hughes, S. H. (1989).
The chicken skeletal muscle alpha actin promoter is tissue-specific in transgenic mice. *Mol. Cell. Biol.* 9: 3785-3792.
- Phan-Dinh-Tuy, F., Tuil, D., Schweighoffer, F., Pinset, C., Kahn, A. and Minty, A. (1988).
The CCAATGG box: a protein binding site common to transcription regulatory regions of the cardiac actin, c-fos and interleukin-2 receptor genes. *Eur. J. Biochem.* 173: 507-515.
- Piette, J., Klarsfeld, A. and Changeux, J.-P. (1989).
Interaction of nuclear factors with the upstream region of the alpha subunit gene of muscle acetylcholine receptor: variations with muscle differentiation and denervation. *EMBO J.* 8: 687-694.
- Pinney, D. F., Pearson-White, S. H., Konieczny, S., Latham, K. E. and Emerson, C. P. (1988).
Myogenic lineage determination and differentiation: evidence for a regulatory gene pathway.
- Pollock, R. and Treisman, R. (1990).
A sensitive method for the determination of protein-DNA binding specificities. *Nucleic Acids Res.* 18: 6197-6204.
- Prywes, R. and Roeder, R. G. (1986).
Inducible binding of a factor to the c-fos enhancer. *Cell* 47: 777-784.
- Prywes, R. and Roeder, R. G. (1987).
Purification of the c-fos enhancer binding protein. *Mol. Cell. Biol.* 7: 3482-3489.

References

- Prywes, R., Dutta, A., Cromlish, J. A. and Roeder, R. G. (1988).
Phosphorylation of serum response factor, a factor that binds to the serum response element of the c-fos enhancer. Proc. Natl. Acad. Sci. USA. **85**: 7206-7210.
- Rudney, 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.
- Renz, M., Neuberger, M., Kurz, C., Bravo, R. and Müller, R. (1985).
Regulation of c-fos transcription in mouse fibroblasts: identification of DNase I hypersensitive sites and regulatory upstream sequences. EMBO J. **4**: 3711-3716.
- Rhodes, S. J. and Konieczny, S. F. (1989).
Identification of MRF-4: a new member of the muscle regulatory factor gene family. Genes Dev. **3**: 2050-2061.
- Rushlow, C. A., Hogan, A., Pinchin, S. M., Howe, K. M., Lordelli, M. and Ish-Horowicz, D. (1989).
The Drosophila hairy protein acts in both segmentation and bristle patterning and shows homology to N-myc. EMBO J. **8**: 3095-3103.
- Sartorelli, V., Webster, K.A., Lassar, A. B. and Kedes, L. (1990).
Muscle-specific expression of the cardiac alpha actin gene requires MyoD1, CArG box binding factor, and SP-1. Genes Dev. **4**: 1811-1822.
- Sassone-Corsi, P., Visvader, J., Ferland, L., Mellon, P. L. and Verma, I. M. (1988).
Induction of the proto-oncogene fos transcription through the adenylate cyclase pathway: characterisation of a cAMP-responsive element. Genes Dev. **2**: 1529-1538.

References

- Sassoon, D., Lyons, G., Wright, W. E., Lin, V., Lassar, A., Weintraub, H. and Buckingham, M. (1989).
Expression of two myogenic regulatory factors myogenin and MyoD during mouse embryogenesis. *Nature* **341**: 303-307.
- Scales, J. B., Olson, E. N. and Perry, M. (1990).
Two distinct Xenopus genes with homology to MyoD are expressed before somite formation in early embryogenesis. *Mol. Cell. Biol.* **10**: 1516-1524.
- Schaffner, W. (1989).
How do transcription factors binding the same DNA sequence sort out their jobs? *Trends In Genetics* **5**: 37-39.
- Schalasta, G. and Doppler, C. (1990).
Inhibition of c-fos transcription and phosphorylation of the serum response factor by an inhibitor of phospholipase C-type reactions. *Mol. Cell. Biol.* **10**: 5558-5561.
- Scheller, R. H., McAllister, L. B., Crain, W. R., Durica, D. S., Posakony, J. W., Thomas, T. L., Bolten, R. J. and Davidson, E. H. (1981).
Organisation and expression of multiple actin genes in the sea urchin. *Mol. Cell. Biol.* **1**: 609-628.
- Schroter, H., Shaw, P. E. and Nordheim, A. (1987).
Purification of intercalater-released p67, a polypeptide that interacts specifically with the c-fos serum response element. *Nucleic Acids Res.* **15**: 10145-10158.
- Schroter, H., Mueller, C. G. F., Meese, K. and Nordheim, A. (1990).
Synergism in ternary complex formation between the dimeric glycoprotein p67^{SRF}, polypeptide p62^{TCF} and the c-fos serum response element. *EMBO J.* **9**: 1123-1130.

