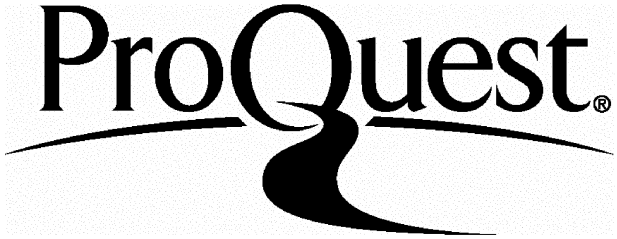
Comparative Analysis of Regulatory Elements of the *Myf-5* Gene

Oliver Coutelle

A thesis presented for the
Degree of Doctor of Philosophy (Ph.D.)
Department of Biochemistry
University College London

1998
National Institute for Medical Research,
Mill Hill, London.



Acknowledgements

The work presented here has been supported by an MRC Studentship. I am grateful to Prof. Peter Rigby for accepting me as a student, and for giving me the freedom to study the pufferfish. I thank Dr Sydney Brenner for letting me work in his Cambridge laboratory and for making available his *Fugu* cosmid library. Particular thanks to Dr. Samuel Aparicio from his lab as a friend and for helpful advice, scientific and otherwise.

My greatest respect to Dr. Dennis Summerbell from our laboratory at NIMR for his brilliant teaching and often needed encouragement, as well as his patience in discussing my transgenic experiments. I shall not forget Dennis' lectures in mouse anatomy which have excited even a rather reluctant student as myself (to a degree) and made me look at my work from a different perspective.

Thanks also for many enjoyable discussions to Dr. Rajeev Gupta and Jonathan Gilthorpe, who shared many of the ups and downs with me. I am grateful to Marie Vandromme, Ed Stanley, Dave Cox, Peter Ashby, Chandi Halai, Debbie Duthie and Jacky Smith. I gratefully acknowledge the technical staff in the SPF unit, particularly Hannah Boyes and Zoe Webster for looking after our mice.

Most of all, I thank my parents, who made all this possible.

Abstract

The work presented here addresses the question of how skeletal muscle formation is initiated in the mouse by dissecting the regulatory mechanisms that control the myogenic regulatory factor *Myf-5*. *Myf-5* is expressed in the dorsal somite from E8, before the other MRFs, *myogenin*, *MRF4* and *MyoD*, become activated. In the mouse *Myf-5* is located 8.5kb downstream of *MRF4*. Previous results have shown that dispersed over the intergenic region and intragenic regions are the regulatory elements involved in directing *Myf-5* expression to the different anatomical subdomains that make up its complete expression pattern. The regulatory element(s) controlling the dorsal somite expression of *Myf-5* is contained in the intergenic region while ventral somite expression depends on elements in the *Myf-5* gene itself. Because of the large size of this region I have isolated the *MRF4* and *Myf-5* genes of the teleost *Fugu rubripes*, which has a genome eight times smaller than that of the mouse. Although synteny is conserved in *Fugu* and the intergenic distance is only 3kb, noncoding sequence including the introns is poorly conserved. Focusing on the *Myf-5* gene itself, sequence comparison between the mouse and human *Myf-5* genes was employed sucessfully to eliminate more than 60% of the intron sequence by identifying conserved regions in the 3'half of each of the *Myf-5* introns which, together with the 3'UTR can activate reporter gene expression in the ventral posterior part of the somites. EMSA analysis with embryonic protein extracts revealed several protein binding regions within the conserved intron fragments and subsequent transgenic analysis showed not only that separate genomic regions control individual anatomical domains of *Myf-5* expression, but that within these regions multiple binding sites are found, adding a further level of complexity to the regulation of *Myf-5*. Analysis of the *Fugu Myf-5* gene in transgenic mice showed remarkable similarities with the expression pattern of *Myf-5* in another teleost, the zebrafish *Danio rerio*. Both are expressed in the presomitic mesoderm, as well as the somites, suggesting that the expression of the *Fugu* transgene is a reflection of its native expression pattern.

Table of Contents

Acknowledgements 2
Abstract 3
List of Figures 9
List of Tables 10
Abbreviations 11

Chapter 1 - Introduction

1. Germlayer Formation in the Mouse 13
2. Germlayer Formation in Teleost Fish 17
3. Mesodermal Patterning 18
4. Somitogenesis 19
5. Somite Patterning 22
    5. 1. Dorsoventral Patterning 22
    5. 2. Sonic Hedgehog and Wnt Signalling 23
    5. 3. Mediolateral Somite Patterning 26
6. Somites in Teleost Fish 28
7. The Myogenic bHLH Factors 31
8. The Myogenic Cascade 32
9. MRF Null Mutations 35
    9. 1. *MyoD* Null Mutants 35
    9. 2. *Myf-5* Null Mutants 36
    9. 3. *Myf-5/MyoD* Double Mutants 36
    9. 4. Myogenin Null Mutants 37
    9. 5. *MRF4* Null Mutants 37
10. What Regulates *MyoD* and *Myf-5*? 38
11. Could Myogenesis Be Repressed? 40
12. The Pufferfish as a Model Organism 42
    12. 1. Genome Size and Complexity 42
    12. 2. The Small Genome of *Fugu rubripes* 43
    12. 3. The *Fugu* Genome Never Acquired 'Junk-DNA' 43
    12. 4. Small Is Beautiful 45
    12. 5. . . . Or is it? 45
    12. 6. Studies on Developmentally Regulated Genes in *Fugu* 46
13. Separate Regulatory Elements Control Sub-Domains of *Myf-5* Expression 47

Chapter 2 - Materials and Methods

1. Solutions and Reagents....49
2. DNA Manipulations....54
2. 1. Phenol Extraction of DNA....54
2. 2. Ethanol Precipitation of DNA....54
2. 3. PEG Precipitation of DNA....54
2. 4. Endonuclease Digestion of DNA....54
2. 5. DNA Fragmentation by Sonication....54
2. 6. Dephosphorylation of DNA....54
2. 7. Repair of DNA Ends....55
2. 8. Oligonucleotide Primer Synthesis....55
2. 9. Agarose Gel Electrophoresis and Documentation....56
2. 10. DNA Fragment Isolation....56
3. DNA Cloning....56
3. 1. DNA Ligations....56
3. 2. Transformation of Competent Cells....57
3. 3. Recombinant Screening....57
4. Isolation of DNA and RNA....58
4. 1. Plasmid Isolation by Alkaline Lysis "Miniprep"3....58
4. 2. Plasmid Isolation by CsCl "Maxiprep"3....58
4. 3. Genomic DNA Isolation from Zebrafish Embryos....59
4. 4. Cosmid DNA Isolation....59
4. 5. Isolation of RNA....59
4. 6. cDNA Synthesis....60
5. Southern Analysis....60
5. 1. Southern Blotting....60
5. 2. Southern Hybridisation Probes....60
5. 3. Southern Hybridisation....61
6. Northern Analysis....61
6. 1. Northern Blotting....61
7. Library Screening....62
7. 1. *Fugu* Cosmid Library Screening....62
7. 1. 1. Sequence Scanning of Cosmid Clones....63
7. 2. Zebrafish cDNA Library Screening....63
7. 2. 2. Filter Preparation and Hybridisation....64
8. Zebrafish *Myf-5* Cloning by Degenerate PCR....65
9. DNA Sequencing....65
9. 1. Sequencing on the ABI 373....65

9. 2. Sequencing on the ABI 377 66
10. Generation and Analysis of Transgenic Mice 66
10. 1. Transgenic Reporter Constructs 66
10. 1. 1. Mouse *Myf-5* Reporter Constructs 66
10. 1. 2. *Fugu Myf-5* Reporter Constructs 69
10. 1. 3. Hybrid *Myf-5* Reporter Constructs 69
10. 2. Preparation of DNA for Microinjection 69
10. 3. Transgenic Procedure 69
10. 4. Transgene Detection by PCR 71
11. Histological Studies 71
12. DNA Electrophoretic-Mobility Shift Assay (EMSA) 72
12. 1. Probe Labelling and Purification 72
12. 2. Preparation of Protein Extracts 72
12. 3. EMSA Conditions 73
13. In Situ Hybridisation of Zebrafish Embryos 73
13. 1. Riboprobe Synthesis 73
13. 2. Preparation of Embryos 74
13. 3. Hybridisation 74
13. 4. Posthybridisation Treatment 74

Chapter 3 - Comparative Sequence Analysis of the Vertebrate *Myf-5* /MRF4 Loci

1. Introduction 76
2. Isolation and Characterisation of *Fugu MRF4* and *Myf-5* 79
2. 1. A *Fugu* Cosmid Contains Both *MRF4* and *Myf-5* 79
2. 2. The Genomic Structures of *Fugu MRF4* and *Myf-5* 82
2. 3. Interspecies Comparison at the DNA Level 85
2. 4. Interspecies Comparison at the Protein Level 85
2. 5. CpG Islands and Repeat Elements 89
3. Pairwise DNA Sequence Analysis Between *Fugu* and Mouse 90
4. Pairwise Comparisons Between Mouse, Human and Bovine 91
5. Summary 95

Chapter 4 - EMSA Analysis of the *Myf-5* Introns

1. Introduction 99
2. Fragment A 102
3. Fragment B 102
4. Fragment C 102
5. Fragment D 102

6. Fragment E 106
7. Fragment F 106
8. Fragment G 106
9. Summary 108

Chapter 5 - Transgenic Analysis of the Regulatory Role of the Mouse *Myf-5* Introns

1. Introduction 109
2. Dissection of Regulatory Elements in *Myf-5* using Transgenic Mice 110
2.1. Construct p12 110
2.2. Construct p1 115
2.3. Construct p2 115
2.4. Construct pUTR 116
2.5. Construct p12UTR 116
2.6. Construct p1UTR 121
2.7. Construct p2UTR 121
2.8. Construct pCFUTRR 126
2.9. Construct pCFH 129
2.10. Construct pCFA 132
2.11. Construct pCFHΔ 132
2.12. Construct pΔAH 132
2.13. Construct pΔCAH 137
2.14. Construct pΔGAH 137
2.15. Construct pCFAΔ 137
2.16. Construct pΔCPH 138
2.17. Construct pΔFPH 138
2.18. Construct pΔGPH 138

Chapter 6 - *Myf-5* in Zebrafish and *Fugu*

1. Introduction 142
2. Cloning of the Zebrafish *Myf-5* homologue 142
3. Expression of Z*Myf-5* in zebrafish embryos 143
4. Transgenic Analysis of *Fugu Myf-5* 147
4.1. Construct *Fugu* p12UTR 147
4.2. Construct *Fugu* p12 152
4.3. Construct pZFM 152
4.4. Construct pZFF 157
4.5. Construct *Fugu* pUTR 157
5. Summary 157

Chapter 7 - Discussion

1. Introduction 161
2. *Fugu MRF4* and *Myf-5* Form a Syntenic Linkage Group 162
3. *Fugu* is Not a Good Model for *MRF4* and *Myf-5* 163
4. Regions in the *Myf-5* Introns Conserved Between Human and Mouse Regulate ..... Ventral Somitic Expression 164
5 Multiple Binding Sites Located in the Mouse *Myf-5* Introns 165
6. Transgenic Analysis Reveals Complex Regulatory Mechanism for *Myf-5*. 166
7. Conserved Regulatory Elements in Both Introns and the UTR of *Myf-5* are ..... Required for Ventral Posterior Somite Expression in the Mouse. 168
8. Separate Elements Appear to Control Intensity and Spatial Distribution of Transgene Expression. 169
9. Fragments D, F and G, but not C, are Involved in Ventral Somitic Expression ..... of *Myf-5*. 170
10. A 'Branchial Arch Element' in the *Myf-5* Gene? 170
11. The 3' UTR of the *Fugu Myf-5* Gene Drives Expression in the Somites and the Presomitic Mesoderm. 171
12. Conserved Element in Intron 1. 174
13. *Myf-5* is Expressed in Lateral Presomitic Cells 174
14. Outlook 176

Appendix

I. Oligonucleotides 179
II. *Fugu MRF4* and *Myf-5* Genes: Complete Coding Sequence 181
III. *Fugu MRF4* and *Myf-5*: Genomic Contig 185
IV. Human *Myf-5* Gene: Complete Coding Sequence 187
V. Zebrafish *Myf-5* Gene: Complete Coding Sequence 187
IV. Mouse *Myf-5* Gene: Complete Coding Sequence 189

Bibliography

..... 191

List of Figures

Figure 1: Early Development of the Mouse Embryo. 16
Figure 2: Epiboly in Teleost Fish. 16
Figure 3: Somite Differentiation. 21
Figure 4: Somite Patterning 25
Figure 5: Somites in Teleost Fish 30
Figure 6: Temporal Expression of the Myogenic Factors in the Mouse 34
Figure 7: The Japanese Pufferfish, *Fugu rubripes.* 44
Figure 8A: Regulatory Elements in the *MRF4/Myf-5* Region. 78
Figure 8B: Construct HMZ17 77
Figure 8C: Construct MFGZ 77
Figure 9A: The Genomic Organisation of the *Fugu MRF4* and *Myf-5* Genes 81
Figure 9B: Selected Restriction Sites of the *Fugu* Region 81
Figure 10: Schematic Comparison of Mouse and *Fugu Myf-5* and *MRF4* 84
Figure 11A: Clustal Alignment: *MRF4* 87
Figure 11B: Phylogenetic Tree: *MRF4* 87
Figure 11C: Clustal Alignment: *Myf-5* 88
Figure 11D: Phylogenetic Tree: *Myf-5* 88
Figure 12A: Dot Plot Analysis of the *MRF4/Myf-5* Region 93
Figure 12B: Dotplot Analysis: *Myf-5* Human vs Bovine 94
Figure 12C: Dotplot Analysis: *Myf-5* Human vs Mouse 94
Figure 13A: Dotplot Analysis: Mouse vs Human *Myf-5* Intron 1 97
Figure 13B: Dotplot Analysis: Mouse vs Human *Myf-5* Intron 2 97
Figure 14A: Size and Location of EMSA Probes 101
Figure 14B: Subdivision of EMSA Fragment D 101
Figure 14C: Origin of the Mouse Tissues for EMSA Extracts 101
Figure 15A: Electrophoretic Mobility Shift Assays: Fragment A 104
Figure 15B: Electrophoretic Mobility Shift Assays: Fragment B 104
Figure 15C: Electrophoretic Mobility Shift Assays: Fragment C 104
Figure 15D: Electrophoretic Mobility Shift Assays: Fragment D 104
Figure 15E: Electrophoretic Mobility Shift Assays: Fragment E 105
Figure 15F: Electrophoretic Mobility Shift Assays: Fragment F 105
Figure 15G: Electrophoretic Mobility Shift Assays: Fragment G 105
Figure 16A-B: Construct p12 114
Figure 16C: Construct HMZ17 114
Figure 17A-C: Construct p1 118
Figure 17D-E: Construct p2 118

Figure 18A-B: Construct pUTR 120
Figure 18C-D: Construct p1UTR 120
Figure 19A-J: Construct p12UTR - Fritz 123
Figure 20A-D: Construct p2UTR 125
Figure 21A-G: Construct pCFUTRR 128
Figure 22A-B: Construct pCFH 131
Figure 22C-D: Construct pCFA 131
Figure 23A: Construct pCFHΔ 134
Figure 23B-C: Construct pCFAΔ 134
Figure 24A-B: Construct pΔAH 136
Figure 24C: Construct pΔCAH 136
Figure 24D: Construct pΔGAH 136
Figure 25A-B: Construct pΔCPH 140
Figure 25C-D: Construct pΔFPH 140
Figure 26A-B: Construct pΔGPH 141
Figure 27A-E: Z*Myf-5* Expression Pattern 145
Figure 28A-H: Construct *Fugu* p12UTR 151
Figure 29: *Myf-5* Homologues: Size Comparison 153
Figure 30: Dotplot: Zebrafish vs *Fugu Myf-5* 154
Figure 31A-C: Construct pZFM 156
Figure 32A-C: Construct *Fugu* pUTR 156
Figure 33A-D: Construct pZFF 159

List of Tables

Table 1: Exon and Intron Sizes of *MRF4* and *Myf-5* 84
Table 2: Summary of EMSA Results for *Myf-5* Intron Fragments 107
Table 3A: Summary of the Mouse Transgenic Results 111
Table 3: Mouse *Myf-5* Reporter Constructs 112
Table 4A: Summary of the *Fugu* and Hybrid Transgenic Results 148
Table 4: *Fugu* and Hybrid *Myf-5* Reporter Constructs 149
Table 5: Size of Exons and Introns of *Myf-5* Homologues 153

Abbreviations

AP	anterior-posterior
ATP	adenosine 5'-triphosphate
bp	base pair
BSA	bovine serum albumin
bHLH	basic -helix-loop-helix protein structure motif
cDNA	complementary DNA
cpm	counts per minute
C-terminus	carboxy terminus
dH_2O	distilled water
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
dATP	deoxyadenosine 5'-triphosphate
dCTP	deoxycytidine 5'-triphosphate
dGTP	deoxyguanosine 5'-triphosphate
dNTP	deoxyribonucleoside 5'-triphosphate
dTTP	deoxythymidine 5'-triphosphate
ddNTP	dideoxyribonucleoside 5'-triphosphate
dpc	days *post coitum*
DTT	dithiothreitol
DV	dorsal-ventral
EDTA	ethylenediaminetetraacetic acid
EMSA	electrophoretic-mobility shift assay
HEPES	N-2-hydroxyethylpiperazine-N'-ethanesulphonic acid
hpf	hours post fertilisation
kb	kilo base (base pairs)
KD	kilo dalton
MBq	mega bequerel
M	molar
Mb	mega base (base pairs)
µM	micro molar
mM	milli molar
µg	micro gram
µl	micro litre
mg	milli gram
MOPS	3-N-morpholinopropanesulphonic acid
MRF	Myogenic Regulatory Factor

MYF	MYogenic Factor
ng	nano gram
ml	milli litre
NP-40	Nonidet P-40
N-terminus	amino terminus
OAc	acetate
OD	optical density
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PMSF	phenylmethylsulphonyl fluoride
RNase	ribonuclease
rpm	revolutions per minute
RT	reverse transcriptase
SSC	sodium chloride/ sodium citrate
SDS	sodium dodecyl sulphate
SPF	specific pathogen free
Tris	2-amino-2-(hydroxymethyl)-1,3-propanediol
TBE	Tris-borate-EDTA buffer
TE	Tris-EDTA buffer
u	units
UV	ultra violet
W	watts
%(w/v)	grams per 100ml
%(v/v)	ml per 100ml
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

Chapter 1

Introduction

An important task in the field of vertebrate embryology is to elucidate the mechanisms by which mesodermal cells acquire their developmental fates. The determination of specific cell fates underlies the most fundamental processes of development. In vertebrates, some of the early steps in determining cell fate take place during gastrulation when cells become committed to each of the three primary germ layers endoderm, ectoderm and mesoderm. Mesoderm specification begins with the ingression of epiblast cells through the primitive streak followed by commitment and allocation to the different mesodermal lineages: midline cells generate axial structures, notochord and prechordal plate, paraxial mesoderm becomes progressively segmented into the somites and also gives rise to the head mesoderm, while lateral mesoderm forms the splanchnopleure and somatopleure. The somites of the paraxial mesoderm give rise to the axial skeleton, trunk muscle, some head and neck muscle, some bones and muscles of the skull and epidermis of the skin. The work presented here is concerned with the regulation of the myogenic transcription factor *Myf-5* and its role in the paraxial mesoderm and in the commitment of cells to the myogenic lineages that give rise to skeletal muscle. Before discussing the process of myogenesis in which mesodermal cells acquire muscle fate, the establishment of the three germ layers is outlined briefly with reference to the different model organisms studied as part of project, including the mouse and teleost fish.

1. Germlayer Formation in the Mouse

Some of the most dramatic stages of differentiation of the embryo take place in the process of gastrulation by establishing the three germ layers. Various aspects of gastrulation have been studied in diverse organisms, including insects, birds, amphibia, fish and mammals. The mouse fate map exhibits topological similarity with those of chick, amphibia and teleost fish, indicating that even though the physical morphology of gastrulation differs between organisms, some of the basic principles are conserved.

Depending on the particular organism studied, however, the actual process of generating the germlayers varies considerably.

Initially the undifferentiated cells of the mouse embryo divide to form a cluster of cells called the morula. Up to the 16 cell stage the morula cells retain their equipotency and can give rise to complete embryos. After further cell divisions the potency of these cells becomes gradually restricted. Tight junctions form between the outer cells while a fluid filled cavity (blastocoel) develops within the embryo. The cells in contact with the outside become trophectoderm while the cells with no contact to the outside become the inner cell mass (ICM). Cells of the ICM at the interface with the blastocoel give rise to the primitive endoderm and cells trapped between the primitive endoderm and the trophectoderm (black) form the epiblast (red) (see Fig. 1, stage E4.5). Around day 5 of development, the blastula stage mouse embryo becomes implanted into the uterus and part of the trophectoderm proliferates and pushes the epiblast ahead of itself (Fig. 1, stage E5.5). The epiblast adopts epithelial character and becomes the primitive mesectoderm surrounding the proamniotic cavity (green) (Fig. 1, stage E6). At around E7 of development the primitive streak is formed from a subset of mesectoderm cells near the interface of the extraembryonic and embryonic halves of the egg cylinder. The epithelial continuity of mesectoderm is lost in the streak as cells delaminate to form the three embryonic germlayers: ectoderm, mesoderm and endoderm. Fate mapping studies of mesectoderm cells using lineage tracers (Gardner 1982, 1983) show that these cells give rise to mesoderm as they ingress through the streak and some of the cells intercalate into the visceral endoderm layer to give rise to embryonic endoderm (Fig. 1, see stage E7) and also to extraembryonic tissues (Gardner and Rossant, 1979). Thus cells from the mesectoderm (epiblast) are thought to contribute to all three germ layers of the developing foetus.

At the anterior end of the streak the node is formed by a group of cells that appears to have organising capacity similar to Hensen's node in the chick embryo and Spemann's organiser in *Xenopus* and is likely to be involved in patterning of axial structures. It has been suggested that the node also contains a proliferative centre from which the notochord and floor plate cells of the neural tube originate (Selleck and Stern, 1991; Sulik *et al.*, 1994). Indeed, transplantation of chick node into zebrafish embryos generates a secondary axis, suggesting that the underlying signals are evolutionarily conserved (Hatta and Takahashi, 1996). Formation of the notochord divides the mesodermal layer into the paraxial mesoderm on either side of the AP axis and marks the onset of neurulation. Further differentiation of the paraxial mesoderm leads to its successive segmentation and the formation of the somites, from which most skeletal muscle is derived. In fish embryos, muscle is also formed from paraxial mesoderm but the process in which paraxial mesoderm is established is not the same.

Figure 1: Early Development of the Mouse Embryo. At around **E4.5** the initially equipotent cell mass divides into epithelial trophectoderm (black) and inner cell mass (red) of the blastula with its fluid filled cavity the blastocoel (yellow).
Around **E5** the embryo implants into the uterus, the trophectoderm of the ectoplacental cone pushes the ICM (epiblast) and the overlaying primitive endoderm (white) inside the egg cylinder. Within the epiblast cells the proamniotic cavity forms (green) at **E6**, followed by the formation of the primitive streak around **E7** at the interface with the extraembryonic ectoderm (striped). Cells delaminating from the mesectoderm (epiblast) ingress through the streak between the mesectoderm and the visceral endoderm and form two new layers the embryonic mesoderm and endoderm and also contribute to the extraembryonic endoderm. At the most anterior end of the streak where the node is formed there is no visceral endoderm overlaying the streak.

Figure 2: Epiboly in Teleost Fish.
2A: Schematic representation of the blastuala stage zebrafish embryo. The enveloping layer (red) is the outermost epidermis of the embryo, enveloping the deep cells (DEL) shown in gray, which ultimately give rise to all three germ layers. At the margin DEL cells leave the blastoderm and contribute to the syncitial (multinucleated) yolk cell (arrows).
2B: At 80% epiboly the synctial layer (blue) moves toward the vegetal (yolk rich) pole. The EVL follows the syncitial layer toward the vegetal pole but remains on the outside. At the margin of the blastoderm (germ ring), the deep cells involute (arrows) and give rise to the hypoblast (shaded gradient) from which the presumptive mesoderm and endoderm form. The outermost, noninvoluting DEL cells become epiblast from which the presumptive ectoderm is formed. Additional cell movement towards the dorsal midline (black arrow) of the embryo leads to a thickening of the germ ring, the embryonic shield (ES).

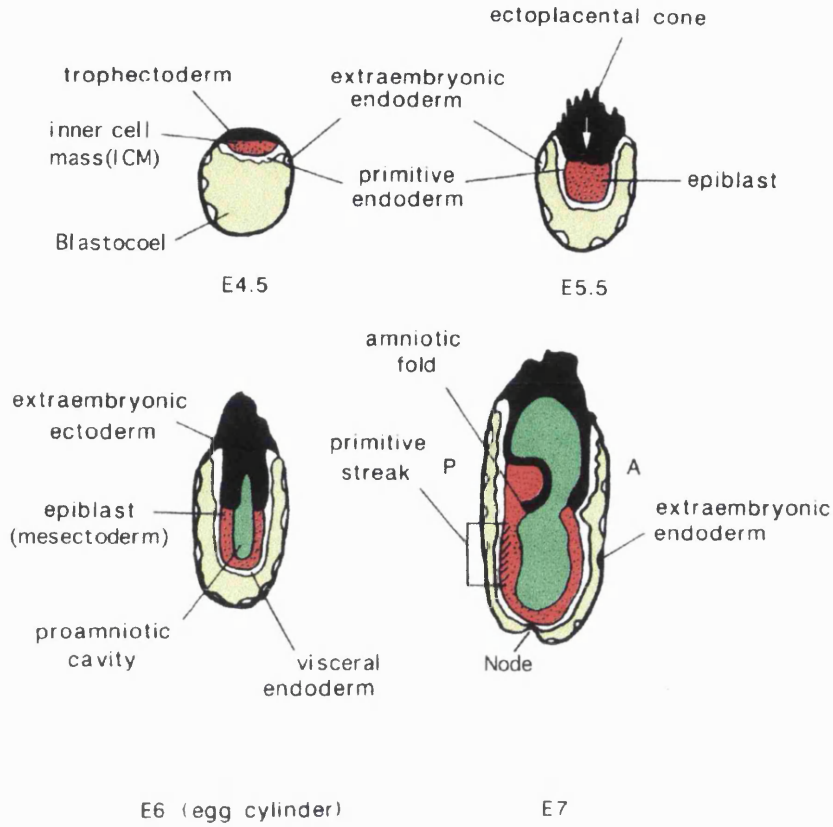


Fig.1: Early development of the mouse embryo, adapted from Hogan et al., 1986.

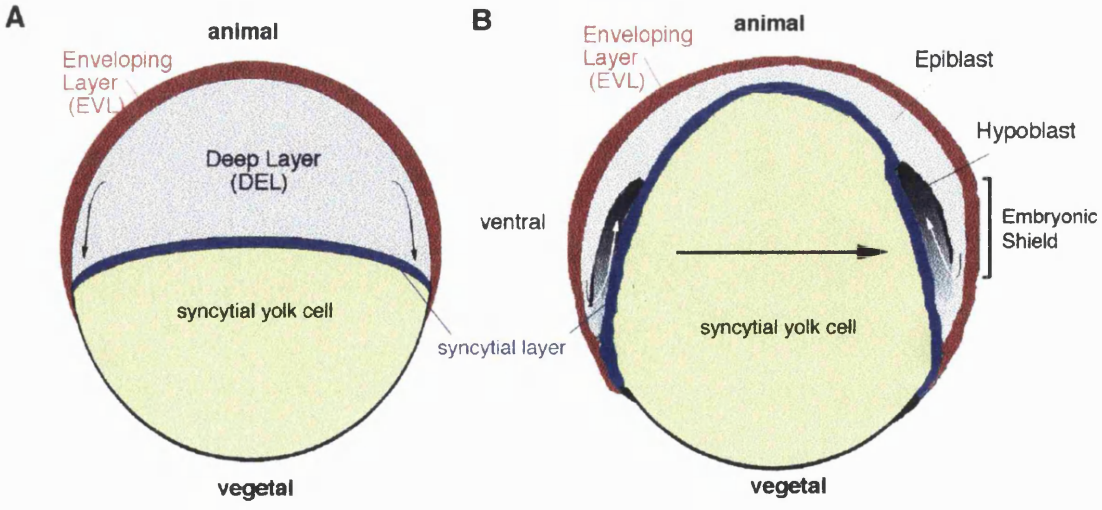


Fig. 2: Epiboly in teleost fish

2. Germlayer Formation in Teleost Fish

The process of gastrulation in amphibia (e.g. *Xenopus*) and teleost fish (e.g. pufferfish, zebrafish) shares many features with higher vertebrates. The most extensively studied teleost is the zebrafish, *Danio rerio*. The main stages of gastrulation in these small translucent embryos can be directly observed.

The fertilised egg at first divides synchronously and these intital cleavages are incomplete (meroblastic) such that the furrows do not pass through the yolk rich region of the egg thus giving rise to a giant uncleaved yolk cell. In zebrafish cleavages generate two populations of blastoderm cells: the enveloping layer (EVL) forming the outer epidermis (periderm) of the embryo (Bouvet, 1976) and the deep cell layer (DEL) beneath the EVL that gives rise to the major ectodermal, mesodermal and endodermal portions of the embryo (Martindale *et al.*, 1987). At the midblastula stage DEL cells located near the margin leave the blastoderm and contribute their nuclei to the syncitial multinucleated yolk cell (Fig. 2A). During gastrulation, blastodermal cells gradually move over the yolk cell towards the vegetal (yolk rich) pole, enveloping the yolk cell which is destined for digestion in the prospective gut of the embryo. This process, for which there is no counterpart in higher vertebrate development, requires spreading of the blastoderm layer and is driven by radial intercalation of blastomeres (epiboly) accompanied by concomitant thinning of the blastoderm. In zebrafish, during this stage extensive cell mixing within the DVL takes place and descendants from single blastomeres move to diverse locations and will eventually give rise to diverse cell types (Kimmel and Warga, 1987). Thus fate maps of pregastrulation zebrafish embryos are difficult to obtain. Involution can be regarded as a similar event as invagination during development of higher vertebrates (Trinkaus, 1988). Involution marks the onset of mesoderm formation and takes place at 50% epiboly, when half of the yolk cell is covered by the enveloping blastoderm. In teleosts involution does not initiate at the dorsal side of the embryo (as it does in amphibia) but begins more or less simultaneously around the circumference of the blastoderm. The involuted internal layer of blastoderm cells forms the hypoblast and the outer blastoderm DEL layer becomes the epiblast (see Fig. 2B). Convergent movement of cells from lateral positions to the dorsal midline of the gastrula leads to formation of the embryonic shield (Fig. 2B), a thickening of the germ ring with some similarity to the primitive streak or the blastopore of amniote embryos and amphibia, respectively. Fate mapping studies in zebrafish have shown that cells from the most lateral (outermost) layer of the epiblast do not involute and these cells give rise to ectodermal derivatives. In contrast, the fate of the involuting cells as described for the streak in amniotes, dependent on when they enter the hypoblast. Cells from the region near the blastoderm margin enter early and give rise to endoderm while cells at more lateral positions give rise to lateral and paraxial mesoderm (Kimmel *et al.*,

1990). The specification of mesodermal fate at the molecular level is a complicated and as yet not fully understood process, in which temporal and spatial distribution of inductive signalling molecules play a major role.

3. Mesodermal Patterning

It is not clear if a homogeneous population of mesoderm cells ever exists, or if cells ingressing through the primitive streak are already predetermined to become one of the various mesoderm derivatives. Indeed, it appears that the first cells ingressing through the primitive streak adopt the fate of lateral mesoderm, while the cells that follow go on to develop into the paraxial mesoderm. Thus, mesodermal fate could be determined by the timing of ingression through the streak alone, but it is likely that additional signalling molecules are involved in patterning of the mesodermal cells. Candidate patterning molecules, some with graded or partially restricted, some with uniform expression along the AP axis of the primitive streak have been found, including peptide growth factors, signalling molecules and transcription factors of the Zinc-finger and homeodomain-families. Their complex expression patterns suggest combinatorial effects and indicate a degree of overlap and redundancy in these molecules (Tam and Trainor, 1994). Since dye labelled cells of the rostral presomitic mesoderm of mouse and chick embryos can give rise to lateral plate mesoderm and endothelium, they can not be totally committed to a particular cell fate until after somite segmentation (Bagnall, 1988; Beddington and Martin, 1989; Selleck and Stern, 1991, 1992; Veini and Bellairs, 1991). Cells may be prepatterned as they ingress through the streak and receive further reinforcing signals depending on the distances from the primitive streak in order to adopt the fate of paraxial or lateral mesoderm. FGF-3, FGF-4 and FGF-5 are known to be expressed in the primitive streak and combinations of FGFs may act to specify the different mesodermal derivatives passing through the primitive streak. However, targeted mutations for each of these genes do not show defects in early gastrulation (Mansour *et al.,* 1993; Hebert *et al.,* 1994; Feldman *et al.,* 1995), probably because they can partly substitute for each other's function. In contrast, homozygous mutations of the FGFR-1 receptor result in severe mesodermal patterning defects (Yamaguchi *et al.,* 1994), suggesting that specific domains of expression of the FGFR-1 receptor rather than the FGF ligands play an essential role in determining the ability of various cells in the streak to respond to overlapping FGF signals. Given the complexity of the patterning mechanisms a combination of both ligand and receptor distribution seems most likely.
Although the mechanisms underlying the fate of cell populations passing through the streak is at present poorly understood, it is clear that epiblast cells entering the primitive streak become allocated to different mesodermal populations. Mesoderm of the midline

gives rise to notochord. Intermediate mesoderm gives rise to the kidney primordium and lateral mesoderm provides much of the mesenchyme involved in development of the viscera and the cartilage of the limb buds enveloped by surface ectoderm. Paraxial mesoderm along both sides of the axial structures gives rise to the somites from which the vertebrae and ribs, skeletal muscle and the dermis of the back are derived.

4. Somitogenesis

Amongst the first segmental structures formed in the mouse are the somites of the paraxial mesoderm. Somites are continuously formed between E8 and E14 at the rostral edge of the presomitic mesoderm and they are displaced anteriorly as the next somite is born. The necessary cell mass is acquired by continuous recruitment of cells to the caudal end of the presomitic mesoderm from different sources including cells ingressing through the primitive streak contributing mainly to anterior somites, while tail bud mesenchyme contributes to posterior somites (Beddington, 1981; Tam and Trainor, 1994). Accordingly, in the absence of primitive streak or caudal tissue, the presomitic mesoderm can only give rise to a limited number of somites (Packard, 1978; Tam, 1986). At the anterior end of the paraxial mesoderm, in the cranial region and at the posterior end of the presomitic mesoderm, additional segmental structures, termed somitomers, have been identified by electron microscopy but according to Tam and Beddington (see Tam and Trainor, (1994)), lineage tracing experiments indicate that the somitomers of the paraxial mesoderm are unlikely to be the direct precursors of the somites. In the mouse, about 65 pairs of somites are sequentially generated in the process of somitogenesis, one pair every 1.5 hours. In zebrafish, roughly 30 somite pairs are formed, starting at 10 hours postfertilisation (hpf), one pair every 20-30 minutes.

In mammals and birds, the newly formed somites are epithelial spheres of cells surrounding mesenchymal cells within a central cavity, the somitocoel (Fig. 3A). The specification of lineages in the somite is linked to the establishment of dorsoventral polarity shortly after segmentation. Reorienting the dorsoventral polarity of the youngest three somites has shown that the dorsoventral axis is not determined until about three hours after segmentation. In contrast, rostrocaudal identity is already determined when the somites are formed (Aoyama and Asamoto, 1988). Following segmentation, the spherical shape of the somites disintegrates into the ventral sclerotome and a dorsal dermomyotome compartment (Fig. 3B). The sclerotome is formed by cells from the

Figure 3: Somite Differentiation.

3A: Early epithelial somite budded off from the presomitic mesoderm, adjacent to the axial structures, neural tube and notochord.

3B: The epithelial somite becomes divided into the dorsal dermomyotome and ventral sclerotome compartments: Following axial signals, the somite breaks open and mesenchymal sclerotome migrates from the ventromedial quadrant of the somite medially, dorsally the epithelial character is retained in the dermomyotome.

3C: Myotome forms as cells move over the ventral and dorsal margins of the dermomyotome. *Myf-5* is first expressed in spatially separated ventral and dorsal myotomes.

3D: The somitic bud derived from the lateral dermomyotome migrates laterally to contribute to the ventrolateral body wall musculature.

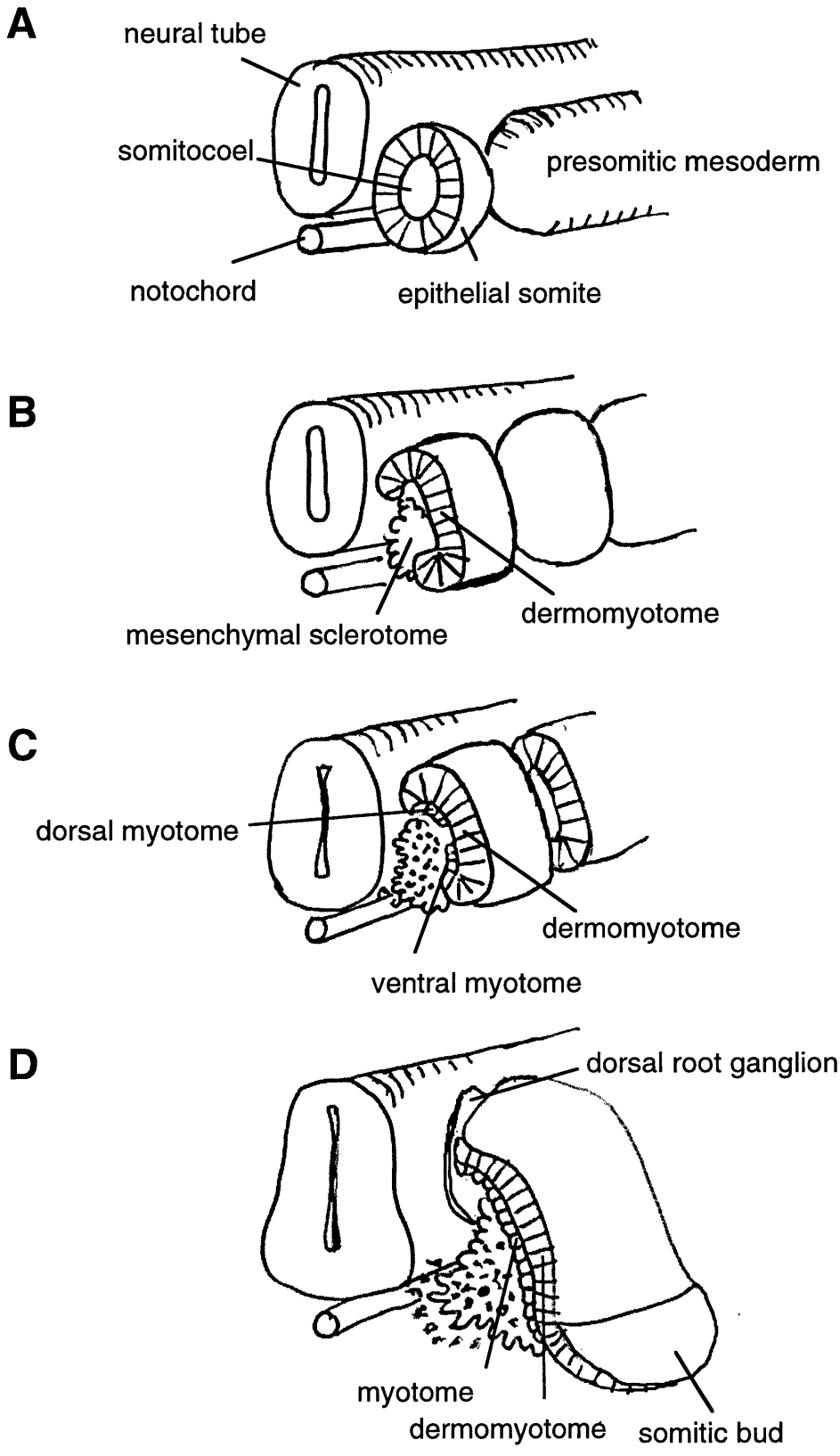


Figure 3: Somite Differentiation

Redrawn and modified from
Spoerle et al., Development 122, 1996

ventral wall of the somite and the somitocoel, giving rise to the precursor population for the skeleton and cartilaginous tissues, while the dorsal compartment retains its epithelial structure as it differentiates into dermomyotome and generates a second layer of myotome mainly from its dorsal medial margin (Fig. 3C). At the lateral margin of the dermomyotome, the somitic bud is formed from which hypaxial muscle precursors are derived (Bober *et al.*, 1994; Daston *et al.*, 1996) (Fig. 3D). The formation of the myotome has recently been studied by dye labelling of prospective myotomal precursors in the dermomyotome of chick embryos (Denetclaw *et al.*, 1997). These studies showed that as the dermomyotome differentiates, the myotome is formed by the cells at the dorsal lip migrating over the entire width of the dorsomedial edge. The myotome extends in the ventrolateral direction, as the myotome fibres progressively lengthen in the anteroposterior direction. Later the dorsal myotome gives rise to the muscle of the vertebrae and the deep muscle of the back, while the ventral part of the myotome forms the body wall muscle and the muscle of the limb. The dermotome differentiates into the dermis of the trunk and tail. Signalling molecules from the axial structures, the surface ectoderm and lateral plate mesoderm play an important role in the patterning and maintenance of these somite compartments as well as in the commitment of cells to the specific myogenic fates.

5. Somite Patterning

5. 1. Dorsoventral Patterning

The first visible differentiation of cell types in the somite is the formation of the dermomyotome dorsolateraly, and the formation of sclerotome on the ventromedial aspect of the somite. In the mouse sclerotome differentiation occurs 6-7 hours after the somite was born at which time the next 4-5 somites have already formed. The timing of somite differentiation is linked to the number of cell divisions the somite has to undergo depending on its axial level (Snow, 1981; Tam 1981; Power and Tam 1993).
Dorsoventral rotation of chick somites shows that polarity along the dorsoventral axis is first observed in somite IV, shortly prior to sclerotome formation (Aoyama and Asamoto 1988; Ordahl and Le Douarin, 1992) whereas rostrocaudal polarity is aquired much earlier. This has been demonstrated by altering both the dorsoventral and rostrocaudal axis at the same time. Rotated somites I-III lost their DV identity but retained their original rostrocaudal polarity, demonstrating that rostrocaudal patterning occurs independently and prior to dorsoventral somite patterning.
Induction and maintenance of dorsal dermomyotome and ventral sclerotome identity requires persistent signalling from axial structures (notochord and neural tube). If the neural tube is removed from the developing embryo, axial musculature derived from the

dorsal half of the somite is missing, while body wall muscle from the ventral half develops normally (Christ *et al.,* 1992; Rong *et al.,* 1992), indicating that separate signals are involved in patterning of the dorsal and ventral half of the somite and that the neural tube is required for dorsal patterning of the somite (Brand-Saberi *et al.,* 1993; Pourquie *et al.,* 1993; Goulding *et al.,* 1994). Similarly, removal of the notochord prior to somite formation results in absence of ventral sclerotome and fusion of the dermomyotome under the neural tube while grafting of ectopic notochord between presomitic mesoderm and surface ectoderm can inhibit dorsal fate, i.e.. dermomyotome differentiation (Pourquie *et al.,* 1993; Goulding *et al.,* 1994), demonstrating that the notochord and the floorplate produce ventralising signals.
These experiments indicate that a morphogenetic gradient might exist along the DV axis, of ventralising signals emanating from the notochord and floorplate and dorsalising signals from the dorsal neural tube and surface ectoderm. Different concentrations of these signals could induce the epithelial somite to give rise to sclerotome ventrally, dermomyotome dorsally and myotome in the middle where both signals compete.

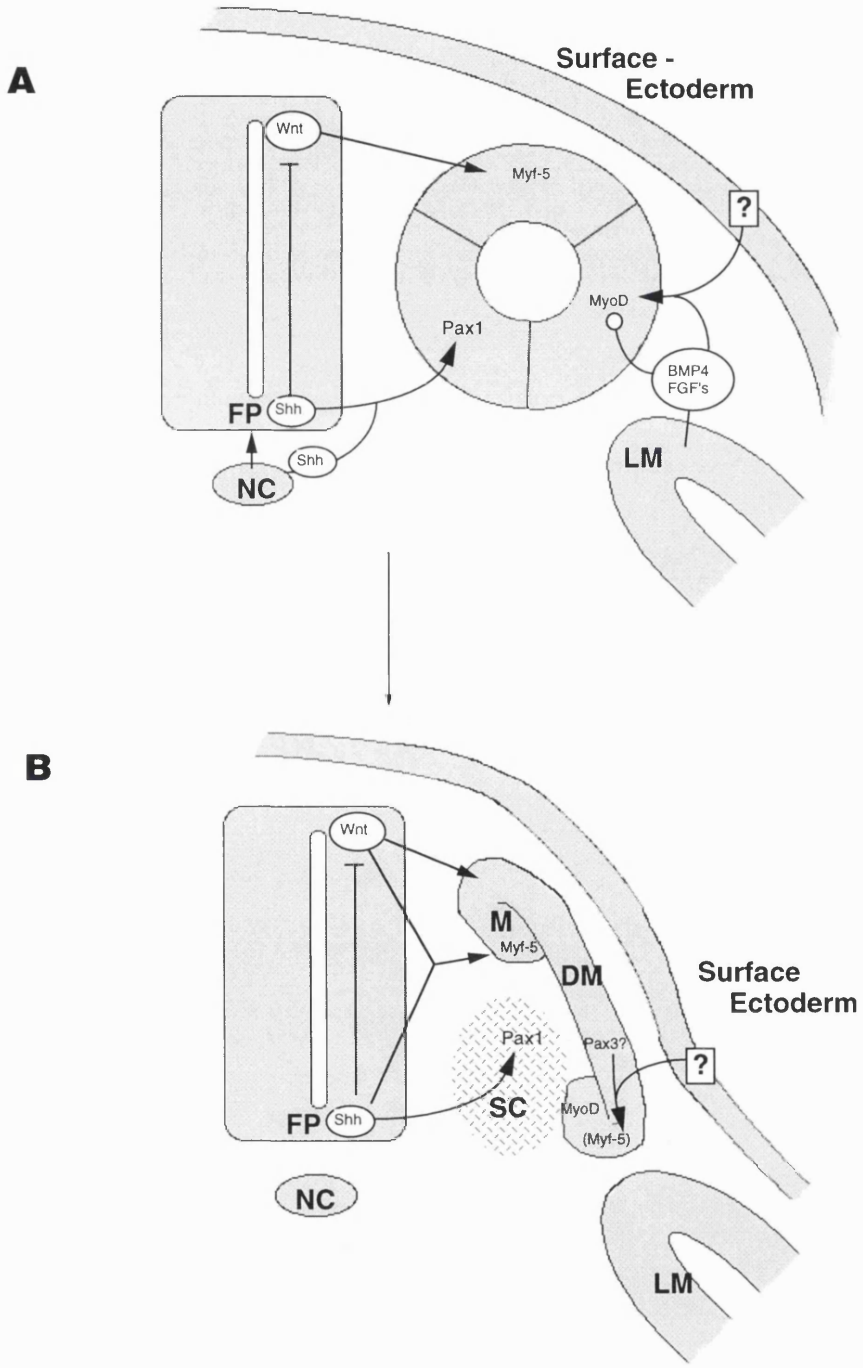
5. 2. Sonic Hedgehog and Wnt Signalling

The dorsalising effect of the neural tube can be mimicked by molecules of the *Wnt* family. *Wnt1, Wnt3* and *Wnt4* genes are expressed in the dorsal neural tube and surface ectoderm and in culture of chick somites I-III induce the myotomal marker *MyoD* (Muensterberg *et al.,* 1995) (Fig. 4A, note, that the roles of *MyoD* and *Myf-5* in birds are reversed compared with the mouse or human and that the diagram in Fig. 4 is a summary of the situation in the mouse embryo). The ability of *Wnt* genes to activate myotomal differentiation appears to be evolutionarily conserved. In *Drosophila,* wingless *(Wnt1)* can activate the *Drosophila* homologue of the myogenic bHLH factors *nautilus* (Couso and Martinez Arias, 1994) and activation of *XMyoD* by *XWnt-8* in *Xenopus* has also been observed (Hoppler *et al.,* 1996). However, experiments in chick have also shown that *Wnt* signals alone are not sufficient to induce myogenesis in the youngest somites I-III and in the presomitic mesoderm. In these tissues a combination of *Wnt1, Wnt3* and sonic hedgehog (shh) is required to induce *MyoD* (Fig. 4) (Muensterberg *et al.,* 1995). Since sonic hedghog is primarily expressed ventrally, first in the notochord and subsequently in the floorplate of the ventral neural tube, the question arises how shh could play a role in dorsal Wnt signalling. Interestingly, sonic hedgehog is a secreted protein whose N-terminal fragment becomes autoactivated and has short and long range signalling functions (Lee *et al.,* 1994; Marti *et al.,* 1995; Roelink *et al.,* 1995). The N-terminal domain can, after nucleophilic attack by the 3ß-OH group of cholesterol, become covalently linked to cholesterol which affects its subcellular distribution (Porter *et al.,* 1996). Shh concentrated in this way in the cell membrane is thought to mediate short range signalling inducing floorplate

Figure 4: Somite Patterning.

Model of the somite patterning signals and their sources in the mouse embryo. The medial and lateral somite halves respond to signals from the dorsal neural tube and lateral plate mesoderm (LM) plus surface ectoderm by expressing *Myf-5* or *MyoD* respectively. The dorsal neural tube signal is likely to be a combination of *Wnt* family members, and the signal from the lateral plate and surface ectoderm can be mimicked by BMP-4. Members of the *FGF* family may also play a role. The action of BMP4 and FGF members is probably to delay terminal differentiation (blunt arrow) in the lateral somite half to allow proliferation and cell migration of hypaxial muscle precursors. Additional positive signals (?) may also be required. Short range signalling of Shh secreted from the notochord (NC) plays a part in floorplate (FP) - induction. Subsequently Shh is also secreted from the floorplate. The ventromedial part of the somite is exposed to high levels of sonic hedgehog (Shh) and develops into sclerotome, expressing Pax-1, a sclerotomal marker. As the somite matures, mesenchymal sclerotome migrates medially and the remaining dermomyotome (DM) begins to differentiate into myotome (M). Initially the dorsal myotome is positive for *Myf-5* and the ventral myotome for *MyoD*, later *Myf-5* is also observed ventrally. The signals activating *Myf-5* and *MyoD* ventrally are not known, but might involve *Pax-3*. Activation of *Pax-3* by surface ectoderm has been demonstrated (Fan and Tessier-Lavigne, 1994).

Figure 4: Somite Patterning



formation in the ventral neural tube. Subsequent long range signalling from the floorplate may be a function of the unmodified N-terminal domain. Long range signalling of Shh is clearly important in sclerotome induction (Fig. 4B). When somites that are already expressing the sclerotomal marker Pax-1 are cultured in the absence of Shh, *Pax-1* expression is gradually lost, but can be restored after incubation with Shh demonstrating that Shh can induce sclerotomal markers in somites that are competent to respond to it (Muensterberg *et al.,* 1995). Remarkably, it appears that another long range function of Shh is to induce competence in somites I-III to respond to dorsalising *Wnt* signals from the dorsal neural tube. Indeed temporary exposure to Shh is required to induce competence in presomitic mesoderm to subsequently respond to *Wnt* signals and express the myotomal marker *MyoD* in chick. Conversely, exposure to *Wnt* signals alone is not sufficient to induce myogenesis in presomitic mesoderm that had not previously been exposed to Shh (Muensterberg *et al.,* 1995).

Thus, not only is a combination of diverse signals required for the formation of the somite compartments, but competence of the somite to respond to such signals is acquired during somite maturation. The effect of ectopic *Shh* expression on sclerotome and myotome formation has been examined in chick embryos and mouse explants and confirms the role of Shh as a ventralising signal in DV patterning (Fan and Tessier-Lavigne, 1994; Johnson *et al.,* 1994). Chick embryos injected with a virus encoding Shh show a dorsal expansion of the *Pax-1* expression domain on the infected side (Johnson *et al.,* 1994), while ectopic expression of *Shh* in the dorsal mesoderm abolished *Pax-3*, a dermomyotomal marker (Fan *et al.,* 1995). Remarkably Shh-induced *Pax-1* activation in these experiments was observed over several somite diameters, and yet *in vivo* the dorsal somite does not adopt sclerotomal fate. This is consistent with a model in which competitive signals from the surface ectoderm and the dorsal neural tube antagonise Shh signals from the floorplate and notochord. Accordingly, high concentrations of Shh secreted from the notochord and floorplate maintain the *Pax-1* expression domain and promote proliferation of the sclerotome, whereas myotome differentiation requires the presence of both sonic hedgehog and Wnt signals (Fig. 4B).

5. 3. Mediolateral Somite Patterning

The medial and lateral halves of the somite probably give rise to separate lineages of muscle cells. Epaxial muscle of the back is derived from the medial half of the somite and hypaxial muscle, including limb and body wall musculature, is derived from the lateral half (Ordahl and Le Douarin, 1992). In culture, cells from the medial half of the presomitic mesoderm of mouse embryos will first express the myogenic bHLH factor *Myf-5* in response to axial signals whereas cells from the lateral half activate *MyoD* following signals from the surface ectoderm (Fig. 4A) (Cossu *et al.,* 1996a) suggesting

that hypaxial and epaxial muscle is derived by separate *MyoD* and *Myf-5* dependent pathways in the lateral and medial somite halves respectively (Ordahl and Le Douarin, 1992).

Medially muscle is formed rapidly from the myotome. In contrast, the lateral somite compartment gives rise to the migratory muscle cell populations of the limb and ventral body wall that are committed to myogenesis but whose differentiation is delayed.

What are the mechanisms activating cell differentiation medially and cell proliferation laterally? It has been shown that medial but not lateral cells require the presence of axial structures for their differentiation, indicating that medial and lateral differentiation of the somites are independently controlled by competing signals (Rong *et al.*, 1992). Interestingly, lateral mesoderm alone fails to activate *MyoD* in presomitic mesoderm, however, surface ectoderm associated with its underlying lateral mesoderm is sufficient to induce *MyoD* laterally (Cossu *et al.*, 1996a). In addition to providing a lateralising signal, lateral plate mesoderm is believed to secrete inhibitory signals that suspend myogenesis in limb muscle precursors of mouse and chick embryos by two days by delaying expression of *MyoD* and *Myf-5* (Fig. 4) (Buckingham, 1992; Pownall and Emerson, 1992; Pourquie *et al.*, 1995). Thus migratory limb muscle precursors are normally prevented from expressing any member of the MRF family until they reach their respective target sites in the limb where they enter myogenesis (Tajbakhsh and Buckingham, 1994; Pourquie *et al.*, 1995).

Candidate signalling molecules mediating these effects of the lateral plate and surface ectoderm include members of the TGFβ and FGF families like bone morphogenetic protein-4 (BMP-4). BMP-4 is a diffusible growth factor expressed primarily in the lateral plate for which it can substitute in patterning of the lateral somite compartment. Consistent with this, lateral tissue grafts expressing virally encoded BMP-4 promote an expansion of lateral somitic cells into more medial domains (Pourquie *et al.*, 1996). Interestingly, signals from the neural tube, probably involving Shh, can counteract the effect of BMP-4, suggesting that mediolateral patterning of the somites, like dorso-ventral patterning, is established by competition between signals emanating from opposite poles along the mediolateral axis. In chick limb buds laterally high concentrations of FGFs and FGFR coincide with suspension of terminal differentiation during migration from the somite to the limb bud (Haub and Goldfarb, 1991; Niswander and Martin, 1992). Following arrival at the target site FGFR levels are down regulated, thus reducing FGFR signalling thereby inducing terminal muscle differentiation. The mechanisms underlying the negative effect of growth factors of the FGF and TGF-β family on myogenesis have been examined by Gerber *et al.*, (1997), who showed that these growth factors interfere with the ability of *MyoD* to remodel the chromatin structure at the *myogenin* locus and block the initiation of *myogenin* transcription. Interestingly, Myf-5 and MyoD are equally capable of remodelling chromatin whereas

myogenin was found to be one order of magnitude less efficient, reflecting the role of *Myf-5* and *MyoD* in early myoblast determination. Such changes induced in the local chromatin structure may influence gene expression within a lineage by determining the access of other lineage specific transcription factors. The myogenic factors share a C/H rich amino acid domain associated with the remodelling function and this domain is also conserved between mammals, birds, amphibia and fish (this study), reflecting the functional role of these factors in lineage restriction and suggesting that the functional mechanism of lineage restriction is also conserved. Other studies indicate that FGF mediated activation of protein kinase C (PKC) *in vivo* and *in vitro* (Davis *et al.*, 1987; Li *et al.*, 1992) plays a role in inhibiting myogenic bHLH proteins. Phosphorylation of a conserved site in the DNA binding domain of myogenin is thought to inhibit DNA binding. Taken together these findings show that the bHLH factors respond to a diverse range of positive and negative regulatory mechanisms.

6. Somites in Teleost Fish

In teleost fish, during gastrulation cells involute at the blastoderm margin and converge to the dorsal side to form the embryonic shield, from which the axial mesoderm is generated. As the axial mesoderm extends it becomes separated from the paraxial mesoderm that gives rise to the somites. In zebrafish about 30 somites are formed between 10 and 24 hpf (hours post fertilisation), giving rise to a new pair of somites every 20-30 minutes (Hanneman and Westerfield, 1989). The myotome is the major component of the somites in zebrafish, while sclerotome is formed by only a small number of cells in the ventromedial region (see Fig. 5A) (Morin-Kensicki and Eisen, 1997). In wild type zebrafish embryos, until about 13 hpf, the somites have the shape of epithelial spheres and subsequently become transformed into chevron-shaped myotomes (Fig. 5B). Somite development in teleosts is characterised by the establishment of three specialised structures: the adaxial cells, the pioneer cells and the horizontal myoseptum (HMS) (Fig. 5B). As in other species, the notochord in zebra fish plays an important role in somite patterning and seems to be required for the formation of these structures. Sonic hedgehog, secreted from the axial mesoderm, has been shown to be involved in the recruitment of adaxial cells to the myotomal lineage and subsequently the muscle pioneers become specified by a different member of the hedgehog family: echidna hedgehog (Currie and Ingham, 1996).

The adaxial cells are first visible in the presomitic mesoderm as large block shaped cells organised in three to five rows adjacent to the notochord (Fig. 5C) (Waterman, 1969; Felsenfeld *et al.*, 1991). Unlike muscle cells of other vertebrates, the adaxial cells of the

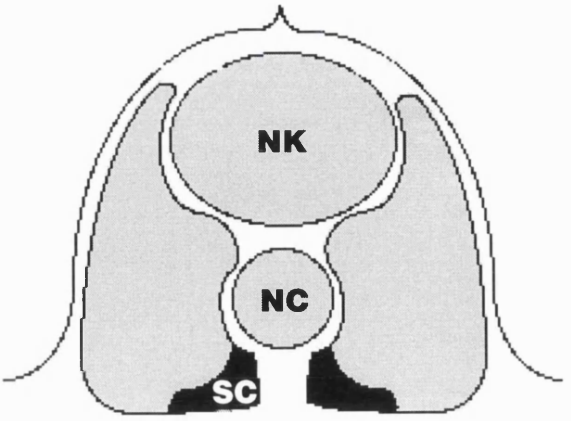
Figure 5: Somites in Teleost Fish.

5A: Schematic transverse section through trunk region of a zebrafish embryo. Compared with other vertebrates the sclerotome (SC) in teleost fish constitutes a very small portion of the somite and is located at its ventromedial margin (black). The myotome makes up most of the somites. Notochord (NC), neural keel (NK). (Redrawn and modified from Morin-Kensicki and Eisen 1997).

5B: Schematic drawing of a lateral view of zebrafish trunk somites. The chevron shape somites are devided in the middle of the myotome by a fibrous sheet, the horizontal myoseptum (HMS) that is formed from the muscle pioneer cells (not shown). The adaxial cells (AD) migrate radially outwards through the myotome to become the most superficial muscle cells in the somite (Morin-Kensicki and Eisen, 1997).

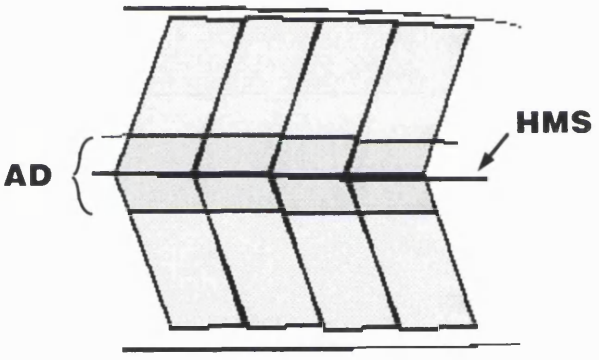
5C: Schematic drawing of the segmental plate in zebrafish. The adaxial cells (AD) are arranged as a sheet between the notochord (NC) and the lateral presomitic cells (LPS). (Redrawn and modified from Devoto *et al.,* 1996).

5A



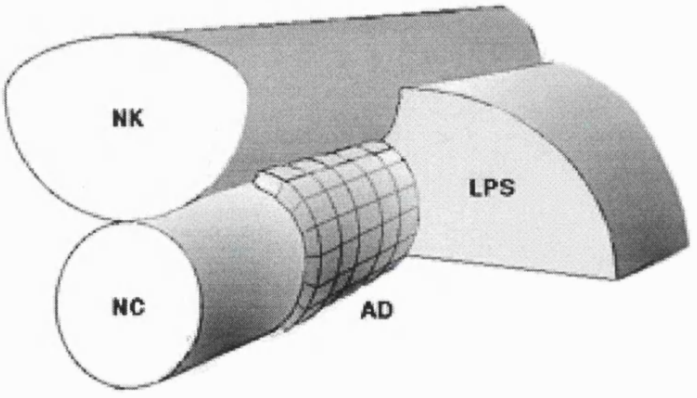
Sclerotome (Sc) in Teleost Fish

5B



Fish Trunk Somites, Lateral View,

5C



Adaxial Cells (AD) and Lateral Presomitic Cells (LPS)

paraxial mesoderm elongate to span the length of the somite and eventually migrate radially through the somite. Following migration they form a monolayer of superficial muscle cells that differentiates into slow muscle fibres (Devoto *et al.*, 1996). A subset of the adaxial cells, the muscle pioneers, extends from the tip of the V-shaped somites along the AP axis (see Fig. 5B). Unlike the other adaxial cells, muscle pioneers do not migrate radially but rather extend from the notochord to the lateral surface of the somite, where, after 28 hpf, they form the horizontal myoseptum (Hatta *et al.*, 1991). The horizontal myoseptum is a fibrous sheet that segregates the hypaxial and epaxial muscle precursors of the somite. Mutations affecting formation of the myoseptum have been designated *you-type* mutations, because they invariable cause curved tails resulting from U-shaped rather than V shaped myotomes (Van Eeden *et al.*, 1996). The paraxial cells lateral to the adaxial cells belong to the lateral presomitic mesoderm (LPS). Vital dye labelling experiments have shown that these cells give rise to fast muscle fibres (Fig. 5C) (Devoto *et al.*, 1996).

In contrast to higher vertebrates, expression of *MyoD* in zebrafish is seen prior to somite formation and is maintained throughout somitogenesis in the adaxial cells along the entire AP axis, including the presomitic mesoderm and two laterally extending bands at the rostral edge of the segmental plate immediately preceding somite formation (Weinberg *et al.*, 1996). If this pattern of expression is a general feature amongst teleost fish and if it is true for other myogenic regulatory factors like *Myf-5* is not yet known. In *Xenopus*, *MyoD* and *Myf-5* are expressed in the unsegmented paraxial mesoderm, but transcripts are found throughout the entire presomitic mesoderm and are not confined to the adaxial cells of the segmental plate (Frank and Harland, 1991; Harvey, 1992; Hopwood *et al.*, 1992). It would be of interest to investigate if in zebrafish *Myf-5*- and *MyoD*-dependent muscle lineages exist in the presomitic mesoderm or in the somites or if zebrafish *MyoD* and *Myf-5* are expressed in overlapping domains.

7. The Myogenic bHLH Factors

Initially research on the specification of muscle focused on the transition from undifferentiated myoblasts to myotubes. Genomic DNA from myoblasts was transfected into fibroblasts and shown to convert the fibroblast cells into myoblasts (Konieczny and Emerson, 1984; Lassar *et al.*, 1986). Similar results were obtained when fibroblasts were treated with 5-azacytidine, an inhibitor of methylation. Based on these data, subtractive hybridisation of cDNA expression libraries from treated and untreated fibroblasts identified the first myogenic factor gene encoding *MyoD* (Davis *et al.*, 1987; Weintraub *et al.*, 1991). Subsequently three further myogenic factors were cloned from mammalian species: *Myf-5* (Braun *et al.*, 1989a), *myogenin* (Edmondson and Olson,

1989; Wright *et al.,* 1989; Salminen *et al.,* 1991) and *MRF4* (Rhodes and Konieczny 1989; Miner and Wold, 1990; Hinterberger *et al.,* 1992) A variety of tissue culture cells transfected with cDNA for each of these proteins can activate the myogenic program, confirming that the bHLH proteins act as myogenic factors (Davis *et al.* 1987; Weintraub *et al.,* 1989; Choi *et al.,* 1990). Myogenic bHLH regulatory genes have also been identified in birds, frogs, sea urchins (Venuti *et al.,* 1991), insects (Michelson *et al.,* 1990), nematodes (Weintraub *et al.,* 1991) and amphioxus (Araki *et al.,* 1996). Most invertebrate species only have a single member of the *MyoD* gene family but it has been shown that the myogenic factors from sea urchin and nematode (Krause *et al.,* 1992) can also activate the myogenic program in mammalian cells, suggesting that the regulatory mechanisms are extremely ancient.

Surprisingly, evidence from *Drosophila* and nematode indicates that even in the absence of the myogenic factors early myogenesis proceeds normally. The most likely explanation is that alternative pathways for myogenic determination and differentiation exist in invertebrates.

Recently a *MyoD* homologue from the ascidian *C. intestinalis*, *CiMDF*, has been cloned (Meedel *et al.,* 1997). Interestingly the *CiMDF* gene is differentially transcribed to produce distinct transcripts *CiMDFa* and *CiMDFb* that have separate functions in myogenesis and overlapping temporal expression, that may distinguish between primary and secondary muscle lineages of ascidians. The presence of E-box motifs in *CiMDF* suggests auto and cross-regulation similar to the vertebrate genes (Meedel *et al.,* 1997). Interestingly, there is little similarity between invertebrate and vertebrate species apart from a highly conserved structural motif, the basic-helix-loop-helix domain of about 60 amino acids, which mediates protein dimerisation and DNA binding. The myogenic factors heterodimerise with ubiquitous E2 proteins (Lassar *et al.,* 1989,1991; Braun *et al.,* 1990; Brennan and Olson, 1990) in order to bind to E-box motifs (CANNTG) that were initially found to be important in the immunoglobulin and muscle creatine kinase enhancers (Buskin and Hauschka, 1989; Gossett *et al.,* 1989; Murre *et al.,* 1989) and are also present in other muscle specific genes including those encoding the myogenic factors themselves. The MRFs can therefore transactivate the expression of muscle structural genes, and auto- or transactivate their own expression (Braun *et al.,* 1989a,b; Edmondson and Olson 1989; Rhodes and Konieczny, 1989; Thayer *et al.,* 1989; Miner and Wold, 1990; Yee and Rigby, 1993).

8. The Myogenic Cascade

In vitro studies on muscle cell lines have shown that the bHLH factors are expressed in a distinct temporal sequence during differentiation. *MyoD* and *Myf-5* are expressed in

myoblasts prior to and after differentiation, while *myogenin* and *MRF4* are expressed only after differentiation (Hinterberger *et al.*, 1991). These results suggested that *MyoD* and *Myf-5* might act early in determining myoblast fate, while *MRF4* and *myogenin* are involved in later differentiation of myoblasts into myotubes. When *Myf-5* is expressed in 10T1/2 cells, myogenesis is associated with expression of *MyoD* but the converse is not true: *MyoD* expression does not lead to activation of *Myf-5* (Braun *et al.*, 1989b), suggesting that *Myf-5* acts upstream of *MyoD* at least *in vitro*. Studies of the expression patterns of the myogenic factors *in vivo* have revealed the temporal sequence of their expression and are in agreement with a transcriptional hierarchy of the myogenic factors (Thayer *et al.*, 1989; Braun *et al.*, 1990; Edmondson *et al.*, 1992; Naidu *et al.*, 1995). In skeletal muscle, *Myf-5* is the first myogenic regulatory gene to be expressed at E8.0, in the dermomyotome of the newly formed somites (Ott *et al.*, 1991). 12 hours later, *myogenin* appears in the myotome of successive somites, (Sassoon *et al.*, 1989) followed by *MRF4* (Bober *et al.*, 1991; Hinterberger *et al.*, 1991) and *MyoD* (Sassoon *et al.*, 1989) (see Fig. 6). *MRF4* is reactivated in a second phase of expression from E14.5 onwards.

The temporal expression of the myogenic factors in the limb differs from that of the trunk musculature. The migratory myoblasts that colonise the limb bud leave the ventrolateral edge of the somite and do not express MRFs until they have reached their destination (Sassoon *et al.*, 1989). Therefore, *Myf-5* is first expressed at E10.5, followed by myogenin and *MyoD* which are coexpressed in the limb myoblasts rather than being sequentially activated as in the somites. *MRF4* mRNA is not detectable in the limb until late in development at E16 (Bober *et al.*, 1991).
There are also species dependent variations in the order and timing of expression of the myogenic factors. In birds, the homologue of *MyoD* instead of *Myf-5* appears first, followed by *myogenin* and *Myf-5* (Pownall and Emerson, 1992). In *Xenopus* and zebrafish, *MyoD* and *Myf-5* are activated in the presumptive mesoderm (Hopwood *et al.*, 1992; Weinberg *et al.*, 1996; this study) whereas myogenic factors are not expressed at significant levels prior to somite formation in the mouse. This raises the possibility that the function of *XMyoD* and *ZMyf-5* may be quite distinct from higher vertebrate species. However, the differences in the temporal expression are less significant if one considers *MyoD* and *Myf-5* as a functionally equivalent pair of genes with a common evolutionary origin (Atchley *et al.*, 1994). *MyoD* and *Myf-5* probably arose from a common gene, as did *myogenin* and *MRF4*, which suggests that the partners in each of the pairs are more closely related and, therefore, might more readily substitute for one another in myogenesis.

Temporal expression of the myogenic factors in the mouse.

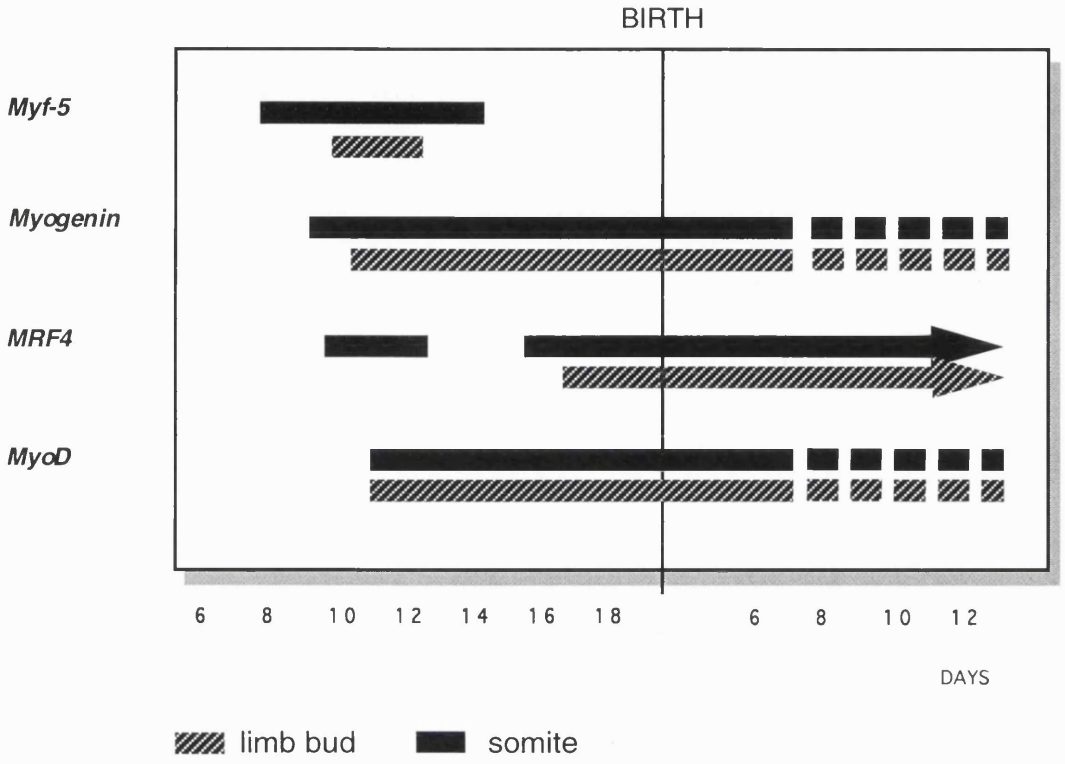


Figure 6:

Temporal Expression of the Myogenic Factors in limb and somites. The myogenic bHLH factors have a distinct temporal expression pattern in limb buds (stippled) and somites (solid) of the mouse embryo. The migratory myoblasts that colonise the limb bud leave the ventrolateral edge of the somite and do not express *MRFs* until they have reached their destination. Nevertheless, *Myf-5* is expressed first, in both somites (at E8.0) and the limb bud (E10.5). 12 hours later, myogenin is activated. In the somites *MRF4* is expressed in an early and late phase interrupted by *MyoD* whereas in the limb the early *MRF4* phase is missing, and myogenin and *MyoD* are coexpressed. In the limb *MRF4* mRNA is not detectable until late in development at E16. Expression of *MyoD* and *Myf-5* (black) converts premyogenic cells into skeletal myoblasts while the expression of myogenin and *MRF4* (gray) is linked to the fusion and differentiation of myoblasts into multinucleated myotubes.

9. MRF Null Mutations

To examine the individual roles of the myogenic factors in myogenesis, targeted null mutations have been introduced into each of the four myogenic regulatory genes and in addition double mutants for *MyoD/Myf-5* (Rudnicki *et al.*, 1993), *MRF4/Myf-5* (Braun *et al.*, 1995), *myogenin/MyoD* and *myogenin/Myf-5* (Rawls *et al.*, 1995) have been obtained by interbreeding these targeted mouse lines. This strategy has been particularly successful in studying myogenesis, because the redundancy between the regulatory factors makes it otherwise impossible to dissect their individual function.

9. 1. *MyoD* Null Mutants

The introduction of a homozygous null mutation in the *MyoD* gene does not produce an embryonic muscle phenotype. Moreover, *MyoD* mutant mice are viable and fertile and indistinguishable from wild type litter mates (Rudnicki *et al.*, 1992) suggesting that *MyoD* and Myf-5 have overlapping functions. However, preferential expression of *MyoD* in fast twitch muscle fibres raises the possibility that subtle fibre type changes occur in the absence of *MyoD* (Hughes *et al.*, 1993). Furthermore, recent evidence shows that satellite cells fail to divide in *mdx/MyoD-/-* double mutant mice, suggesting that *MyoD* is required for satellite cell activation during muscle fibre regeneration (Megeney *et al.*, 1996). Clearly, the functional overlap of *Myf-5* and *MyoD* is limited since a single *Myf-5* allele can not rescue the *MyoD* phenotype and results in lethality due to the apparent reduction in skeletal muscles, whereas a single copy of *MyoD* is sufficient to rescue *Myf-5* knockout mice. It has been suggested that *MyoD* may be able to recruit more cells possibly from different precursor populations to the myogenic lineage than *Myf-5* (Braun and Arnold, 1996). Assuming that competition exists between cell lineages alternatively determined by *MyoD* and *Myf-5*, the upregulation of *Myf-5* in *MyoD* deficient mice can be explained as an expansion of the *Myf-5* myogenic cell lineage. According to this model, the *MyoD* cell lineage seems to be enlarged at the expense of the *Myf-5* cell lineage in the wild type situation, suggesting that *MyoD* expression is either more stable or more responsive to environmental cues compared with *Myf-5*. The different expression domains of *MyoD* and *Myf-5* within the somite might support such a model (Braun and Arnold, 1996). The most interesting observation in *MyoD* mutant mice is that *Myf-5* mRNA levels are elevated two-fold, suggesting that *Myf-5* may substitute for the absence of *MyoD* in the development of skeletal muscle. This is supported by the finding that the levels of *MRF4* and *myogenin* transcripts are normal in these mice (Rudnicki *et al.*, 1992).

9. 2. *Myf-5* Null Mutants

Myf-5 mutant mice survive to birth but in contrast to the *MyoD* null mice they die perinatally due to respiratory failure because the distal portion of the ribs necessary for insertion of the diaphragm fails to develop (Braun *et al.*, 1992). In the *Myf-5* mutant mice skeletal myofibres are normal although myotome formation is delayed until *MyoD* is expressed, suggesting that in the absence of *Myf-5, MyoD* can activate the *myogenin* promoter but that *myogenin* can not substitute for either *Myf-5* or *MyoD* function in myotome formation. In agreement with this, expression of *MyoD* is unaffected in *myogenin* knockouts, indicating that *MyoD* is independent of *myogenin* and can not substitute functionally for *myogenin*. The rib phenotype is unexpected because the sclerotomal precursor cells from which the ribs develop do not express significant levels of *Myf-5* (Smith *et al.*, 1994), suggesting that the defect is the result of the delay in myotome formation which might interfere with cell signalling (Grass *et al.*, 1996). Indeed, recent analysis of *Myf-5* null mice suggested that activation of *MyoD* in the trunk of these mice might be mediated by *Pax-3* and is delayed compared with wild type animals (Tajbakhsh *et al.*, 1997). If sclerotomal precursors are only receptive for a patterning signal during a limited early period before the epithelial somite compartmentalises, the delayed expression of *MyoD* compared with *Myf-5* may preclude such signalling and be responsible for the observed phenotype.
Alternatively, the failure of *MyoD* expression to rescue the rib defect of *Myf-5* deficient mice may be due to the fact that *MyoD* is mainly expressed in the dorsal part of the somite possibly too far away to allow signalling to sclerotome to form the rib-blastema (Grass *et al.*, 1996).

9. 3. *Myf-5/MyoD* Double Mutants

When double null mutants for *MyoD* and *Myf-5* were generated by interbreeding *Myf-5* and *MyoD* mutant mice, they failed completely to develop skeletal myoblasts and were not viable confirming that either *MyoD* or *Myf-5* is required for the determination or survival of the myoblast lineage (Rudnicki *et al.*, 1993). Recently *Pax-3* has been implicated in the regulation of *MyoD* and *Myf-5* (Maroto *et al.*, 1997; Tajbakhsh *et al.*, 1997). If *Pax-3* was required for *MyoD* expression, double mutant mice for *Pax-3-/-* (splotch mice) and *Myf-5* should mimic the phenotype of *MyoD/Myf-5* double knockout mice and fail to produce myoblasts. Indeed in the limbs and the trunk of *Splotch/Myf-5-/-* mutants, MyoD protein and transcripts were essentially absent and myoblasts were not detected. However, expression in the head was apparently unaffected, suggesting that *MyoD* expression in the head is independent of *Pax-3* and thus regulated differently from the trunk (Tajbakhsh *et al.*, 1997).

9. 4. *Myogenin* Null Mutants

Two independent null mutations have been generated for *myogenin* (Hasty *et al.,* 1993; Nabeshima *et al.,* 1993), both showing similar results. *Myogenin* null mice die perinatally like the *Myf-5* mutants due to respiratory failure resulting from the absence of a functional diaphragm and the associated muscle deficiency (Hasty *et al.,* 1993). Mice lacking a functional *myogenin* gene have normal numbers of myoblasts but severely reduced myofibres resulting in a drastic muscle phenotype. Thus *myogenin* is required for normal myoblast fusion and differentiation *in vivo.*

Surprisingly, myoblasts of *myogenin* null mice readily fuse with wild type myoblasts in chimaeric mice containing a mixture of *myogenin* null and wild type cells (Myer *et al.,* 1997). However, due to the lack of *myogenin* these myoblasts do not activate significant levels of muscle specific proteins (e.g.. MCK) and fail to differentiate and fully rescue the *myogenin* null phenotype. These experiments suggest that *myogenin* controls the expression of a secreted fusion promoting ligand required for myoblast fusion (Myer *et al.,* 1997). In *myogenin* mutant mice the late phase of *MRF4* expression after E14 is not observed, suggesting that *myogenin* is required for this late phase of *MRF4* expression. However, the presence of residual myofibres in myogenin null mice would suggest that myogenin is not involved in the early phase of *MRF4* expression.

9. 5. *MRF4* Null Mutants

Three independent *MRF4* null mutations have been generated (Braun and Arnold, 1995; Patapoutian *et al.,* 1995; Zhang *et al.,* 1995) but their phenotypes are surprisingly variable, probably due to variations in the targeting constructs.

The null mutation generated by Braun and Arnold (1995) is the most severe and leads to a significant reduction in *Myf-5* transcription, effectively representing a phenocopy of the *Myf-5* null mutant, with defects in the formation of the distal portion of the ribs resulting in perinatal death due to respiratory failure (Braun and Arnold, 1995). In the Zhang knockout,*Myf-5* expression is not significantly affected, and this is the only mutation for which homozygous mice are viable, although a slight rib defect is also observed, this is distinct from that of the *Myf-5* null mutants. Adult muscle is apparently normal and *myogenin* mRNA is elevated, suggesting that *myogenin* is substituting for the function of *MRF4.* The third mutant generated by Patapoutian *et al.* (1995) has an intermediate phenotype. This allele also produces a rib defect similar to that of the Zhang mutation, but with sufficient severity to result in inviability. In contrast to the Braun and Arnold mutation, where *Myf-5* transcripts are not detectable, the Patapoutian allele shows significantly reduced expression of *Myf-5* in the somites; sclerotomal defects are more severe and the rib abnormalities result in embryonic lethality. Additional defects in early myotome formation are also observed, albeit with apparently little effect on subsequent myotome differentiation. At present it is difficult to ascertain whether the

absence of MRF4 is directly responsible for the phenotype, because due to the proximity of the two genes (less than 7kb), *cis*-effects of the *MRF4* null allele on *Myf-5* gene can not be excluded (Olson *et al.*, 1996). In compound heterozygous animals carrying either the intermediate or the weakest *MRF4* null allele on one chromosome, and a *Myf-5* knockout allele on the other chromosome, *Myf-5* expression was down regulated by a *cis*- mechanism. The compound heterozygote presented with intensified rib defects and increased mortality, supporting the notion that *cis*-acting interactions between *Myf-5* and *MRF4* may play a significant role in regulating expression of these genes in the early myotomes of wild type embryos (Yoon *et al.*, 1997).

Taken together, the data from the timing of expression of the myogenic factors, the knockout mice and from transfection experiments, suggest that myogenesis is a developmental cascade in which *Myf-5* and *MyoD* recruit pluripotent mesodermal stem cells to the myogenic lineage while their subsequent differentiation into myocytes depends on *myogenin*. Interbreeding the different mutant mouse lines confirmed a certain degree of overlap in the functions of *MyoD* and *Myf-5* while also showing that the function of *myogenin* is restricted to the control of myoblast differentiation and does not overlap with early functions of *MyoD* and *Myf-5*. Consistent with the putative role of *Myf-5* and *MyoD* in early myoblast determination and *myogenin* in myocyte formation, double null mutant *myogenin/Myf-5* and *myogenin/MyoD* mice show the combined phenotype of each of the individual mutations. Mice die perinatally from rib defects and share the myoblast deficiency of the *myogenin* mutation (Rawls *et al.*, 1995). The role of *MRF4* is at present not fully understood but it is possible that *MRF4* has essentially redundant functions in myogenesis. Whether *MRF4* function in adult muscle is linked to stress, ageing, disease or degeneration/ regeneration remains to be seen (Olson *et al.*, 1996).

10. What Regulates *MyoD* and *Myf-5*?

The search for putative signals upstream of *Myf-5* and *MyoD* has recently become focused on the paired-box transcription factor *Pax-3*. The expression of *Pax-3*, both prior to somitogenesis in the paraxial mesoderm and after somite formation in the ventrolateral dermomyotome is consistent with a role in patterning of muscle progenitors (Goulding *et al.*, 1994). However, the most obvious feature of homozygous *Pax-3* mutants (splotch), is a deficiency of limb muscle that is caused by the failure of the limb muscle progenitors to migrate to the prospective limb although muscles of the back and body wall are also affected (Franz *et al.*, 1993; Goulding *et al.*, 1994, Tajbakhsh *et al.*, 1997). The limb muscle defect in splotch mice is clearly the result of failed cell migration

rather than differentiation as lateral halves of somites of splotch mice were able to differentiate when transplanted to chick limb buds (Daston *et al.,* 1996). In fact, down regulation of *Pax-3* coincides with upregulation of the myogenic regulatory factors in the limb (Williams and Ordahl, 1994) which would suggest that *Pax-3* inhibits rather than activates myogenic bHLH factors in migratory limb muscle progenitor cells.
Nevertheless a number of observations suggest that *Pax-3* is involved in the activation of the myogenic factors *Myf-5* and *MyoD* in hypaxial muscle precursors derived from the lateral dermomyotome. Firstly, double mutant mice for *Pax3* and *Myf-5* mimic the phenotype of *MyoD* and *Myf-5* double knockout mice, and do not produce myoblasts because *MyoD* fails to be activated in the trunk, suggesting that *Pax-3* is normally required for *MyoD* expression in the trunk and that *Pax-3* acts upstream of *MyoD* (Tajbakhsh *et al.,* 1997). Secondly, overexpression of mouse *Pax-3* in chick somite explants is sufficient to activate *MyoD,* even in nonmuscle precursors, including the neural tube (Maroto *et al.,* 1997). Thirdly, in the absence of axial structures, surface ectoderm normally induces the expression of *Pax-3* and *MyoD* in presomitic mesoderm (Fan and Tessier-Lavigne, 1994) but fails to do so in explants from splotch mice, suggesting that in the lateral somite, surface ectoderm normally activates *MyoD* through a *Pax-3* dependent pathway. Furthermore, in splotch explants, *MyoD* expression was also reduced in the presence of axial structures, suggesting that the *Myf-5* -dependent activation of *MyoD* is also affected in *Pax-3* mutants (Tajbakhsh *et al.,* 1997). Finally, *Pax-3* deficiency results in a shortening and disorganisation of the *Myf-5* positive somitic bud of interlimb somites from which much of the hypaxial musculature is derived (Tajbakhsh *et al.,* 1997), implicating Pax-3 in *Myf-5* expression.
If *Pax-3* is able to activate *MyoD* expression *in vivo*, the question arises whether *MyoD* activation involves the earliest myogenic factor *Myf-5* or if *Pax-3* can activate *MyoD* independently. The data from the Splotch/*Myf-5* double null mutants seems to suggest that Pax-3 can activate *MyoD* in a *Myf-5* independent pathway. However, it appears that both *Myf-5*-dependent and -independent pathways for *MyoD* activation exist in different muscle lineages. Evidence that *MyoD* expression in most of the trunk does at least normally depend on *Myf-5,* comes from the recently observed (and previously not recognised) delay of *MyoD* expression in *Myf-5* null mutants (Tajbakhsh *et al.,* 1997). In these mutants *MyoD* expression is delayed by about 24 hours until *MyoD* becomes activated by Pax-3. It appears that only in the absence of *Myf-5* will *MyoD* become activated by Pax-3.
Another question is why *Pax-3* in the dermomyotome (Goulding *et al.,* 1994) fails to activate *MyoD*. The fact that *Pax-3* is expressed at high levels in the neural tube but does not activate myogenesis unless over-expressed, suggests that there are inhibitory signals that can suppress the *Pax-3*-dependent myogenic pathway. Similar inhibitory signals might suppress myogenesis in *Pax-3* expressing limb muscle progenitors until they

reach the correct destination in the limb bud and might also inhibit myogenesis in the presomitic mesoderm where *Pax-3* is also expressed. The presence of negative elements in the *Pax-3* promoter has recently been inferred, based on the observation that a *LacZ* reporter driven by 1.6 kb of *Pax-3* promóter sequence was expressed in a ventrally expanded domain compared to a reporter construct driven by 14kb of upstream sequence or the wild type expression pattern (Natoli *et al.,* 1997). It is presently unclear whether Pax-3 activates *MyoD* directly or indirectly or how *Pax-3* might be involved in the ventrolateral expression of *Myf-5*. Further evidence from coculture experiments of paraxial mesoderm in the presence of Wnt and Shh showed that activation of *Pax-3* and *Pax-7* occurs together with *Myf-5*, suggesting that *Myf-5* is indeed activated in a *Pax-3*-independent pathway or significantly faster than *MyoD* (Maroto *et al.,* 1997).
Interestingly, dorsomedial muscle progenitors of Splotch/*Myf-5 -/-LacZ* mutant mice express *LacZ* normally from the *Myf-5* promoter in the head and (some) neck muscles and their precursors, indicating that *Myf-5* expression depends on *Pax-3* only in the ventrolateral domain of interlimb somites (Tajbakhsh *et al.,* 1997).
However, both splotch and splotch/*Myf-5* double null mice exhibit severe neural tube defects leading to *spina bifida* and exencephaly, and it can not be ruled out that these defects may also affect dorsal somite patterning. Precedence for this is seen in openbrain *(opb)*, a mouse mutant with severe defects in dorsal neural tube closure. In *opb* mutants dorsal somite development is abnormal and results in loss of epaxial musculature although hypaxial muscle was apparently unaffected (Spoerle *et al.,* 1996).

11. Could Myogenesis Be Repressed?

In some vertebrates like *Xenopus* and zebrafish myogenesis is initiated prior to somitogenesis. In these species members of the myogenic bHLH family are first expressed in the presomitic mesoderm and continue to be expressed in the differentiating somite. However, in higher vertebrates myogenic factors are not expressed at significant levels in the presomitic mesoderm. These observations raise the question as to whether in higher vertebrates myogenesis might be actively repressed in the presomitic mesoderm or if the necessary induction signals from the environment are missing or if competence to respond to such signals must first be acquired.Because low levels of transcripts of the first myogenic factor *Myf-5* have been detected in the presomitic mesoderm of mice using RT-PCR analysis (Kopan *et al.,* 1994) and in *LacZ/Myf-5* knock-in mice at E9 (Cossu *et al.,* 1996a), it seems unlikely that the upstream signals initiating myogenesis are absent from the segmental plate. A more favourable model is one where myogenesis is repressed or threshold dependent. In agreement with this, low level myogenic bHLH gene expression has been reported as unstable in the absence of axial signals (Bober *et*

al., 1994). On the other hand, both somitic and presegmental plate cells have been found to differentiate into either chondrocytes or myocytes when grown in low density culture (George-Weinstein *et al.,* 1994). These results might indicate that cell cell interactions within the paraxial mesoderm act to repress differentiation in the absence of positive signals. According to this model, signals from the axial structures, or the surface ectoderm, may relieve the intrinsic inhibitory signal in the intact somite (Muensterberg *et al.,* 1995).

What could the nature of the intrinsic inhibitory signals be? A number of candidate molecules that are expressed in the presomitic mesoderm have been suggested to inhibit myogenesis. In *Drosophila* notch is activated by the ligand delta and mediates inhibition of neural and muscle precursor cells. The expression of the mouse notch1 receptor in the presomitic mesoderm is consistent with its suggested function as a repressor of the myogenic factors (Conlon *et al.,* 1995). However, homozygous mouse mutants for notch1 fail to produce a muscle phenotype, probably due to redundancy amongst the notch family members (Swiatek *et al.,* 1994). In 3T3 tissue culture cells and frog cells, the constitutively active intracellular domain of mouse notch1 has been shown to repress myogenesis probably by competing with *Myf-5* and *MyoD* for a putative co-factor that would allow them to activate the muscle program (Kopan *et al.,* 1994). The identity of the co-factor is not yet known. However, since forced heterodimers of E12 and MyoD are also inhibited by notch, it is unlikely to be the ubiquitous E12 protein, with which both Myf-5 and MyoD heterodimerise to activate transcription (Kopan *et al.,* 1994). How might notch inhibit myogenesis? It has been shown that notch association with MyoD prevents nuclear import of MyoD, probably by specifically masking the nuclear localisation signal, a mechanism that is also used by other factors like I-mfa to maintain sclerotome identity by preventing the sclerotome cells from switching on myogenic factors and adopting myoblast fate (Chen *et al.,* 1995). Interestingly in *Drosophila* notch mediated repression of *nautilus* appears to be relieved by wingless, the homologue of vertebrate Wnts (Couso and Martinez Arias, 1994; Ranganayakulu *et al.,* 1996). Perhaps Wnt members could also relieve notch mediated repression of MyoD in the medial somite, where shh molecules are secreted by the lateral neural tube (Cossu *et al.,* 1996b).