References

- Schwartz, R. and Rothblum, K. (1980).
Regulation of muscle differentiation: isolation and purification of chick actin messenger ribonucleic acid and quantitation with complementary deoxyribonucleic acid probes. *Biochemistry* **19**: 2506-2514.
- Shang, M., Dougan, S. T., McFadden, G. and Greenberg, M. E. (1988).
Calcium and growth factor pathways of c-fos transcriptional activation require distinct upstream regulatory sequences. *Mol. Cell. Biol.* **8**: 2787-2796.
- Shani, M. (1986).
Tissue-specific and developmentally regulated expression of a chimeric actin-globin gene in transgenic mice. *Mol. Cell. Biol.* **6**: 2624-2631.
- Shani, M., Zevin-Sonkin, D., Sayol, O., Carmon, Y., Katcoff, D., Nudel, U. and Yaffe, D. (1981).
The correlation between the synthesis of skeletal muscle actin, myosin heavy chain and myosin light chain and the accumulation of mRNA sequences during myogenesis. *Dev. Biol.* **86**: 483-492.
- Shaw, P. E., Schroter, H. and Nordheim, A. (1989).
The ability of a ternary complex to form over the serum response element correlates with serum inducibility of the human c-fos promoter. *Cell* **56**: 563-572.
- Shirakata, M., Nabeshima, Y., Konishi, K. and Fujii-Kuriyama, Y. (1988).
Upstream regulatory region for inducible expression of the chicken skeletal myosin alkali light chain gene. *Mol. Cell. Biol.* **8**: 2581-2588.
- Slack, J. M. W., Dale, L. and Smith, J. C. (1984).
Analysis of embryonic induction by using cell lineage markers. *Phil. Trans. R. Soc. Lond. B.* **307**: 331-336.

References

- Smith, D. P. and Old, R. W. (1991).
Xenopus laevis Oct-1 does not bind to certain histone H2B promoter octamer motifs for which a novel octamer-binding factor has high affinity. *Nucleic Acids Res.* 19: 815-821.
- Smith, J. C. (1989).
Mesoderm induction and mesoderm-inducing factors in early amphibian development. *Development* 105: 665-677.
- Smith, J. C. and Slack, J. M. W. (1983).
Dorsalisation and neural induction: Properties of the organiser in Xenopus laevis. *J. Embryol. Exp. Morph.* 78: 299-317.
- Smith, J. C., Dale, L. and Slack, J. M. W. (1985).
Cell lineage labels and region-specific markers in the analysis of inductive interactions. *J. Embryol. Exp. Morph.* 89 Supplement : 317-331.
- Southern, E. M. (1975).
Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98: 503-517.
- Sternberg, E.A., Spizz, G., Perry, W. M., Vizard, D., Weil, T. and Olson, E. N. (1988).
Identification of upstream intragenic regulatory elements that confer cell-type restricted and differentiation specific expression on the mouse muscle creatine kinase gene. *Mol. Cell. Biol.* 8: 2896-2909.
- Strehler, E., Periasamy, M., Strehler-Page, M. and Nadal-Ginard, B. (1985).
Myosin light chain 1 and 3 gene has two structurally distinct and differentially regulated promoters evolving at different rates. *Mol. Cell. Biol.* 5: 3168-3182.

References

- 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 characterisation of sarcomeric actin genes expressed in Xenopus laevis embryos. J. Mol. Biol. 187: 349-361.
- Sudarwati, S. and Nieuwkoop, P. D. (1971).
Mesoderm formation in anuran Xenopus laevis (Daudin). Wilhelm Roux' Archiv. 166: 189-204.
- Tapscott, S. J., Davis, R. C., Thayer, M. J., Cheng, P.-F., Weintraub, H. and Lassar, A. B. (1988).
MyoD: a nuclear phosphoprotein requiring a myc homology region to convert fibroblasts to myoblasts. Science 242: 405-410.
- Taylor, M., Treisman, R., Garrett, N. and Mohun, T. (1989).
Muscle-specific (CAG) and serum responsive (SRE) promoter elements are functionally interchangeable in Xenopus embryos and mouse fibroblasts. Development 106: 67-78.
- Taylor, S. M. and Jones, P. A. (1979).
Multiple new phenotypes induced in 10T $\frac{1}{2}$ and 3T3 cells treated with 5-azacytidine. Cell 17: 771-779.
- Thisse, B., Stoetzel, C., Gorostiza-Thisse, C. and Perrin Schmidt, F. (1988).
Sequence of the twist gene and nuclear localisation of its protein in endomesodermal cells of early Drosophila embryos. EMBO J. 7: 2175-2183.