12. The Pufferfish as a Model Organism

12. 1. Genome Size and Complexity

It has been estimated that unicellular protozoae, nematodes and flies have 12000-14000 genes (Lewin, 1994; Nüsslein-Volhard, 1994; Miklos and Rubin, 1996). The human and mouse genomes are estimated to have between 50000 and 100000 genes (Antequera and Bird, 1993; Fields *et al.,* 1994). Using thymidine-^{3}H-labelled DNA digested with methylation sensitive restriction enzymes, Antequera and Bird estimated the total number of CpG islands in the mouse and human genome at around 40000. By sequencing 152 complete transcription units containing a transcription start site and flanking sequence (ca 0.2% of the human genes) they obtained an estimate of the fraction of genes associated with CpG islands of around 50% for a total gene number of about 76000 (Antequera and Bird, 1993; Fields *et al.,* 1994). Other estimates from expressed sequence tags (ESTs) representing approximately half of the human protein coding genes estimated the total gene number at about 65000 (Fields *et al.,* 1994). Therefore, a six fold increase in gene number has occurred between invertebrates and vertebrates. This increase is thought be the result of at least two whole genome duplications (Holland *et al.,* 1994), in addition to partial chromosome duplications and gene duplications resulting in large gene families (Lundin, 1993). The evolution of the myogenic factors also supports a model in which an ancestral gene was tandemly duplicated to give a linked pair followed by chromosome or whole genome duplication giving rise to two linked pairs on different chromosomes, one of which was subsequently split by chromosomal rearrangements (Atchley *et al.,* 1994). Both of the duplications must have occurred before teleost fish diverged from land vertebrates, early in vertebrate evolution, because the genomic organisation of the myogenic factors in teleosts is conserved (this study). It is conceivable that duplication of regulatory genes aided the evolution of new developmental pathways. Further support for this is given by the existence in *Drosophila* of only one copy of many of the genes involved in cell signalling and patterning, like the *Ras, Raf, notch* and *MRF* genes or the *Hox* cluster, for which vertebrates have at least three members.

Genome size varies enormously between vertebrate species, for example, the genome of some salamander species is up to twenty times larger than that of man. Similar observations have been made in plants, where genome size is even more variable amongst related species. Although many organisms with very large genomes are polyploid and the variation in haploid genome size is probably significantly smaller, it is apparent that morphological complexity is not directly correlated with an increase in genome complexity or size of the genome. The great increase in genome size between vertebrates and invertebrate species poses significant problems for the isolation and functional analysis of vertebrate genes. The ideal model would be a vertebrate organism with a small but highly conserved genome.

12. 2. The Small Genome of *Fugu rubripes*

The smallest vertebrate genomes have been reported for the tetraodontoid teleosts ('bony fish'), amongst which the Japanese pufferfish *Fugu rubripes* (see Fig. 7) has been studied in some detail. In the late sixties, tetraodontoid fish were studied by Hinegardner and Rosen (1972) who, on the basis of reassociation kinetics, estimated that the haploid genome of 0.4pg contained 380Mb of sequence, compared to 3000Mb in man. Since genome size estimates by physical measurement can be inaccurate, Brenner *et al.* (1993) have recently confirmed the small genome size of the pufferfish by two independent methods. One of the strategies included sequence analysis of random plasmid clones containing sonicated *Fugu* DNA. The coding potential for the obtained sequences was estimated by comparison with existing entries for other species in the SWISSPROT database. These studies indicated that *Fugu* exons were highly conserved and that a fraction of 0.791% of the total nucleotide sequence was coding. To compare this value to that expected for human, the coding fraction of the human genome was estimated by construction of a nonredundant database of coding sequences called RATMAN which contains only one copy for each vertebrate gene known. In other words, this database contains a single representative of each of the presently known genes found in any of the vertebrate species. RATMAN contained 3.09Mb of (coding) sequence which by comparison represents 0.103% of the total human genome. Thus, a database search with random human genomic sequence against RATMAN should identify 0.103% as coding sequence. Provided that the total number of genes between vertebrate species is approximately the same, this would suggest that the *Fugu* genome is 0.791/0.103=7.68 times smaller (ca 400Mb) than the human genome at around 3000Mb. Similar results were obtained by screening an unamplified phage library with single copy probes. An average of 24625 phage clones with an insert size of around 16.4kb had to be screened to identify a positive clone, indicating a genome size for *Fugu* of 24625x16.4kb= 403Mb, some 7.6 times smaller than human.

12. 3. The *Fugu* Genome Never Acquired 'Junk-DNA'

Based on the estimated size of the *Fugu* genome, one might expect the ancestral genome of primitive fish to be generally small. Surprisingly, however, only the pufferfish and its relatives have retained the archetypal genome size during evolution, while many independent genome expansions must have occurred to account for the large genome size of most fish species today. Although this might seem unlikely, the alternative, that primitive fish had a much larger genome comparable to that of higher vertebrates today, would imply that *Fugu* has undergone several rounds of genome compression, a

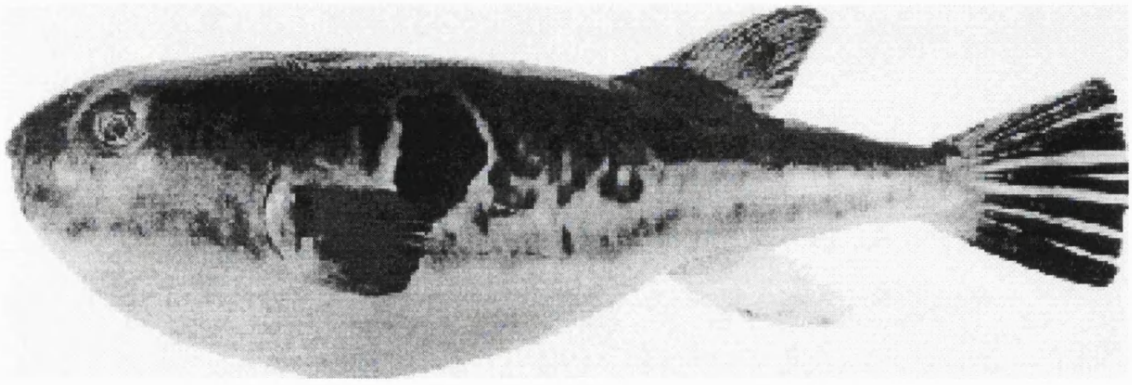


Fig. 7: The Japanese Pufferfish, *Fugu rubripes* has one of the smallest known vertebrate genomes.

possibility not favoured by Brenner and colleagues who suggest that the *Fugu* genome has never acquired the 'junk' DNA that is found in mammalian genomes. Evidence for these assumptions comes from the random *Fugu* sequence data. In the 127831bp of random *Fugu* sequence, the most abundant repeat, GT-AC, occurred 30 times comprising 1050bp in total, the average distance between GT-AC repeats in *Fugu* is (127831-1050)/30= 4.2kb compared with 30kb in human (determined for chromosome 16). Since the *Fugu* genome is estimated to be 7.68 times smaller than the human genome (see above) this suggests that the total number of repeat elements in *Fugu* is the same as in higher vertebrates and this means that genome expansion in mammals occurred between the repeats. Since the positional conservation of the repeat elements is only 30%, repeats are not ancient and the expansion in mammals is a relatively recent event. Unfortunately, this argument does not offer any explanation as to why the *Fugu* genome remained small when most other genomes expanded. Whatever the mechanisms involved, one of the most useful properties of *Fugu* is the small size of its introns.

12. 4. Small Is Beautiful . . .

With a modal (most frequently observed) value of 80bp, the size of *Fugu* introns is reduced by a factor corresponding to the reduction of the genome as a whole. Since *Fugu* genes are fully functional they are likely to contain the necessary control elements in a compressed form, which would make them ideal for identification of regulatory elements, particularly where these are spread over large regions between genes or involve long range regulatory elements that control entire gene families. There is presently too little data to show conclusively how much linkage conservation exists between *Fugu* and mammalian species, especially over large distances, but Trower *et al.* (1996) have demonstrated that a 12.4kb segment of a *Fugu* cosmid containing three genes corresponds to a region of more than 600kb in human, indicating a high degree of synteny and compression between *Fugu* and higher vertebrates. In agreement with this, it has recently been reported that the diploid *Fugu* genome comprises 44 very small chromosomes (Miyaki *et al.*, 1995).

12. 5. . . . Or is it?

If the usefulness of the *Fugu* model beyond its small genome is considered, and specifically how it lends itself to functional studies, there are considerable limitations. The long generation time and large size of *Fugu* are prohibitive for genetics or embryology. Although *Fugu* fish are farmed commercially in Japan, as a culinary delicacy, export of live fish is restricted and consequently resources are very limited. Without *Fugu* embryos, gene expression studies *in situ*, or the generation of transgenic *Fugu* fish are not feasible. For these reasons the *Fugu* model is likely to remain a genomic rather than a true biological model, and it will not replace other model

organisms like mouse, fly and zebrafish. However, another pufferfish *Tetraodon fluviatilis,* a close relative of *Fugu rubripes,* appears to be a more manageable species and has recently been proposed as a suitable model (Crnogorac-Jurcevic *et al.,* 1997). Surprisingly, in his original study, Hinegardner (1968) already reported the genome of *T. fluviatilis* as the smallest of the 300 species of teleost fish examined (including *Fugu*), which poses the question as to why *Fugu rubripes* was the model of choice for Brenner and colleagues. There are several advantages of *T. fluviatilis* over *Fugu rubripes*, most importantly that the former is a fresh water puffer that grows to only 17cm in length (compared to the 50 cm or so for *Fugu*) and can be maintained and bred in a conventional aquarium according to Sterba and Krapp (see Crnogorac-Jurcevic *et al.,* 1997). Initial studies on the *cytochrome-b* and *Wnt* gene sequence confirm the close relationship between *T. fluviatilis* and *Fugu* with more than 80% nucleotide and amino acid identity. Furthermore, these studies show that the high degree of sequence similarity is likely to allow *T. fluviatilis* homologues of most *Fugu* genes to be cloned rapidly by cross hybridisation of *Fugu* probes to *T. fluviatilis* libraries or amplification by degenerate PCR.

12. 6. Studies on Developmentally Regulated Genes in *Fugu*

Irrespective of the particular pufferfish species, the most interesting targets of sequence comparisons with other vertebrates are the developmentally restricted genes because they are the most likely to show regulatory conservation, and amongst these the *Hox* complexes of vertebrates are probably the most conserved in regulation and genomic organisation. Recently the degree of *Hox* gene conservation in *Fugu* has been examined but surprisingly the results revealed considerable organisational differences between *Fugu* and other vertebrates (Aparicio *et al.,* 1997). For example, in the *Hox-d* complex of *Fugu* five genes are missing compared with the mammalian or zebrafish *Hox-d* clusters. Furthermore, all of the group 7 genes and a number of other orthologues are absent in *Fugu.* Also new paralogous *Hox* genes (group 2 genes) have been found in *Fugu*, but not any other vertebrate. The overall size of the *Fugu Hox a, c* and *d* complexes is smaller than their mammalian counterparts, whereas the *Hox-b* complex is slightly larger than in the mouse, indicating significant evolutionary divergence between the *Fugu* and mammalian *Hox* clusters. Despite these apparent differences, previous studies on the *Hox* genes have identified conserved regulatory elements in *Fugu* that were shown to be functional in transgenic mice (Marshall *et al.,* 1994; Aparicio *et al.,* 1995). Specifically, Marshall *et al.* identified two *Hox-b1* enhancer elements conserved between *Fugu* and chick, of which one contained a retinoic acid response element, while Aparicio *et al.* demonstrated that reporter constructs containing conserved *Hox-b4* sequence elements from *Fugu* were able to set the correct boundaries of *Hox-b4* expression between rhombomeres 6/7 in the hindbrain of transgenic mice. To further

evaluate the usefulness of *Fugu* for the identification of functional regulatory elements, it is important to examine the extent of sequence conservation of other developmentally regulated genes, like the myogenic regulatory factors, between *Fugu* and mammals. Parallel comparisons with other teleosts, such as zebrafish, or amphibian models might also be informative for the identification of such elements. However, the small size of the *Fugu* genome should allow for the highest degree of stringency in such comparisons. Indeed comparisons between mouse, chick, and zebrafish sequences of the *Hox-d11* loci have revealed conserved elements involved in the control of *Hox-d11* expression in the tetrapod limb, despite the fact that the zebrafish region was not significantly smaller than the equivalent region in the mouse. Transgenic mice carrying the full fish *Hox-d11* construct as well as hybrid mouse-fish *Hox-d11* constructs showed that the fish sequences were able to drive expression in the mouse forelimbs although the fish elements required the presence of the mouse promoter to do so (Beckers *et al.*, 1996). The ability to test the function of conserved elements in transgenic mice provides the necessary assay to evaluate the *in vivo* role of the candidate regions, and locate binding sites of the relevant transcription factors. In this study the usefulness of *Fugu* for the identification of regulatory elements controlling expression of the myogenic factor *Myf-5* is examined.

13. Separate Regulatory Elements Control Sub-Domains of *Myf-5* Expression

Previous studies on the transcriptional regulation of the *myogenin* gene have established that the complete spatio-temporal expression pattern of *myogenin* can be recapitulated in transgenic mice by only 133bp of promoter sequence containing binding sites for a bHLH factor and a protein of the MEF-2 family (Cheng *et al.*, 1993; Yee and Rigby, 1993). The temporal expression of the myogenic factors, and evidence from the knockout mice, suggested that *Myf-5* is the bHLH factor involved in *myogenin* activation. We therefore turned our attention to the transcriptional regulation of *Myf-5*. Unlike *myogenin*, with its well defined regulatory region, *Myf-5* is subject to complex regulation, which is likely to result from the diversity of signals that *Myf-5* expressing cells have to respond to. In vertebrates *Myf-5* and *MRF4* form a syntenic linkage group. Transgenic mice expressing *LacZ* under the control of the entire region, extending over 13kb from the *MRF4* gene through to the 3' untranslated region (3'UTR) of *Myf-5*, and deletion variants thereof, show that several regulatory elements are located within the intergenic region as well as the *MRF4* and *Myf-5* genes themselves. These elements autonomously drive expression in discrete subdomains of the wild type *Myf-5* expression pattern as defined by *in situ* hybridisation (Summerbell *et al.*, unpublished). The identified elements behave as enhancers both in the context of a heterologous or the

endogenous promoter. In combination they reproduce much, but not all of the wild type expression pattern of *Myf-5*. Notably, the regulatory elements that control expression of *Myf-5* in the limbs are absent from the more than 13 kb of sequence examined. The large size of the region has made the isolation of discrete regulatory elements difficult. In previous studies on the *Hox* cluster, regulatory elements have been identified on the basis of their evolutionary conservation in other species (Marshall *et al.,* 1994; Aparicio *et al.,* 1995; Beckers *et al.,* 1996). For this reason we wanted to investigate the usefulness of established vertebrate model organisms for the identification of regulatory elements controlling *Myf-5*. This approach is based on three assumptions; first, that due to functional constraints the regulatory elements are evolutionarily conserved; secondly, that elements identified in this way can respond to the endogenous transcription factors and third; that the regulatory elements can do so in isolation from their native environment and reproduce (at least) most of the wild type expression pattern in the context of a transgenic reporter.

In this study I show that linkage between *MRF4* and *Myf-5* extends to the pufferfish *Fugu* but that with the exception of the coding regions the *Fugu* sequence is not visibly conserved, and does not allow candidate regulatory elements to be predicted on the basis of sequence conservation. I therefore turned to traditional transgenic methods to identify and characterise the regulatory regions in the mouse *Myf-5* gene itself. Previous studies from our laboratory have shown that *Myf-5* expression in the ventral posterior margin of the dermomyotome depends on regulatory elements located in a region containing the *Myf-5* gene with its 3' UTR. I show here that regulatory elements in each of the two introns and in the UTR are necessary for expression in the ventral posterior margin of the somites. Results from our laboratory indicated that, at least in the case of *Myf-5*, expression in particular anatomical subdomains is controlled by separate regulatory elements. My data suggest that even for these subdomains multiple dispersed regulatory elements may be required to control expression, adding a further level of complexity to the regulation of *Myf-5* expression.

Chapter 2

Materials and Methods

1. Solutions and Reagents

Solutions were made up from laboratory grade reagents and sterile distilled water (dH_2O) and stored at room temperature unless stated otherwise.

Ammonium Persulphate (APS) 10% (w/v)

10 % (w/v) APS (Kodak 11151).

Ampicillin 1000x stock

100 mg/ml ampicillin (sodium salt, Sigma).

Avertin

100 % (w/v) 2,2,2-tribromoethanol dissolved in tertiary amyl alcohol. Diluted to 2.5 % (w/v) in distilled water. Stored in dark.

Bandshift Buffer 2 x

100 mM Tris-HCl (pH 7.9), 12 mM $MgCl_2$, 0.4 mM EDTA, 2 mM DTT, 30 % (v/v) glycerol. Stored in aliquots at -20°C.

BHI Medium

3.7 % (w/v) brain-heart infusion medium (DIFCO). Autoclaved.

Church & Gilbert Hybridisation Solution

1 mM EDTA, 500 mM Na_2HPO_4 (pH 7.2), 7 % (w/v) SDS.

Church & Gilbert Washing Solution

40 mM Na_2HPO_4 (pH 7.2), 1% (w/v) SDS, 1 mM EDTA.

CY Medium

10 g/l casamino acids, 5 g/l yeast extract, 50 mM NaCl, 25 mM KCl, pH 7.0 adjusted with HCl, autoclaved.

Deoxyribonucleoside Triphosphates (dNTPs) 10x

2.5 mM dATP, dCTP, dGTP and dTTP (Ultrapure, Pharmacia).

Embryo Extraction Buffer

10 mM Tris-HCl (pH 8.0), 0.1 M EDTA pH 8.0, 0.5 % (w/v) SDS, 20 µg/ml preboiled RNAse A

Embryo Fixative

Mouse: Mirsky's fixative (National Diagnostics). Made from 10x concentrate and 10x buffer (National Diagnostics) in distilled water. Stored at 4°C.

Zebrafish: 4% (w/v) paraformaldehyde (PFA) in PBS made up fresh, heated to 65°C and stirred until completely dissolved, then cooled on ice prior to use.

Embryo Fuchsin Staining Solution

0.1 % (w/v) acid fuchsin (Sigma) in PBS, pH 7.2 - 7.4.

Embryo Staining Solution

5 mM $K_3Fe(CN)_6$, 5 mM $K_4Fe(CN)_6 \cdot 3H_2O$, 0.01 % (w/v) Na-Deoxycholate, 2 mM $MgCl_2$, 0.02 % (w/v) NP40, 1 mg/ml X-Gal, in PBS.

Embryo Washing Solution

0.02 % (w/v) NP40 made up in PBS and stored at 4°C.

Ethidium Bromide

10 mg/ml stock solution stored at 4°C in the dark.

G50 gel filtration mix

5 % (w/v) slurry of Sephadex™ G50 superfine (Pharmacia) prepared in STE (50 mM NaCl, 10 mM Tris-HCl (pH 7.4), 1 mM EDTA) autoclaved, Triton X-100 added to 0.05 % (v/v) and stored at 4°C.

HEPES, 1M pH 7.5

23.83 g HEPES (Sigma H-3375), dH_2O to 100 ml, pH 7.5 adjusted with potassium hydroxide (KOH) stored at 4°C.

In Situ Hybridisation Solutions I-III

Solution I

4xSSC adjusted to pH 4.5 with citric acid, 1 % (w/v) SDS, 50 % (v/v) formamide, prewarmed to 65°C.

Solution II

500 mM NaCl, 10 mM Tris-HCl (pH 7.5), 0.1 % (v/v) Tween20

Solution III

2xSSC (pH 4.5), 50 % (v/v) formamide, made up fresh.

Hybridisation buffer

55 % (v/v) formamide, 10xSSC, 56 mM EDTA, 100 mg/ml heparin (Boehringer Mannheim); 0.2 % (v/v) Tween (Sigma); 50mg/ml yeast t-RNA (Sigma).

IPTG

Isopropyl β-D-thiogalactopyranoside (Sigma), 1M stock solution.

Kanamycin Sulphate (Kan):

Stock of 5 mg/ml: 500 mg kanamycin (Boehringer Mannheim 106 801) dH_2O to 100 ml. Added at final conc. 20 mg/ml to media.

L-Agar plates

LB containing 1.5% (w/v) bacto-agar (DIFCO). Autoclaved and cooled to 55°C. Poured as plates (ampicillin added for selective plates) and dried at room temperature for 2-3 days. Stored at 4°C.

Lambda Dilution Buffer

100 mM NaCl, 20 mM $MgSO_4$, 50 mM Tris-HCl (pH7.5) made up with dH_2O autoclaved, add gelatin to 0.01 % (v/v)

L-Broth (LB)

1 % (w/v) bacto tryptone (DIFCO), 0.5 % (w/v) bacto yeast extract (DIFCO) and 0.5 % (w/v) NaCl. Autoclaved, stored at 4°C.

Loading Buffer for ABI Sequencing Reactions

5:1 deionized formamide: 25 mM EDTA with 50 mg/ml blue dextran.

Loading buffer for Agarose Gel Electrophoresis

0.25 mg/ml Orange G (Sigma) in 30 % (v/v) glycerol.

Micro-extraction Buffer

20 mM HEPES (pH 7.8), 450 mM NaCl, 200 mM EDTA, 500 mM DTT, 1 µg/ml aproteinin, 1 µg/ml eupeptin, 100 µg/ml PMSF and 1 µg/ml pepstatin A. Made fresh from stock solutions.

Microinjection Buffer

10 mM Tris-HCl (pH 7.4), 0.1mM EDTA. Made up using high quality, endotoxin free water (W1053, Sigma). Filter sterilised through a 0.22 µm filter. Stored in aliquots at -20°C.

MOPS Gel

1 % (w/v) agarose, 35 % (v/v) formaldehyde in 1x northern running buffer.

Northern Running Buffer

0.2 M MOPS, 10 mM EDTA, 50 mM NaAcetate (pH 5.2), made up in 0.1 % (v/v) DEPC, adjusted to pH 7.0 with NaOH and autoclaved.

NTMT

100 mM NaCl, 100 mM Tris-HCl (pH 9.5), 50 mM $MgCl_2$, 0.1 % (v/v) Tween 20

PFA (Paraformaldehyde)

Made up fresh. 4 % (w/v) paraformaldehyde dissolved in PBS, by heating gently, avoiding toxic fumes.

PBS (Phosphate buffered saline)

137 mM NaCl, 2.7 mM KCl, 1.4 mM KH_2PO_4,4.3 mM $Na_2HPO_4 \cdot 7H_2O$

PEG - Mix

26.2 % (w/v) PEG 8000 (polyethylene glycol), 6.6 mM $MgCl_2$, 0.6 M sodium acetate (pH 5.2).

Phenol, TE Saturated

Ultrapure phenol (Amresco) equilibrated to pH 7.8 - 8 with 10 mM Tris-HCl (pH 8.0), 1 mM EDTA. Stored under 100 mM Tris-HCl (pH 8), 0.02 % (w/v) β-mercaptoethanol at 4°C in the dark.

Phosphate Buffer

1M $NaHPO_4$ adjusted to pH 7.2 with 80 % (w/v) H_3PO_4.

Plasmid Solution I: Resuspension Buffer

25 mM Tris-HCl (pH 8.0), 50 mM glucose, 10 mM EDTA.

Plasmid Solution II: Lysis Buffer

0.2 M NaOH, 1% (w/v) SDS.

Plasmid Solution III: Neutralisation Buffer

3 M potassium acetate pH 4.8 (with acetic acid). Autoclaved.

Polymerase Chain Reaction (PCR) Buffer 10x

100 mM Tris-HCl (pH 8.3), 500 mM KCl, 15 mM $MgCl_2$.

RNA Extraction Buffer

25 g Guanidinium thiocyanate, 29.3 ml dH_2O, 1.76 ml 0.75 M NaCitrate, pH7, 2.64 ml 10 % (w/v) Sarcosyl, 38 µl β-mercaptoethanol, stored at -70°C.

RNA Sample Buffer

50 % (v/v) deionised formamide, 18 % (v/v) formaldehyde, 0.1 % (v/v) ethidium bromide (10 mg/ml), 0.1volume of 10x northern running buffer.

RNase A stock (250x)

10 mg/ml RNase A in 10 mM Tris-HCl (pH 7.5), 15 mM NaCl. Boiled for 15 minutes to inactivate DNase contamination.

Sodium Acetate (NaOAc)

3 M Sodium Acetate (pH 5.2) adjusted with glacial acetic acid.

Southern Denaturing Solution

400 mM NaOH, 600 mM NaCl.

Southern Neutralizing Solution

500 mM Tris-HCl (pH7.0), 1M NaCl.

SSC (Standard Saline Citrate) 20x

3 M NaCl, 0.3 M $Na_3citrate{\cdot}2H_2O$, adjusted to pH 7.0 with 1M HCl.

TBE (Tris-Borate EDTA) Electrophoresis Buffer 1 x

89 mM Tris, 89 mM boric acid, 2 mM EDTA.

TBS (10x)

1.5 M NaCl, 250 mM KCl, 250 mM Tris-HCl (pH 7.5).

TBST

1xTBS, 0.1 % (v/v) Tween 20

TE (Tris EDTA) Buffer

10 mM Tris-HCl (pH 8.0), 1 mM EDTA. Autoclaved.

Top-Agarose

0.6 % (w/v) agarose in L-Broth, autoclaved and kept at 55°C immediately before use.

Transcription buffer for T3, T7 RNA polymerases

200 mM Tris-HCl (pH 8.0), 40 mM $MgCl_2$, 10 mM spermidine, 250 mM NaCl.

X-Gal 4 % (w/v)

40 mg/ml 5-bromo-4-chloro-3-indolyl-β-D-glactoside (X-Gal) made up in dimethylformamide (DMF) , stored at -20°C in the dark.

2. DNA Manipulations

2. 1. Phenol Extraction of DNA

An equal volume of TE-saturated phenol was added to an aqueous DNA sample. The mixture was vigorously vortexed, prior to centrifugation at 13000 x g at room temperature for 5 minutes. The upper, aqueous layer was carefully removed, avoiding the phenol interface and an equal volume of chloroform was added followed by vortexing and centrifugation as before. The upper aqueous, layer was carefully removed and ethanol precipitated.

2. 2. Ethanol Precipitation of DNA

One tenth volume of 3M NaOAc (pH 5.2) and 2.5 volumes of 100% ethanol were added to the DNA sample which was incubated for 10-30 min at -70°C before the DNA was recovered by centrifugation at 13000 x g at room temperature before rinsing the pellet with 70 % (v/v) ethanol. Where ligations were precipitated prior to electro-transformation, 0.5 ml of 20 mg/ml glycogen (Boehringer Mannheim) were added per 10ml ligation as a carrier and then ethanol precipitated as before.

2. 3. PEG Precipitation of DNA

PEG precipitation was used mainly for purification of PCR products from oligonucleotides and dNTPs prior to sequencing and for precipitation of reactions, where buffer changes were necessary. PEG also precipitates residual enzymes from the reaction. Briefly, an equal volume of PEG-mix was added to the reaction, which was incubated for 5 min at room temperature and centrifuged at 13000 g for 5-10 min. The usually invisible pellet was washed with 70 % (v/v) ethanol prior to resuspension in TE.

2. 4. Endonuclease Digestion of DNA

Restriction enzyme digestions were performed in the manufacturer's buffers according to manufacturer's instructions. The minimum reaction volume was 10 µl with a maximum DNA concentration of 100 ng/µl.

2. 5. DNA Fragmentation by Sonication

Random DNA fragmentation for subcloning was achieved by sonication (Deininger, 1983) using a High Intensity Ultrasonic Processor (Vibra Cell VC300, Sonics & Materials).

2. 6. Dephosphorylation of DNA

5' Phosphates were removed from digested vector DNA using calf intestinal alkaline phosphatase (CIAP, Boehringer Mannheim) to avoid self-ligation of vector ends specifically where α-complementation could not be used to screen for recombinants (eg.

when generating reporter constructs). Between 1 and 10 units of CIAP were added to dephosphorylate 50 pmoles 5' terminal phosphorylated vector ends by incubation at 37°C for 30 minutes. The phosphatase was inactivated either by addition of EDTA (50mM final concentration) and incubation at 85°C for 20 minutes, followed by PEG precipitation or phenol extraction.

2. 7. Repair of DNA Ends

(i) 5' Overhangs

5' overhangs were filled in by incubation with 1 U of Klenow fragment of DNA polymerase I (Boehringer Mannheim) in the presence of 50 mM dNTPs at room temperature for 20 minutes.

(ii) 3' Overhangs

3' overhangs were removed by the 3'- 5' exonuclease activity of T4 DNA Polymerase (Boehringer Mannheim). One µg of DNA was incubated with one unit of enzyme in the presence of 50 mM dNTPs at 12°C for 30 minutes.

(iii) PCR Products

PCR products that were to be cloned were usually amplified with *Pfu (Pyrococcus furiosus)* polymerase (Stratagene) which possesses a more than ten fold greater fidelity than Taq polymerase. The estimated fidelity for Pfu is 1.6×10^{-6} errors/base (Lundberg *et al.*, 1991) compared to Taq's 1.1×10^{-4} errors/bp (Tindall and Kunkel, 1988). An additional advantage of Pfu is that unlike Taq it gives rise to blunt-ended products. Amplification products generated by Taq polymerase have 3' overhangs which were polished with either T4 DNA Polymerase or Klenow as above.

(iv) Sonicated DNA Fragments

Sonication products for "shot gun cloning" were end-repaired using Klenow enzyme as above.

2. 8. Oligonucleotide Primer Synthesis

Oligonucleotides were chosen to have a GC content of about 50 %. Repetitive motifs and stretches of any one nucleotide were avoided. Most oligonucleotides were at least 17 bp in length with annealing temperatures of 50°C or more. After oligonucleotide synthesis, the base- and phosphorous protecting groups must be removed to ensure that the oligonucleotide is biologically active. To deprotect the DNA, the oligonucleotide was incubated in ammonium hydroxide (NH_4OH) for 5-12 hours at 55 °C or for 1 hour at 85°C. DNA was recovered by ethanol precipitation, and the concentration was assayed by measuring the spectrophotometric absorbance of a diluted aliquot at 260 nm. The concentration was calculated, assuming that for single stranded oligonucleotides 1 OD equals 20 µg/ml and that the average molecular weight per deoxyribonucleotide is 325 g/mol. Some oligonucleotides were purchased from Oswel.

2. 9. Agarose Gel Electrophoresis and Documentation

DNA fragments were generally separated by electrophoresis through 0.7-1.5 % (w/v) agarose in 1 x TBE containing 0.25 µg/ml ethidium bromide. Small DNA fragments of less than 200 bp in size were efficiently resolved on 2 % (w/v) low melting point agarose gels (Seaplaque, Flowgen).

The products were visualised under ultra-violet light and photographed using a Polaroid Land Camera fitted with a Wratten filter (No. 25, Kodak) and Polaroid 667 black and white film or video captured using the UVP gel documentation system (GDS 7500) and printed on a Mitsubishi video copy processor.

2. 10. DNA Fragment Isolation

Fragments were size separated by agarose gel electrophoresis, excised from the gel and recovered using the QiaexII Gel Extraction kit (Qiagen) following the manufacturer's instructions. For microinjection constructs, more extensive purification was applied: The DNA fragments were run into a 0.6 % (w/v) low melting point agarose gel (Seaplaque, Flowgen), excised from the gel and incubated for 1-2 hours at 42°C with 5U gelase enzyme (Epicentre Technologies) to dissolve the agarose. Subsequently the DNA was recovered by phenol extraction followed by two extractions with chloroform, and finally ethanol precipitation. The DNA was taken up in microinjection buffer and diluted appropriately prior to pronuclear injection.

3. DNA Cloning

3. 1. DNA Ligations

Ligations containing 1-5 ng/µl of linearized vector DNA and a 2-3 fold molar excess of insert DNA were incubated with 1 unit of T4 DNA ligase (Boehringer Mannheim) in 1x supplied ligation buffer for 45 min -2 hours at 25°C or overnight at 16°C. The DNA ligase was inactivated by incubation at 70°C for 10 min. For blunt end ligations, vector self-ligation was largely eliminated by redigestion of the ligation with the endonuclease that had been used to linearise the vector prior to transformation. For electro transformation, salt ions, which may adversely affect the conductivity of the reaction were removed from the ligations by ethanol precipitation in the presence of 10 µg of glycogen (Boehringer Mannheim). Usually one half of the ligation was used to transform competent cells (see below).

3. 2. Transformation of Competent Cells

The electroporation method of transforming *E. coli* can produce efficiencies greater than those achieved with the best chemical methods and was used throughout (Bower *et al.*,

1988). 25 µl of *Epicurian Coli* Electroporation-competent cells (Stratagene) were mixed with 10-50 ng of ligation products and added to a Gene Pulser®/ E coli Pulser cuvette™ (Biorad). Electroporation was carried out using a Biorad GenePulser using the recommended conditions (1.7 Volts, 0.25 mFarrad, 200 Ohms). The transformation products were immediately resuspended in 500 µl BHI and incubated shaking for 30 min at 37°C before plating on L- agar plates containing ampicillin (10 µg/ml). Where appropriate, 40 µl of 4 % (w/v) X-Gal and 5 µl of 1 M IPTG were added per agar plate to allow blue/white screening. Plates were incubated upside down, overnight, at 37°C.

3. 3. Recombinant Screening

i) α-Complementation

The use of bacterial strains deficient for the α–peptide of ß–galactosidase allows recombinant bacterial colonies to be screened as *LacZ* negatives on agar plates containing the lac operon inducer IPTG and the ß-galactosidase substrate X-Gal. In general, the gene for the α- peptide is inactivated in recombinant plasmids, resulting in the formation of white colonies on plates containing X-Gal. Nonrecombinant bacterial colonies complement the host deficiency by expressing the α-subunit of the ß-galactosidase gene and metabolise X-Gal thus forming a blue precipitate.

ii) PCR Screening

Individual bacterial colonies were picked into a 96 well microtiter culture dish (Sero Well®, Bibby Sterilin Ltd.) containing 100 µl BHI (supplemented with ampicillin: 100 $\mu g \cdot ml^{-1}$) per well.

After 3-12 hours of incubation at 37 °C, a 96 pin replicator (hedgehog) was used to transfer approximately 0.5µl of medium from the culture dish to a 96 well thermostable thermowell™ plate (Costar) containing 20 µl of PCR-mix (0.1 mM dNTPs, 1x PCR buffer, PCR primers each at 0.5 µM concentration and 0.5 units of Amplitaq® (Perkin Elmer) made up with distilled water). The bacteria are lysed in the PCR machine by incubation at 95°C for 90 seconds, followed by 32 cycles of denaturation at 95°C for 30 seconds, primer annealing for 30 seconds at 45-65°C (depending on the primer) followed by primer extension at 73°C. One minute extension was allowed for each kb of amplification product. An aliquot of the PCR products was examined by gel-electrophoresis. Where necessary the PCR products were PEG precipitated and sequenced or analysed by restriction endonuclease digestion.

4. Isolation of DNA and RNA

4. 1. Plasmid Isolation by Alkaline Lysis "Miniprep"

Overnight bacterial cultures (2 ml) were pelleted in a microfuge at 13 000 x g for 2 minutes. The pellets were resuspended in 200 µl of plasmid solution I, lysed with 200 µl of solution II and neutralised with 200 µl of solution III. Samples were then extracted by shaking with 500 µl of phenol:chloroform and centrifuged (13 000 rpm for 2 minutes). The aqueous phase (500 µl) was transferred to a fresh tube, precipitated with isopropanol and pelleted at 13 000 rpm for 10 minutes at RT. Plasmid DNA was washed with 70 % (v/v) ethanol, air dried and resuspended in 30 µl of TE containing 40 µg/ml RNase A. 2-5 µl was used for restriction analysis.
For more extensive manipulations, DNA was prepared from 250 ml overnight cultures by the alkaline lysis method and further purified by equilibrium centrifugation in a CsCl-ethidium bromide gradient (see Sambrook *et al.*, 1989).

4. 2. Plasmid Isolation by CsCl "Maxiprep"

An individual bacterial colony was picked to inoculate 250 ml of BHI supplemented with ampicillin at 100 $\mu g \cdot ml^{-1}$. The culture was incubated at 37 °C for 12-16 hours shaking vigorously. A glycerol stock of a 1 ml aliquot of the culture was prepared by addition of 500 µl of 50 % (w/v) glycerol and stored at -80°C. From the remaining culture bacteria were harvested by centrifugation (8000 rpm, 20 minutes, 20°C) in a Sorvall RC-5B centrifuge fitted with a GSA rotor, then resuspended in 20 ml of plasmid solution I. Cells were lysed by gentle inversion with 40 ml of solution II and the lysate neutralised with 30 ml of plasmid solution III. After centrifugation (GSA- 7 rotor, 10 000 rpm, 15 minutes, 4°C) the supernatant was filtered through 2 layers of medical gauze (Johnson & Johnson) and the plasmid DNA precipitated with 0.6 volumes of isopropanol. The previous centrifugation step was repeated and the DNA pellet rinsed with 70% (v/v) ethanol, dried and resuspended in 5 ml of TE. 4.75 ml of this solution were transferred to a universal container (Sterilin) and 5.1 g of CsCl and 250 µl of ethidium bromide (10 mg/ml) were added. After 5 minutes at room temperature, a precipitate had formed which was pelleted in a Sorvall RC-6000B centrifuge (15000 rpm, 5 minutes, 20°C). The supernatant was transferred to a polyallomer ultracentrifuge tube (Beckman Quick-Seal #342412) and heat sealed carefully. Supercoiled plasmid DNA was banded by centrifugation in a Beckman VTi 80 rotor (70000 rpm, 4 hours, 20°C). Banded DNA was removed from the centrifuge tube with a syringe and extracted with 2-5 volumes of water-saturated butanol to remove the ethidium bromide. CsCl was removed by diluting to 2.5 volumes with distilled water followed by alcohol precipitation at room temperature. The DNA concentrations were determined from spectrophotometric

absorbance of an aliquot at 260nm (assuming that for double stranded DNA 1 OD_{260} = 50 µg/ml).

4. 3. Genomic DNA Isolation from Zebrafish Embryos

Genomic DNA from 12-14 somite stage zebrafish embryos was prepared by the following method. Embryos were allowed to grow at 29 °C for 36 hours until they had reached the appropriate developmental stage. The zebrafish were placed into a sterilin tube and stirred up repeatedly in order to remove damaged embryos and unfertilised eggs which tend to settle more slowly and can be removed with the supernatant. Approximately 300 embryos were frozen on dry ice and used later for genomic DNA preparation, the remaining embryos were used for *in situ* hybridisation.
Prior to use the frozen embryos were thawed on ice, resuspended in 1ml of embryos extraction buffer and macerated. The preparation was incubated at 37°C for 1hour to allow RNA to be hydrolysed. In all subsequent steps care was taken not to shear the DNA by vortexing or pipetting. Proteinase K was added to 100µg/ml followed by incubation at 50°C for 3 hours. DNA was extracted twice with phenol saturated with 0.5M Tris-HCl (pH 8.0), 10mM EDTA followed by two extractions with $CHCl_3$.
The extracted DNA was precipitated with 0.1 volume of 3M Na acetate, 1mM EDTA pH 7.0 and 0.54 volumes of isopropanol and dissolved in 300µl of 0.2 x SSC overnight. The optical density at 260 nm was measured and the DNA concentration determined. A digest of 10 µg DNA in restriction buffer with and without *Eco*R1 produced a continuous smear, with a prominent satellite band around 1kb, with *Eco*R1 and without the enzyme a tight band at limit mobility (>23Kb) was apparent.

4. 4. Cosmid DNA Isolation

A cosmid culture was streaked out onto L-agar plates supplemented with kanamycin and incubated overnight at 37 °C. A single cosmid clone was picked and used to inoculate 5-200 mls of BHI containing 100 µg/ml kanamycin. Following incubation overnight at 37 °C with shaking, DNA was prepared using the Wizard™ Miniprep system (Promega) according to the manufacturer's protocol. Larger quantities of cosmid DNA were prepared using Qiagen Midipreps following the manufacturer's instructions.

4. 5. Isolation of RNA

RNA from zebrafish was isolated by a modification of the method of Chomczynski and Sacchi (1987). About 400 zebrafish embryos between 14 and 22 somite stage were used for RNA preparation. The embryos were resuspended in 1 ml of RNA extraction buffer and vortexed vigorously until they started to decompose. 0.1 volumes of 2M NaAcetate pH 4 were added and vortexed. The preparation was then extracted with an equal volume of unbuffered phenol, followed by choloroform extraction. Two volumes of

ethanol were added for precipitation on dry ice following centrifugation at RT for 15 min at 13000g before the pellet was rinsed with 70% (v/v) ethanol, air dried briefly and resuspended in RNase free ddH_2O.

4. 6. cDNA Synthesis

Reverse transcription PCR (RT-PCR) was used to clone the bHLH domain of zebrafish *Myf-5*. Total RNA from whole zebrafish *Danio rerio* (36 hours) was a kind gift from Linda McNaughton (Division of Developmental Neurobiology, NIMR). First strand cDNA synthesis was carried out by annealing 6.5 pMoles of dT_{24} primer to 1 µg of RNA by incubation at 65 °C for 10 min allowing the reaction to cool down slowly to room temperature before supplementing with 4 µl of 5x RT buffer (Boehringer), 1 µl RNase inhibitor (Promega), 2 µl of 2 mM dNTPs (Boehringer) and H_2O to 20 µl followed by incubation at 42°C for 5 min before adding 1.5 µl avian reverse transcriptase (Boehringer). Incubation was continued for a further 40 min at 42°C. The cDNA was stored at -20°C.

5. Southern Analysis

5. 1. Southern Blotting

The alkaline Southern blotting method was used to transfer DNA from agarose gels to positively charged nylon membranes (GeneScreenPlus®, Du Pont) for subsequent Southern hybridisation (Southern, 1975). For transfer of large DNA fragments (>3kb) the agarose gel was depurinated in 0.25 M HCl for 10 min and then soaked in neutralising solution for 10 min (Wahl *et al.*, 1979). A transfer tray was set up as described by Sambrook *et al.*, (1989). Blotting was performed for at least two hours or overnight. The DNA was fixed to the nylon membrane by exposure of the "DNA-side" to ultraviolet irradiation (254nm) for 30 seconds in a Stratalinker® (Stratagene). The membrane was subsequently rinsed in neutralisation solution for 10 min and stored at 4°C for hybridisation to a radiolabelled probe.

5. 2. Southern Hybridisation Probes

DNA fragments to be used as Southern hybridisation probes were generated by PCR amplification. Labelling was performed using the Klenow fragment of *E. coli* DNA polymerase I, either to extend a probe specific primer or by random hexamer primed DNA synthesis (Prime-a-gene labelling kit, Promega). Because the Klenow Fragment lacks 5'-3' exonuclease activity, labelled nucleotides incorporated during hexamer labelling are not subsequently excised as monophosphates. In the labelling reaction dCTP was substituted with a ß-particle emitter (α–^{32}P)-dCTP thus generating highly labelled DNA. The radionucleotides were purchased from Amersham with a half life of

14.3 days, and specific activity of 3000 Ci·mmol^{-1}. The probes obtained ranged in size from 75 to 600 nucleotides with a median size range of 200-400 nucleotides. The products of the labelling reactions were subjected to a purification step in order to remove unincorporated radionucleotides and judge the efficiency of the labelling reaction. Briefly, a spin dialysis column was assembled in a 1.5 ml Eppendorf tube by addition of 1ml of Sephadex CL6B resin (Pharmacia) over a bed (200 µl) of glass beads (glass beads for gas chromatography, BDH #15031). Two holes were introduced near the bottom of the tube, then the column was placed inside a universal container (Sterilin) and centrifuged at 500 x g for 2 minutes. A collection tube was put in place, the probe was added to the column and the centrifugation repeated. The column was rinsed with 20 µl of TEN and centrifuged again to collect the probe. Prior to use, all probes were heat denatured at 95°C for 2 minutes and snap-chilled on ice to prevent reannealing.

5. 3. Southern Hybridisation

The nylon hybridisation membranes prepared by Southern blotting were prehybridised in hybridisation solution at 50°C for 15 min (60 min when used for the first time) before the denatured probe was added and hybridisation was continued for 2-12 hours. Hybridisation temperatures were chosen according to the length and specificity of the probe:

i) Oligonucleotide probes were hybridised at a temperature that was 10°C below the Tm of the oligonucleotide.

ii) Specific probes of more than 60 bp length were hybridised at 65 °C.

iii) Cross species probes of more than 60 bp in length were hybridised at 50°C.

After hybridisation the nylon membrane (GeneScreenPlus®, Du Pont)) was rinsed in Church and Gilbert washing solution several times for 10 min until background hybridisation had been sufficiently eliminated. The stringency of the washing steps was successively increased by raising the temperature by 5 °C after each washing step, starting at 10°C below the hybridisation temperature. Progress of the washing steps was monitored with a Geiger counter, or exposure of the membrane to a phosphoimager screen (Molecular Dynamics) or autoradiographic film (Kodak XAR).

6. Northern Analysis

6. 1. Northern Blotting

RNA samples were prepared by precipitating an appropriate amount of RNA in the presence of 0.1 volumes of 4M LiCl and 2 volumes of ethanol, supplemented with 1/50 volume of 250 mM EDTA, at -70°C for 30 min. The pellet was rinsed in 70% (v/v) ethanol, before being resuspended in 10 µl of RNA sample buffer followed by incubation at 65 °C for 5 minutes. Typically 5 µg of total RNA or 1-2 µg of polyA

selected RNA per lane were loaded on a 1% "MOPS gel". Electrophoresis was carried out in 1x northern running buffer (made up with dH_2O, rather than DEPC-H_2O) at 10 V/cm for several hours. Bromophenol blue loading dye was added to one of the lanes to monitor the progress of electrophoresis. Following electrophoresis, the gel was photographed with a ruler aligned with the wells prior to being soaked in 20x SSC for 15 min. A northern transfer tray was set up, as described by Sambrook *et al.* (1989) for Southern blotting, except that 20x SSC was used as the transfer buffer. GeneScreen Plus (Du Pont) Transfer Membranes were prewetted in 20x SSC prior to use. Blotting was performed overnight. The membrane was rinsed in 2x SSC before the RNA was fixed to the nylon membrane using a Stratalinker® (Stratagene) as recommended by the manufacturer. The membrane was stored at -20°C until hybridisation to a radiolabelled probe.

6. 2. Riboprobes for Northern Hybridisation

Plasmid clones containing suitable cDNA fragments were prepared by maxiprep as described. 10 µg of plasmid DNA were linearised with a suitable restriction enzyme, extracted once with phenol and twice with chloroform, before ethanol precipitation. The pellet was rinsed in 70 % (v/v) ethanol before resuspending in RNase-free ddH_2O. 1µg of linearised plasmid DNA was used in a 20 µl labelling reaction, supplemented with 2µl of 5mM ribonucleotides (ATP, CTP, GTP, Boehringer) , 1 µl RNase inhibitor (Promega), 2 µl transcription buffer (10x), 2 µl RNA polymerase (T7, T3 depending on which strand of the template was transcribed), 5 µl (α-^{32}P)-UTP (Amersham) with a specific activity of 400 Ci·mmol^{-1}. The labelling reaction was incubated at 57°C for 2 hours. Following the labelling reaction, the DNA template was hydrolysed by addition of 0.5 µl of DNaseI and incubation at 37°C for 30 minutes. Finally the reaction product was ethanol precipitated and resuspended in dH_2O prior to use.

7. Library Screening

7. 1. *Fugu* Cosmid Library Screening

A *Fugu* genomic cosmid library was kindly provided by Dr G. Elgar from Dr. S. Brenner's laboratory in Cambridge. The library had been constructed in the lawrist 4 vector and consisted of an ordered array of 38016 clones spotted onto two high density filters. Both filters were screened by hybridisation with mouse *Myf-5* cDNA and *MRF4* genomic probes. Hybridisation was carried out at 50°C according to the protocol described by Church and Gilbert (1984). For a secondary round of screening, positive cosmid clones were picked into TY medium containing 20mg/ml kanamycin and grown overnight, before DNA was prepared using Wizard Minipreps® (Promega). An aliquot of the miniprep cosmid DNA was digested with *Eco*R1 and the digestion products

resolved by agarose gel electrophoresis. The DNA was transferred to a nylon membrane (GeneScreenPlus®, Du Pont) by Southern blotting, and probed with a labelled PCR product spanning the mouse *MRF4* helix-loop-helix domain. Only four cosmids hybridised to the HLH-PCR probe. A subset of these cosmids was identified again using a different secondary screening strategy: 1 µl of miniprep DNA from the positive first round cosmids was spotted onto three separate gridded Nylon Filters (Hybond N®, Amersham). The DNA was allowed to dry, then soaked first in denaturing solution for 3 min and then in neutralising solution, before being baked onto the filter at 80°C for 2 hours. Each of the three filters was probed at 50°C with one of the following probes: the *Myf-5* cDNA probe, the *MRF4* genomic probe, or the labelled HLH specific PCR product. Two candidate cosmids were identified in this way.

7. 1. 1. Sequence Scanning of Cosmid Clones

A medium size DNA preparation was made from a cosmid clone using the Plasmid-Midiprep-Kit (Qiagen). The cosmid DNA was fragmented by sonication, the ends were polished with Klenow, and ligated to *Eco*RV cut pBS II KS + vector (Stratagene). More than 800 subclones were picked into 96 well culture dishes and PCR amplified as described (Rosenthal *et al.,* 1993). The products were separated by gel electrophoresis, blotted and probed with the HLH specific PCR fragment. Hybridisation positive subclones were sequenced and database searches revealed homology to the *MRF4* and *Myf-5* homologues in the SWISSPROT database.

7. 2. Zebrafish cDNA Library Screening

7. 2. 1. Plating a Phage Library

A cDNA library of post-somitogenesis zebrafish constructed by Robert Riggleman and Karthryn Helde in Lambda ZAP II (Stratagene) was distributed by David Jonah Grunwald (Dept. of Human Genetics, University of Utah, USA). The titre was approximately $9x10^8$ pfu/ml and had recently been determined by counting plaque forming units (pfu) obtained from serial dilutions of the library. Lawn cells were prepared by picking a single colony of XL-1 blue (Stratagene) cells into a starter culture, which was used to inoculate 500 ml of L-broth, supplemented with 0.2% (w/v) Maltose and 10 mM $MgSO_4$, The culture was grown to saturation and cells pelleted by centrifugation at 2000g for 5 minutes at room temperature. The pellet was resuspended in 50 ml of 10 mM $MgSO_4$, and storable for several days at 4°C for subsequent plating steps. 10 mls of cells were infected with $9x10^5$ pfu of the library and incubated with shaking at 37°C. After 30 minutes 25 ml of molten top agarose (50°C) were added and quickly poured over prewarmed L-plates (without antibiotic). The top agar was allowed to set and plates incubated at 37°C upside down for 12-14 hours. In total $1.8x10^6$ pfu were plated onto two 22 x 22cm L-agar dishes (BioAssay Dishes, NUNC, Denmark)

equivalent to 18 pfu·mm^{-2} although examination of the plates 12 hours later suggested that the density of the phage plaques was about 5 times lower than that.

7. 2. 2. Filter Preparation and Hybridisation

Four sheets of nylon membrane (Hybond N+, Amersham), two for each plate, were cut to squares of 22x20 cm. The edges were marked with a whole puncher and labelled with pencil. The membranes were carefully applied onto the surface of the phage plate such that almost the entire surface of the plate was covered. The plaques contain unpackaged recombinant DNA which binds to the filter and can be denatured, fixed and hybridised. The position of the membrane on the plate was outlined on the back of the plate. After 1 minute, the membrane was lifted from the phage plate and placed with the DNA side up onto a sheet of Whatmans 3MM soaked thoroughly in denaturing solution. After 5 minutes, the membrane was place onto a second sheet of Whatmans 3MM soaked in neutralising solution where it remained for another 5 minutes before finally being submerged in 2x SSC. A duplicate filter was prepared from the same plate by applying a new membrane to the surface of the phage plate, this time for 2 minutes, before lifting it into denaturing solution, followed by neutralisation solution and 2x SSC as before. Secondary phage screening was carried out in the same way except that smaller round dishes (13 cm diameter) were used with pre-cut circular membranes (Hybond N+, Amersham). The phage DNA was UV cross linked to the membranes using a Stratalinker™ (Stratagene) as recommended by the manufacturer. A suitable probe was obtained by RT- PCR using degenerate primers as described below and labelled using the prime-it random primer labelling kit (Stratagene) to a specific activity of about 3x 10^9 dpm/µg. The nylon membranes were prehybridised in Quickhybe™ (Stratagene) for 2 hours, before the denatured probe was added. After overnight hybridisation, the filters were washed at high stringency in 0.1% (w/v)SDS, 0.1xSSC with increasing temperature of up to 75°C and air dried briefly before exposure to autoradiographic film (Kodak) overnight. Double positive plaques were identified by aligning the autoradiographs of the duplicate filter sets. The positive phage were then identified on the original phage plate and stenciled out with the base of a large pipette tip. The agarose plug was resuspended in 1ml of lambda dilution buffer and kept at 4°C. 1 µl of this phage stock was diluted in 100 µl of lambda dilution buffer. Of this solution 0.5 and 5 µl were used to infect 3 ml of lawn cells, incubated as before and plated with 6 ml of top agarose. Duplicate filters were lifted, and screened as before.

8. Zebrafish *Myf-5* Cloning by Degenerate PCR

Using an aliquot of zebrafish cDNA (see cDNA Synthesis for details) the bHLH domain of zebrafish *Myf-5* was amplified by PCR using a hemi-nested approach with degenerate oligonucleotides. Two degenerate oligonucleotides Zebra1 and Zebra2 were designed on the basis of sequence conservation in the bHLH domain and paired in low stringency PCR with *Fugu* Myf-5 Rev primer (see Appendix I). The first round amplification was carried out with Zebra1 and Myf-5 Rev. The first three cycles of thermocycling were performed at low annealing temperature 45°C for 1min and primer extension at 67°C for 3 min followed by more stringent amplification at 52°C with 3 minute extensions at 73°C for 29 more cycles. No amplification products were visible on agarose gels after this PCR. An aliquot of the reaction product was amplified in a second PCR using the nested Zebra2 primer and Myf-5-Rev. Amplification was carried out using 30 cycles of 95°C 30sec, 52°C 1min, 72°C 3min. The product was clearly visible after agarose gel electrophoresis and was eluted, blunted and cloned into the *Eco*RV site of pBS-KS (Stratagene). A total of 15 recombinant clones (apparently of identical size) was sequenced with primers M13 F and M13 R in both orientations. The sequence obtained appeared to fall into three homology groups, for each of which a consensus was generated using the DNA Star package (DNA Star Inc.) and compared to existing SWISSPROT and GENBANK database entries. The consensus sequence showed some 70% homology with both mouse *Myf-5* and Japanese quail in two blocks of 50 and 70bp respectively. Comparison with *Fugu Myf-5* however revealed 80% and 95% identity at the nucleotide- and amino acid level, respectively, suggesting that it was zebrafish *Myf-5*.

9. DNA Sequencing

Nucleotide sequences were determined by a variation of the dideoxynucleotide chain termination method (Sanger *et al.*, 1977) using Taq-dye-terminator® cycle sequencing chemistry (Perkin Elmer) on automated sequencing instruments ABI373A and ABI377 (Perkin Elmer).

9. 1. Sequencing on the ABI 373

Approximately 1 µg of either mini- or maxi-prep DNA or PEG precipitated PCR product was used as template in the cycle sequencing reactions using the Prism®Dye Terminator-sequencing kit (Perkin Elmer) following the manufacturer's recommended conditions. Unincorporated dye terminators were removed from the cycle sequencing products by a gel filtration step as described by Rosenthal and Jones (1992). The reactions were

resolved on 6% (w/v) polyacrylamide gels (7M urea) at 60 watts for 10 to 12 hours. Shotgun sequencing of cosmid subclones was carried out as described (Rosenthal *et al.*, 1993). Shotgun clones were sequenced from both ends and sequences were assembled into contigs using the Seqman module of the DNA Star package (DNA Star Inc.) for Macintosh® computers. Gaps were closed by primer walking.

9. 2. Sequencing on the ABI 377

Due to the increased sensitivity of the instrument and improved chemistry, only 400 ng of double stranded plasmid DNA or 200 ng of PEG precipitated PCR product were used as templates in the cycle sequencing reaction using the Amplitaq FS, Prism Dye Terminator kit (Perkin Elmer), according to the manufacturer's instructions. Unincorporated dye terminators were removed by ethanol precipitation prior to resuspension in 6 µl formamide loading buffer. 2 µl of the reaction were loaded onto a 4 % (w/v) denaturing polyacrylamide gel and resolved by electrophoresis for 7 hours at 1000 volts in 1x TBE buffer. Sequence analysis was carried out using the DNA Star™ software package (DNA Star Inc.) as before.

10. Generation and Analysis of Transgenic Mice

10. 1. Transgenic Reporter Constructs

The reporter plasmid used (pHLS-I) was kindly provided by J. Gilthorpe of our laboratory, containing the hsp68-*LacZ*-SV40 poly(A) reporter gene (Whiting *et al.*, 1991) in a pUC (Pharmacia) vector backbone.

10. 1. 1. Mouse Myf-5 Reporter Constructs.

Constructs involving the mouse Myf-5 gene were based on sequence data compiled by Summerbell and Halai from our laboratory. The genomic sequence of mouse *Myf-5* is shown in Appendix VI with annotations relating to the cloning steps presented here.

p12 (both introns): A PCR product was generated with 1U of *Pfu* polymerase (Stratagene) using primers In1f and In2r and 10ng of pMEEK (Summerbell *et al.*, unpublished) and cloned into the *Msc*I site upstream of the hsp68 minimal promoter driving *LacZ* to give **p12f** and p12r. Only p12f was used to make transgenic animals.

p1 (intron 1 construct): p12r was digested with *Bgl*II at a unique site in the vector immediately upstream of the insert and a second site within the insert, just 30 bases from the 5' end of intron 2 , thus liberating all but 30 bases of intron 2 , leaving intron 1 and exon 2 of *Myf-5* in the reverse orientation upstream of the hsp68 promoter driving *LacZ*.

p2 (intron 2 construct): p12 was digested with *Bgl*II at a unique site in the vector immediately upstream of the insert and a second site in the insert just 30 bases from the 5' end of intron 2, thus liberating the first intron and exon 2 from p12, leaving intron 2 in the forward orientation upstream of hsp68 driving *LacZ*.

pUTR (3' UTR construct): pMEEK was digested at a *Pst*I site located 11 bases from the 5' end of exon 3 of *Myf-5* and an *Nhe*I site located downstream of the *Myf-5* 3' UTR to give a 1.7kb *Pst*I/*Nhe*I restriction fragment that was gel purified, and cloned into *Pst*I/*Xba*I digested reporter plasmid pHLSI upstream of the hsp68 promoter driving *LacZ.*

p12UTR (Both introns + the UTR construct): An *Msc*I/*Nhe*I restriction fragment starting at an unique *Msc*I site in intron 1 of *Myf-5*, and extending to the same *Nhe*I site that delimited the 3' UTR in the pUTR construct, was cloned into *Msc*I/*Xba*I digested p1.

p1UTR (intron 1 + 3'UTR construct): Digestion of p12 with *Bgl*II liberated a fragment containing Intron 1, exon 2 and the first 30bp of intron 2 of *Myf-5.* This restriction fragment was cloned into a unique *Bgl*II site of pUTR upstream of the 3' UTR. This reporter was linearised for injection by a *Sal*I/*Bam*HI (partial) digest.

p2UTR (intron 2+UTR construct): p12UTR was digested with *Not*I to remove the vector backbone, followed by restriction at a unique *Msc*I site 41bp upstream of exon 2 to remove all but 41bp of intron 1 prior to injection.

pCFUTRR (5' deletion variant of p12UTR in which 330bp of intron 1 were deleted): a PCR product was generated using *Pfu* Polymerase (Stratagene) with primers CF and UTR-Reverse and cloned into the *Msc*I site of pHLSI upstream of the hsp68 minimal promoter driving *LacZ.* Only the forward orientation was used to produce transgenic animals by pronuclear injection.

pCFH (3' deletion construct of pCFUTRR up to *Hind*III): pCFUTRR was digested with *Hind*III in the multiple cloning site (MCS) of pHLSI and near the 3' end of the 3'UTR of *Myf-5* thus removing approximately 467bp from the 3' end of the insert to generate pCFH.

pCFA (3' deletion of 693bp to *Ase*I site in the 3'UTR derived from pCFH): pCFH was linearised with *Hind*III followed by partial digestion with *Ase*I which cuts in the 3' UTR and in the ampicillin gene of the vector, allowing the undesired digestion products to be eliminated by selection on ampicillin. Partial digestion products were first examined by gel electorphoresis, suitable fractions were pooled, blunted and religated.

pCFHΔ (internal deletion construct derived from pCFH): pCFH was digested at unique *Avr*II and *Nsi*I sites to remove a 337bp fragment containing 63bp of non-conserved sequence from the 3' end of intron 1, exon 2 (75bp) and the non-conserved 5' half of intron 2 (199bp).

pCFAΔ (internal deletion construct derived from pCFA): pCFA was digested at unique *Avr*II and *Nsi*I sites to remove exon 2 and flanking intron sequences as for pCFHΔ. Due to the absence of suitable restriction sites, an intermediate construct pAH was designed from which the following constructs were derived.

pAH was designed to contain only the UTR sequence between the *Ase*I and *Hind*III sites of pCFH. Sequences 5' of the *Ase*I site in the UTR were removed from pCFH as a *SalI/Ase*I (partial) fragment. *SalI* digestion in the polylinker was followed by partial digestion with *Ase*I with sites in the ampicillin resistance gene of the vector and the 3' UTR of *Myf-5*. The restriction products were blunt end ligated, transformed and selected on ampicilin plates.

pΔAH (intron fragments C, D, F and G added to pAH): PCR amplification between primers CF (kinased) and INIR on pCFHΔ was used to generate a suitable template of the intron sequences, followed by *Pst*I digestion to remove the In1rev primer. The product was then blunted and cloned into the blunted and phosphatased *Pst*I site of pAH to give pΔAH.

pΔCAH (intron fragments D, F and G added to pAH): PCR amplification between primer DF (kinased) and In1rev on pCFHΔ was used to generate a suitable template of the intron sequences, followed by *Pst*I digestion to remove the In1rev primer. The product was then cloned blunt into the *Pst*I site of pAH as above to give pΔCAH.

pΔGAH (intron fragments C, D and F added to pAH). PCR amplification between kinased primers CF and FR on pCFHΔ was used to generate a suitable template of the intron sequences. The product was then cloned blunt into the *Pst*I site of pAH as above to give pΔGAH.

pΔCPH (derived from construct pCFHΔ but lacking fragment C): To obtain a suitable vector, the *Myf-5* intron sequence was removed from pCFHΔ by digestion with *Pst*I followed by religation. The insert was generated by PCR with primers DF (kinased) and IN2R followed by *Pst*I digestion and blunt ending prior to cloning in the blunted phosphatased *Pst*I site of the vector. Clones in the forward orientation were selected by PCR screening of recombinants (primer FR + vector primer: PRI-F).

pΔGPH (derived from construct pCFHΔ but lacking fragment G): The vector was prepared from pCFHΔ as for pΔCPH. The insert was generated by PCR with primers CF and FR (both kinased). The product was blunted and cloned into the blunted phosphatased *Pst*I site of the vector. Recombinants in the forward orientation were selected by PCR screening with In2rev/ PRI-F.

pΔFPH. Fragment D was PCR amplified with primers DF and In1rev, and cloned into the *Eco*RV site of pBS-KS followed by insertion of fragment G (amplified using primers IN2R and kinased GF) into the *Avr*II site (within fragment D). The resulting insert was PCR amplified with primers DF and In1rev, digested with *Pst*I , blunted and cloned into pCFHΔ from which the intron sequences had been removed by *Pst*I digestion. Recombinant clones were identified by PCR screening with primers DF and IN2R.

10. 1. 2. Fugu Myf-5 Reporter Constructs

***Fugu* p12UTR**: A 10kb *Hin*dIII cosmid subclone was digested with *Sty*I to liberate a 2kb fragment containing the entire *Myf-5* gene and the putative 3' UTR. The *Sty*I fragment was cloned into the *Msc*I site of the pHLS reporter plasmid upstream of the hsp68 promoter, driving *LacZ*.

***Fugu* pUTR**: *Fugu* p12UTR was digested upstream and downstream of the insert using *Pvu*II and *Xba*I, to liberate a fragment containing exon 3 and the 3' UTR of *Fugu Myf-5*. The vector backbone was cut at the same time with *Cla*I to allow good separation of the *Pvu*II/*Xba*I fragment for gel purification. The isolated fragment was then cloned into *Msc*I/*Xba*I cut pHLS, upstream of hsp68 driving *LacZ* as before.

***Fugu* p12**: The *Fugu* introns were PCR amplified with primers flanking intron 1 (MC-B) and intron 2 (MC-R) (see Appendix I for sequence), and cloned into the *Msc*I site upstream of the hsp68 minimal promoter driving *LacZ* to give *Fugu*-p12. Only the forward orientation was used to make transgenic animals.

10. 1. 3. Hybrid Myf-5 Reporter Constructs

pZFM: (zebrafish introns on mouse 3' UTR): A PCR fragment extending from the 3' end of exon 1 to the 5' end of exon 3 of zebrafish *Myf-5* containing the two introns with the intervening second exon was amplified by nested PCR: first with primers ZM1 and ZM1R, then with nested primers Tail1 and Tail2 each with *Nsi*I restriction sites at the 5' ends (see Appendix I for primer sequence). The recipient vector was generated by liberating the mouse introns from pCFH (see above) by *Pst*I digestion, followed by gel purification. The PCR product was digested with *Nsi*I and cloned into this *Pst*I cut pCFH derived vector to yield pZFM.

pZFF: (zebrafish introns on *Fugu* 3' UTR): As for pZFM the zebrafish *Myf-5* introns were PCR amplified with primers Tail1 and Tail2 and digested with *Nsi*I. The resulting product was cloned into the *Pst*I site of the remaining polylinker of *Fugu*pUTR.

10. 2. Preparation of DNA for Microinjection

Digestion of the reporter construct (10-50µg) with *Not* I liberated the vector backbone and yielded linear fragments of DNA which were purified from agarose gels using the QiaexII gel purification system (Qiagen). Transgenic mice were produced by microinjecting the DNA (1-5 mg/ml in microinjection buffer) into the pronuclei of fertilised one-cell mouse embryos (Hogan *et al.*, 1994).

10. 3. Transgenic Procedure

Throughout these experiments (CBA x C57B10) F1 mice were used as embryo donors, stud and vasectomized males, pseudopregnant females, and breeding females. Animals

were supplied by the SPF Breeding Unit at the NIMR and were maintained in a constant environment on a 24 hour light-dark cycle (05.00-21.00 hrs light period). All regulated procedures performed were licensed under the Animals Scientific Procedures Act 1986. Twelve female mice (4-week old) were superovulated (Hogan *et al.,* 1994) by injection with 5 IU/0.1ml of pregnant mare's serum (PMS), between 15.00-16.00 hrs. A second intraperitoneal injection of 5 IU/0.1ml of human chorionic gonadotrophin (hCG) was administered 46 hours later. Hormones were obtained from Intervet Laboratories as Folligon (PMS) and Choluron (hCG), resuspended at 50 IU/ml in sterile PBS and stored at -20°C in aliquots. Following hCG injection each female was placed in a cage with a stud-male, then removed the following morning and checked for a vaginal plug. Oviducts from plugged females were removed and the eggs were dissected out into ambient temperature modified M16 instant mouse embryo medium (Specialty Media, Cat. M3-010P-5F). Cumulus cells were removed by transferring the embryos to M16 medium containing 300 µg/ml hyaluronidase for 2-5 minutes. After washing thoroughly in M16 to remove any debris, embryos were transferred to M2 medium (Specialty Media, Cat. MR-015P-5D) microdrop-cultures under mineral oil (Sigma #M8140), equilibrated in an incubator at 37°C with 5 % (v/v) CO_2. DNA constructs were microinjected into either one of the two pronuclei of a one-cell embryo, contained in a droplet of M16 medium under oil. Generally, embryos were microinjected in batches of 20-30, washed through equilibrated M2 and returned to the incubator. The microinjection setup consisted of a Nikon Diaphot-TMD inverted microscope fitted with Nomarski optics combined with Leitz-E micromanipulators and mounted on a vibration-free table.

Embryos which survived the microinjection procedure were reimplanted into the oviducts of pseudopregnant plugged females (Hogan *et al.,* 1994). Recipient mothers were prepared by mating 6-8 week old females in natural oestrus with vasectomised males and were supplied by the SPF Breeding Unit at the NIMR. 12-15 embryos were transferred into each oviduct of a female, anaesthetised by intraperitoneal injection with 350 µl of 2.5 % (w/v) avertin or 300 µl hypnorm/hypnovell. This time point was arbitrarily defined as 0.5 days *post coitum* (d*pc*) or E0.5.

The embryos that developed were harvested at the appropriate time points and stained for ß-galactosidase activity. Transgenic embryos carrying *LacZ* reporter constructs were analysed for ß-galactosidase expression as described (Yee and Rigby, 1993). After staining, embryos were post-fixed in Mirsky's fixative (National Diagnostics) for 60 minutes at room temperature. Embryos were stored in 70% (v/v) ethanol at 4°C.

In order to visually enhance underlying anatomical features, some embryos were counterstained in fuchsin stain solution for 2 hours, then washed: first in PBS for 1 hour, followed by PBS plus 70 % (v/v) ethanol (1:1) for 20 minutes, and finally 70 % (v/v) ethanol (Summerbell, unpublished).

10. 4. Transgene Detection by PCR

Transgenic embryos or animals were identified by PCR amplification of a transgene-specific fragment (450bp) spanning the junction of the *LacZ* gene and SV40 poly (A) signal, as described (Gilthorpe and Rigby in press, 1997). Briefly, tail biopsies (0.5cm) were taken from mice at 3 weeks of age or placental tissue was retained following the dissection of F_0 embryos. The samples were placed in 500 µl of 1 x proteinase K digestion buffer containing proteinase K at a final concentration of 100 µg/ml and incubated at 55°C overnight. Prior to PCR the cellular debris was pelleted by centrifugation and PCRs were performed on 1 µl of the digested sample in a total reaction volume of 20 µl overlaid with mineral oil in a 96-well plate (Thermowell, Costar) and a PTC-100 thermal cycler (MJ Research, Genetics Research Instrumentation). Each reaction contained 1x PCR buffer (1.5mM $MgCl_2$), 1x dNTPs (250µM), 50ng of each primer, 1.5u of *Taq* DNA polymerase (Amplitaq, Perkin Elmer Cetus) and 1µl of digested tissue sample. The amplification reaction conditions were; 94°C for 3 mins, 28 cycles of 94°C for 30 secs, 55°C for 30 secs., 72°C for 30 secs., followed by 72°C for 2 mins. PCR products were resolved by agarose gel electrophoresis.

11. Histological Studies

Embryos were wax-sectioned by W. Hatton of the Sectioning and Histology Service at the NIMR. Blue staining embryos stored in 70% (v/v) ethanol were dehydrated through 85% (v/v), 95% (v/v) and 2 changes of absolute ethanol. Dehydration steps were performed at room temperature for 30-60 minutes according to the size of specimen. After clearing (2 x 30 minute changes of Histoclear, National Diagnostics) at room temperature, embryos were impregnated with paraffin wax (Fibrowax, BDH) by incubating at 60°C with 1:1 (v/v) Histoclear:wax, followed by 3 changes of wax for 30-60mins each. Specimens were transferred to disposable plastic moulds containing fresh molten wax, oriented as required and allowed to cool. Wax blocks were stored desiccated at 4°C. Sections were cut at 6µm thickness and collected on TESPA coated slides (Rentrop *et al.,* 1986), dewaxed with changes of Histoclear (2 x 5 minutes), and serially passed through 3 changes of absolute, 1 x 95% (v/v), 1 x 70% (v/v) ethanol and 1 x distilled water (1 minute each). Sections were then counterstained with eosin (0.5% (w/v) in 25%(v/v) ethanol) for 10-30 seconds. Excess eosin was removed by 2 x 10 second rinses in distilled water and dehydrated through 70% (v/v), 95% (v/v) and 2 x absolute ethanol (15 seconds each). Sections were then cleared through four changes of Histoclear (2-3 mins) and mounted under glass cover slips with D-PX mountant (BDH #360294H).

12. DNA Electrophoretic-Mobility Shift Assay (EMSA)

12. 1. Probe Labelling and Purification

Homogenous double stranded full length probes were generated by labelling of oligonucleotides or PCR products.

i) Oligonucleotide Probes:

Complementary oligonucleotides (2pmol each) were labelled in a 10 µl reaction containing 1x T4 kinase buffer, 6 µl (γ^{32}-P)-ATP (5000Ci/mmol. Redivue, Amersham) and 10 units of T4 Polynucleotide kinase (Boehringer Mannheim) for 30 minutes at 37°C. The reaction was heated to 70°C for 20 minutes to inactivate the enzyme and the oligonucleotides annealed at room temperature for 1 hour to overnight.

ii) PCR Probes:

PCR products were PEG precipitated to remove "cold" nucleotides prior to labelling. 2 pmoles of PCR product were heat denatured and then incubated at 37 °C for 30 minutes with 5 µl of (α-^{32}P)-dCTP (3000 Ci/mmol, Redivue, Amersham) in the presence of 1 µl of Klenow polymerase (Boehringer Mannheim) in 1x supplied buffer, supplemented with 30 µM each dATP, dGTP, dTTP and 20 pM of each of the two PCR primers.

Prior to use, both types of probe were separated from single-stranded DNA and unincorporated radiolabelled nucleotides on a native polyacrylamide gel. Gels (6-12% (w/v) dependent on probe size) were prepared from a stock solution of 30% (w/v) acrylamide/1.034 % (w/v) bisacrylamide in 1x TBE (Easigel, Scotlab) and run at 4-8 W for 3-5 hours. After electrophoresis, the gel was exposed to Kodak XAR film for 10 minutes and the resulting autoradiograph used as a template for the excision of the probe-containing gel slice. The probe was eluted overnight in 200 µl of distilled water at room temperature, transferred to a fresh tube and stored at -20°C. The specific activity of freshly labelled probe was approximately 1000cpm/fmol and the probe was usable for up to 2 weeks. The recovery of the labelled probe was generally 50-75 % by this method, as judged by comparing the cpm emitted by the eluted versus non-eluted probe. Specific competitor DNA for oligonucleotide probes was prepared by annealing 1 µg of each unlabelled complementary oligonucleotide in a 10 µl volume, followed by dilution to an appropriate concentration.

12. 2. Preparation of Protein Extracts

Some of the extracts from cell lines and embryos were kindly provided by Dr Marie Vandromme of this laboratory. The method used was that described by Schöler *et al.* (1986) for the preparation of whole cell micro-extracts except that the sonication step was replaced with three cycles of freeze-thawing. Briefly, approximately 10^7 cells were pelleted and resuspended in 1x micro-extraction buffer and transferred to an ethanol-dry

ice bath until frozen. Thawing was achieved by incubation at 30°C. Samples were microcentrifuged (14 000rpm for 10 minutes at 4°C) and the supernatant aliquotted, assayed for protein concentration and stored at -80°C.
Protein concentration was determined using the Bradford micro-assay method (Bradford, 1976), by comparison against known amounts of standard protein (BSA, Sigma). Samples were diluted to a final volume of 800 µl in distilled water, 200 µl Bio-Rad Protein Assay reagent added and the contents mixed by inversion. The samples were allowed to stand at room temperature for 10 minutes before reading the absorbance at 595nm.

12. 3. EMSA Conditions

DNA binding reactions were set up in a 20 µl volume and contained: 1x binding buffer, 1 µg of sonicated salmon sperm DNA, specific competitor (where indicated) and 10-20µg of whole cell extract. After incubation on ice for 15 minutes, 1 µl (10fmol) of probe was added and the incubation continued for an additional 10 minutes at 30°C. Samples were kept on ice before being loaded onto a 5% (w/v) polyacrylamide gel. Electrophoresis was performed using BRL vertical gel electrophoresis apparatus (fitted with a recirculating pump) in 0.5 x TBE (200 volts for 2 hours at 4°C). The gel was dried under vacuum at 80°C and exposed to Kodak XAR or Biomax film with an intensifying screen at -80°C.

13. *In Situ* Hybridisation of Zebrafish Embryos

13. 1. Riboprobe Synthesis

10 µg of CsCl purified plasmid DNA were linearised by digestion with the appropriate restriction enzymes in a 100 µl reaction for 2 hours, typically using 5-10 units of enzyme per µg of DNA. The DNA was recovered by phenol/chloroform extraction followed by ethanol precipitation and resuspended in 20 µl of dH_2O. The transcription reaction was carried out according to the RNA-labelling-kit (Boehringer Mannheim) using 1 µg (2 µl) of the linearised DNA according to the manufacturer's recommended conditions. After incubation at 37°C for 2 hours, the DNA template was removed by addition of 40units of RQ1 DNase (Promega) and incubation was continued for 20 minutes. The reaction was stopped by addition of EDTA (pH 8.0) to a final concentration of 20 mM and the RNA precipitated with 0.1 volumes of 4 M LiCl, and 75 µl of prechilled ethanol for 20 minutes on dry ice. The RNA was recovered by centrifugation (13000 rpm for 10 minutes), the pellet rinsed in 70 % (v/v) ethanol and resuspended in 50 µl of RNase free dH_2O. The integrity and quantity of the RNA

preparation was checked by brief agarose gel electrophoresis and aliquots were stored at -70°C.

13. 2. Preparation of Embryos

Fish embryos were obtained from the animal facility at NIMR and allowed to grow at 29°C to the desired developmental stage. The embryos were placed in a 50 ml tube (FALCON) and separated from debris and unfertilised or dead eggs by rinsing several times in dH_2O. Following each rinse, the embryos were allowed to settle briefly and the floating debris was decanted or pipetted off. The cleaned up embryos were fixed in 4 % (w/v) ice-cold PFA (made up in PBS) and could be stored at -20°C for up to several weeks at this stage. Before use, the required number of embryos was placed in PBT, manually dechorionated under the dissecting microscope and equilibrated in methanol for at least 20 minutes at -20°C. The embryos were then rehydrated by 3 minute washes in a series of 75 %, 50 %, 25 % (v/v) methanol in PBT before being taken up in PBT. (Proteinase K treatment used for mouse embryos was omitted). To reduce background staining, embryos were bleached by incubation in 6% w/v hydrogen peroxide (made up in PBT) for 1 hour at room temperature, washed three times for 5 minutes in PBT and refixed in 4% (w/v) PFA for 20 minutes. Following the fixation, embryos were washed twice for 5 minutes in PBT, then prehybridised in hybridisation solution for at least 1 hour at 60-70°C. At this stage embryos can be stored conveniently at -20°C for up to several weeks.

13. 3. Hybridisation

Zebrafish embryos were placed in netwells (Costar 3477) fitting 12 well dishes (Costar 3512) to minimise handling and incubated in hybridisation buffer at 65°C - 70°C depending on probe length for at least 1hour. Prior to use 1 µg of riboprobe were incubated for 5 minutes at 80°C in hybridisation buffer to remove secondary structure. The netwell containing the embryos was then transferred to new hybridisation buffer and 150 ng of riboprobe were added to the hybridisation solution. Hybridisation was continued at the prehybridisation temperature overnight in a sealed plastic box humidified by placing a tissue soaked in 50 % (v/v) formamide in the box.

13. 4. Posthybridisation Treatment

Following overnight hybridisation, the embryos were washed for 20 min each in prewarmed hybridisation buffer (65°C) supplemented with RNaseA (100 µg/ml) then TBST at room temperature for one hour including 0.5mg/ml phosphatase inhibitor (Levamisole). Subsequently, embryos were preblocked in MABT 1M maleic acid pH 7.5, 0.25 M NaCl, 1 % (v/v) Tween20, 10 % (v/v) heat treated sheep serum (Sigma) for 3-4 hours at room temperature. Blocking serum was removed and a 1:200 dilution of

preabsorbed sheep anti-digoxygenin- alkaline-phosphatase conjugated Fab fragments (Boehringer Mannheim) in MABT was added and incubated rocking overnight at 4°C. The following morning, the embryos were washed for 5 minutes in NTMT at room temperature and then transferred to siliconised glass histology dishes. Residual buffer was removed, and the staining reaction started by addition of 1ml of fresh substrate NTMT containing 1.5 mg/ml NBT (Boehringer Mannheim) and 3.25 mg/ml BCIP (Boehringer Mannheim). Progress of the reaction was monitored at 30 minute intervals; at the desired intensity the substrate was removed by washing in NTMT for 5 minutes and terminated by incubation of embryos in 1xPBT pH 5.5 overnight at 4°C. Following incubation, the embryos were post fixed in Mirsky's fixative and stored at 4°C.

Chapter 3

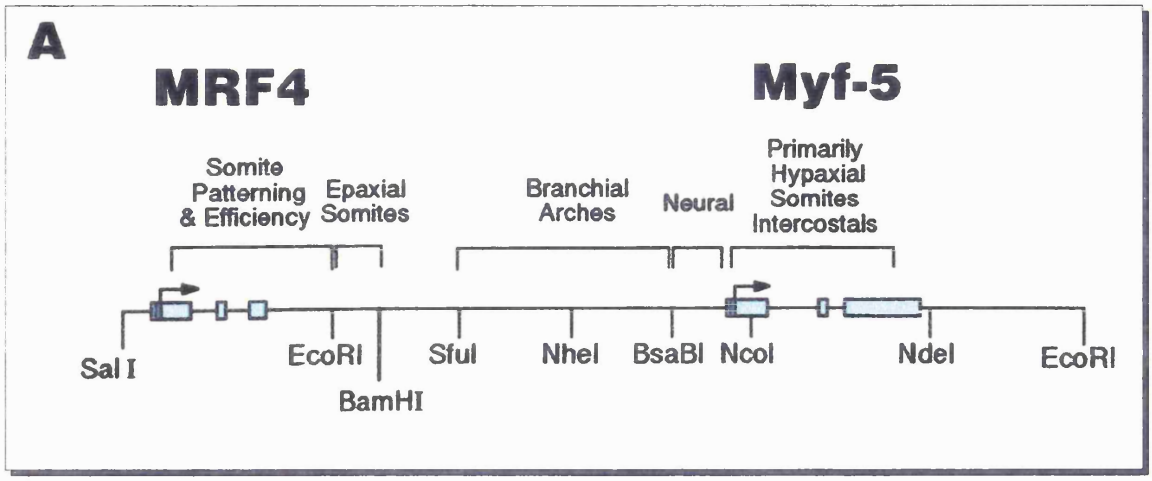
Comparative Sequence Analysis of the Vertebrate *Myf-5 /MRF4* Loci

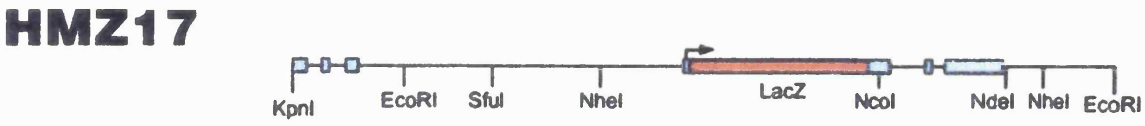
1. Introduction

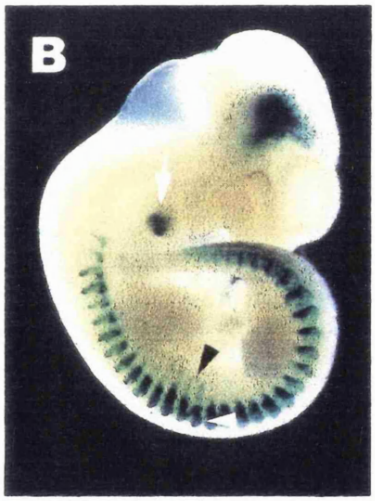
The myogenic basic-helix-loop-helix transcription factors, *Myf-5*, *myogenin*, *MRF4* and *MyoD*, share the ability to activate the myogenic program in certain cell types in culture (Davis *et al.*, 1987; Sassoon *et al.*, 1989; Weintraub *et al.*, 1989; Wright *et al.*, 1989; Choi *et al.*, 1990). *In vivo*, their temporal expression together with results obtained from null mutations in each of the factors and combinations of these, suggest that skeletal myogenesis is initiated through the sequential activation of the myogenic factors in a hierarchical manner (Thayer *et al.*, 1989; Braun *et al.*, 1990; Edmondson *et al.*, 1992; Naidu *et al.*, 1995). In the mouse, the first myogenic factor to be expressed is *Myf-5* at E8 (Ott *et al.*, 1991). To ultimately understand how myogenesis is initiated, our laboratory is studying the control of *Myf-5* with a view of identifying the regulatory factors that activate its transcription. Our studies indicate that in contrast to *myogenin*, whose regulatory elements are contained in 133bp of promoter sequence (Yee and Rigby, 1993), the transcriptional regulation of the *Myf-5* gene is complex and involves a number of dispersed elements in the 8kb region upstream of *Myf-5*, possibly involving the adjacent *MRF4* gene as well as the *Myf-5* introns and the 3' UTR of the *Myf-5* gene. A schematic representation summarising the regulatory regions identified so far is shown in Figure 8A. A *LacZ* reporter gene construct (HMZ17), containing the entire region from the *MRF4* gene to the 3' UTR of *Myf-5* reproduces much of the endogenous *Myf-5* expression pattern, with conspicuous absence of expression in the limbs (see Fig. 8B, Summerbell *et al.*, unpublished). Progressive deletion of part of the intergenic region led to successive loss of some of the expression domains, thus delineating the positions of the regulatory elements required for reporter

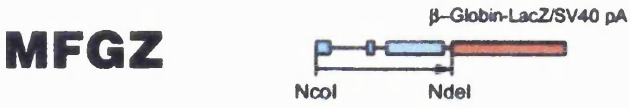
Figure 8: Regulatory Elements in the *MRF4/Myf-5* region.

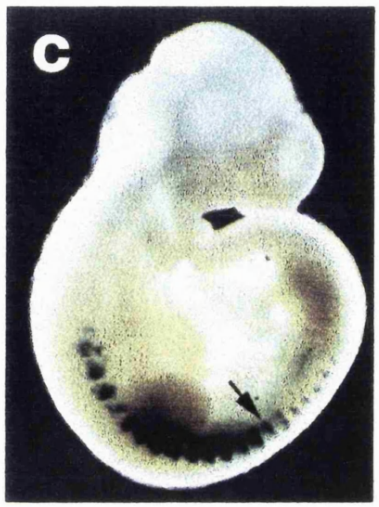
Fig. 8A: Schematic representation of the mouse genomic region containing the *MRF4* and *Myf-5* genes. The exons are depicted as light blue boxes. Regulatory regions that have been shown in transgenics to activate transcription in distinct anatomical domains of the mouse embryo are marked by brackets and annotated. Note that regulatory elements controlling ventral somite expression are separated by more than 7kb of DNA from the elements controlling dorsal somitic expression.
Fig. 8B: Expression pattern obtained at E10.5 with **HMZ17** containing the entire region from the *MRF4* gene to the 3' UTR of *Myf-5* . The transgene in this construct is under the control of the endogenous *Myf-5* promoter and reproduces much of the endogenous *Myf-5* expression pattern, except for expression in the limbs (Summerbell *et al.*, unpublished). Expression in the interlimb somites includes the dorsal somite margin (white arrowhead) and ventral somitic bud (black arrowhead) and branchial arch expression (white arrow).
Fig. 8C: Expression pattern obtained at E10.5 with MFGZ containing the *Myf-5* gene with its 3' untranslated region upstream of a heterologous *LacZ* reporter. This construct reproduces expression in the ventral (arrow) but not the dorsal somite compartment (compare with HMZ17 in Fig. 8B).


A
MRF4
Myf-5
Somite Patterning & Efficiency
Epaxial Somites
Branchial Arches
Neural
Primarily Hypaxial Somites Intercostals
Sal I
EcoRI
BamHI
SfuI
NheI
BsaBI
NcoI
NdeI
EcoRI


HMZ17
KpnI
EcoRI
SfuI
NheI
LacZ
NcoI
NdeI
NheI
EcoRI


B


MFGZ
β-Globin-LacZ/SV40 pA
NcoI
NdeI


C

gene expression in these anatomical domains. This analysis showed that the dorsal and ventral somite compartments are controlled by distinct regulatory regions. Elements involved in regulating the hypaxial somite compartment mapped to the *Myf-5* gene itself and the expression pattern produced by these elements is shown in Fig. 8C. Note the absence of dorsal somitic expression. Several discrete regulatory sites have also been mapped in the intergenic region which activate reporter gene expression in the epaxial domain of the somites and the branchial arches (data not shown).
Because several studies have previously been successful in identifying regulatory elements controlling *Hox* gene expression by sequence comparisons between the chick and mouse genes and their homologues in teleost fish (*Fugu* and zebrafish) (Marshall *et al.*, 1994; Aparicio *et al.*, 1995; Beckers *et al.*, 1996) we pursued a similar strategy to identify and characterise regulatory elements involved in the regulation of the *Myf-5* gene using sequence comparisons between the mouse and *Fugu*.. The small size of the *Fugu* genome would suggest that the intergenic region should be significantly reduced in size, thus allowing regulatory elements to be identified rapidly.

2. Isolation and Characterisation of *Fugu MRF4* and *Myf-5*

2. 1. A *Fugu* Cosmid Contains Both *MRF4* and *Myf-5*

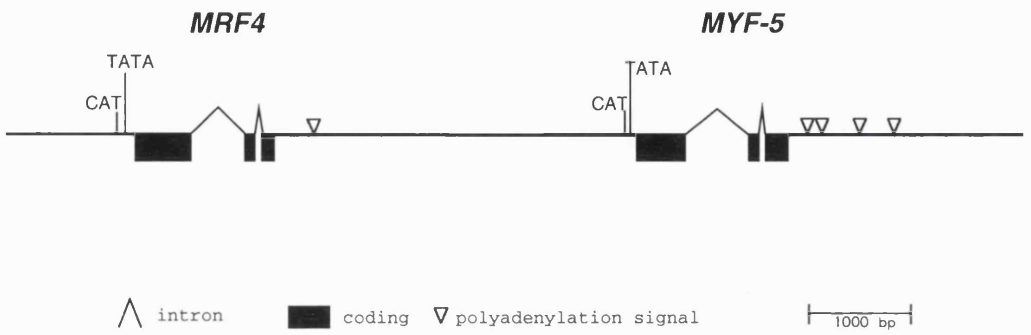
A *Fugu rubripes* genomic cosmid library consisting of an ordered array of 38016 clones representing approximately four-fold coverage, was screened by hybridisation with a mouse *Myf-5* cDNA probe and a genomic *MRF4* probe (for details see Chapter 2 section 7.1.). Briefly, screening with both probes produced a total of 69 positive cosmids. 26 cosmids hybridised to the *Myf-5* cDNA probe and 43 cosmids hybridised to the *MRF4* genomic probe, no double positives were identified at this stage. In a second round of screening, Southern blots of the positive clones were probed with a PCR product spanning the conserved *MRF4* helix-loop-helix (HLH) domain amplified with primers HLH-For and HLH-Rev (see Appendix I). This domain is highly conserved in both *MRF4* and *Myf-5,* and to a lesser degree in the remaining myogenic factors and other members of the HLH-family of transcription factors. Only four cosmids hybridised to the HLH-probe. In order to confirm these results further, dot blot filters containing DNA of all 69 cosmids from the first round of library screening were re-screened with all three probes, the *Myf-5* cDNA probe, the *MRF4* genomic probe, and the HLH specific PCR product. In this screen two of the cosmids were positive for all three probes. To determine whether the cosmids contained the *MRF4* and *Myf-5*

Figure 9A: The genomic organisation of the *Fugu MRF4* and *Myf-5* genes.

A gridded *Fugu* cosmid library was screened with mouse probes for *MRF4* and *Myf-5*. A positive cosmid was isolated and a 10kb *Hin*dIII fragment containing both the *Fugu MRF4* and *Myf-5* genes was sequenced. The *Fugu* genes have the same organisation as the *MRF4* and *Myf-5* genes in other vertebrates. The first exon contains the bHLH domain and is highly conserved. Both *Fugu* genes are not significantly reduced in size, although the intergenic region is smaller than in other vertebrates. Exons are indicated as filled boxes, introns as brackets between exons. Putative CAT and TATA boxes as predicted by the neural network program GRAIL; polyadenylation signals are indicated as triangles.