References

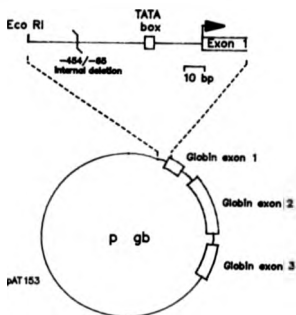
- Tobin, S. L., Zulanf, E., Sanchez, F., Craig, E. A. and McCarthy, B. O. (1980).
Multiple actin related sequences in the Drosophila melanogaster genome. Cell **19**: 121-131.
- Treisman, R. (1985).
Transient accumulation of c-fos RNA following serum stimulation requires a conserved 5' element and c-fos 3' sequences. Cell **42**: 889-902.
- Treisman, R. (1986).
Identification of a protein binding site that mediates transcriptional response of the c-fos gene to serum factors. Cell **46**: 567-574.
- Treisman, R. (1987).
Identification and purification of a polypeptide that binds to the c-fos serum response element. EMBO J. **6**: 2711-2717.
- Villares, R. and Cabrera, C. V. (1987).
The acheate-acute gene complex of D. melanogaster: conserved domains in a subset of genes required for neurogenesis and their homology to myc. Cell **50**: 415-424.
- 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. USA. **75**: 1106-1110.

References

- Vandekerckhove, J. and Weber, K. (1978c).
The amino acid sequence of actin from chicken skeletal muscle actin and chicken gizzard smooth muscle actin. *FEBS Letts.* **102**: 219-222.
- Vandekerckhove, J. and Weber, K. (1979).
The complete amino acid sequence of actins from bovine aorta, bovine heart, bovine fast skeletal muscle and rabbit slow skeletal muscle. *Differentiation* **14**: 123-133.
- 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. Biochem.* **113**: 595-603.
- Vandekerckhove, J. and Weber, K. (1981b).
Diversity of expression of non-muscle actin in amphibia. *J. Mol. Biol.* **152**: 413-426.
- Walsh, K. and Schimmel, P. (1987).
Two nuclear factors compete for the skeletal muscle actin promoter. *J. Biol. Chem.* **262**: 9429-9432.
- Walsh, K. and Schimmel, P. (1988).
DNA binding site for two DNA binding factors is important for expression in muscle cells. *Mol. Cell. Biol.* **8**: 1800-1802.
- Walsh, K. (1989).
Cross binding of factors to functionally different promoter elements in the c-fos and skeletal actin genes. *Mol. Cell. Biol.* **9**: 2191-2201.

References

- Wang, X.-M., Tsay, H.-J. and Schmidt, J. (1990).
Expression of the acetylcholine receptor delta subunit in differentiating chick muscle cells is activated by an element that contains two 16-bp copies of a segment of the alpha-subunit enhancer. *EMBO J.* **9**: 783-790.
- 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. USA.* **73**: 2018-2022.
- Wilson, C., Cross, G. S. and Woodland, H. R. (1986).
Tissue-specific expression of actin genes injected into Xenopus embryos. *Cell* **47**: 589-599.
- Wright, W. E. (1984).
Control of differentiation in heterokaryons and hybrids involving differentiation defective myoblast variants. *J. Cell Biol.* **98**: 436-443.
- Wright, W. E., Sassoon, D. A. and Lin, V. K. (1989).
Myogenin, a factor regulating myogenesis, has a domain homologous to MyoD. *Cell* **56**: 607-617.
- Yaffe, D., Nudel, U., Mayer, Y. and Neuman, S. (1985).
Highly conserved sequences in the 3' untranslated regions of mRNAs coding for homologous proteins in distantly related species. *Nucleic Acids Res.* **13**: 3723-3737.
- Yutzey, K. E., Kline, R. L. and Konieczny, S. F. (1989).
An internal regulatory element controls Troponin I gene expression. *Mol. Cell. Biol.* **9**: 1397-1405.
- Ziff, E. B. (1990).
Transcription factors: a new family gathers at the cAMP response site. *Trends In Genetics* **6**: 69-72.

A.**B.**