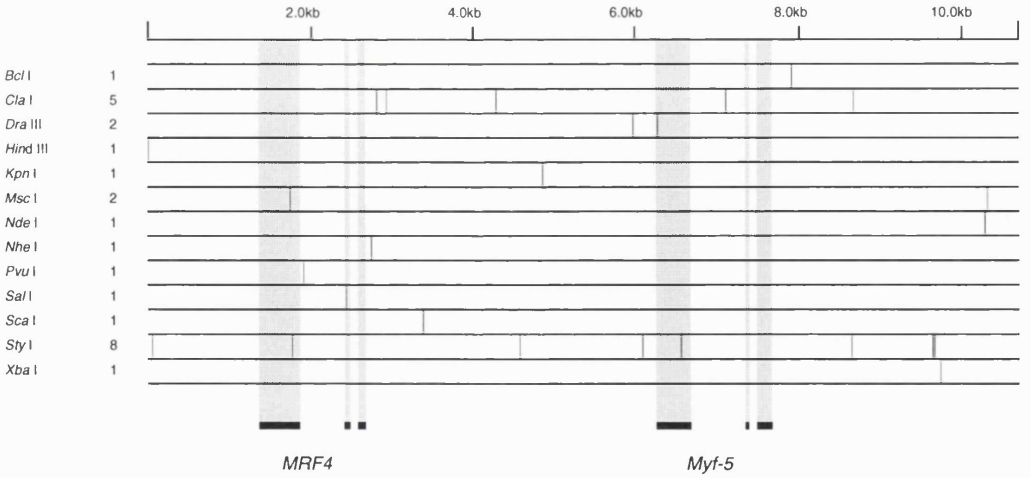
Figure 9B: Selected restriction sites in the *Hin*dIII cosmid subclone containing the *Fugu MRF4* and *Myf-5* genes. Restriction sites are based on sequence information. The complete genomic sequence is presented in Appendix II.

Figure 9A:



The genomic organisation of the *Fugu MRF4* and *Myf-5* genes and regulatory sites as predicted by GRAIL

Figure 9B:



Selected Restriction sites in the 10kb *Hind*III cosmid subclone

homologues, one of the isolated cosmids was sonicated and the products subcloned into plasmids. The inserts of approximately 800 subclones were PCR-amplified and screened by Southern hybridisation to the HLH probe. DNA sequencing of the hybridisation positive subclones revealed a high degree of homology in the SWISSPROT database with either *Myf-5* or *MRF4* entries, suggesting that the cosmid contained both the *Fugu MRF4* and *Myf-5* homologues. Additional subclones of this cosmid were sequenced and identified two additional potential genes, with homology to the *C. elegans* transposable element TCB2 transposase and the human oxysterol binding protein, although these were not characterised further.

2. 2. The Genomic Structures of *Fugu MRF4* and *Myf-5*

To isolate a smaller, more manageable subclone of the cosmid, containing both the *Fugu MRF4* and *Myf-5* genes, oligonucleotides specific for *Fugu MRF4* and *Myf-5* were designed on the basis of the partial sequence obtained from the shotgun phase. The oligonucleotides were endlabelled and used as Southern hybridisation probes on 6-cutter digests of the cosmid DNA (see Chapter 2, section 5.3.). Both oligonucleotide probes hybridised to a 10kb *Hin*dIII fragment indicating that the *Fugu MRF4* and *Myf-5* genes are closely linked. The identified 10kb *Hin*dIII fragment was gel purified, cloned and sequenced using a combination of the shotgun and primer walking approaches (see Chapter 2, Section 9). A total of approximately 415 individual sequence reads was assembled into 38 contigs. Gaps between the contigs were closed by 48 primer walks, until all contigs merged. The final contig is represented schematically in Appendix III and a summary of the *Fugu* region with a restriction map is shown in Fig. 9A and 9B. The complete genomic nucleotide sequence is also shown in Appendix II. The length of the final contig was 10676 bp, with 7.4 fold average coverage. The positions of the *MRF4* and *Myf-5* genes and their general genomic structure in the 10kb contig was initially determined by performing blastn searches (Altschul *et al.*, 1990) with subfragments of the contig using the NCBI database. Matches of the *Fugu* sequence obtained with entries in the database were aligned along the contig to position the genes. The exact exon-intron structure was subsequently determined by sequencing *Fugu Myf-5* and *MRF4* muscle cDNA amplified with hemi-nested primer pairs (see Appendix I for oligonucleotides). The organisation of the *Fugu* genes is very similar to that of the mouse *MRF4* and *Myf-5* genes and the other myogenic regulatory genes (Fig. 10A). Both have 3 exons separated by two introns. The first, and largest, exon in both genes is the most highly conserved, owing to the characteristic basic-helix loop-helix (bHLH) domain shared by all of the myogenic factors. The two helix domains mediate dimerisation of the MRF proteins with E2 proteins while the basic domain facilitates DNA binding to E-box motifs in target genes. Exon 2 as in the mouse genes is tiny and

Figure 10:

Schematic comparison of the genomic region containing the genes for *Myf-5* and *MRF4* of mouse and *Fugu*. The intergenic region in *Fugu* is less than half the size of the mouse. The overall length of the *Fugu* genes is not significantly reduced. Exons are represented as grey boxes, and introns as solid lines. The 3' untranslated region up to the first putative polyadenylation signal is represented by a blue box. Corresponding regions of the *Myf-5* and *MRF4* genes in the two species are shadowed in light grey. The size difference of the intergenic region is indicated by the arrows.

Table 1: Comparison of the size of the mouse, human and *Fugu* exons and introns of *MRF4* and *Myf-5*. In the *Fugu* the overall length of the *MRF4* and *Myf-5* genes is slightly smaller than in the mouse. This is due to a fivefold reduction of intron 2 in both genes. In *Fugu* intron 1 of *Myf-5* is not significantly reduced and intron 1 of *Fugu MRF4* is even 85% larger than the mouse intron. For both genes, the length of the coding region is very similar in *Fugu* and mouse.

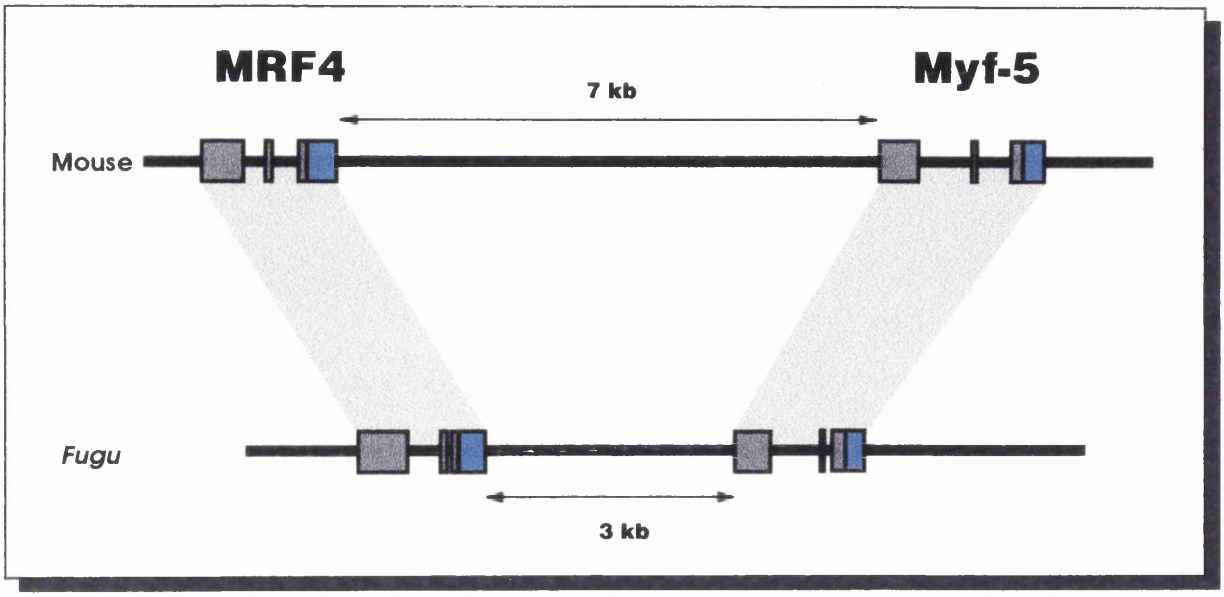


Figure 10: The genomic organisation of the region containing the MRF4 and Myf-5 genes in *Fugu* and Mouse.

MRF4	Mouse	Human	*Fugu*
Exon1	543	522	521
Exon2	90	93	90
Exon3	114	123	108
Intron1	283	n/a	524
Intron2	320	n/a	73

Myf-5	Mouse	Human	*Fugu*
Exon1	500	500	452
Exon2	75	76	72
Exon3	188	212	196
Intron1	708	793	634
Intron2	435	427	85

Table 1: Exon and Intron sizes of the MRF4 and Myf-5 genes in Mouse, Human and *Fugu*

relatively poorly conserved and the third exon contains a serine rich region that is conserved among the other myogenic regulatory factors across vertebrate species.
The translation start codon of exon 1 was assigned by comparison with the corresponding mouse exon and the prediction made by the neural network program GRAIL (Überbacher *et al.*, 1991). GRAIL also predicted a CAT and TATA box in the promoter at the expected distance upstream of the initiation codon (see Fig. 9A and Appendix II for nucleotide sequence). The putative termination codon of exon 3 was assigned in a similar fashion. GRAIL also predicted the stop codon and potential polyadenylation sites in the 3' UTR.

2. 3. Interspecies Comparison at the DNA Level

To facilitate sequence comparisons with the mouse, a cosmid containing the mouse *MRF4* and *Myf-5* genes that had previously been isolated by Dr. S. P. Yee of our laboratory was sub-cloned and sequenced by Dr. D. Summerbell and C. Halai from our laboratory (see Appendix VI for the genomic sequence of mouse *Myf-5*).
In *Fugu*, mouse and human the *MRF4* and *Myf-5* genes form a syntenic linkage group. *MRF4* is located upstream of the *Myf-5* gene and both genes are arranged in the same transcriptional orientation. The start codons of the *Fugu* genes are separated by less than 5 kb compared with nearly 9kb in the mouse. A comparison of the genomic structure and size differences of the mouse and *Fugu* genes is illustrated in Fig. 10. As expected, the overall length of the *Fugu* genes is smaller than that of the corresponding mouse homologues. Interestingly the two genes are not equally compacted; while *Fugu MRF4* is only 3% smaller than the mouse gene, the *Fugu Myf-5* gene is 25% smaller than in the mouse. As the coding regions are very similar in length, the reduced overall size of the genes is due to the remarkably reduced length of the second intron in both *Fugu* genes; it is only one fifth of the size of the corresponding mouse intron. Table 1 shows a summary of the exon and intron sizes for *Fugu*, mouse, human. Despite the compact nature of the *Fugu* genome, the first intron of *MRF4* is almost 85% larger than the equivalent mouse intron, and intron 1 of *Fugu Myf-5* is only slightly smaller than in the mouse.
The same splice acceptor consensus sites are used in the *Fugu* and mouse *MRF4* and *Myf-5* genes. In both species the splice acceptor is CAG except for intron 2 of *Myf-5* which has a TAG consensus.

2. 4. Interspecies Comparison at the Protein Level

The overall sizes of the conceptual translation products of the *Fugu Myf-5* and *MRF4* genes are very similar compared to the mouse homologues. The *Fugu MRF4* gene encodes a 241 amino acid protein compared with 242 amino acids in the mouse. The *Fugu Myf-5* gene encodes a 240 amino acid protein compared with 255 amino acids in

Figure 11:

Clustal alignment of the conceptual translation products of the *Fugu MRF4* (**Fig. 11A**) and *Myf-5* genes (**Fig. 11C**) with homologues of other species as indicated. Amino acid residues that are conserved in at least two species are shown on a coloured background. Residues of the bHLH domain are shown on red (basic domain) or green (helix I and II) background. The positions of the exon and intron boundaries are indicated by triangles. Significant conservation is seen in the bHLH domain required for dimerisation and DNA binding and in a serine rich region in exon 3. The basic domain is almost completely conserved between all species. Upstream of the basic domain lies a C/H rich domain associated with the ability of *Myf-5* and *MyoD* to remodel chromatin structure ([Gerber *et al.*, 1997]). A phylogenetic tree of sequence convergence is shown in **Fig. 11B** for MRF4 and **Fig. 11D** for Myf-5. The *Fugu* sequence in each case represents the most ancient ancestor. The branching order for MRF4 follows the known evolutionary branching pattern, fish, amphibians, birds, mammals, interestingly the branching order for Myf-5 is different despite the close linkage between *MRF4* and *Myf-5*.

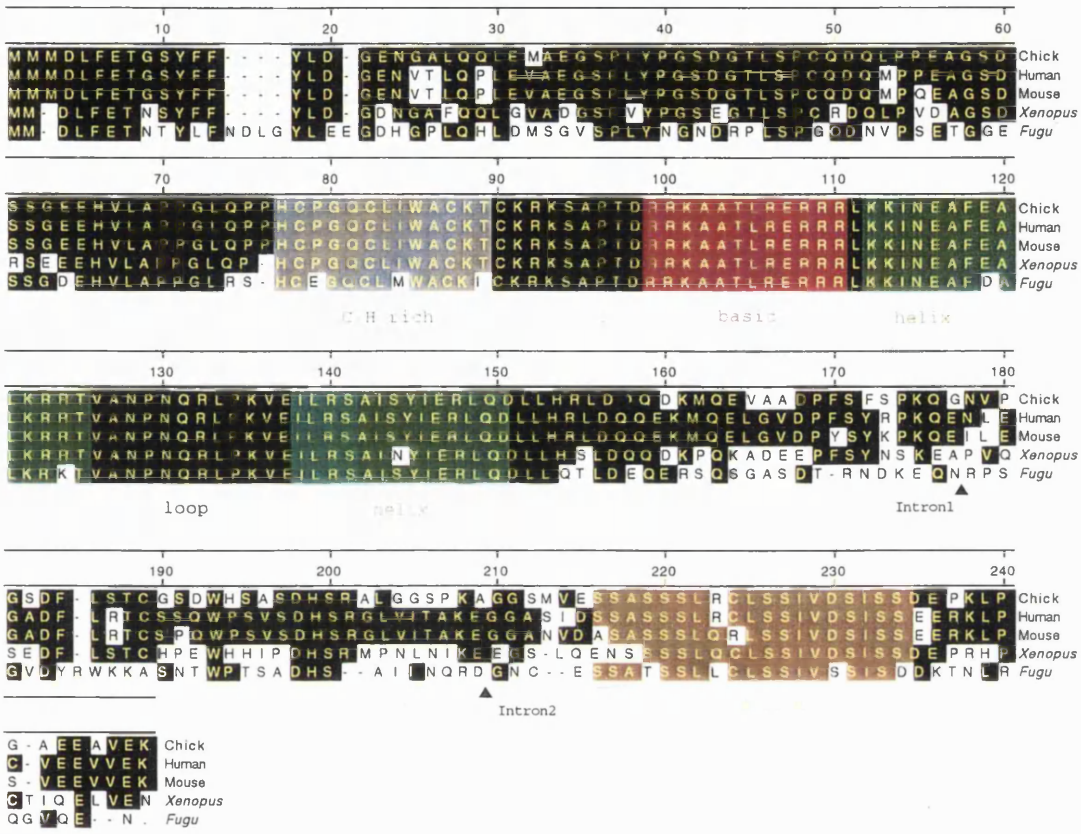
References:

MRF-4: Mouse: Miner, J.H. & Wold, B. (1990). Medline Identifier (MI): 90138943. Human: Braun,T. *et al.*, (1990). MI: 90183982. Chicken: Fujisawa-Sehara *et al.*, (1992). MI: 92250560. Xenopus: Jennings, C.G. (1992). MI: 92164857

MYF-5: Human: Braun T. *et al.*, (1989a). MI: 89251600. Bovine: Barth J.L. *et al.*, (1993). MI: 93273229. Mouse: Buonanno,A. *et al.*, (1992). MI: 92158662. Chicken: Saitoh, O. *et al.*, (1993). MI: 93281401 Xenopus: Hopwood, N.D. *et al.*, (1991). MI: 91372152.

Clustal Alignment: MRF4

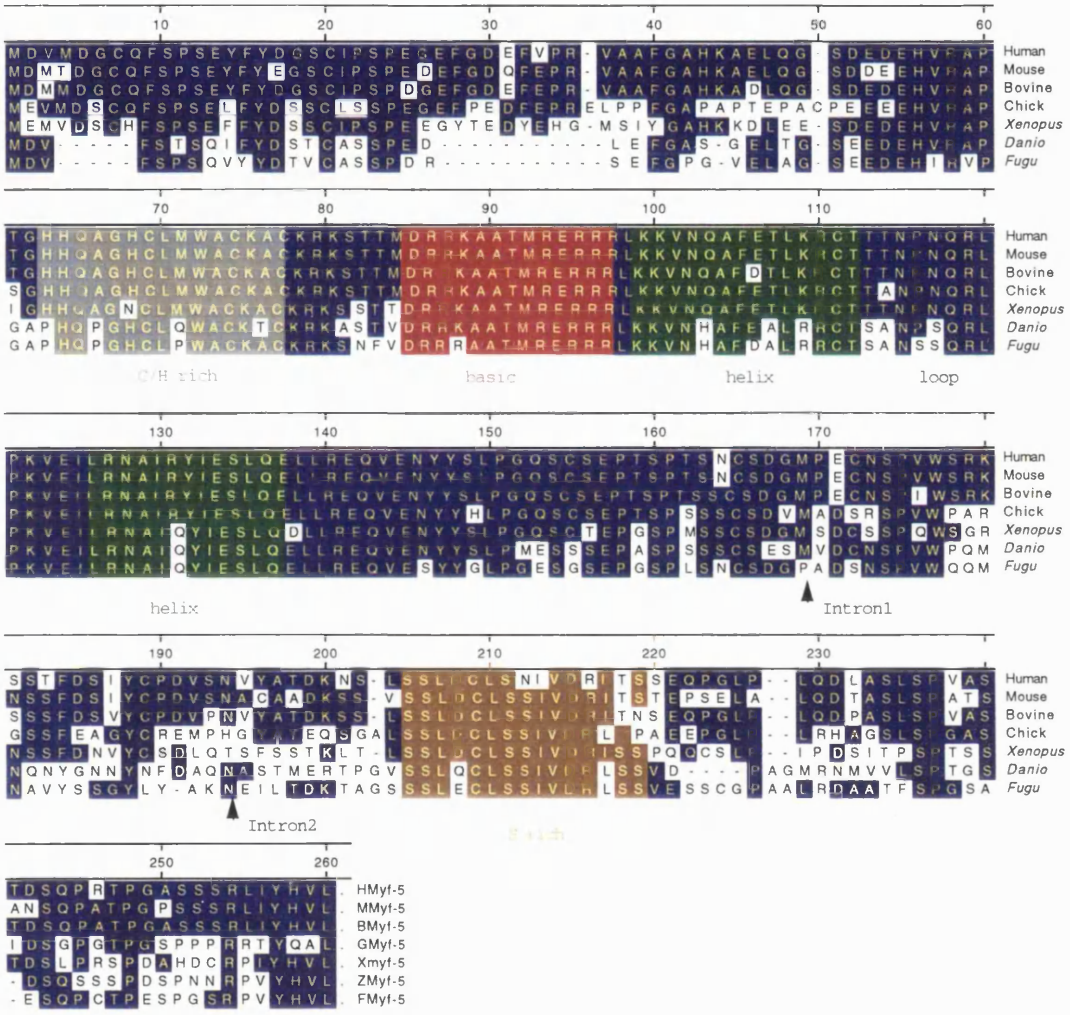
11A



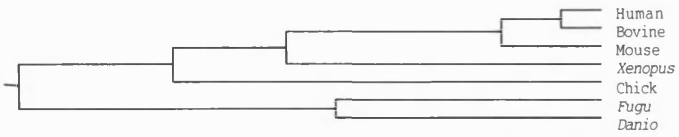
11B

Clustal Alignment: Myf-5

11C



11D



mouse. The overall pairwise identity between the mouse and *Fugu* genes at the protein level reaches 56% for *Myf-5* and 60% for *MRF4*. Figure 11 shows an alignment of the *Fugu* MRF4 and Myf-5 proteins with the mouse, human, chick, bovine, zebrafish and *Xenopus* homologues and illustrates the regions of sequence divergence and conservation amongst these species (see figure legend for references).

The protein homology is particularly strong over the basic-helix-loop-helix domain where it rises to above 80% (Fig. 11A and C). Nearly all of the amino acid changes in the bHLH domain are conservative substitutions. The two amphipathic helix domains of both *Fugu MRF4* and *Myf-5* are well conserved but more strikingly the basic domain of *Fugu MRF4* is completely identical amongst species as diverse as mouse, chick, *Xenopus* and *Fugu*. The basic domain of *Fugu Myf-5* is also completely conserved except for a single conservative amino acid substitution (Fig. 11C). Upstream of the basic domain lies a C/H rich domain that is found in mammals, birds, amphibia and fish (this study) and is thought to be associated with the ability of *Myf-5* and *MyoD* to remodel chromatin structure (Gerber *et al.*, 1997). Further towards the 5' end of the gene, sequence conservation gradually breaks down in all of the species, even those as closely related as human and mouse, while in *Fugu* and zebrafish *Myf-5* many of the N-terminal amino acids are noticeably absent (Fig. 11C). An additional serine-rich region in exon 3 is shared between all four myogenic factors and is highly conserved between mammals and amphibia.

To examine the ancestral relationship of the *Myf-5* and *MRF4* genes, a phylogenetic tree was constructed using the DNA Star software package (DNA Star Inc.). It is interesting that, while *MRF4* follows the known evolutionary branching pattern, with fish diverging before amphibia, followed by birds and finally mammals (Fig. 11B), the *Myf-5* gene does not adhere to this order, despite the close physical linkage. In the case of *Myf-5*, the fish gene diverged first but then birds follow before amphibia, and mammals (Fig. 11D). In mammals the temporal order of expression of the myogenic factors is *Myf-5* followed by *myogenin*, followed by *MRF4* and finally *MyoD*. Interestingly, in birds the *MyoD* homologue is expressed first (Pownall and Emerson, 1992) and substitutes for most of the functions associated with *Myf-5* in vertebrates. Its similarity with *MyoD* may thus be responsible for the altered position in the evolutionary branching order.

2. 5. CpG Islands and Repeat Elements

In mammals the CpG dinucleotide occurs at only ~20% of the frequency that would statistically be expected based on its GC content. However, this under-representation is lifted in CpG rich islands often associated with house keeping genes (Gardiner-Garden and Frommer, 1987). The 14172bp mouse fragment contains CpG dinucleotides at a third of the statistically expected frequency expected for its GC content of 42%

(expected: 625; CpGs observed: 192 CpG dinucleotides). The equivalent 10681bp fragment of *Fugu* has a GC content of 48% matching closely the estimate of 44% by Brenner *et al.*(1993) for the whole genome and supports the observation that compared with mammalian genomes, the *Fugu* genome has a remarkably uniform GC distribution (The *Fugu* Landmark Mapping Project, http://fugu.hgmp.mrc.ac.uk).
The *Fugu* region contains CpG dinucleotides at 70% of the expected frequency, (expected: 635, observed: 433) more than twice as frequently as in the mouse region and matching other estimates for the whole *Fugu* genome of 60% (Elgar *et al.*, 1996). For comparison, the GC dinucleotides in the mouse and *Fugu* fragments both occur at 93% of the statistically expected frequency. It has been shown that methylation in *Fugu* occurs at the same CpG dinucleotide as in mammals, but the 'tiny' fraction is absent in *Fugu* and other fish species (Saccone *et al.*, 1993; Aissani and Bernardi, 1991; Elgar, unpublished). Thus although the CpG dinucleotide is less suppressed in *Fugu* compared with mammals, there appear to be fewer restriction sites, suggesting that the CpG dinucleotides in *Fugu* might not be as heavily methylated compared with mammals (Elgar, 1996). This is supported also by the observation that unlike in human where CpG islands are frequently associated with the 5' ends of genes, in *Fugu* CpGs appear to be more randomly distributed and are absent from many house keeping genes ([Elgar, 1996]).
In the mouse sequence a number of repetitive elements with between 10 and 37 repeat units were found in both introns of *MRF4* and the intergenic region. No repeats were found in the *Fugu* sequence, consistent with the observation that repetitive DNA is highly clustered and constitutes less than 10% of the *Fugu* genome (Brenner *et al.*, 1993).

3. Pairwise DNA Sequence Analysis Between *Fugu* and Mouse

Pairwise sequence alignment between the mouse and *Fugu* regions containing the *MRF4* and *Myf-5* genes was carried out to find conserved regulatory elements. The results of this alignment are shown in a dot plot diagram in Figure 12A. Matches of at least 75% identity over a 20bp window appear as diagonal lines in the diagram. The bHLH region of exon 1 of both genes appears on the diagonal, indicating some degree of conservation between *Fugu* and mouse. Interestingly, the bHLH regions of both genes not only share significant sequence homology between *Fugu* and mouse (Fig. 12A, red lines), but also between each other (Fig. 12A, green lines). The observed cross similarity is the result of functional conservation amongst the myogenic bHLH family members. However, no significant homology was found in the introns and in the

untranslated regions or in the entire intergenic region between *MRF4* and *Myf-5*. The poor conservation of the large intergenic region and the introns was unexpected in view of the fact that the *Fugu MRF4* and *Myf-5* genes have been maintained as a linked pair through millions of years of evolution and the genomic organisation in *Fugu* and other species was found to be very similar. It can not be excluded that conserved protein binding sites exist in *Fugu*, that are too divergent to be readily recognisable in sequence comparisons or that different regulatory mechanisms operate between mammals and fish. To test these possibilities, alignments between more closely related mammalian species were performed.

4. Pairwise Comparisons Between Mouse, Human and Bovine.

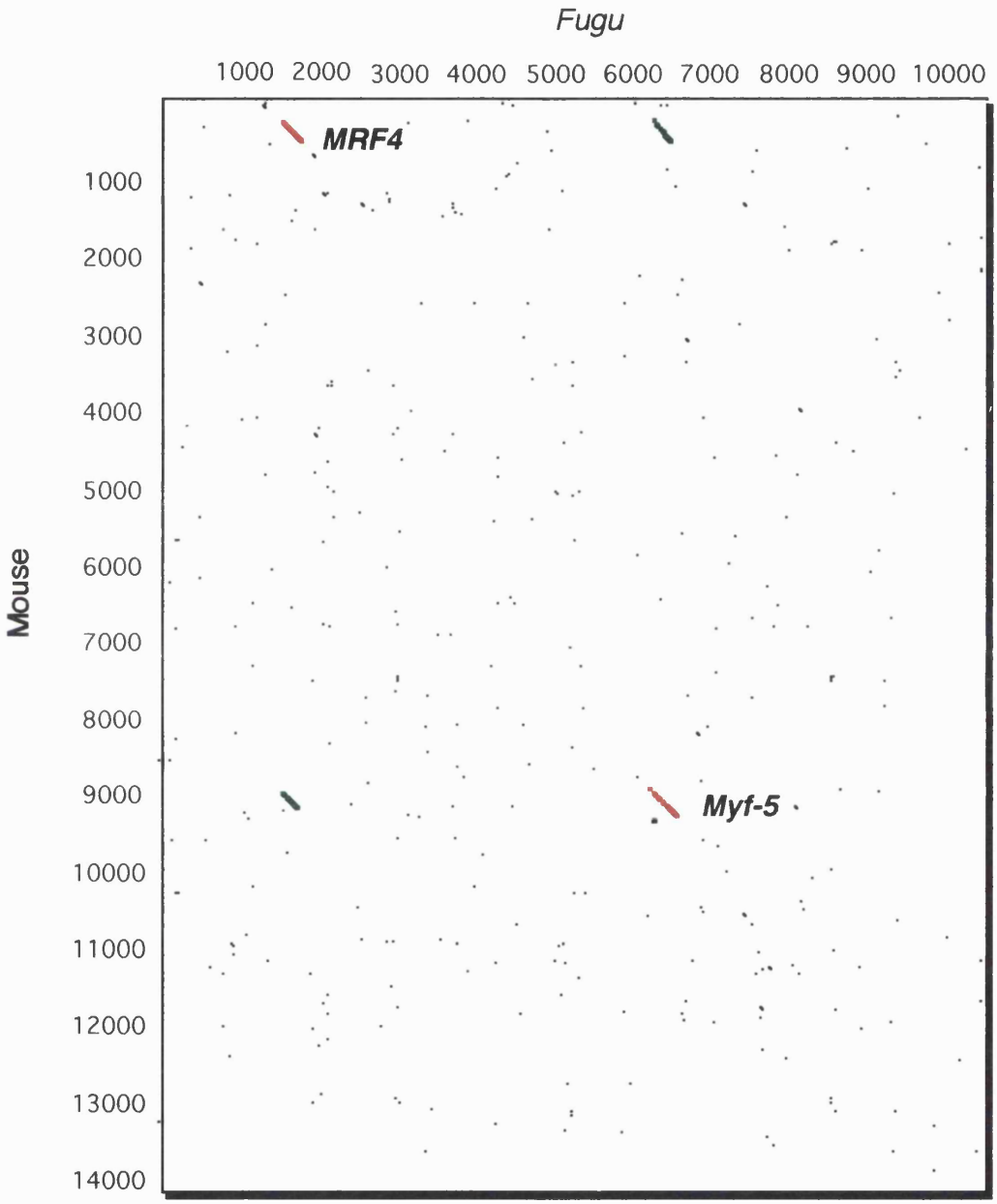
Due to the lack of sequence conservation between the *Fugu* and mouse region, I turned my attention to the *Myf-5* gene itself. Data from our laboratory had previously shown that the *Myf-5* gene contained regulatory elements involved in controlling the ventral expression domain of *Myf-5* in the somites (see Fig. 8C). Sequence information including the introns and the 3' UTR was already available for bovine and mouse *Myf-5*. For additional comparisons the intron sequence of the human *Myf-5* gene was determined: Briefly, the nucleotide sequence of the human *Myf-5* introns was determined by sequencing genomic PCR products amplified from human genomic DNA with pairs of primers derived from exon 1 and exon 3 of mouse *Myf-5* (In1for and In2rev see Appendix I), spanning the intervening introns and exon 2. The intron boundaries of the human gene were determined by comparison with the published cDNA sequence (Braun *et al.*, 1989b). The size of the introns and exons is given in table 1 and the complete genomic sequence of human *Myf-5* is presented in Appendix IV.
As expected, comparisons between *Fugu* and either bovine or human *Myf-5* failed to show conserved sequence blocks outside of the coding region (not shown). Therefore two-way sequence comparisons were carried out between the human *Myf-5* gene and the mouse and bovine homologues respectively. The dot plot analysis is shown in Figure 12B and C and revealed substantial sequence homologies, both in the coding and noncoding regions. In agreement with the relationships established for the protein sequence in the phylogenetic tree in Fig. 11D, the genomic sequence of human *Myf-5* more closely resembled bovine *Myf-5* than mouse *Myf-5*. Between the human and bovine sequences a high degree of conservation (>80%) over the entire length of the exons and the introns of *Myf-5* was found with no obvious bias in any particular region (note contiguous diagonal in Fig. 12C). Unfortunately, this high level of similarity

Figure 12A: Dot Plot Analysis of the *MRF4/Myf-5* Region
Two way comparison of the genomic region comprising the *MRF4* and *Myf-5* genes from *Fugu* and mouse. Despite syntenic conservation of the *MRF4* and *Myf-5* genes in *Fugu*, no significant matches are found outside of the bHLH regions. Regions with at least 75% similarity over a 20bp window appear as lines on the diagonal. The bHLH regions of exon 1 of *MRF4* and *Myf-5* show a significant degree of conservation not only between *Fugu* and mouse (red), but also between each other (green).

Figure 12B and C: Two way comparison of the human *Myf-5* genomic sequence with the bovine (**B**) or mouse homologues (**C**). Regions of similarity appear as lines on the diagonal. The exon - intron structure is superimposed onto the diagram to allow regions of similarity to be aligned with the genomic organisation. The human intron sequence shows significantly more homology to the bovine homologue than the mouse homologue. Between human and mouse, the 3' halves of the introns are significantly conserved, whereas the 5' halves are not. Exons are well conserved, in mouse, human and bovine.

Dotplot Analysis: *MRF4-Myf-5* region

Fig. 12 A: *Fugu* vs Mouse



Percentage: 75; Window: 20

Dotplot Analysis: Myf-5

Fig 12B: Human vs Bovine

Percentage: 75; Window: 15; Min Quality: 1

Seq1(1>3634)	Seq2(1>3570)
human	bovine
(1>3634)	(1>3570)

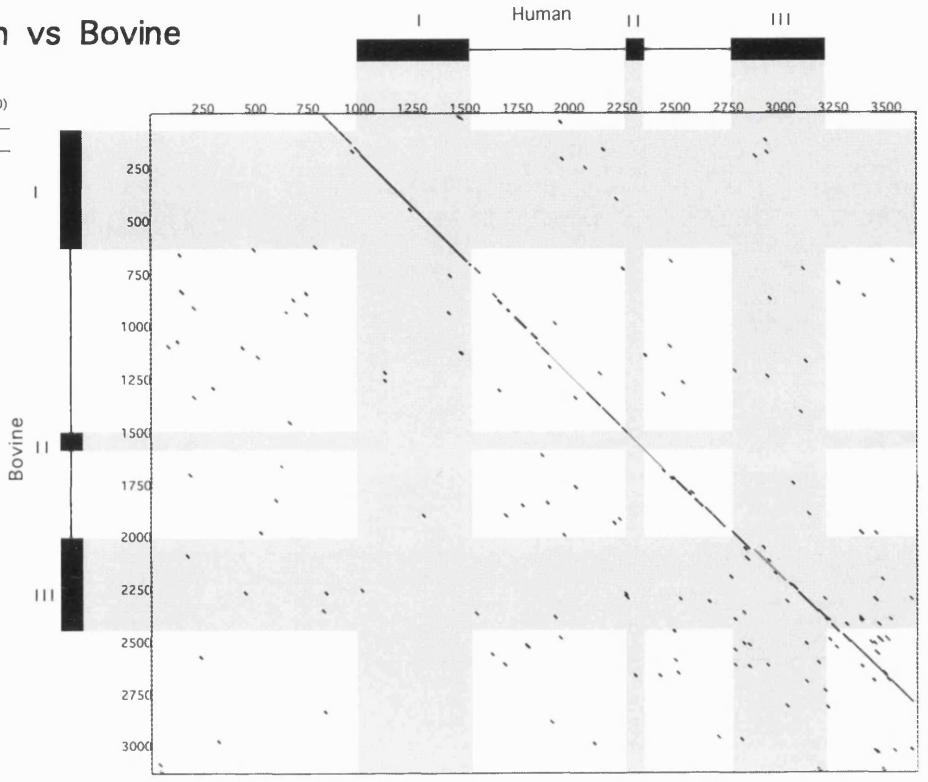
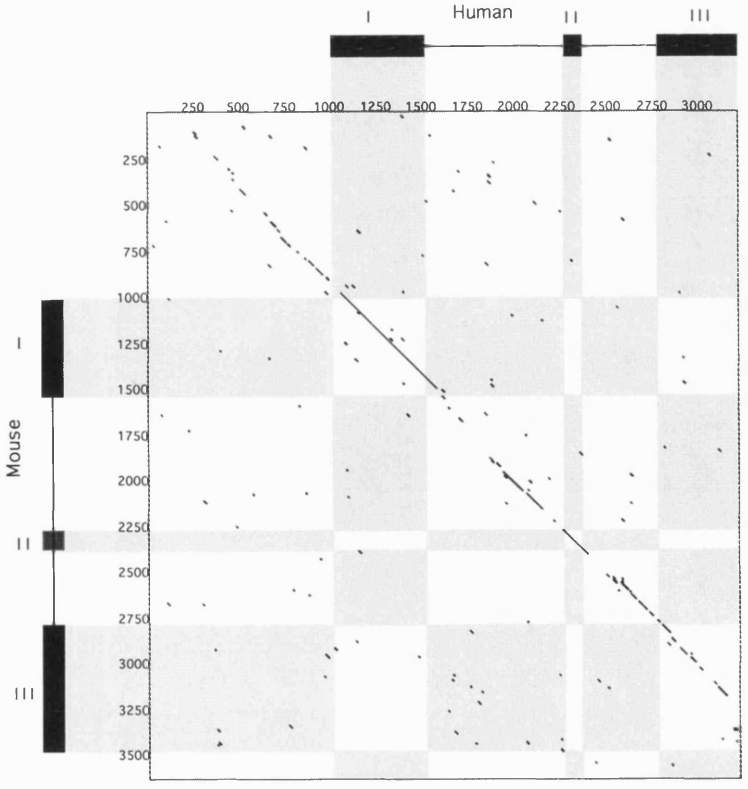


Fig 12C: Human vs Mouse

Percentage: 75; Window: 15; Min Quality: 1

Seq1(1>3214)	Seq2(1>3634)
mouse	human
(1>3214)	(1>3634)



precluded identification of specific conserved elements just like the complete lack of sequence conservation observed previously between *Fugu* and mouse *Myf-5*. This result illustrates the fine balance between sequence divergence and conservation required for meaningful comparison. In this regard dot plot analysis of the human and mouse *Myf-5* genes proved more informative.

Figure 12C shows the two way sequence comparison between the mouse and human *Myf-5* genes. The diagonal forms a solid line in the coding region but breaks down considerably in the 5' half of each intron. A more detailed comparison of the introns sequence is shown in Figure 13A and B. In contrast to the 5' half, the 3' half of each of the introns has remained highly conserved suggesting that they may harbour regulatory sites. The 3' half of the first intron of mouse *Myf-5* shares more than 78% similarity with the human or bovine sequence over a continuous stretch of 300bp compared with less than 40% identity outside of this region (Fig. 13A). Similarly, in the second intron the 3' half shares 75% homology compared with less than 40% in the 5'-half (Fig. 13B). Below the dot plot diagram in Figure 13A and B the nucleotide sequence of the major conserved sequence blocks in the 3' half of the introns is shown.

5. Summary

The *MRF4* and *Myf-5* genes of the pufferfish *Fugu rubripes* form a syntenic linkage group that has been maintained for 430 MYrs since teleosts started to evolve. The *Fugu MRF4* and *Myf-5* genes have the same genomic organisation as other myogenic factors. The bHLH domain in the first exon shares more than 80% identity with the mouse and has been shown to be necessary and sufficient to induce myogenesis (Davis *et al.*, 1990). The precise sequence of the basic region is essential for transactivation of target promoters and myogenic conversion of fibroblasts in tissue culture (Weintraub *et al.*, 1991; Davis and Weintraub, 1992; Schwarz *et al.*, 1992), which is reflected by its nearly complete conservation in *Fugu* and diverse species such as birds, amphibia and mammals. Similarly, the amphipathic helix domains required for dimerisation with E2 proteins (Murre *et al.*, 1989) show mainly conservative amino acid substitutions, and a cysteine-histidine-rich region upstream of the basic domain that is thought to mediate changes in chromatin structure required by Myf-5 and MyoD for the activation of their target genes is also highly conserved. This high degree of conservation would suggest that the mechanism by which these proteins act is extremely ancient.

Although the *Fugu MRF4* and *Myf-5* genes are separated by less than 3 kb compared to 7kb in the mouse, one might have expected the genes to be as close as 1kb considering the size of the *Fugu* genome. Despite the relatively modest size reduction of the region in *Fugu* there is no significant sequence conservation in the intergenic region or the introns

Figure 13 A and B: Two way comparison showing the conserved regions of the human and mouse *Myf-5* introns 1 and the relative positions of the bandshift fragments A, B,C & D tested. Regions of similarity appear as lines on the diagonal. In both introns the 3' half shows a high degree of sequence conservation, whereas the 5' half does not. Below the diagram, the nucleotide sequence of the major conserved sequence block in the 3' half of the intron is shown.

Dot Plot Analysis: Mouse vs Human *Myf-5* Introns

Fig 13A: Intron 1: Conserved Regions

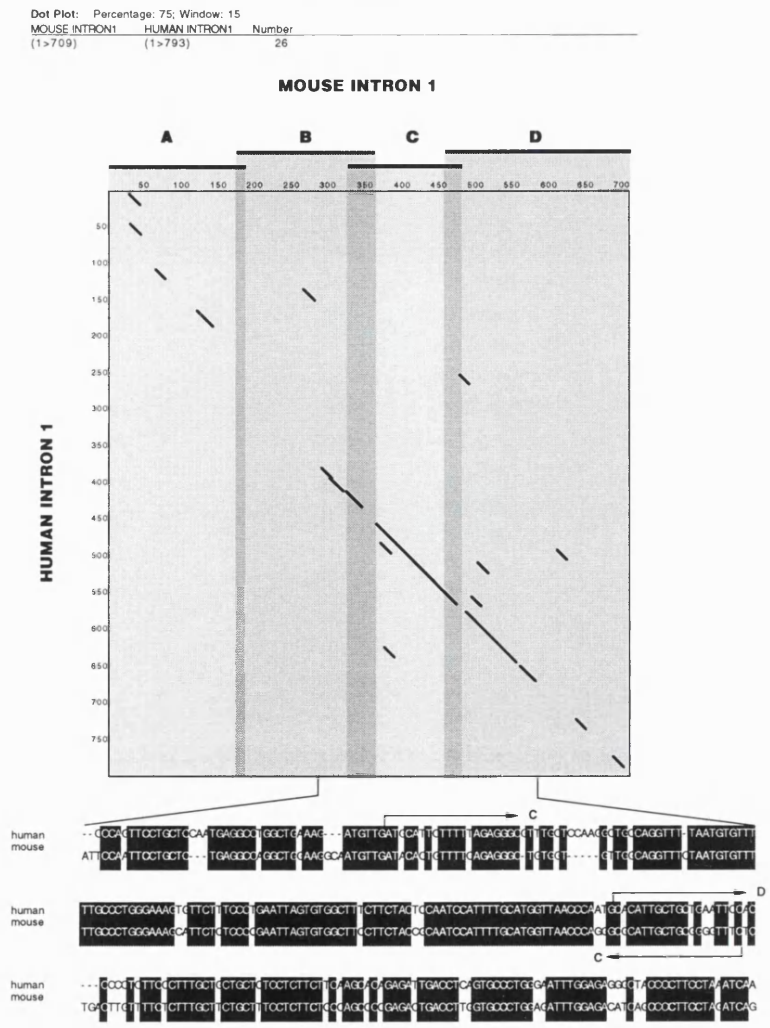
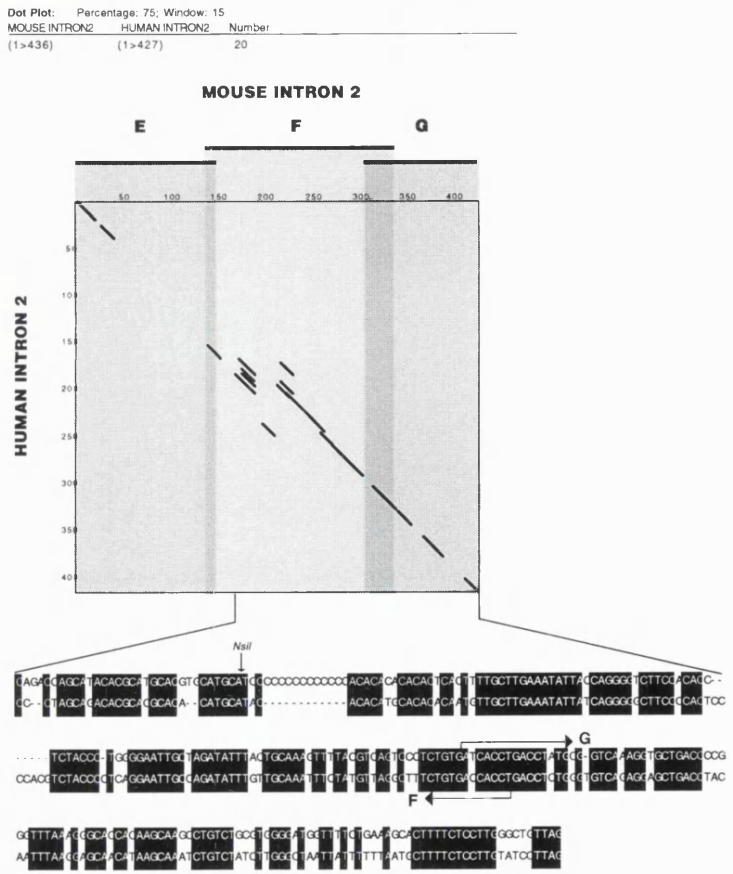


Fig 13B: Intron 2: Conserved Regions



of *MRF4* or *Myf-5*. In fact, the *Fugu Myf-5* and *MRF4* genes are nearly the same size as the mouse genes. The reduction in the size of the *Fugu* introns is unequal, indicating that the distribution of putative regulatory elements in *Fugu* and mouse might be different. Since the myogenic factors have split up very early in vertebrate evolution, they have diverged significantly in species as distant as *Fugu* and mouse, precluding identification of conserved regulatory elements. Comparisons of the *Fugu Myf-5* or *MRF4* genes with the homologues in mouse, bovine and human revealed little similarity outside of the coding regions. Therefore *Fugu* is probably not a good model for vertebrate myogenic genes. The most informative sequence comparison of *Myf-5* was between mouse and human. In both introns of *Myf-5*, a remarkable bias in sequence conservation in the 3' half compared with the 5' half was apparent. While the 3' halves of both introns were highly conserved, showing more than 75% identity, the 5' halves shared less than 50% sequence identity, suggesting that the second half of each of the introns contains functionally conserved elements. Similar sequence comparisons between the mouse and bovine *Myf-5* genes produced almost identical results, suggested the presence of regulatory sites in the conserved intron regions, a possibility that was examined further in electrophoretic mobility shift assays (EMSA)

Chapter 4

EMSA Analysis of the *Myf-5* Introns

1. Introduction

Data from our laboratory had previously shown that the *Myf-5* gene contained regulatory elements involved in controlling the ventral expression domain of *Myf-5* in the somites (see Fig. 8C). Sequence comparisons between mouse and human revealed a high degree of sequence conservation in the 3' halves of the *Myf-5* introns suggesting the presence of regulatory sites. To facilitate the identification of putative transcription factor binding sites, the ability of the intron regions to associate with proteins from mouse embryonic extracts *in vitro* was tested. Electrophorectic mobility shift analysis (EMSA) was performed with sub-fragments of the *Myf-5* introns using various embryonic extracts (see Chapter 2, section 12 for details). Briefly, adjacent overlapping intron fragments spanning both introns of *Myf-5* were generated by the polymerase chain reaction (PCR). Figure 14A shows the primer positions and the length of the seven products (A-G) generated. The source of the embryonic tissues used in these extracts is illustrated in Figure 14C. Details concerning the preparation of the embryonic protein extracts from these tissues can found in Chapter 2, section 12.2. The dissections required considerable dexterity and were carried out by Dr Summerbell of our laboratory. Each of the seven fragments was tested for its ability to specifically bind proteins from *Myf-5* expressing and nonexpressing embryonic cell extracts as well as F9 embryonal carcinoma cell (EC) extracts, with the intention of obtaining additional information about the tissue distribution of the binding factors.

The EMSA results are summarised in Figure 15. Surprisingly, no significant differences were found in bandshift experiments with somite extracts compared with other tissues, suggesting that the binding activities are ubiquitous and/or that minor contamination of the embryonic tissues eliminated the intended tissue specificity. Although six of the seven fragments bound proteins in the EMSA assays, it became clear in subsequent transgenic analysis (see Chapter 5), that only three of the fragments (D, F and G) were required to drive expression in the ventral posterior somite domain, suggesting that under the *in vitro* conditions and considering the large sizes of the fragments, some of

Figure 14A: Size and location of the probes used in mobility shift assays. Adjacent and overlapping probes (A-G) for EMSA were generated by PCR. Each probe (depicted by double headed arrows) overlapped with the adjacent fragment by at least 15 bp. The length is shown below each fragment, the PCR primers are shown as arrow heads above (See Appendix I for primer sequences).

Figure 14B: Oligonucleotides subdividing Fragment D of intron 1 into 5 overlapping adjacent fragments were used in bandshift analysis using E10.5 whole embryo extracts. Fragments D1-D3 formed specific complexes whereas fragments D4 to D5 did not, indicating that the most likely binding site lies in the 110bp region at the 5' end.

Figure 14C: Origin of the mouse tissues used to make whole cell protein extracts for EMSA. The anatomical regions depicted were dissected from E9.5 embryos and processed as described in Chapter 2 section 12.2 to obtain information about tissue specificity of the binding activities. The somites and arches express the highest level of *Myf-5* whereas head neck and flank were expected not to express significant levels of *Myf-5* however; no significant differnces were observed between these extracts in EMSA.

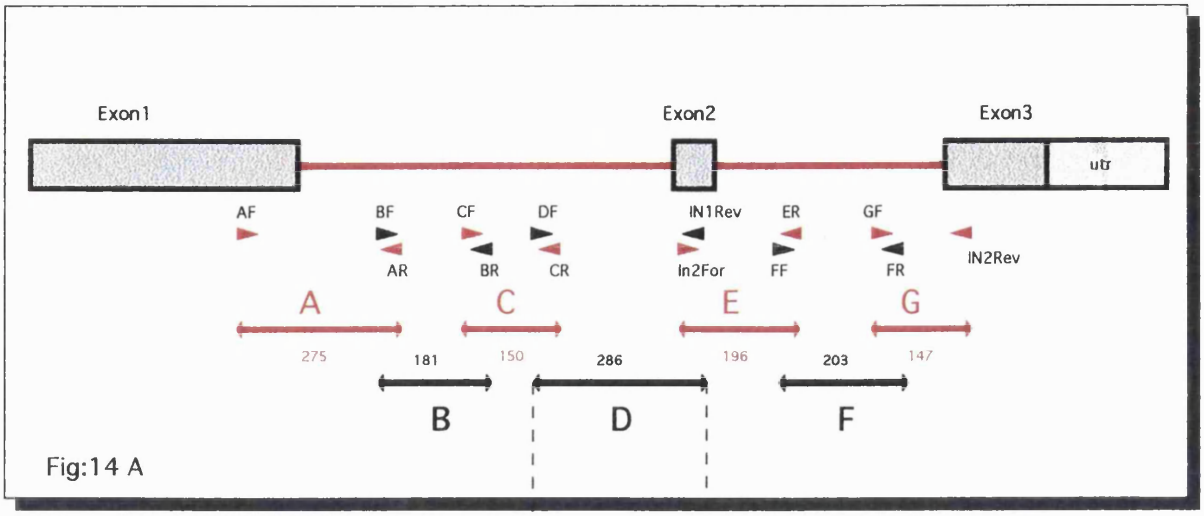


Fig:14 A

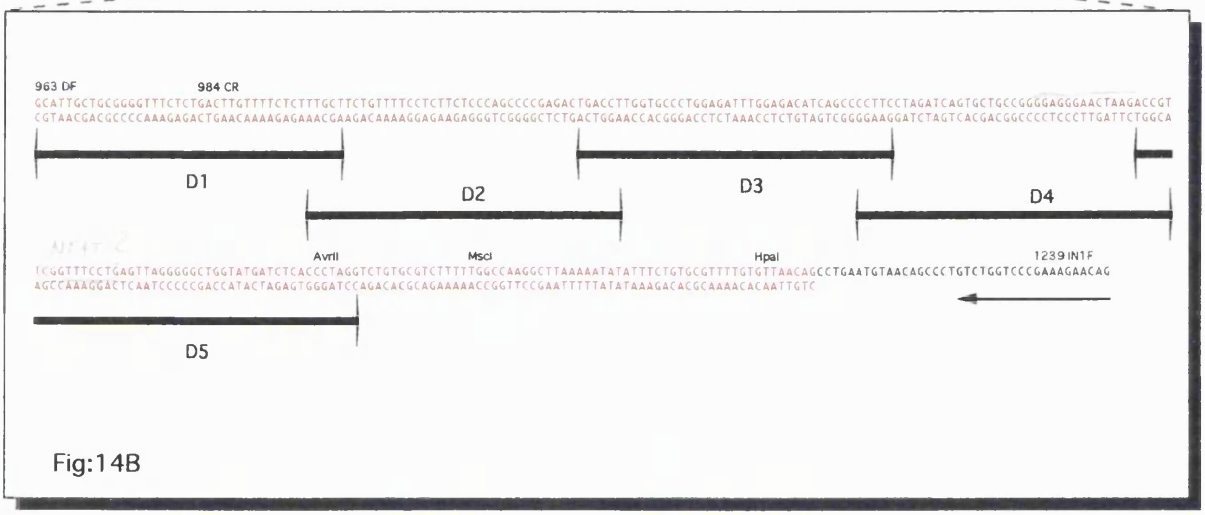


Fig:14B

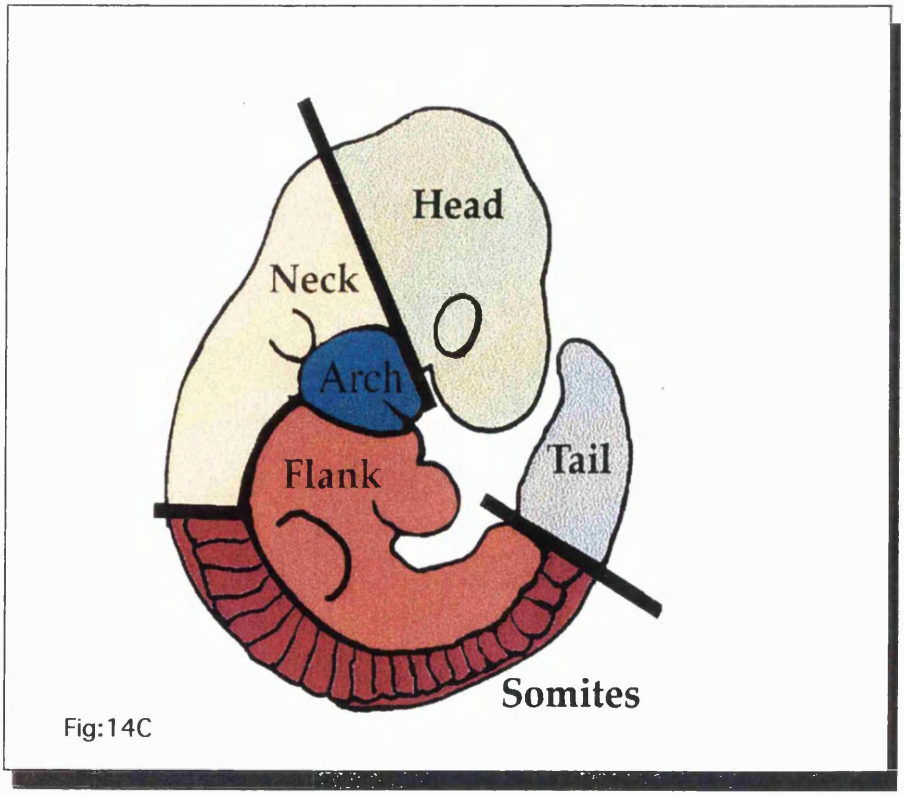


Fig:14C

the binding activities were non-specific or may be involved in regulatory functions other than the regulation of the ventral posterior expression domain of *Myf-5*. The mobility shift data are summarised in the following:

Fragment A at the 5' end of intron 1 is poorly conserved between mouse and human (Fig. 13A) and although mobility shifts were observed with somite and branchial arch extracts (Fig. 15A lanes 2 and 3), these were not efficiently competed by 200 fold molar excess of unlabelled fragment A (compare lanes 9 and 10) indicating that the observed binding activities were probably non-specific.

Fragment B also lies in the nonconserved region of intron 1 (Fig. 13A) but showed a clearer binding pattern with one major and several minor binding activities with F9 EC cell nuclear extracts (Fig. 15B, lane 2), extracts from the flank and somites (lanes 4 and 5) and whole embryo nuclear extracts (lane 6). The strongest signal was obtained with whole embryo nuclear extracts suggesting that the binding activity may be enriched in the nucleus (compare lane 6 with lanes 2, 4 and 5). The DNA-protein complex is competed efficiently in the presence of 50- to 200-fold molar excess of unlabelled probe (compare lane 6 with lanes 7-9) suggesting that the binding activity is specific.

Fragment C falls in the conserved region of intron 1 sharing more than 78% identity between mouse and human (Fig. 13A). Two major complexes formed with F9 EC cell extracts and different embryonic extracts (Fig. 15C, lanes 2-7). Interestingly, a third, faster migrating complex was formed only with F9 nuclear extracts (lanes 2 and 8) but this complex was not efficiently competed with cold competitor and is probably non-specific (compare lanes 8 and 9). Of the remaining two bands the slower complex appeared to be specific as it could be competed with 100-fold excess of unlabelled fragment C (compare lanes 8 and 9) while the other complex was not affected by competition (compare lanes 10 and 11).

Fragment D , located at the conserved 3' end of intron 1, produced a complex banding pattern with various whole embryo extracts and somite extracts that was difficult to interpret (data not shown). Therefore five subfragments D1-D5 (see Fig. 14B) were tested in bandshift assays with E10.5 whole embryo extracts. Fragments D1-D3 formed specific complexes that could be competed in the presence of both 50- or 100-fold excesses of unlabelled fragment D1 (lanes 3 and 4), D2 (lanes 9 and 10) or D3 (lanes 15 and 16) respectively, but not by non-specific oligonucleotide competitor (NS) at the same concentration (lanes 5 and 6, 11 and 12 and 17 and 18). In contrast, fragments D4 and D5 failed to bind specific proteins, as the binding activity was not efficiently competed by either specific or non-specific competitor in 50- to 100-fold molar excess (lanes 21 to 24 and 27 to 30), suggesting that tight but non-specific binding was observed.

Figure 15: Electrophoretic Mobility Shift Assays (EMSA).

Mouse *Myf-5* intron fragments (see Fig 14. for details) were radioactively labelled and used as probes in EMSA as described in Chapter 2, section 12.3. Whole cell extract from embryonic tissues was used as indicated (see Fig. 13C for details): WE = whole embryo E10.5, A = Arch, F = Flank, H = Head, N = Neck, S=Somites. Nuclear extracts from F9-EC cells were a gift from Dr. Vandromme of this laboratory. NE = nuclear extracts from E10.5 embryos prepared by the method described by Nicolas and Goodwin (1993). Specific competitor was used in molar excess as indicated. FP = free probe. Please refer to text for details of interpretation.

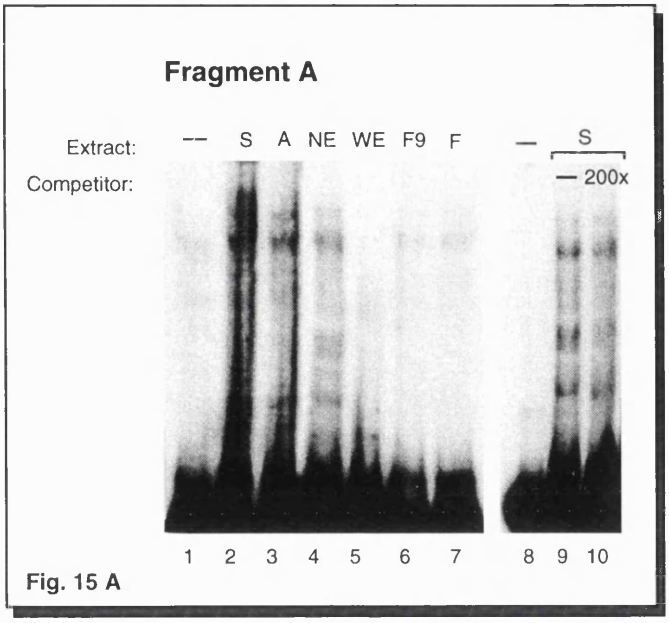


Fig. 15 A

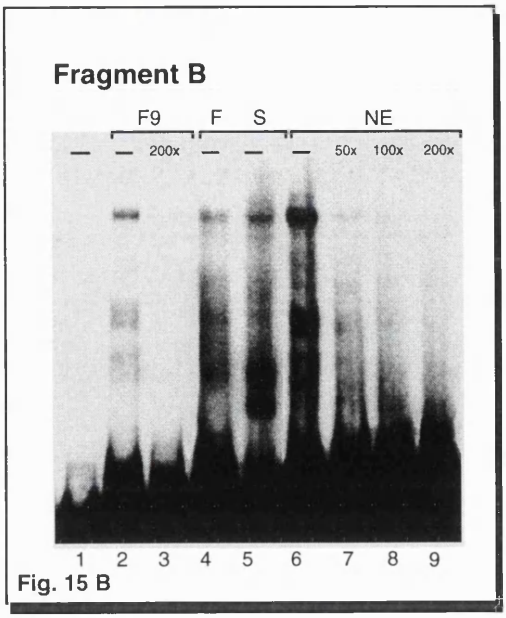


Fig. 15 B

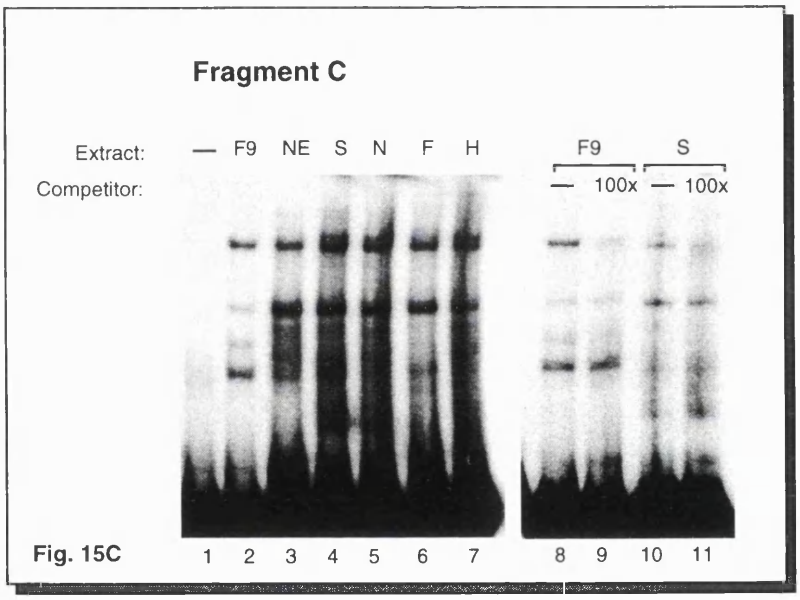


Fig. 15C

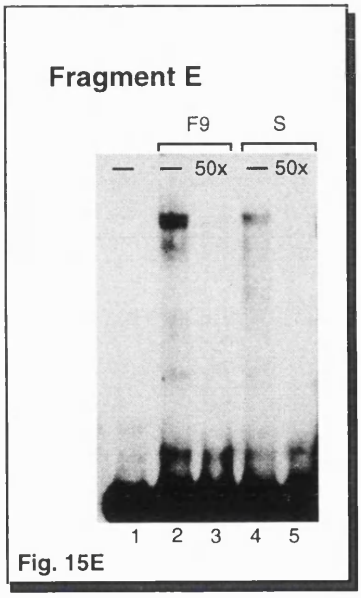


Fig. 15E

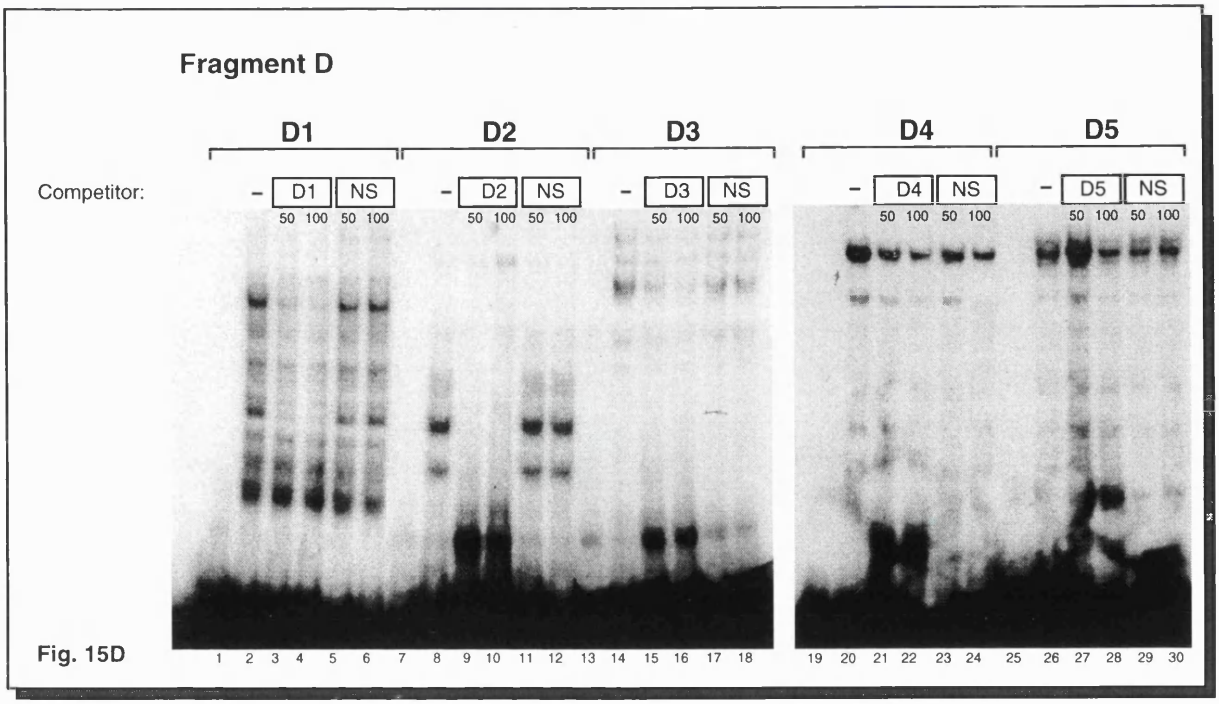


Fig. 15D

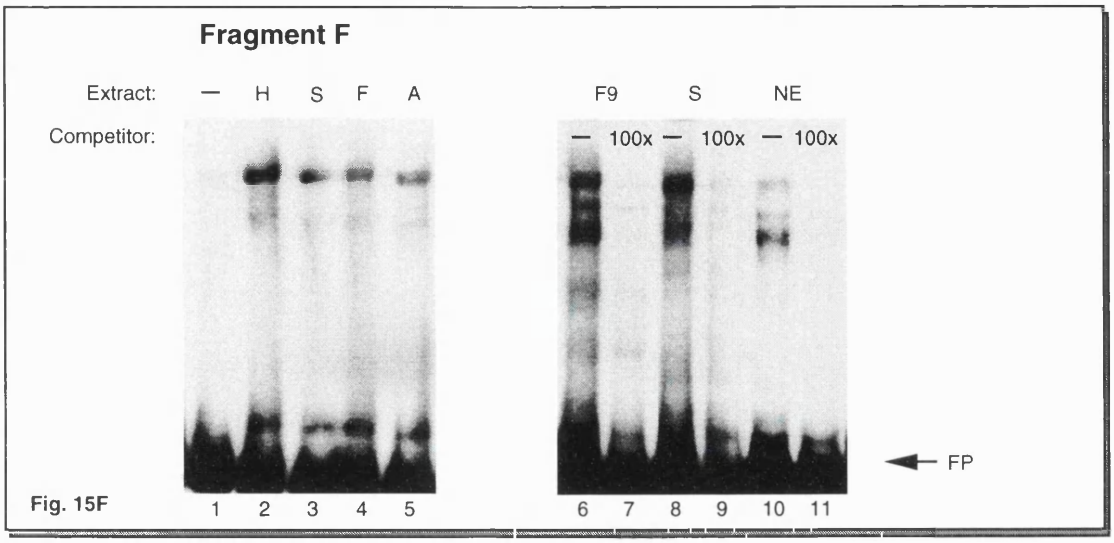


Fig. 15F

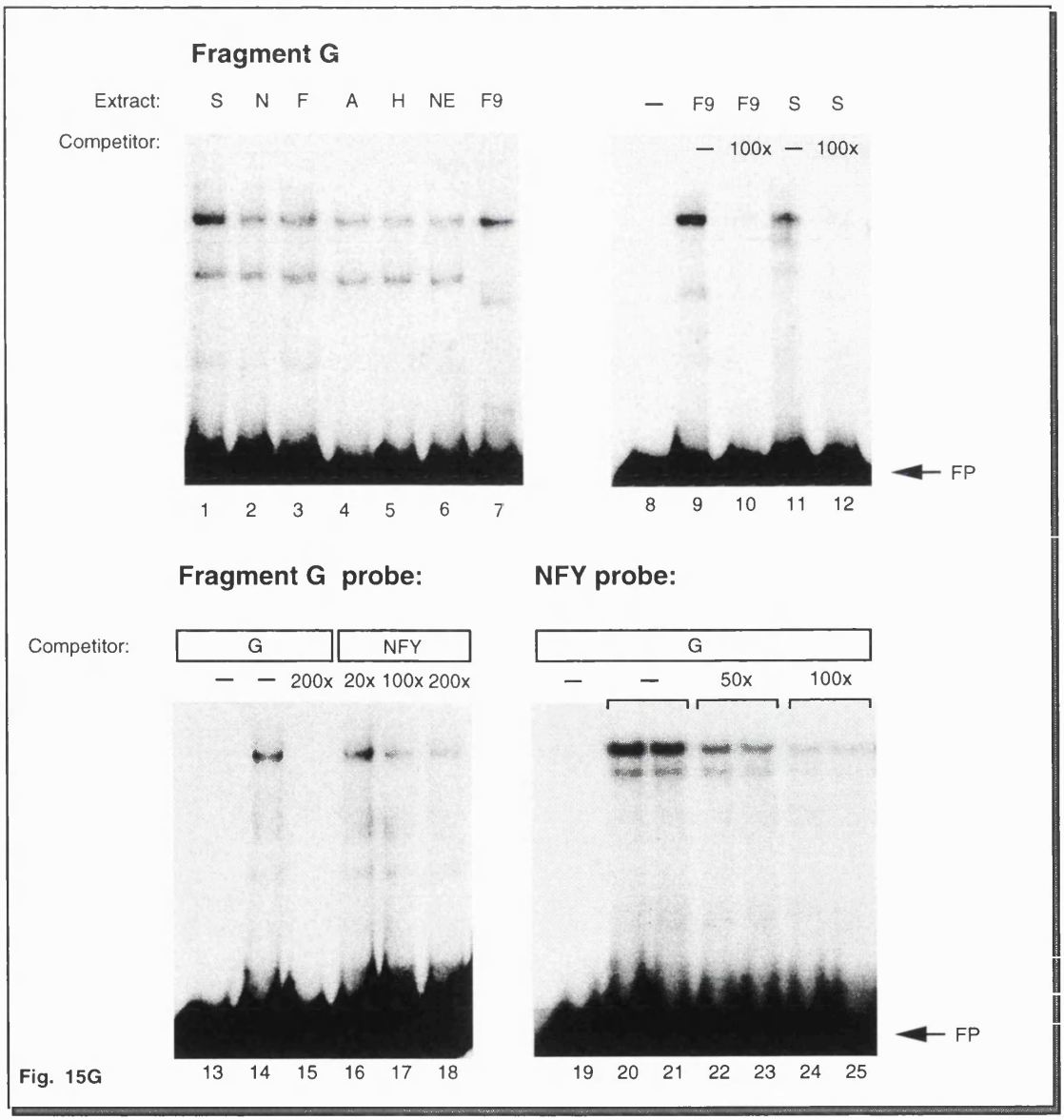


Fig. 15G

Fragment E is located in the nonconserved region at the 5' end of intron 2 (Fig. 13B). One major binding activity was observed with F9 and somite extracts (Fig. 15E lanes 2 and 4). The activity was specifically competed by preincubation with 50-fold molar excess of unlabelled fragment E, suggesting the presence of a specific high affinity binding site (compare lanes 2 and 3 and lanes 4 and 5).

Fragment F comprises 203 bp of the conserved central part of intron 2 (Fig. 13B). The same major band was observed with different whole cell embryonic extracts including those from the head, somite, flank and arch (Fig. 15F, lanes 2 to 5) as well as nuclear extracts from F9 cells and whole embryos (lanes 6 and 10). When nuclear extracts were used, additional faster migrating bands were observed that were absent or less abundant in the whole cell extracts (compare lanes 6 and 10 with 1 to 5). In whole embryo nuclear extract the faster band represented the main binding activity (lane 10), suggesting that this binding activity is enriched in the nucleus. Regardless of the extract used, all of the binding was competed by preincubation with 100-fold molar excess of unlabelled fragment F suggesting that the complexes formed were specific (compare lanes 6, 8 and 10 with 7, 9 and 11). Fragment F is well conserved in mouse and human (see Fig. 13B), particularly in the 3' half, however, no known binding sites could be identified by sequence comparison with the TRANSFAC database. Deletion of fragment F in transgenic assays abolished ventral posterior expression in the somites, indicating that the binding activities are involved in the regulation of the transgene (see Chapter 5, compare construct pΔFPH construct pΔCPH).

Fragment G comprises the conserved 3' end of intron 2 (see Fig. 13B). Two major binding activities were observed with this probe when it was incubated with different embryonic cell extracts including: somite, neck, flank, arch, head, whole embryo nuclear extract and F9 embryonal carcinoma cell extracts (Fig. 15G, lanes 1 to 7). With F9 extracts a different faster migrating complex formed that was not seen with any of the embryonic extracts (compare lane 7 with 1 to 6). Although little variation in intensity was observed for the two bands with different embryonic extracts, the slow band in the somite and F9 extracts was relatively more intense (Fig. 15, compare lane 1 and 7 with lanes 2-6). Protein binding from F9 and somite extracts was completely competed by preincubation with 100 fold excess of unlabelled fragment G (compare lanes 9 and 11 with 10 and 12) indicating a high degree of specificity. Closer examination of the sequence of fragment G revealed the presence of a potential NFY binding site matching the NFY consensus in 8 out of 9 bases (CCATCCCCA the underlined C is normally a T). NFY is a heteromeric transcription factor (Hooft van Huijsduijnen, R.*et al.*, 1990) that has been shown to be identical with the transcription factor HoxTF, involved in the regulation of *Hox* genes (Gutmann *et al.*, 1991).

Summary of the Bandshift Data

Intron1

Fragment	Bandshift Result	
A	nonspecific binding	—
B	specific binding (F9, somites and others)	+
C	multiple bands, one specific (F9, somites and others)	+
D1	multiple bands, one specific (whole embryo extract)	+
D2	two specific activities (whole embryo extract)	+
D3	multiple bands, one specific (whole embryo extract)	+
D4	nonspecific binding	—
D5	nonspecific binding	—

Intron 2

Fragment	Bandshift Result	
E	specific binding (F9 and somites)	+
F	specific binding (F9 and others)	+
G	specific binding (F9 and others) potential NFY binding site	+

Table 2: Summary of the EMSA Results for *Myf-5* Intron Fragments
Specific binding activities were observed for probes B to G (+) but not A (-). The specificitiy of binding was assessed by competion with unlabelled probe see Fig. 8A-G for details. Note, that fragment D is subdivided into fragments D1 to D5 of which D4 and D5 failed to bind specific protein (-). Please refer to text for details.

Studies from our laboratory showed that NFY plays a role in the regulation of the *Hox-b4* gene, where it is necessary for all of the somitic and lateral mesodermal expression and part of the neural expression pattern observed in transgenic mice. Interestingly, an NFY site has also been found in the promoters of the *MyoD* and *myogenin* genes (Gilthorpe, unpublished; Gutman *et al.,* 1994), suggesting that NFY might be involved in the regulation of myogenic regulatory factors. To test whether the NFY site in the *Myf-5* introns was involved in protein binding, competition experiments with a NFY-specific oligonucleotide, corresponding to the *MyoD* promoter region -628 to -602 (Zingg *et al.,* 1991), at 20-, 100- and 200-fold molar excess (Fig. 15G, lanes 16 to 18) were carried out. The *MyoD*-(NFY) - oligonucleotide did compete, although less efficiently than fragment G itself (compare lane 15 with lane 18). Similar results were obtained by testing the ability of fragment G to compete for protein binding when the *MyoD*(NFY)- oligonucleotide was used as a probe. Fragment G in 100 fold molar excess competed efficiently for protein binding to the NFY oligonucleotide (Fig. 15G, compare lanes 20 and 21 with 22 to 25).

9. Summary

In conclusion, the results of the EMSA experiments suggest that multiple binding sites for proteins from embryonic extracts exist in the introns of *Myf-5*. As no significant difference was observed between various tissue-specific embryonic extracts, most of these binding activities appear to be ubiquitously distributed amongst embryonic tissues. Most of the complexes formed were also observed with nuclear F9 cell extracts although on two occasions (fragments C and G) the banding pattern was slightly altered. The Bandshift results are summarised in table 2. Although specific binding activities were found in all of the fragments (except A), only a subset of these appears to be required for the regulation of *Myf-5* in transgenic mice. Subsequent analysis of the *Myf-5* introns using *LacZ*-reporter constructs showed that deletion of the 'positive' bandshift fragments C and E had no apparent effect on the ability of transgenic constructs to reproduce the complete expression pattern in transgenic mice. However, the participation of the deleted fragments in the regulation of *Myf-5* can not be completely excluded because the wild type expression pattern of *Myf-5* is complex and the possibility of redundancy between the intron elements has not been sufficiently examined. Ultimately, to assess the function of regulatory elements *in vivo*, transgenic analysis of the *Myf-5* introns aimed at identifying a minimal region that is sufficient to direct *LacZ* transgene expression to the ventral posterior domain of the somites is necessary. The results of such an analysis are presented below.

Chapter 5

Transgenic Analysis of the Regulatory Role of the Mouse *Myf-5* Introns

1. Introduction

Previous studies from our laboratory established that a *LacZ* reporter gene construct (HMZ17) containing the intergenic region between *MRF4* and *Myf-5* including the two genes themselves (totalling some 14kb) could reproduce much of the endogenous *Myf-5* expression pattern in transgenic mice (see Fig. 8B and also Fig. 16C). The main features of this construct (HMZ17) include *LacZ* expression in the ventral (arrow, Fig. 16C') and dorsal parts of the somites (white arrow, Fig. 16C"), the branchial arches (white arrowhead, Fig. 16C) the trigeminal ganglion, the intercostal and abdominal muscles, the neural tube and head mesenchyme. Results from deletion constructs indicated that the complete expression pattern of HMZ17 appears to arise through combined transcriptional regulation by a number of separate control regions responsible for expression in individual anatomical subdomains. Several discrete regions, probably containing autonomous elements controlling expression in the branchial arches, the neural tube, the dorsal and ventral margins of the somites have been mapped by our laboratory (see summary in Fig. 8A). The dorsal somite expression domain was lost when 2.2kb of sequence were deleted from the 5' end of the 14.2kb HMZ17 construct (data not shown). In contrast the ventral domain of the somites was lost when the *Myf-5* gene near the 3' end of HMZ17 was deleted, suggesting that elements controlling expression in the ventral part of the somites were located in the *Myf-5* gene itself (data not shown). This was confirmed when the *Myf-5* gene including its 3' UTR was shown to direct expression of a *LacZ* reporter gene to the ventral portion of the somites (see Fig. 8C, MFGZ). These experiments demonstrated that ventral and dorsal expression domains in the somites are regulated by distinct control elements, separated by at least 4kb of sequence; the dorsal element being located in the intergenic region and the ventral

element within the *Myf-5* gene and/or the 3' UTR. To refine the elements in the *Myf-5* gene controlling the ventral somite pattern more closely, I combined the results of the comparative sequence analysis and electrophoretic mobility shift assays to identify candidate regions of the *Myf-5* gene and tested their ability to drive reporter gene expression to the expected *Myf-5* expression domains in transgenic mice.
Although pairwise comparisons between *Fugu* and mouse had failed to identify conserved sequence elements, similar comparisons between mouse and human revealed significant differences in conservation between the 5' and 3' halves of the *Myf-5* introns suggesting potential regulatory regions in the conserved 3' halves of both introns. Electrophoretic mobility shift assays on fragments A to G from the two introns of *Myf-5* showed that except for fragment A, all of the bandshift probes were able to bind proteins from embryonic and cell line extracts *in vitro*. To obtain information about the *in vivo* role of these candidate regulatory regions, the binding sites need to be tested functionally in transgenic mice. Suitable constructs were generated by placing deletion fragments of the *Myf-5* gene upstream of a minimal heat shock promoter plasmid driving the *LacZ* gene. In the absence of any enhancer elements, the wild type reporter construct introduced into the pronucleus of transgenic mice is essentially silent, although basal expression of *LacZ* is occasionally observed in the ventral neural tube ([Rossant *et al.*, 1991; Whiting *et al.*, 1991; Gilthorpe, pers.comm. my observation]). Putative enhancer elements inserted in the reporter are expected to drive expression of the *LacZ* transgene in transgenic mice to the target sites of the enhancer.

2. Dissection of Regulatory Elements in Myf-5 using Transgenic Mice.

At first, it was important to test if the *Myf-5* introns without the 3' UTR including the intervening exon were sufficient to direct expression of the reporter gene to the ventral posterior margin of the somite or if additional sequences were required. A *LacZ* reporter construct containing the introns of *Myf-5* but lacking the 3' UTR was made and tested. (Details of the design of the constructs are given in Chapter 2, section 10). A map of the transgenic reporter constructs and a summary of the results is shown in table 3 and table 3A, respectively.
p12 (introns 1 and 2 construct) see Fig. 16A and B: Of five transient transgenic animals two expressed the reporter gene. Both embryos showed *LacZ* staining in the somites, however, the expression patterns were inconsistent and

Table 3A: Summary of the Mouse Transgenic Results

construct	Exp/Tg	S	TG	DRG	A	NT	Comment
p12	2 / 5	+ / E	+ / E	−	+ / −	+ / −	variable expression
p1	7/ 11	− / E	− / E	+ / −	− / E	+	variable ectopic
p2	4/ 6	−	−	−	−	+	
p12UTR	Line	+	+	+	+	+ / −	highly consistent somitic
pUTR	5/ 5	−	−	−	−	+	
p2UTR	6/ 6	−	+	+	+	+ / −	consistent nonsomitic
p1UTR	2/ 6	−	−	−	−	+	
pCFUTRR	Line	+	+	+	+	+	strong somitic expression
pCFH	2/ 2	+	+	+ / −	+	+ / −	
pCFA	4/ 4	(+)	+	+ / −	+ / −	+	very faint somitic
pCFHΔ	3/ 4	+	+	+ / −	+	+ / −	strong somitic expression
pCFAΔ	2/ 2	(+)	+ / −	+ / −	−	+ / −	very faint somitic
pΔAH	3/ 6	+ / −	+	+ / −	−	+ / −	variable expression
pΔCAH	2/ 5	−	+	+	+	+ / −	
pΔGAH	2/ 5	−	−	−	−	+	
pΔCPH	2/ 3	+	+	+ / −	+	+ / −	
pΔGPH	2/ 4	−	−	−	−	+	
pΔFPH	2/ 7	−	−	−	−	+	

Table 3A: Summary of the expression patterns obtained with the constructs indicated. **Exp/Tg** shows the number of animals expressing the transgene out of the total number of transgenic animals obtained (as detected by PCR). **S** indicates ventral posterior somitic expression, **TG** trigeminal ganglion, **DRG** dorsal root ganglion, **A** branchial arch and **NT** Neural tube expression. (+) shows whether or not (-) a construct produced the correct expression or if expression was ectopic (E). Combinations of these indicate that variable results were obtained.

Table 3: The genomic structure of the *Myf-5* region is shown at the top. Exons are depicted as gray boxes, introns as red lines, and the 3' UTR is shown in light gray and boxed. Vertical lines show the position of common restriction sites or primer binding sites amongst different constructs or in the genomic sequence. PCR- primers used to generate suitable fragments are indicated by labelled arrows, v-shaped brackets indicate deletions. Each construct was made by cloning the indicated region of the *Myf-5* gene upstream of the minimal *hsp68* promoter driving the *LacZ* gene with an SV40 polyadenlyation site. A + or - sign indicates whether correct expression was obtained, +/- indicates that the construct produced inconsistent or very weak expression.

Table 3: Mouse *Myf-5* reporter constructs

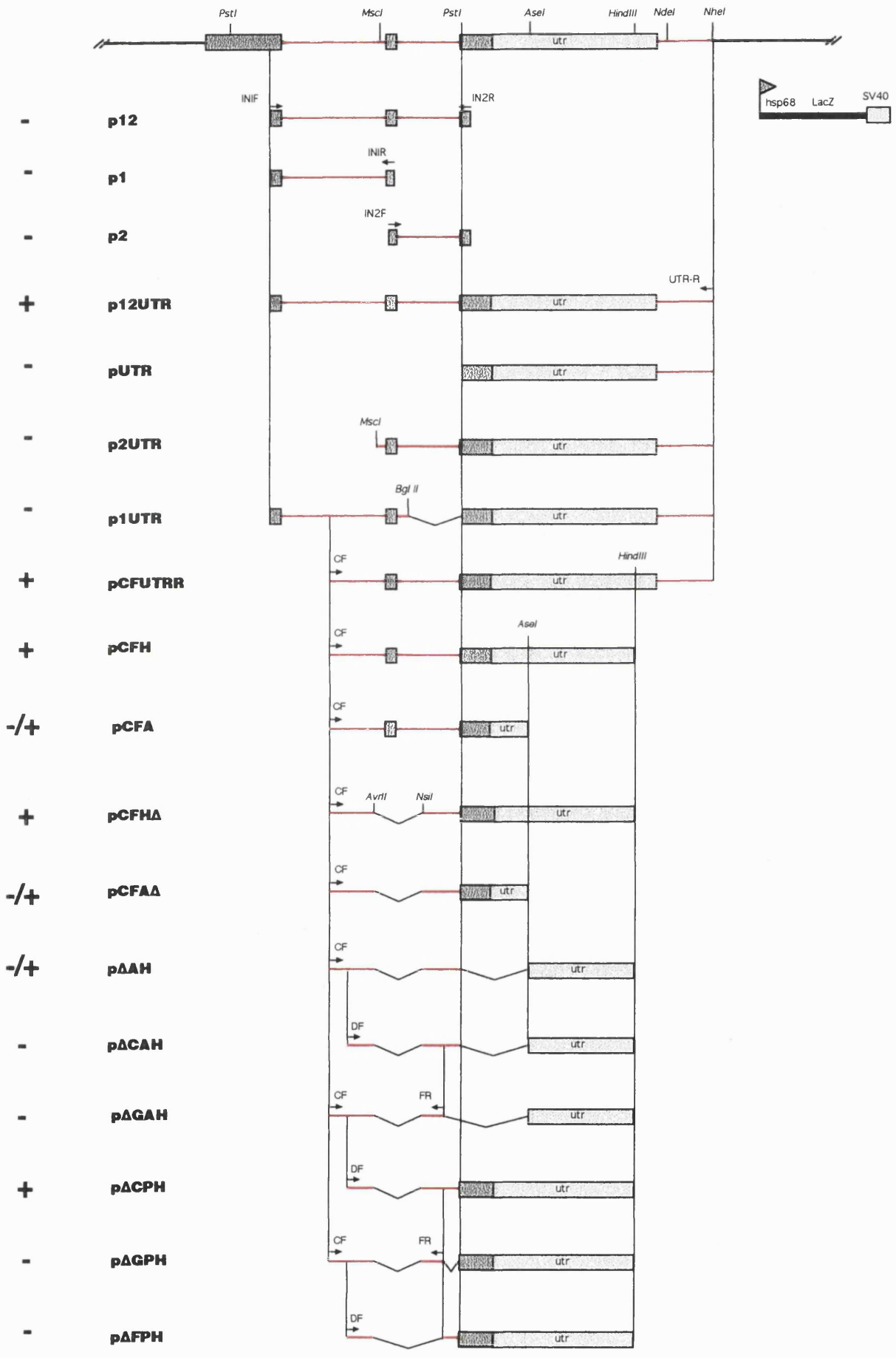
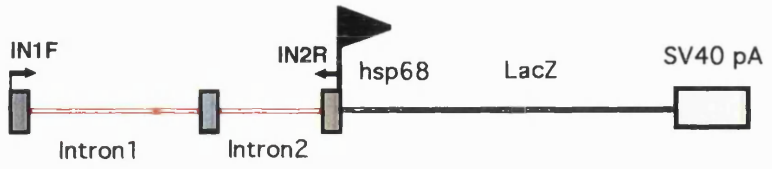
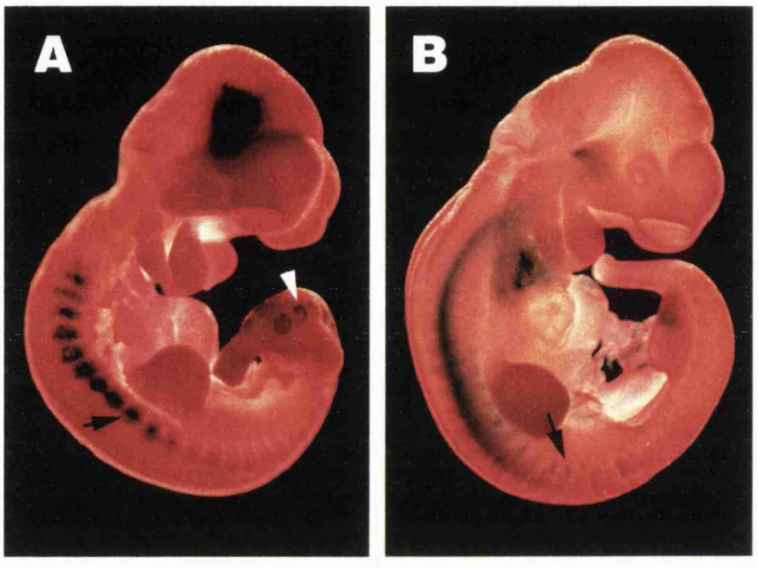
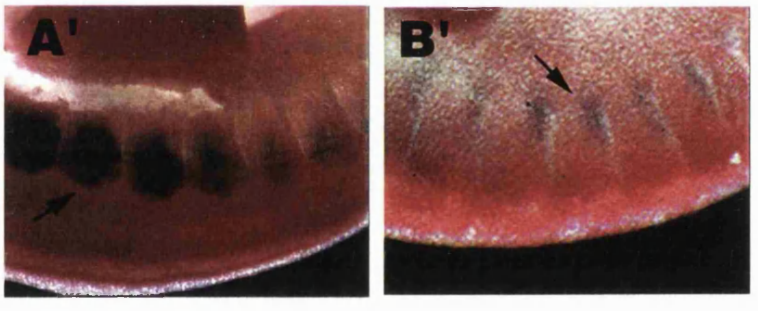


Fig. 16: Map of the p12 construct containing the *Myf-5* introns upstream of the hsp68 promoter driving *LacZ*. β-galactosidase expression in two p12 transgenic embryos at E10.5 stained with X-Gal (16A) and (16B), and an HMZ17 transgenic animal of the same age (16C). Close-ups of the same embryos are also shown (16A', 16B', 16C' and 16C").
Blue staining is localised in the somites of both animals in 16A and 16B but is not consistent. Close ups show that expression in the thoracic somites is mainly dorsal (arrow) in 16A' and ventral posterior (arrow) in 16B'. However, the pattern in 16B' resembles that obtained with HMZ17 (arrow) in 16C', see text for details.

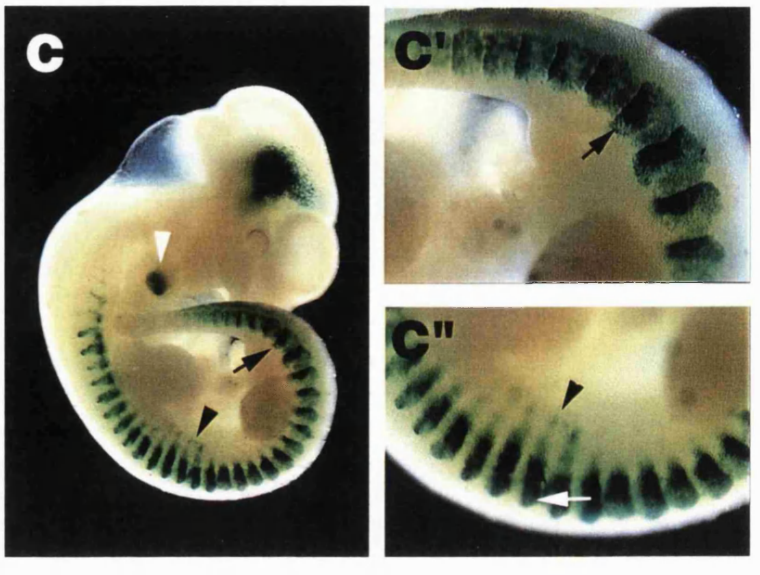
p12







HMZ17



showed little or no overlap (compare Fig. 16A and B). The embryo in Fig. 16B shares a similar ventral somitic expression pattern with HMZ17 (arrow, Fig. 16C) but the level of expression in the somites was noticeably weaker. Additional ectopic staining was seen in the ventral neural tube, the cranial nerves, and the hind limb bud. Somitic staining was observed throughout the thoracic somites in the ventral posterior margin of the dermomyotome (arrow, Fig. 16B') including the tail somites, where staining is not typically observed in MFGZ embryos.

In contrast, Fig. 16A shows a transient transgenic animal with *LacZ* expression in areas that are clearly not part of the HMZ17 expression pattern. Unlike the ventral expression observed in HMZ17 which is mainly in the dermomyotome the expression seen in Fig. 16A is in the dorsomedial part of the myotome (arrow, Fig. 16A'), confined to the occipital and thoracic somites. Expression is absent from the more posterior thoracic and lumbar somites. Further differences are found in the tail somites. Notably, expression is observed at the anterior rather than the posterior edge seen in HMZ17 (compare white arrowhead, Fig. 16A with arrow Fig. 16C').

To examine if either of the two introns of *Myf-5* was responsible for the somitic expression seen in Fig. 16B, reporter constructs with either intron 1 or intron 2 driving *LacZ* from the hsp68 promoter were tested in transgenic mice.

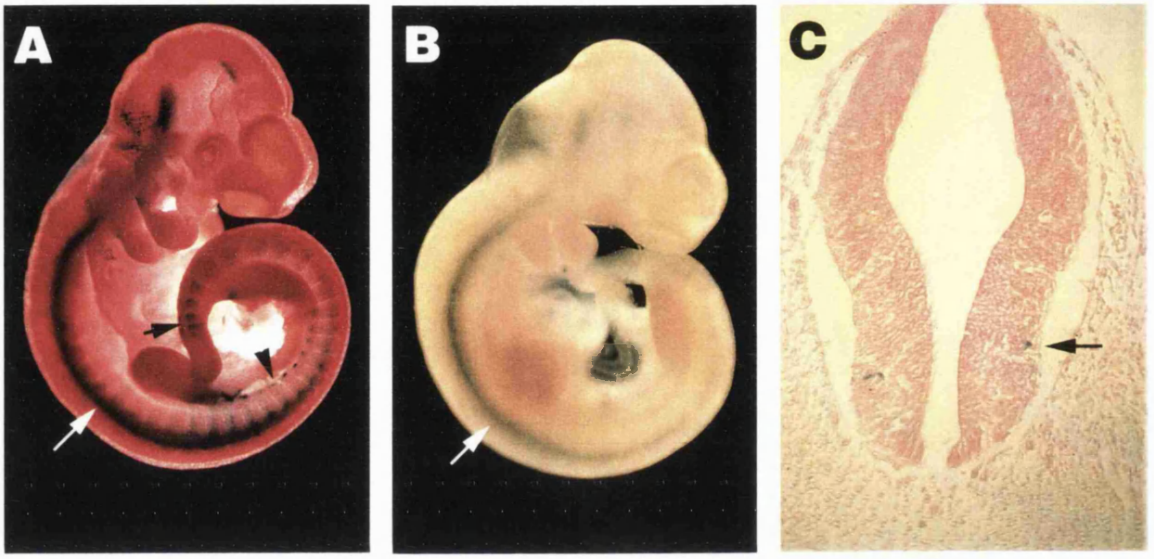
The first intron (**p1**-construct) gave *LacZ* expression in seven out of eleven transgenic animals (see Fig. 17 A-C). All of the embryos showed weak but consistent *LacZ* staining in the ventral neural tube (white arrows, Fig. 17A,B) and in only one transgenic animal was additional expression observed. In this embryo, the head mesoderm, the ventral somitic bud (arrow head) of the thoracic somites, and the tail somites (arrow) were stained (Fig. 17A). In contrast to the expected ventral posterior expression pattern, staining in the tail was confined to the anterior edge of the somites, similar to that seen with p12 (compare with Fig. 16A). Unusually, the most caudal margin of expression in the tail extended into the presegmental plate (arrow, Fig. 17A). Similar results were obtained with the second intron (**p2**construct, Fig. 17D). Of six transgenic animals, four expressed the *LacZ* transgene and all showed faint staining in the ventral neural tube (arrowhead) similar to the pattern seen with p1 before. This was confirmed in transverse sections of both p1 and p2 embryos, showing *LacZ* expressing cells near the lateral edges of the ventral third on both sides of the neural tube (arrow, Fig. 17C, E). No other expression domains for p2 were observed. Similar ventral neural tube expression has previously been reported to be ectopically induced by the heat shock promoter hsp68, used in all of these constructs (Joyner *et al.* 1987; Buckingham, 1994). It is therefore not possible to decide if the neural tube expression represents the true expression domain of the DNA tested, or merely an artefact of the construct. However, the observed expression is an indication that the construct integrates in a chromosomal region that is permissive for reporter gene expression.

Taken together, both the first (p1) and second intron (p2) of *Myf-5* failed to activate reporter gene expression in the ventral somites in transgenic mice. In combination, both introns (p12) produced a somitic but inconsistent expression pattern, strongly suggesting that additional or independent elements in the 3' UTR were required. To test these possibilities, it was first investigated if elements in the 3' UTR would either be necessary or sufficient to reproduce the ventral somite expression pattern.
Pronuclear injection of **pUTR** (3' UTR construct, see Fig. 18A and B) yielded 5 transgenic embryos, which again showed no somitic expression but consistent faint staining in the ventral neural tube (arrowheads), similar to that observed in p1 and p2 transgenic mice (compare Fig. 18A and B with Fig. 17B and D). No other expression domains were found, indicating that the 3' UTR does not contain self-sufficient regulatory elements that could activate *LacZ* expression in the ventral somite compartment and that elements of both the introns and the 3' UTR may be required.
Since none of the constructs tested thus far reliably reproduced the ventral posterior somite pattern, it was necessary to exclude the possibility that the hsp68 promoter used was in some way interfering with the regulatory elements tested. Previous results from our laboratory using a minimal β-globin promoter construct had shown that the *Myf-5* gene together with the 3' UTR (see Fig. 8C, construct MFGZ) contained all the necessary elements for ventral posterior somite expression. Therefore in the absence of any interference, an equivalent construct employing the hsp68 promoter should produce a very similar pattern of expression.
Indeed, transient transgenic animals obtained with **p12UTR** (introns + UTR construct, see Fig. 19) exhibited the expected ventral somitic expression pattern, closely resembling previous results with MFGZ and other constructs (Summerbell *et al.*, unpublished). For detailed analysis, three independent p12UTR lines were raised. No noticeable differences in the pattern of *LacZ* expression were apparent between the lines except for slight variations in the level of expression which was most intense in the line termed Fritz60. Reporter gene expression in Fritz60 was first visible in ventral cells of the thoracic somites around E9 (arrowhead, Fig. 19A) which is approximately 24 hours after *Myf-5* transcripts can first be detected in the primitive somites before differentiation into sclerotome and dermomyotome (Ott *et al.*, 1991). In subsequent stages the intensity of expression increased rapidly throughout the thoracic somites and became confined to the ventral posterior margin before the 20 somite stage (arrowhead, Fig. 19B).
Additional expression was transiently observed from E10 in the core of the hyoid arch (arrowhead, Fig. 19C') contributing to the facial musculature and the trigeminal ganglion (TG, Fig. 19C), with branches progressively extending into the maxillary and mandibular regions (compare Fig. 19C with D, arrowheads). Transverse sections (Fig. 19J) through the thoracic somites show that β-galactosidase staining is confined to the ventral edge of the dermomyotome and

Fig.17: β-galactosidase expression of transgenic p1 (17A) and (17B), and p2 embryos (17D) at E10.5 stained with X-Gal, and transverse wax sections (17C) and (17E) through the neural tube of animals in 17B and 17D respectively. Maps of the p1 and p2 constructs are also shown. Both the first (p1) and second intron (p2) of *Myf-5* activate *LacZ* expression in the ventral neural tube (arrows in 17C and 17E) but failed to activate reporter gene expression in the ventral somites. The animal in 17A shows additional ectopic expression, see text for details.

p1


IN1F
IN1R
hsp68
LacZ
SV40 pA
Intron1


A
B
C

p2


IN2F
IN2R
hsp68
LacZ
SV40 pA
Intron2

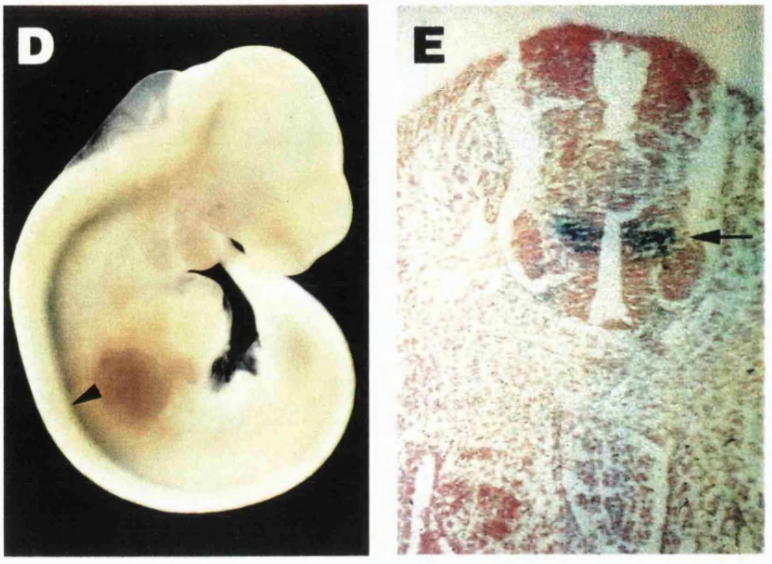
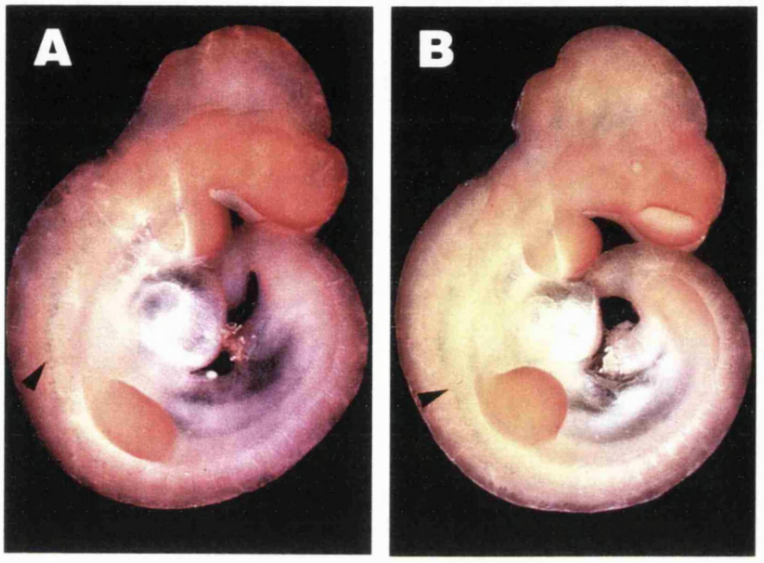
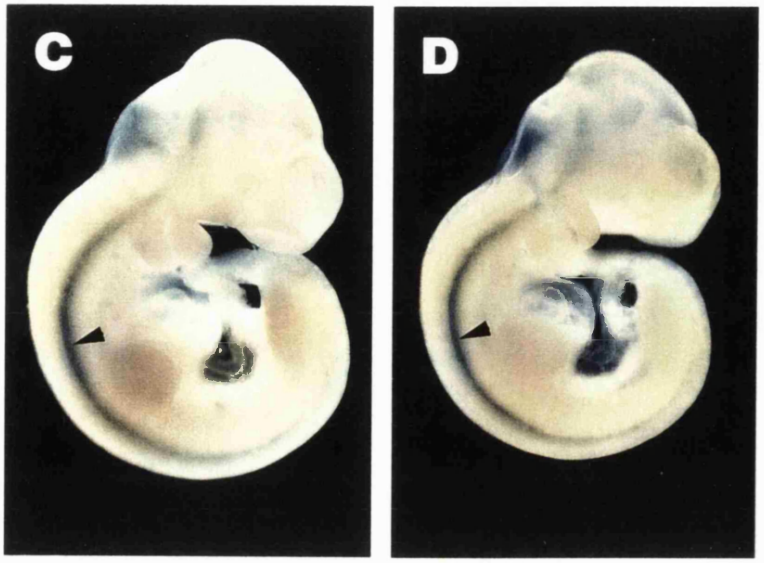

D
E

Fig. 18: Transgenic **pUTR** embryos (18A) and (18B) showed no somitic expression but faint staining in the ventral neural tube (arrowheads), similar to that observed in p1 and p2 transgenic mice (compare Fig. 18A and B with Fig. 17B and D) indicating that elements from both the introns and the 3' UTR may be required for ventral somite expression.
(18C) and (18D): Transgenic **p1UTR** embryos combining intron 1 with the 3' UTR also failed to direct expression to the somites, suggesting elements in intron 2 are necessary for the somitic expression pattern.

pUTR



p1UTR



myotome (Fig. 19H) while coronal section (Fig. 19G) shows that the staining is most intense at the posterior edge (P) of the somites (Fig. 19F). In the mature somites, at the level of the forelimb and anterior to that, *LacZ* expression in the dermomyotome became progressively weaker and increased myotomal staining was observed. After the 30 somite stage the intensity of somitic expression was generally declining although additional staining was noticeable in the somitic bud (sb) posterior to the forelimb at the most ventral margin of the somites (arrow, Fig. 19D). Additional staining in dorsal neural crest cells, the ventral neural tube and the limb buds was also frequently observed at this stage. At E12.5 faint staining was seen in muscle groups ventral to the somite boundary (black arrowhead), and stripes of body wall muscle (arrow) between the limbs and shoulder muscle (white arrowhead) were visible (Fig. 19E). Weak staining was now also observed deep within the hind limb (HL) and in the digits of the forelimb (FL) (Fig. 19E'). No staining was found in the tail somites.

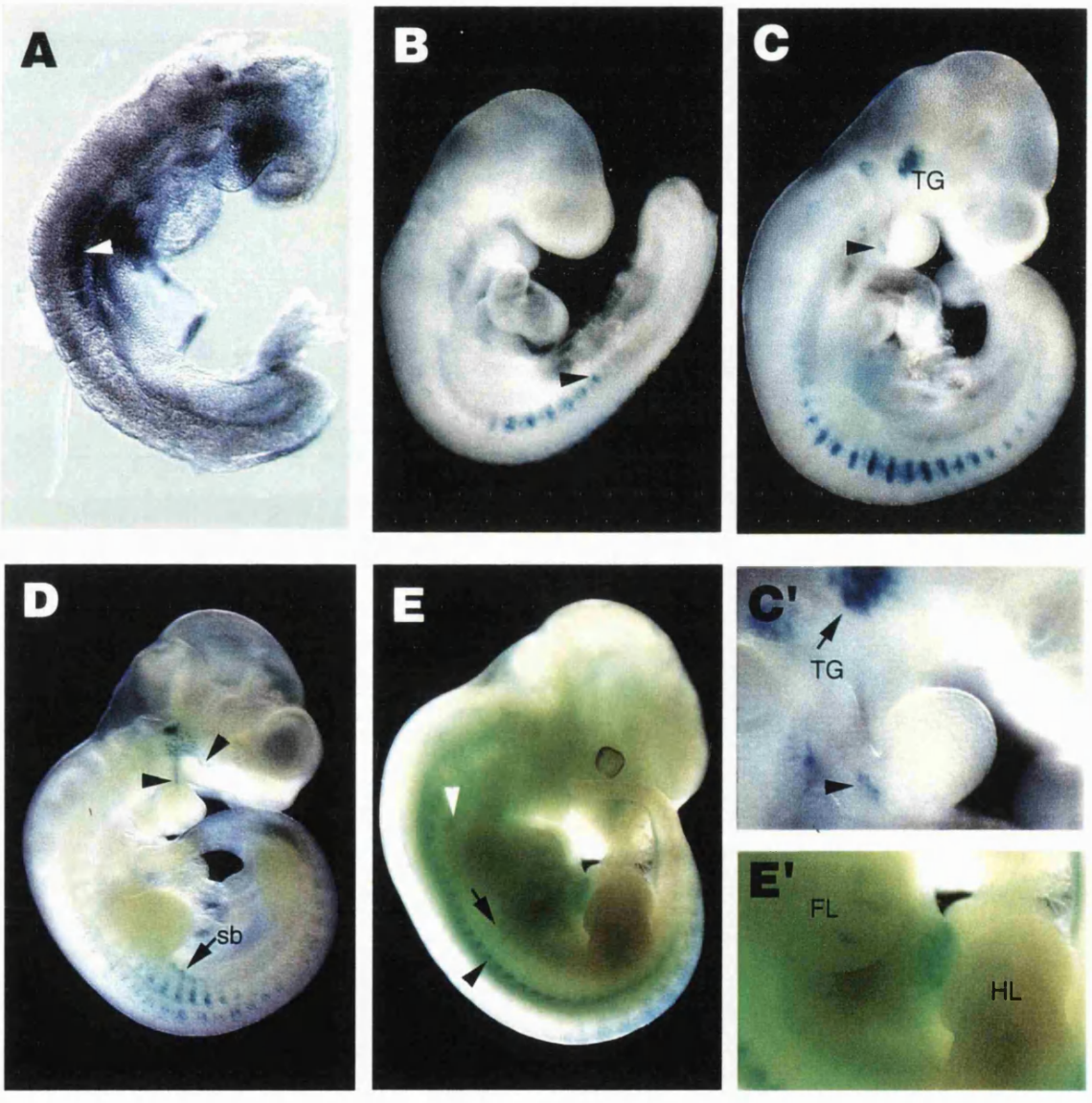
The pattern and efficiency of transgene expression in p12UTR showed that in combination, the *Myf-5* introns together with the 3' UTR can drive expression of the reporter gene from the hsp68 promoter to the ventral posterior margin of the dermomyotome in a pattern closely resembling previous results with MFGZ (Summerbell *et al.*, unpublished). Furthermore, this pattern seems to be controlled by at least two necessary but not sufficient regulatory elements. One or more of these elements appeared to be located in the 3' UTR of *Myf-5* and the other element(s) somewhere in the *Myf-5* introns. To examine whether in combination with the 3' UTR, either intron 1 or intron 2 could recapitulate this pattern, reporter constructs with either intron 1 or intron 2 positioned upstream of the 3' UTR were tested.

p1UTR (intron 1 + 3' UTR construct): Pronuclear injection of p1UTR yielded 6 transgenic embryos of which two expressed the *LacZ* transgene (Fig. 18C and D). Both of these embryos showed staining exclusively in the ventral neural tube in a pattern that appeared to be very similar to that obtained with p1, p2, and pUTR. No other expression domains were observed, suggesting that intron 2 was necessary for the somitic expression pattern obtained with p12UTR.

To test if intron 2 in combination with the 3' UTR was sufficient for correct ventral posterior expression, the first intron was removed from p12UTR to give **p2UTR** (intron 2+UTR construct). Six transient transgenic animals were obtained with *LacZ* expression in the dorsal root ganglion (DRG), and the trigeminal ganglion (TG, Fig. 20). The neural crest staining extended from the level of the cervical somites to the thoracic somites. No staining was observed in the somites of any of the transgenics, suggesting that the deleted first intron contained necessary elements for ventral somitic expression.

Fig. 19: Transgenic embryos of a **p12UTR** line called 'Fritz' stained for β-galactosidase activity at different stages of development.
This line demonstrated that the pattern previousely obtained with MFGZ using the globin promoter (see Fig. 8C) can be faithfully reproduced if not enhanced by using the hsp68 promoter. A map of the construct is shown. The 3' end of the construct is delimited by an *Nhe*I site in the 3' UTR. The embryos obtained with p12UTR show a very consistent expression pattern, which closely resembles MFGZ embryos, with characteristic staining in the ventral posterior margin of the dermo myotome including the somitic bud *(sb)*. Some ectopic staining is observed in the trigeminal ganglion (TG), and the dorsal root ganglia in the older embryos. Temporal expression is shown in embryos at E9 (19A), at E9.5 (19B), at E10 (19C), at E10.5 (19D), at E12.5 (19E).
Fig. 19F shows a coronal section of a Fritz embryo at E10 with staining in the posterior margin (P) of thoracic somites. A Transverse section of a Fritz embryo at E10 is shown in Fig. 19H with ventral somitic staining and ectopic expression in the notochord (NC). Figures 19G and 19J show schematic representations of the structures seen in 19F and 19H. See text for details.

p12UTR - FRITZ



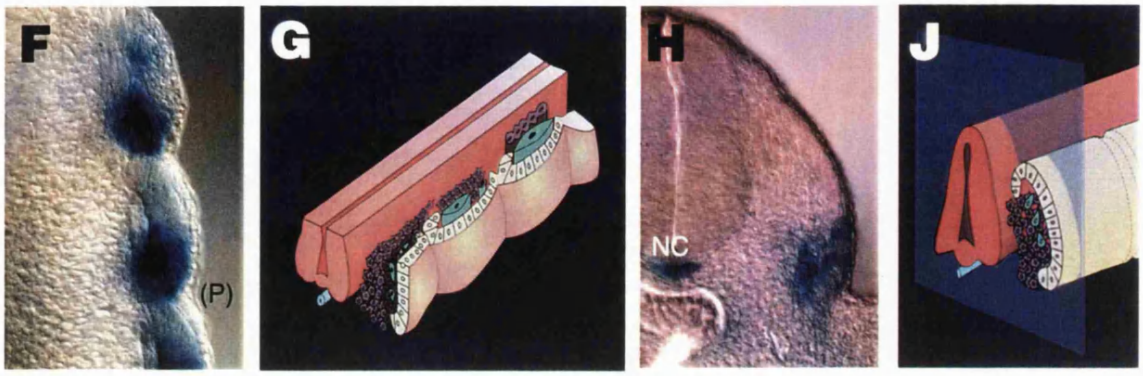
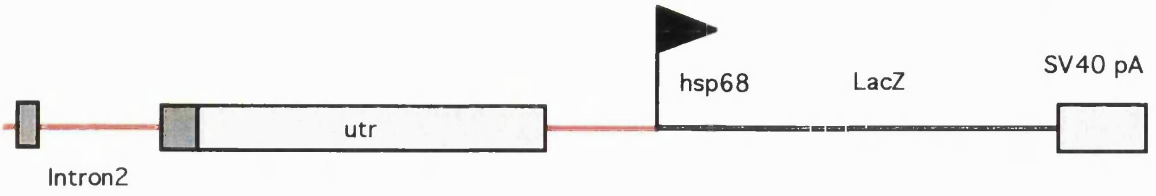
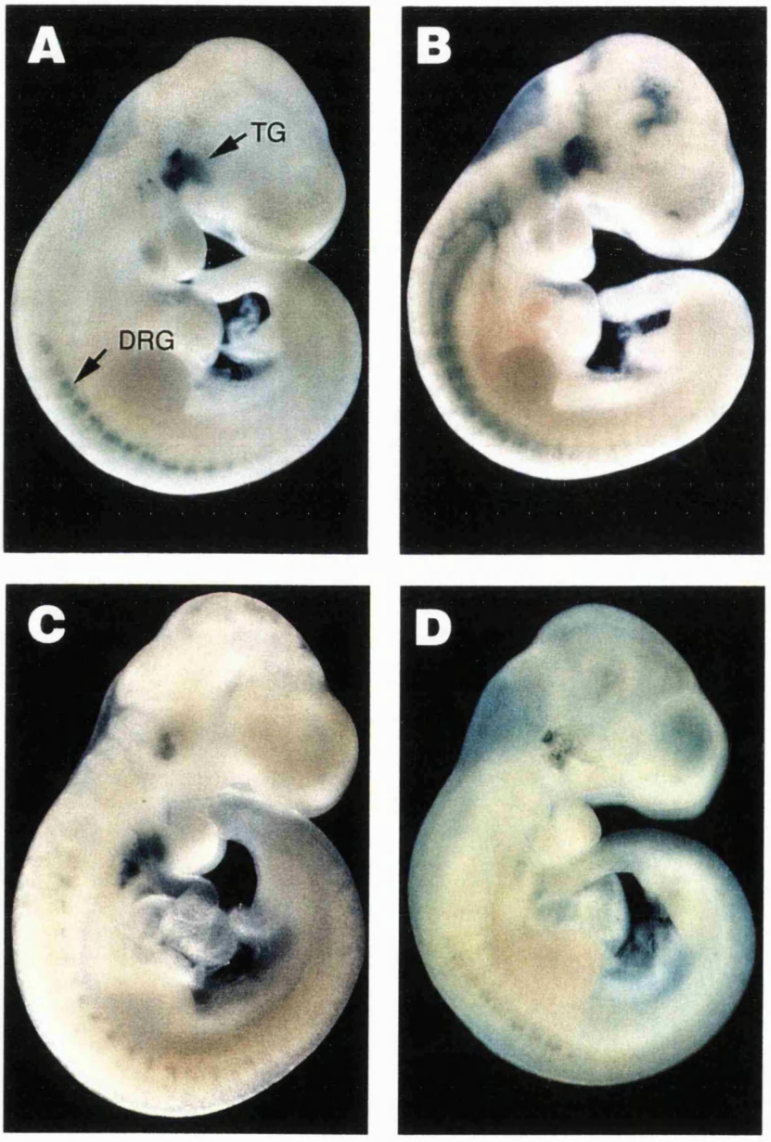


Fig. 20: Transgenic embryos of **p2UTR**, combining intron 2 and the 3' UTR of *Myf-5*. All of the transgenics showed staining in the trigeminal ganglion (TG) and in the dorsal root ganglia (DRG), but failed to activate the reporter gene in the somites.

p2UTR



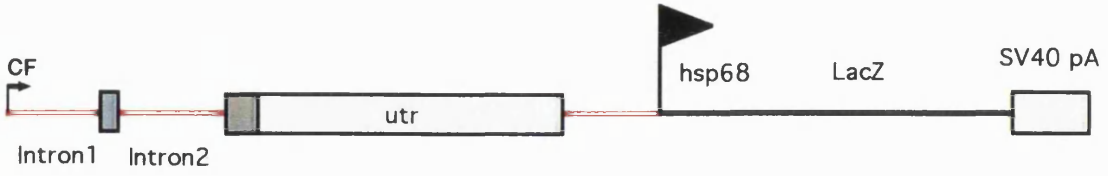


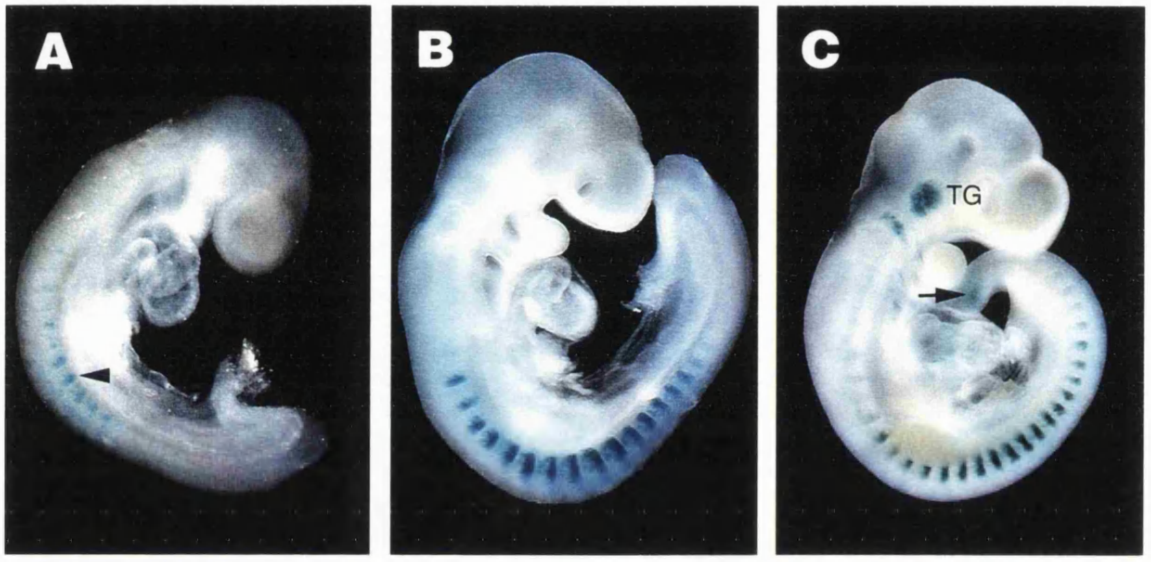
Taken together the results suggested that deletion of either intron or the UTR abolished the ventral somitic expression observed with p12UTR. The most likely explanation is that both introns and the 3' UTR contain necessary but not sufficient elements involved in ventral somitic expression indicating a complex scenario where distinct elements in intron 1 and intron 2 as well as the 3' UTR interact to control expression in the posterior, ventral margin of the dermomyotome.

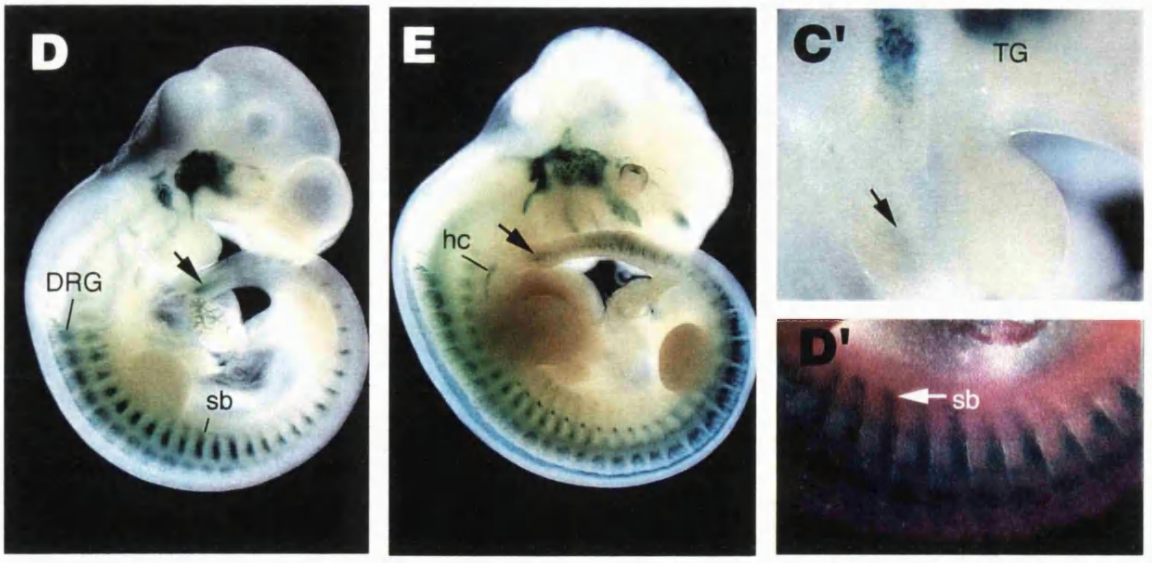
Since neither intron nor the 3' UTR appeared to be dispensable, terminal deletions from the 5' and 3' ends of the p12UTR construct were made to identify the smallest functional regulatory region. Based on the sequence comparisons between the mouse and human *Myf-5* genes that had revealed a high degree of conservation of the 3' half but not the 5' half of each of the two introns, a derivative of p12UTR lacking the poorly conserved 5' half of the first intron was made. 330bp of the 5' end of intron 1 were deleted from p12UTR to give **pCFUTRR**(see Fig. 21). Three independent transgenic lines were generated by pronuclear injection. Although small differences in the levels of expression were observed, the pattern of expression was very similar in these lines. The line shown in Fig. 21 (CFUTRR221) is one of the strongest expressing, and the pattern resembles closely that of p12UTR. However, the levels of expression were increased and extended to more anterior and posterior somites at all stages (compare panels A to E in Fig. 21 and Fig. 19). Staining in CFUTRR embryos was first observed at E9 in the ventral posterior domain of the thoracic somites (arrowhead, Fig. 21A) and was clearly more intense than in p12UTR transgenic mice (compare Fig. 19A and Fig. 21A). Over the next 12-24 hours, expression levels increased further, until uniformly strong staining was seen at around E9.5 (18 somites) in the ventral posterior quadrant of all but the youngest 2 to 3 somites (Fig. 21B). The most intense expression remained at the ventral posterior margin of the dermomyotome of the thoracic somites but additional faint staining could also be seen across nearly the entire width and length of thoracic somites. Transverse sections show that at around E10.5 *LacZ* expression was not only observed in the dermomyotome (DM) but increasingly found in the myotome (M) of the thoracic somites (Fig. 21F) confined to the posterior edge (P) of the somite boundary (Fig. 21G). In contrast to p12UTR embryos, from around E10 (24 somites) *lacZ* expression was seen in all of the somites along the AP axis (compare panels C to E in Fig. 21 and 19) and both CFUTRR and p12UTR embryos shared similar staining in the trigeminal ganglion (TG) and the second branchial arch. This agrees with the observations in *Myf-5/lacZ* knock-in mice that expression in this arch is more intense than in the other arches (Tajbakhsh *et al.,* 1996). Although generally reporter gene expression in CFUTRR221 animals was strong, staining in the arches was weaker than in other CFUTRR lines and less intense than in p12UTR (compare panel C' in Fig. 21 with that in Fig. 19). However, in one of the other CFUTRR lines expression in both the second and the third arch was observed

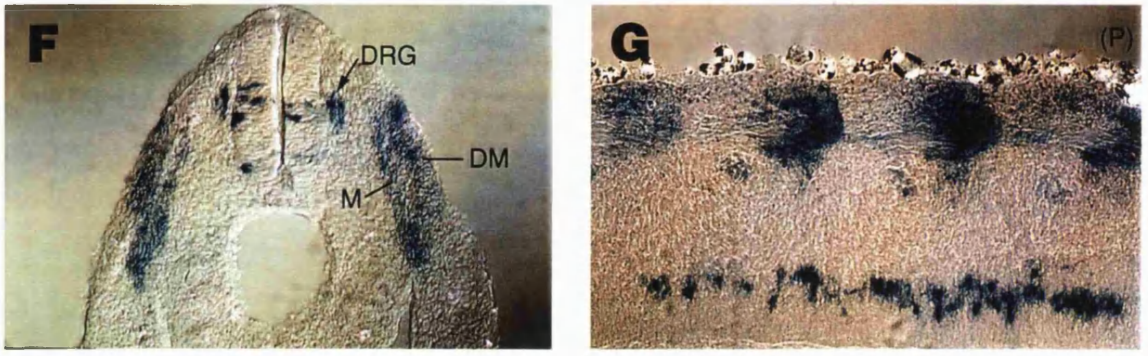
Fig. 21: Transgenic embryos of a **pCFUTRR**- line (221) stained for β- galactosidase activity at different stages of development. For this construct 330 bp of the nonconserved 5' end of intron 1 were deleted from p12UTR without compromising the the pattern of expression compared with the parent construct p12UTR (Fig. 19). The temporal sequence of expression is shown in embryos: at E9 (21A), at E9.5 (21B), at E10 (21C), at E10.5 (21D) and at E12 (21E) . Somitic staining is first seen at E9 and increases in intensity in the ventral posterior compartment. From E10.5 somitic staining between the limbs can be seen in a closeup in the ventral somitic bud (sb) (see 21D'). Additional staining is found in the dorsal root ganglia (DRG), the trigeminal ganglia (TG) , the arches (see arrow in 21C' and at E12 hypoglossal chord (hc). Fig. 21F shows a transverse section at E10 with ventral somitic staining not only in the dermomyotome (DM) but increasingly found in the myotome (M) of the thoracic somites. Fig. 21G is a coronal section at E10 showing staining in the posterior margin (P) of thoracic somites. See text for details.

pCFUTRR


CF
Intron1
Intron2
utr
hsp68
LacZ
SV40 pA


A
B
C
TG


D
DRG
sb
E
hc
C'
TG
D'
sb


F
DRG
DM
M
G
(P)

(not shown). An unusual feature of CFUTRR221 was that *LacZ* expression extended into the presegmental region of the tail from E10 onwards (arrow, Fig. 21C, D, E). However, none of the other CFUTRR-lines showed similar expression in the presomitic mesoderm, indicating that this is likely to result from interactions with elements outside of the transgene at the integration site. From E10.5 somitic staining between the limbs can be seen in the ventral somitic bud (sb) (Fig. 21D') similar to that observed in p12UTR embryos (Fig. 19D). Interestingly *LacZ* expression in the somitic bud coincides with the reported pattern of Pax-3 expression at this stage, consistent with a role of Pax-3 in the activation of *Myf-5* in this domain (Tajbakhsh *et al.,* 1997). In contrast with p12UTR, between the 25 and 30 somite stage, CFUTRR embryos developed more pronounced staining in the nervous system, including the dorsal root ganglia (DRG, Fig. 21D and F), the ventral neural tube (Fig. 21F), and the trigeminal ganglion (TG), (compare panels D and E in Fig. 21 and Fig. 19). At E12 staining is also observed in muscle cells of the hypoglossal chord (hc) that contribute to the muscles of the tongue (Fig. 21 E) and has also been described in the *LacZ*-knockin mice by Tajbakhsh *et al.* (1996).

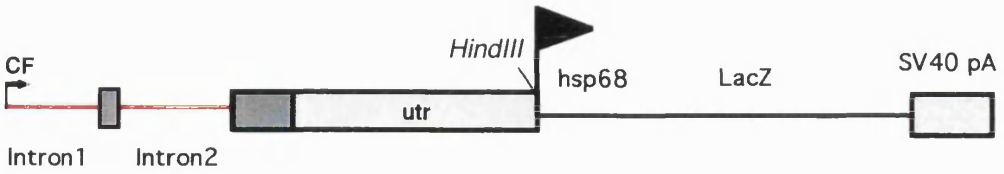
Consistent with the notion that regulatory elements are likely to be evolutionarily conserved, the pattern obtained with CFUTRR demonstrated that using sequence comparisons between mouse and human, nonconserved regions from the first intron of *Myf-5* could be identified and removed without affecting the key features of transgene expression compared with p12UTRR embryos, including expression in the ventral posterior dermomyotome (and myotome) of the thoracic somites, the branchial arches and trigeminal ganglion. To narrow down further the regulatory region from the 3' end, 467 bp of the untranslated region were removed by digestion at a *Hin*dIII site in the 3' UTR to yield **pCFH**. Figures 22A and B show that the expression pattern of two transient transgenic animals obtained with this construct was similar to that of CFUTRR (compare with Fig. 21C) with comparable expression in the ventral posterior domain of the thoracic somites (arrow, Fig. 22A"). Both transgenic mice also show staining in the trigeminal ganglion (TG) and in the core of the first and second branchial arch (double headed arrows, Fig. 22A' and B). Although both embryos show ventral posterior expression in the thoracic somites, the animal in Fig. 22B has expression in most of the somites along the AP axis and additional staining in neural crest cells, the neural tube and the forelimb (fl) and hind limb (hl) buds, probably as an integration site effect.

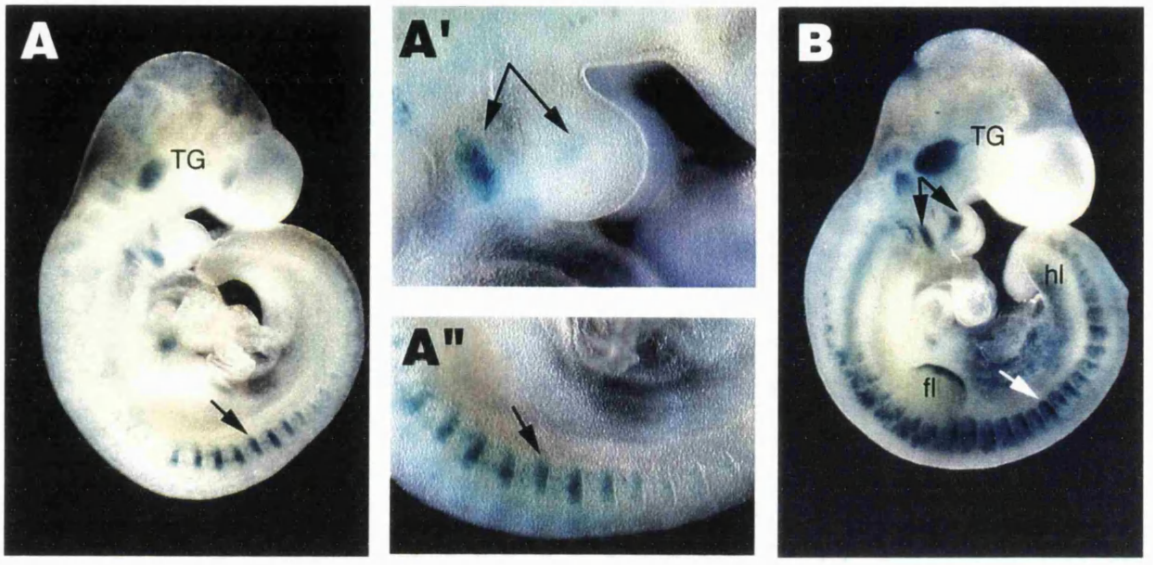
The results show that the region downstream of the *Hin*dIII site in the 3' UTR did not carry essential elements and deletion appeared not to compromise the expression pattern of the reporter gene. To test if further 3' deletions could be made while retaining the expression pattern, approximately 700bp were removed from the 3' end by digestion at

Fig. 22: Transient transgenic embryos of **pCFH** stained with X-Gal show that the region downstream of the *Hin*dIII site in the 3' UTR is not required for ventral somitic expression. pCFH is a derivative of pCFUTRR removing some 500 bp of the 3' end of the 3' UTR but retaining similar ventral somitic expression, compare 22A and 22B with Fig. 21C. A close-up is seen in 22A" showing ventral posterior expression in the thoracic somites (arrows in 22A" and 22B). Additional staining includes the trigeminal ganglion (TG) and the first and second branchial arch (double headed arrows in 22A' and 22B).

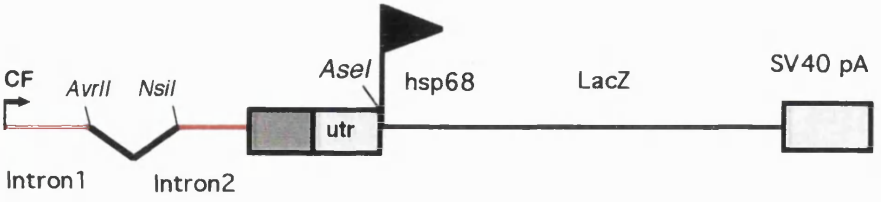
Figures 22C and 22D: Additional 700 bp were removed from the 3' end by digestion at a unique *Ase*I site in the 3' UTR to give **pCFA**, and resulted in patchy expression in isolated cells of the dermomyotome (arrow heads) although still at the ventral posterior margin of the thoracic somites, arrowheads in close-ups 22C' and 22D'. Thus, separate regulatory elements might control the intensity and spatial distribution of transgene expression, the former located in the *AseI*/*Hin*dIII fragment of the 3' UTR, the latter most likely in the introns of the *Myf-5* gene.

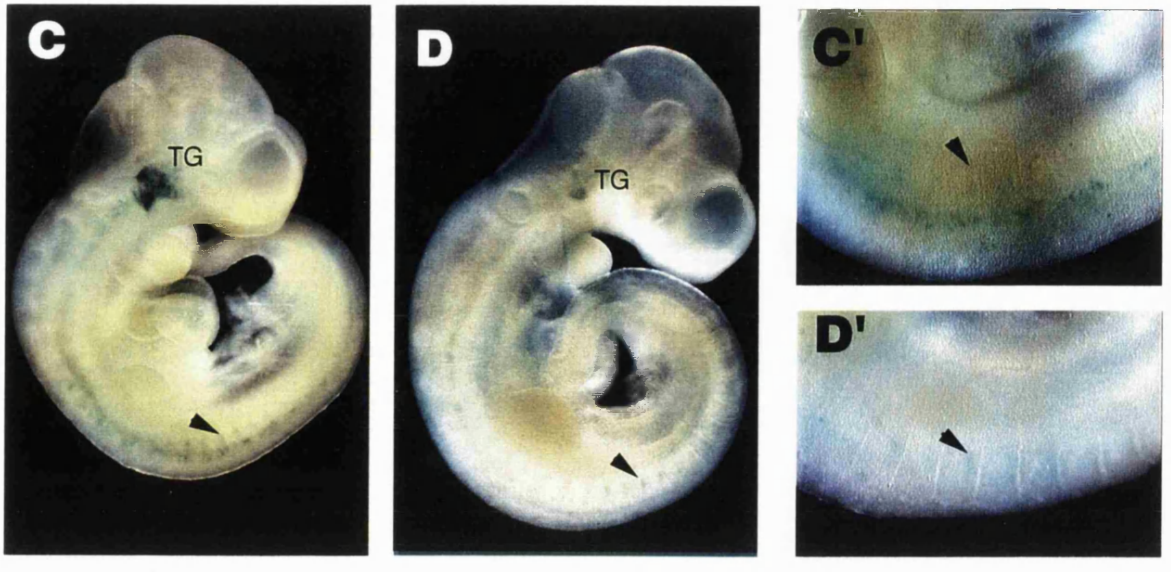
pCFH


CF
HindIII
hsp68
LacZ
SV40 pA
utr
Intron1
Intron2


A
TG
A'
A"
B
TG
hl
fl

pCFA

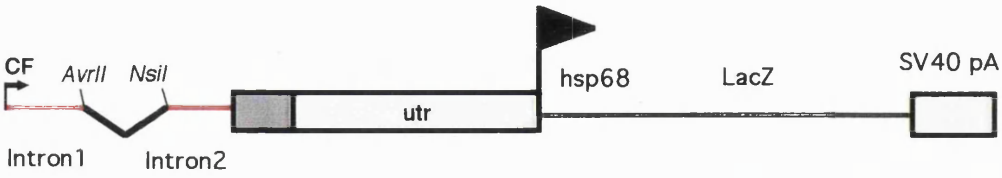

CF
AvrII
NsiI
AseI
hsp68
LacZ
SV40 pA
utr
Intron1
Intron2

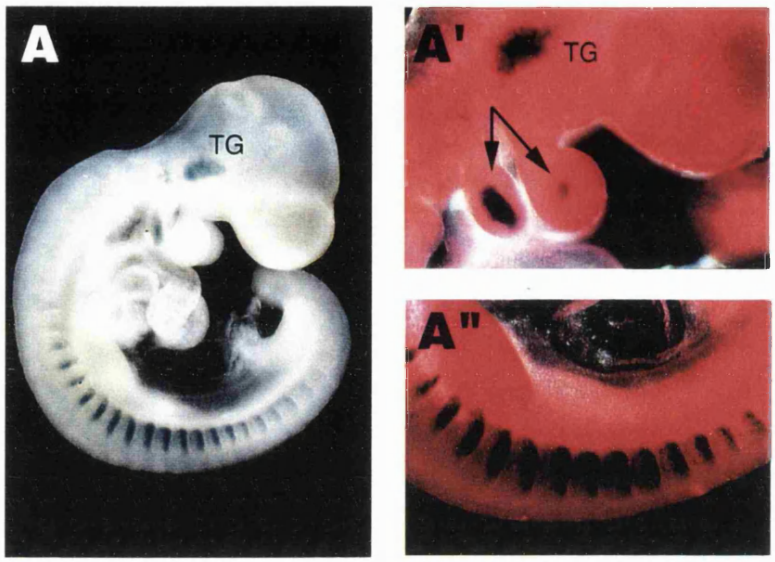

C
TG
D
TG
C'
D'

a unique *Ase*I site in the 3' UTR to give **pCFA**(see Fig. 22 C and D). This construct appeared to have retained the ability to drive *LacZ* expression to the ventral posterior domain of the somites as well as most of the other expression domains seen in p12UTR and CFUTRR embryos, however, the levels of expression were drastically reduced (Fig. 22C, D). Only patchy expression in isolated cells of the dermomyotome (arrow heads) was seen at the ventral posterior margin of the thoracic somites (arrowheads, Fig. 22C' and D'). In agreement with the previous constructs, further faint staining was also observed in the other expression domains, including the ventral neural tube, the second branchial arch and the trigeminal ganglion but the levels of expression were significantly reduced in comparison to pCFH, the parent construct. These results suggested that elements influencing the level of transgene expression are located between the *Ase*I and *Hind*III sites in the 3' UTR of the *Myf-5* gene. When deleted, the levels of reporter gene expression drop dramatically although residual activity in the ventral posterior domain of the somites as well as the other expression domains was retained, indicating that separate regulatory elements might control the intensity and spatial distribution of transgene expression, the former located in the *Ase*I/*Hin*dIII fragment of the 3' UTR, the latter most likely in the introns of the *Myf-5* gene. Following the successful elimination of nonconserved 5' sequence from the first intron, the nonconserved 5' half of the second intron was deleted by removing a 339bp *Avr*II/*Nsi*I fragment from pCFH yielding the construct **pCFHΔ**(see Fig. 23A). The deleted fragment also contained the second exon of the *Myf-5* gene and removed all of the nonconserved sequence from the introns. As expected, deletion of the nonconserved 5' half of the second intron in pCFHΔ did not compromise the pattern of expression of the transgene, which was remarkably similar to that of pCFH its parent construct (compare Fig. 23A with Fig. 22A), and also resembled that of Fritz and CFUTRR (compare with Fig. 19 and Fig. 21). Intense expression in the ventral posterior dermomyotome was observed (Fig. 23A"), with additional expression in the trigeminal ganglion (TG) and the first and second branchial arches (Fig. 23A', double headed arrow), similar to pCFH embryos. Thus, the nonconserved regions of both of the introns and the intervening second exon are not required for the correct expression of the transgene. In other words, the 3' half of both introns in combination with the 3' UTR is sufficient to drive reporter gene expression to the correct ventral posterior somite domain and includes expression domains for the trigeminal ganglion and the branchial arches. If the role of the conserved intron sequences is to direct the transgene to the ventral posterior dermomyotome, and the role of the *Ase*I-*Hin*dIII fragment of the 3' UTR is to increase the general level of expression, then the region upstream of the *Ase*I site up to the second intron may not be required. To examine this possibility, this region was deleted from pCFHΔ to yield **pΔAH** (see Fig. 24A and B).

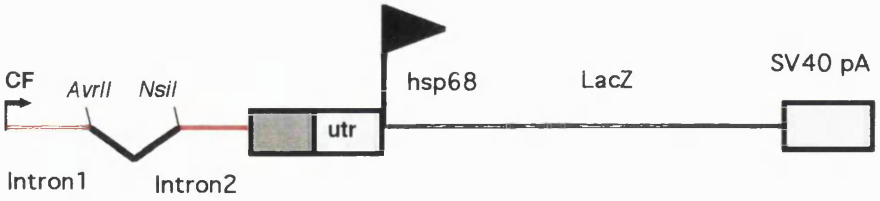
Fig. 23: Transient transgenic embryo of **pCFHΔ** stained with X-Gal showing that the second exon of *Myf-5* and the nonconserved regions between the *Avr*II and *Nsi*I sites in the introns are not required for ventral somitic expression. Intense expression in the ventral posterior dermomyotome is seen in pCFHΔ similar to pCFH its parent construct, compare 23A" with Fig. 22A. Expression in the trigeminal ganglion (TG) and branchial arches, double headed arrow in (23A') has also been retained.
Figures 23B and 23C: **pCFAΔ**.is a deletion variant of pCFHΔ terminating at the *Ase*I site in the 3'UTR. Residual expression was found in the ventral posterior margin of the dermomyotome of most of the thoracic somites, (arrowheads in close-ups 23B' and 23C'), suggesting that an efficiency element (downstream of the *Ase*I site) had been lost but that the somite control elements in the introns had been retained.

pCFHΔ


CF
AvrII
NsiI
hsp68
LacZ
SV40 pA
utr
Intron1
Intron2


A
TG
A'
TG
A"

pCFAΔ


CF
AvrII
NsiI
hsp68
LacZ
SV40 pA
utr
Intron1
Intron2

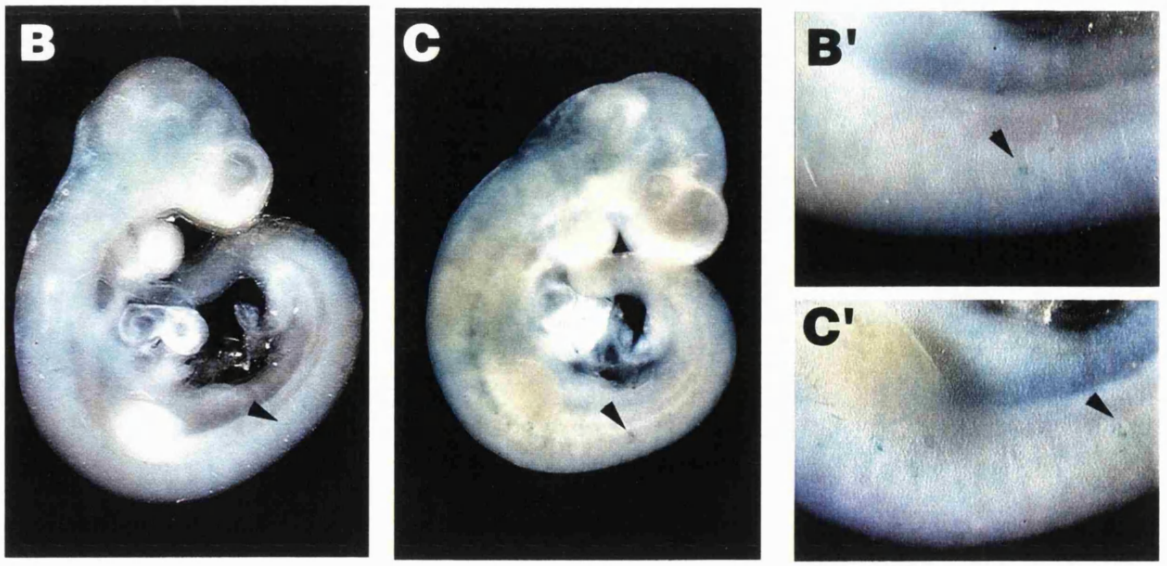
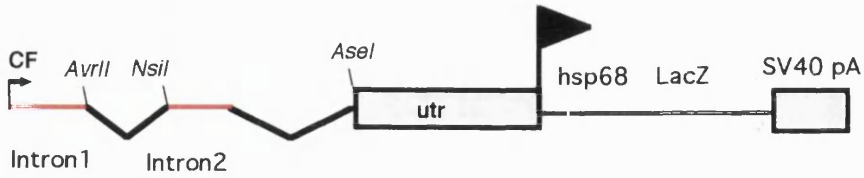
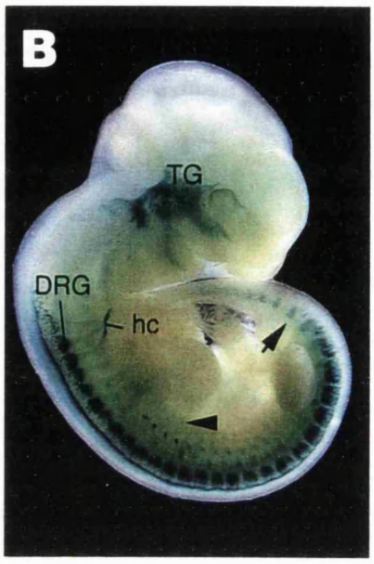
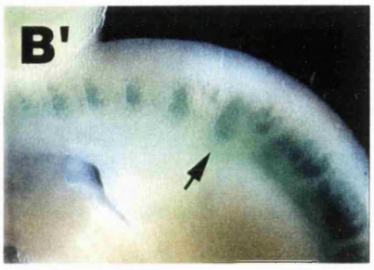

B
C
B'
C'

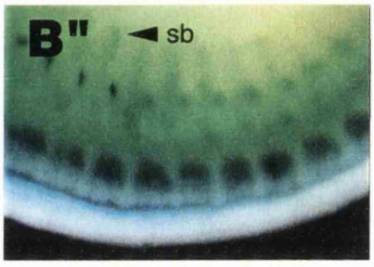
Fig. 24: Transient transgenic embryos of **pΔAH** stained with X-Gal.The 5' end of the 3' UTR was deleted, while retaining the intron elements as in pCFHΔ. The expression pattern obtained was inconsistent, compare 24A and 24B.
While 24B showed expression in the ventral somitic bud (sb in 24B") the ventral posterior dermomyotome, arrow in 24B', dorsal root ganglia (DRG), trigeminal ganglion (TG) and the hypoglossal chord (hc), compare 24B with Fig. 21E; this result was not reproducible. A more typical result is shown in (24A).
Figures 24C and 24D: pΔCAH and **pΔGAH** are further deletion derivatives of pΔAH lacking the conserved fragments C and G of the first and second intron respectively. Neither construct produced a somitic expression pattern.(24C) pΔCAH retained the expression in the branchial arches (arrowhead) the dorsal root ganglia (DGR) and trigeminal ganglion (TG) . pΔGAH only showed faint ventral neural tube staining (arrowhead, 24D).
The lack of reproducible somitic expression in any of these constructs suggests that elements upstream of the *Ase*I site in the 3' UTR are required.

pΔAH

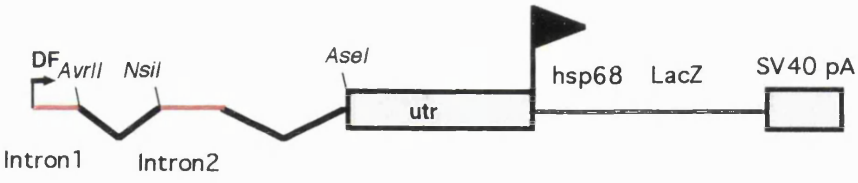




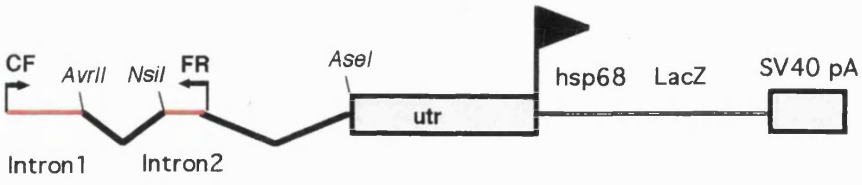


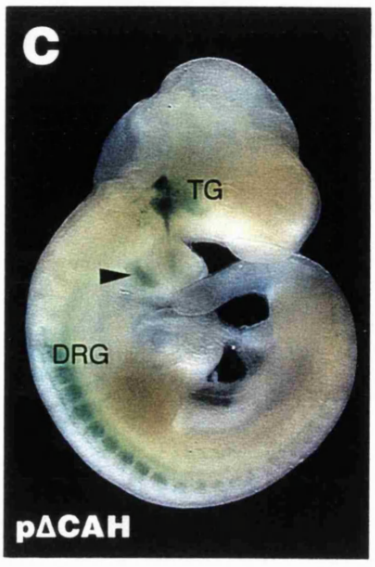


pΔCAH



pΔGAH





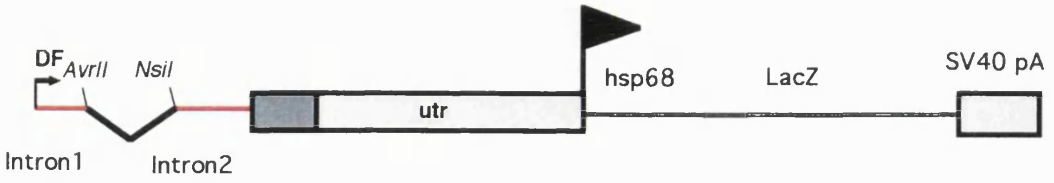
The first of the transient transgenic animals obtained by pronuclear injection showed the expected expression pattern in the ventral somitic bud (sb) (Fig. 24B") at the level of the thoracic somites as well as expression in the ventral posterior dermomyotome of more posterior somites (arrow, Fig. 24B'). In complete agreement with similar stage CFUTRR embryos, staining was seen in the dorsal root ganglion (DRG), trigeminal ganglion (TG) and the hypoglossal chord (hc) (compare Fig. 24B with Fig. 21E). On the basis of this result two derivatives **pΔCAH** and **pΔGAH** with additional deletions removing the conserved fragments C and G (see Fig. 11A and B) from the first and second intron respectively, were tested and at the same time two more transgenics were generated with pΔAH. Figures 24A, C and D show one of two expressing transgenics obtained for each construct, pΔAH, pΔCAH and pΔGAH. Surprisingly, none of these showed somitic expression. Interestingly, pΔCAH retained the expression in the branchial arches (arrowhead), the dorsal root ganglia (DGR) and trigeminal ganglion (TG) (Fig. 24C), whereas expression of pΔAH and pΔGAH constructs was only seen in the ventral neural tube (arrowhead, Fig. 24D) or the trigeminal ganglion (TG, Fig. 24A), respectively. In total, 16 transgenic animals were generated with these three constructs and only the one in Fig. 24B expressed the transgene in the expected pattern in the somites, probably by co-opting enhancer elements at the integration site. This would mean that vital elements had been deleted by removing the 3' UTR region upstream of the *Ase*I site, and is consistent with the proposed role of the 3' UTR as a general element involved in regulating the intensity of expression but not its spatial pattern. To test the role of the deleted 3' UTR region, a derivative of pCFHΔ was generated in which the 3' UTR sequence downstream of the *Ase*I site had been deleted to give **pCFAΔ**. Although this construct showed drastically reduced expression levels comparable with pCFA before, it was encouraging to find that the residual expression was found in the ventral posterior margin of the dermomyotome of most of the thoracic somites (Fig. 23B' and C'). This was consistent with the results of pCFA and the idea that an efficiency element (downstream of the *Ase*I site) had been lost but that the somite control elements in the introns had been retained. Some additional faint expression was observed in the ventral neural tube and neural crest cells (Fig. 23C). In combination, the results of deleting the two halves of the 3' UTR - upstream of the *Ase*I site (pCFA and pCFAΔ) and downstream of the *Ase*I site (pΔAH), indicated that both the 5' and 3' halves of the untranslated region are required for the efficient expression of the transgene. Deletion of the downstream half resulted in drastically reduced expression levels, whereas deletion of the upstream half abolished all somitic expression. Therefore the downstream half appears to contain an efficiency element that controls the levels of expression. This raises the question as to what function the upstream half of the 3' UTR has, and why its deletion compromises all expression? The most likely explanation is that the spacing between the somite control elements in the introns and the efficiency

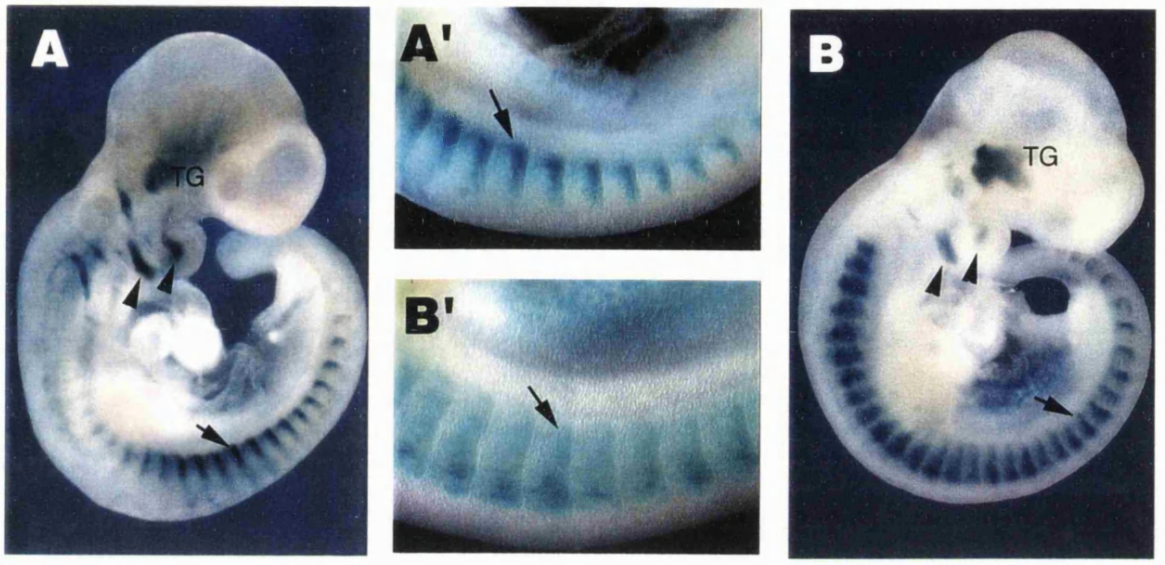
element in the UTR is critical. This could easily be tested by substituting the upstream half of the 3' UTR with an unrelated piece of DNA. However, further studies concentrated on the task of characterising the intronic regulatory elements in more detail. To delimit more closely the functional elements in the introns, the 5' and the 3' halves of the remaining conserved region of both introns was individually deleted from pCFHΔ. Deletion of fragment C (see dot plot analysis, Fig. 13A) of the conserved region of intron 1 yielded **pΔCPH** (see Fig. 25A). Transient transgenics expressing this construct showed intense ventral posterior staining in the somites (arrow) resembling the pattern of its parent construct pCFHΔ (compare Fig. 25A' and B' with Fig. 23A"), suggesting that the deleted fragment was not required for the ventral posterior somite pattern, despite the apparent evolutionary conservation in mouse and human, and the ability to bind proteins from embryonic extracts in EMSA (see Fig. 15C). As before staining was also observed in the first and second branchial arches (Fig. 25A, B arrowheads) and the trigeminal ganglion (TG), indicating that the elements driving expression in these domains have also been retained in pΔCPH. Together with the data showing that elements of both introns and the UTR are required for correct transgene expression, the results indicate that at least one of the necessary elements is located at the 3' end of intron 1 in the conserved region of fragment D (Fig. 13A). More detailed analysis of this fragment by EMSA had identified the most likely position of the protein binding site(s) in the 110 bp at the 5' end (Fig. 15D) narrowing down the candidate region even further.

Using a similar approach to delimit the relevant regions of the remaining conserved sequence of intron 2, the 5' half corresponding to bandshift fragment F (see Fig. 15F and Fig. 11B) was deleted from the context of pΔCPH to give **pΔFPH**. Transient transgenics expressing this construct showed only faint ventral neural tube expression (arrow) but no somitic expression pattern (Fig. 25 C and D), suggesting that the deleted fragment F contained regulatory elements necessary for ventral somitic expression. Similar results were obtained when the 3' half comprising fragment G (Fig. 11B and Fig. 15G) was deleted from pΔCPH to give **pΔGPH**. Only ventral neural tube expression was observed (Fig. 26A and B, arrow heads) suggesting that in agreement with their evolutionary conservation between mouse and human, and their ability to bind proteins in EMSA, both fragments F and G of the second intron harbour essential regulatory elements.

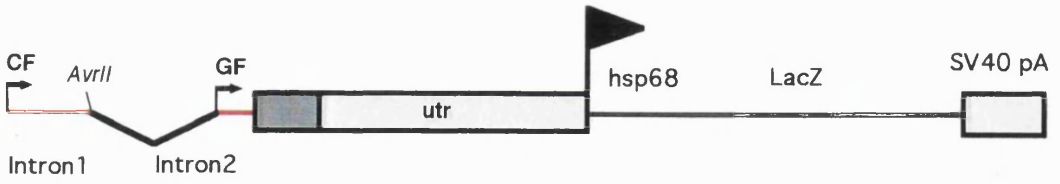
Fig. 25: Transgenic embryos of **pΔCPH** showing that fragment C of the conserved region of intron 1 is not involved in the ventral somitic expression domain.
Figures 25A' and 25B': Closeup of the thoracic somites showing intense ventral posterior staining (arrows) resembling the pattern of its parent construct pCFHΔ (compare with Fig. 23A"). Figures 25A and 25B show that expression was retained in the first and second branchial arches (arrowheads) and the trigeminal ganglion (TG).
Figures 25C and 25D: Transient transgenic embryos of **pΔFPH** showing the effect of deleting fragment F from the second intron (see Fig. 13B). Only faint ventral neural tube expression (arrow) but no somitic expression was observed, suggesting that the deleted fragment F contained regulatory elements necessary for ventral somitic expression. Ectopic expression in the ventral neural tube was observed (arrows).

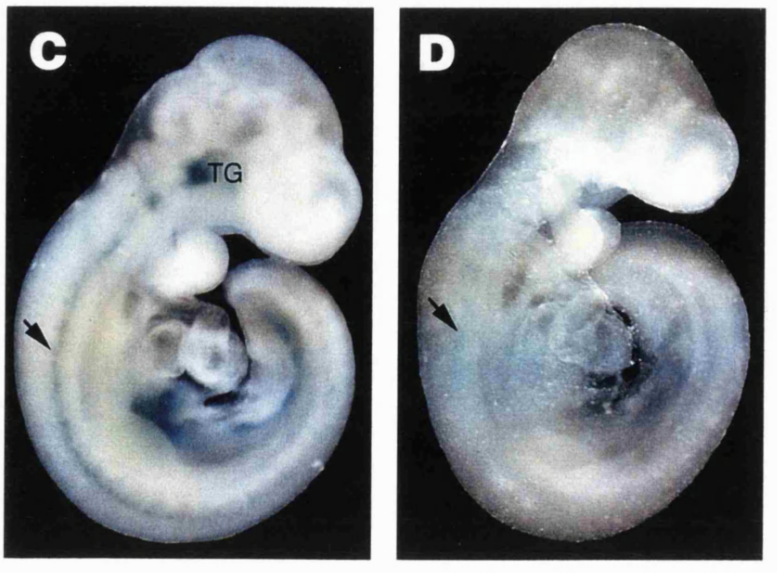
pΔCPH


DF
AvrII
NsiI
hsp68
LacZ
SV40 pA
utr
Intron1
Intron2

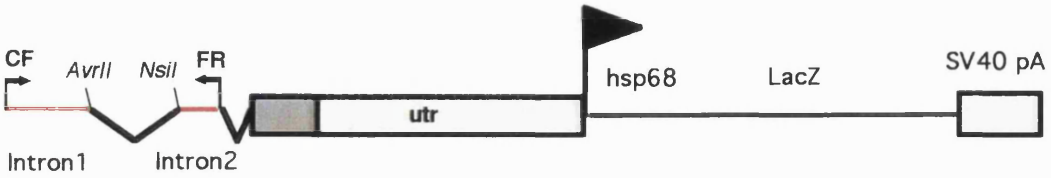

A
TG
A'
B'
B
TG

pΔFPH


CF
AvrII
GF
hsp68
LacZ
SV40 pA
utr
Intron1
Intron2


C
TG
D

pΔGPH



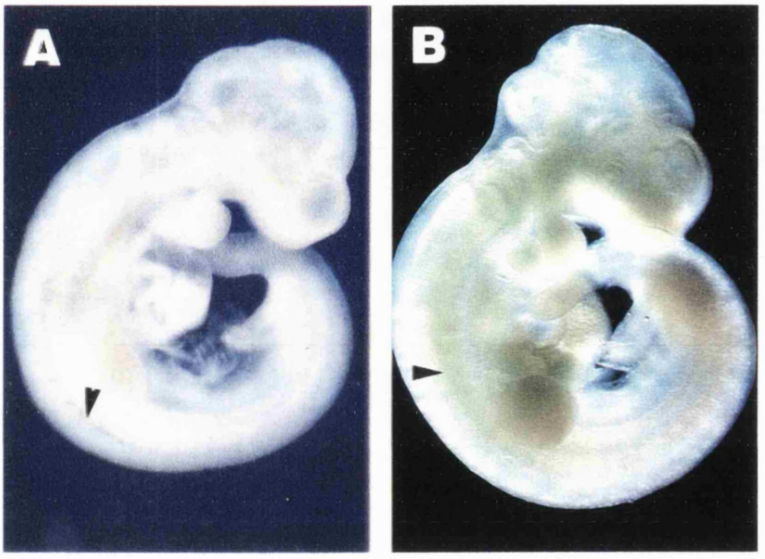


Figure 26

A and B: Transient transgenic embryos of pΔGPH showing the effect of deleting fragment G from the first intron (Fig. 13B). Only faint ventral neural tube expression (arrow) but no somitic expression is observed, suggesting that the deleted fragment G contained regulatory elements necessary for ventral somitic expression. Ectopic expression in the ventral neural tube was also observed (arrowheads).

Reporter Gene Expression from Transgenic Lines

Most of the transgenic data presented derived from transient transgenic animals which were generated by pronuclear injection and sacrificed for analysis. However, at several stages it was necessary to obtain more detailed information regarding the dynamic nature or histology of the expression pattern. For this purpose, transgenic lines were raised for some of the crucial constructs by mating transgenic animals identified by tail diagnosis with wild type partners. The offspring was analysed for transgene expression as described (Chapter 2, Section 10.3.). Homozygous transgenic animals are currently being raised by interbreeding germline transmitters in order to maintain these lines for further studies.

In total, five lines were raised representing four different constructs: p12UTR, pCFUTRR (2 separate lines), pCFHΔ, and pZFF. The pattern of transgene expression in all of these lines is generally an accurate reflection of the pattern illustrated for the transient transgenic animals.

The expression pattern of the line termed **Fritz (p12UTR)** is as presented and discussed in this chapter (see Fig. 19 and text) and represented a starting point for the subsequent deletion constructs including pCFUTRR. Two lines were obtained with this construct and the pattern for line 221 is shown in Fig. 21. The second **pCFUTRR** line has a very similar expression pattern to line 221 except that the staining in the tip of the tail (arrows in Fig. 21 C, D, E) is absent from similar stage embryos of this line (data not shown). In comparison with Fritz embryos, both pCFUTRR lines are more strongly expressing in the somites but also have more prominent expression in the trigeminal and dorsal root ganglia, especially in older animals (around E12). The **pCFHΔ** line shows a pattern similar to that presented in this chapter (see Fig. 23), including *LacZ* expression in the ventral posterior domains of the somites, the arches and in the trigeminal ganglion as described. However, the expression levels of this line are weaker than those presented for the transient transgenic animal in Fig. 23. More detailed analysis of pCFHΔ animals is in progress.

The **ZFF** line harbours a hybrid transgene, consisting of the zebrafish introns and the *Fugu* UTR of *Myf-5* upstream of the *hsp68* promoter driving *lacZ* (see Chapter 6, Fig. 33). The line animals show a similar expression pattern to that presented in Fig. 33. The typical expression in the tail mesoderm is observed in the line animals although somewhat variable staining is seen in the skin which was also observed in the transient animals.

Chapter 6

Myf-5 in Zebrafish and *Fugu*

1. Introduction

The apparent absence of sequence conservation in the MRF4 /Myf-5 region between mouse and *Fugu* raised questions about the role of *Myf-5* as a myogenic factor in fish, especially since *Myf-5* homologues from other teleost fish species have not previously been studied. It would therefore be important to examine if the expression pattern of *Myf-5* in teleost fish was consistent with a role in myogenesis and if *Myf-5* would be expressed in the paraxial mesoderm. Since *Fugu* embryos are not readily available for such studies, a related species was used to study the *Myf-5* expression pattern by *in situ* hybridisation. The *Myf-5* homologue of the zebrafish, *Danio rerio* was cloned and its expression pattern was examined in zebrafish embryos. Pairwise sequence comparisons between the genomic sequence of mouse and zebrafish *Myf-5* were performed with the aim of identifying conserved blocks of sequence. In addition, the *Myf-5* genes of the two teleost species, *Fugu* and zebrafish, were compared to investigate if conserved fish specific regulatory elements can be found.

2. Cloning of the Zebrafish *Myf-5* homologue

A nested set of degenerate oligonucleotides Zebra1 and Zebra2 (see Appendix I) derived from the bHLH domain of the first exon of *Myf-5* was designed on the basis of sequence conservation between *Myf-5* and other vertebrate homologues and paired in low stringency RT-PCR with *Fugu* Myf-5 Rev primer (see Chapter 2, section 8 for details). The first round products were re-amplified in a second PCR using the nested Zebra2 primer and Myf-5-Rev (see Appendix I for primer sequence). The PCR products were cloned and sequence analysis revealed about 70% homology with both mouse and Japanese quail *Myf-5* in two blocks of 50 and 70bp respectively, while 80% identity was found with *Fugu Myf-5* at the nucleotide level and 95% similarity at the amino acid level suggesting that the amplification product was zebrafish *Myf-5*. A full length cDNA

clone was obtained by screening a zebrafish cDNA library in bacteriophage lambda with the PCR product (see Chapter 2, section 7.2 for details). The cDNA insert of the isolated phage clone was sequenced and the intron sequence was determined from genomic zebrafish DNA. The complete nucleotide sequence is presented in Appendix V. The conceptual translation of the coding region was aligned with the *Myf-5* homologues of other vertebrates (see Fig. 11C, Chapter 3). The most likely ATG codon was located 131nt from the 5' end of the phage insert generating an open reading frame of 238 amino acids compared with 240 aa in *Fugu*. The zebrafish *Myf-5* protein (ZMyf-5) closely resembles *Fugu* Myf-5. 70% of the amino acid are identical. Although in general amongst vertebrates the homology at the 5' end of the *Myf-5* gene breaks down, strong similarity is maintained between the two fish species including a gap of 12-15 amino acids near the N-terminus compared with the other vertebrate species (see Fig. 11C, Chapter 3). The bHLH domain of *ZMyf-5* is nearly identical to that of *Fugu* and very similar to that of the other vertebrate species examined, underlining its conserved role in heterodimerisation and DNA binding. Based on the high degree of sequence homology with the *Myf-5* homologues the gene was termed *ZMyf-5* and its expression pattern in zebrafish was determined by *in situ* hybridisation to zebrafish embryos.

3. Expression of *ZMyf-5* in zebrafish embryos

To examine the spatial and temporal pattern of *ZMyf-5* expression, *in situ* hybridisations were performed on whole zebrafish embryos using digoxigenin labelled *ZMyf-5* and *ZMyoD* (a kind gift of Derek L. Stemple, Division of Developmental Biology, NIMR) probes. Consistent with *ZMyf-5* representing the zebrafish homologue of the myogenic factor *Myf-5*, transcripts were found in the paraxial mesoderm and the developing somites of zebrafish embryos. Similar to frogs, but in contrast to mouse and chick, expression of both *ZMyf-5* and *ZMyoD* was found also in the presomitic mesoderm and thus precedes myogenic differentiation. The expression pattern of *ZMyf-5* is shown in Figure 27.

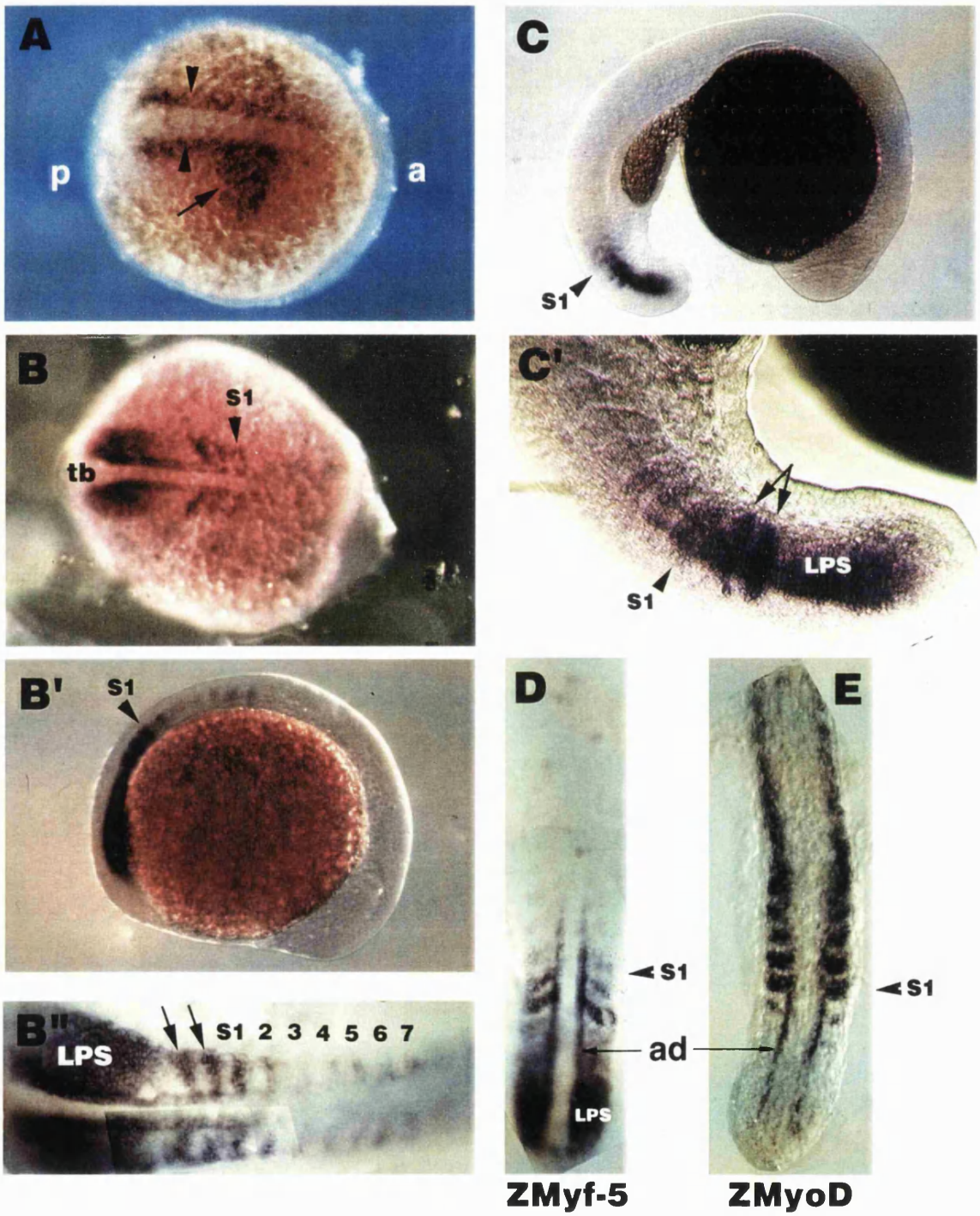
In cleavage and blastula stage embryos no *ZMyf-5* or *ZMyoD* transcripts were detected. The first *ZMyf-5* transcripts were observed in the dorsal hypoblast of gastrula stage embryos at about 80% epiboly before the embryonic shield begins to elongate along the anteroposterior axis (Kimmel *et al.,* 1993). At 90% epiboly, *ZMyf-5* transcripts were found in two stripes (arrowheads, Fig. 27A) on each side of the embryonic shield including the prospective adaxial cells adjacent to the notochord which also expressed *ZMyoD* (see Fig. 27A and Weinberg *et al.,* 1996 Fig. 2). In addition, *ZMyf-5,* unlike *ZMyoD* was also expressed in broad triangular shaped domains extending laterally from the adaxial cells (arrow, Fig. 27A). These cells later form the lateral presomitic cells

Fig. 27: Expression pattern of *ZMyf-5* in zebrafish embryos

Localisation of *ZMyf-5* (27A-D) and *ZMyoD* transcripts (27E) at different developmental stages of zebrafish embryos. 27A: dorsal view of a presegmentation embryo at 90% epiboly, 27B shows a dorsal view, 27B' a lateral view and 27B" a flatmount at 10-12 hpf (6-7 somites). 27C shows a lateral view and 27C' a close-up phase contrast of the caudal somites and the presegmental plate. Figures 27D and E show flatmounts of 24 hpf (30 somites) embryos probed with *ZMyoD* or *ZMyf-5*, respectively. *ZMyf-5* expressing cells were found in the paraxial mesoderm and developing somites consistent with *ZMyf-5* representing the homologue of the myogenic factor *Myf-5* in zebrafish. *ZMyf-5* transcripts first appear in two stripes (arrowheads, 27A) and in broad triangular shaped domains on on each side of the embryonic shield. At the onset of segmentation, *ZMyf-5* transcripts form lateral projections from the notochord in the first 6-7 somites (S1-S7, 27B, 27B', 27B") and in two additional stripes spaced at segmental interval at the rostral edge of the unsegmented paraxial mesoderm (arrows, in 27B"). Neither *ZMyf-5* (27B) nor *ZMyoD* (27E) are highly expressed in the tail bud (tb).

Figures 27D and E, *ZMyf-5* transcripts are found in the lateral presomitic cells (LPS) adjacent to the adaxial cells (ad) while *ZMyoD* is not. At this stage only the most recently formed somites expressed *ZMyf-5* followed by the two stripes at the rostral edge of the unsegmented presomitic mesoderm (double headed arrow, 27C') and the block of lateral presomitic cells (LPS) that continued to express *ZMyf-5* (27C and D).

ZMyf-5 Expression Pattern



(LPS) of the segmental plate (Fig. 27B and B" and schematic in Fig. 5C, Chapter 1). Prior to somite formation the expression domains of both *ZMyf-5* and *ZMyoD* extend further along the A-P axis and became slightly narrower towards the midline. The anterior boundary of *ZMyf-5* did not seem to extend as far as that of *ZMyoD* at any stage. Neither *ZMyf-5* nor *ZMyoD* were highly expressed in the tail bud (tb) (Fig. 27B and 27E) which develops after closure of the germ ring and gives rise to the tail mesoderm. At the onset of segmentation, (10 to 12 hpf) both *ZMyf-5* and *ZMyoD* transcripts were simultaneously observed as lateral projections from the notochord in the first 6-7 somites (S1-S7) and in two additional stripes of expression spaced at segmental interval at the rostral edge of the unsegmented paraxial mesoderm prior to somite formation (arrows, Fig. 27B"). In the presomitic mesoderm expression of *ZMyoD* was confined to a pair of narrow longitudinal rows of cells (adaxial cells) adjacent to the notochord whereas strikingly *ZMyf-5* transcripts were also found in the lateral presomitic cells (LPS) adjacent to the adaxial cells (Fig. 27D and Weinberg *et al.*, 1996, Fig. 2D). Interestingly, while *ZMyoD* is expressed with more or less uniform intensity in the somites along the anteroposterior axis and weakly in the presomitic mesoderm, *ZMyf-5* expression was strongest in the caudal stripes of the presomitic mesoderm, and weakest in somites 3-5, with intermediate levels of expression in somites 1-2 and 6-7 (compare Fig. 27B" and Fig. 27E or Weinberg *et al.*, 1996, Fig. 2D). Within the somites, *ZMyf-5* expression was slightly more intense at the lateral edge than medially, particularly in somites 4-7 (Fig. 27B"). In contrast the strongest *ZMyoD* expression was seen medially (Weinberg *et al.*, 1996 Fig. 3C). This could indicate a role of *ZMyf-5* in early differentiation in zebrafish somites which takes place from the medial to lateral edge (Blagden *et al.*, 1997). Although somites are continuously formed every 20 to 30 minutes (Hanneman and Westerfield, 1989), expression of *ZMyf-5* became progressively confined to the youngest somites and the two stripes in the unsegmented paraxial mesoderm, indicating that the *ZMyf-5* message in the somites might be short-lived. Note how the intensity of staining drops in somites 1 and 2 of Fig. 27B" to 27D. Only small quantities of mRNA remains in somites formed 1-2 hours previously (compare with somites 3-7 in Fig. 27B" and D). At the 24 somite stage only the most recently formed somites expressed *ZMyf-5* followed by the two stripes at the rostral edge of the unsegmented presomitic mesoderm (double headed arrow, Fig. 27C') and the block of lateral presomitic cells (LPS) that continued to express *ZMyf-5* (Fig. 27C and D). The expression of *ZMyf-5* thus seemed to disappear progressively from differentiated somites, suggesting that its main role in zebrafish is in lineage determination rather than differentiation. In agreement with this, the expression of *myogenin* in zebrafish has previously been shown to follow that of *MyoD* with a delay of approximately 3 hours (Weinberg *et al.*, 1996). Consistent with this is the finding that zebrafish *myogenin* is only expressed in fully segmented somites (Weinberg *et al.*,

1996) in a subset of *ZMyoD* and *ZMyf-5* expressing cells, suggesting, that besides *MyoD* and *Myf-5* other factors influence the expression domain of *myogenin* in zebrafish.

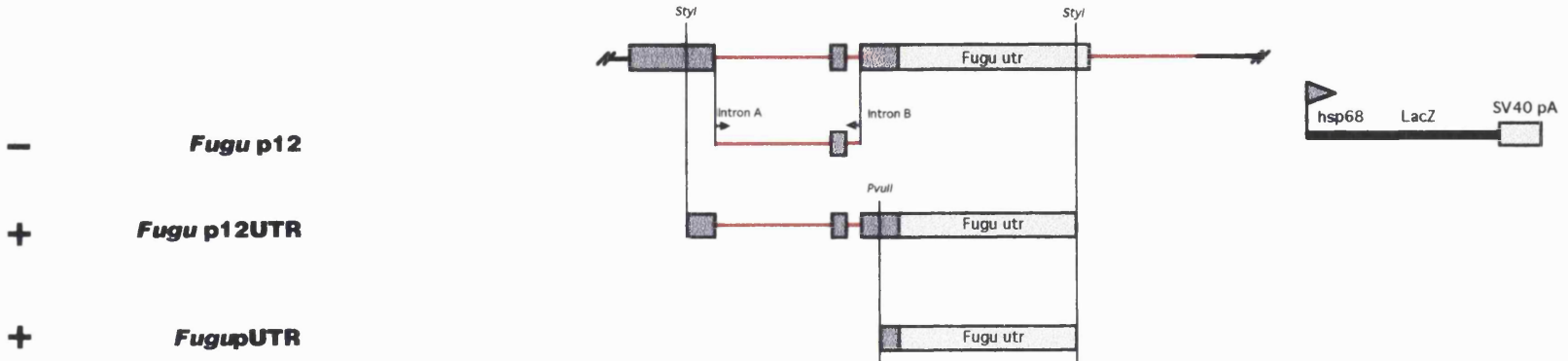
4. Transgenic Analysis of *Fugu Myf-5*

Sequence comparisons between the mouse and *Fugu Myf-5* genes failed to reveal conserved sequence blocks outside of the coding region. However, conservation of synteny in *Fugu* argues strongly for functionally conserved mechanisms in the control of this myogenic factor. If such shared elements with similar functions do exist, their identification in *Fugu* could be valuable for the detailed characterisation of similar regulatory sites in the mouse. Transgenic studies on the mouse *Myf-5* gene indicated that distinct regulatory elements involved in ventral somitic expression exist in both of the *Myf-5* introns as well as the 3' UTR. To test if the equivalent *Fugu* sequences could drive reporter gene expression to similar anatomical domains, a *Fugu* version of p12UTR comprising the *Fugu Myf-5* gene with its 3' UTR was used to generate transgenic mice. A summary of the constructs made and the results is shown in table 4 and table 4A, respectively.

Pronuclear injection of ***Fugu* p12UTR** produced a total of six transgenic animals all of which expressed the reporter gene (Fig. 28A to F). Five of these animals showed a similar pattern of staining in the caudal somites and the ventral somitic bud (sb) resembling that seen in older embryos with the functional mouse transgene (compare with p12UTRR (Fig. 19), CFUTRR (Fig. 21), CFH (Fig. 22)). Unexpectedly intense staining was also seen throughout the presomitic mesoderm of the tail (unlabelled arrows in Fig. 28A to F), a feature that is not shared with the equivalent mouse p12UTR construct (compare Figs. 28 and 19). Transverse sections through the tail of a *Fugu* p12UTR embryo show that the transgene was expressed throughout the entire presomitic mesoderm in the tip of the tail (excluding the neural tube (NT) and posterior diverticulum of the gut (gut), reminiscent of the expression in amphibia and fish (Fig. 28H). Expression was seen in the sacral and lumbar somites and in the somitic bud (sb) between the limbs (see Fig. 28C and F). Transverse sections through the somites showed that most expression was located in the ventral dermomyotome (vDM, Fig. 28G), although stained cells were found along the entire medial edge rather than the caudal boundary of the somites as is characteristic for the equivalent mouse *Myf-5* construct. Less intense staining was found in the rest of the dermomyotome and in the myotome of the caudal somites in which the transgene was expressed regardless of their exact position along the AP axis. From E10 onwards, embryos showed increasing

Table 4: *Fugu* and hybrid *Myf-5* reporter constructs

Fugu Constructs

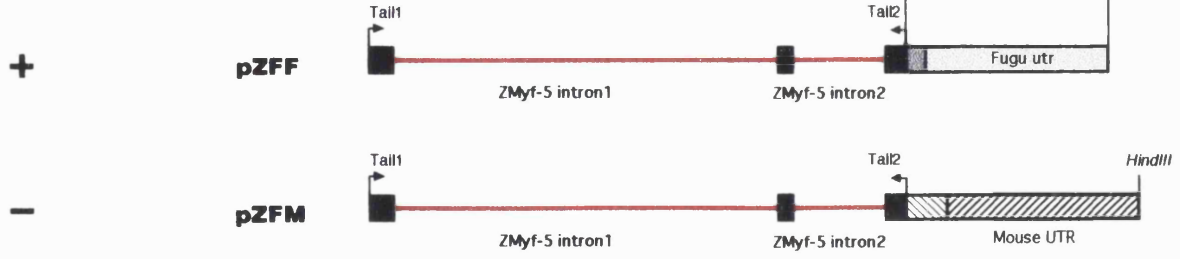


\- ***Fugu* p12**

\+ ***Fugu* p12UTR**

\+ ***Fugu*pUTR**

Zebrafish Hybrid Constructs



\+ **pZFF**

\- **pZFM**

Table 4A:

Summary of the *Fugu* and Hybrid Transgenic Results

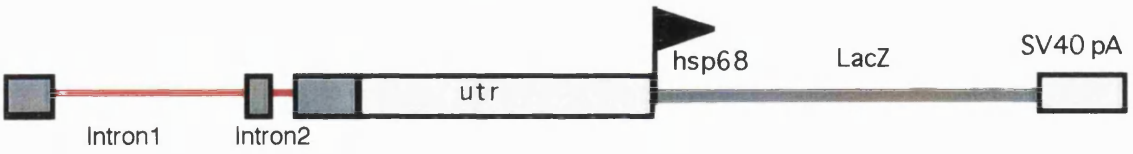
construct	Exp/Tg	TS	TG	DRG	A	Skin	Comment
***Fugu*p12**	1 / 10	–	–	–	–	+	isolated cells only
***Fugu*p12UTR**	6/ 6	+	+ /–	+ /–	+	+	
***Fugu*pUTR**	3/ 12	+	+ /–	+ /–	+	+	
Hybrid constructs							
pZFF	2/ 4	+	+ /–	+ /–	+	+	
pZFM	4/ 11	–	–	+	+ /–	–	consistent nonsomitic

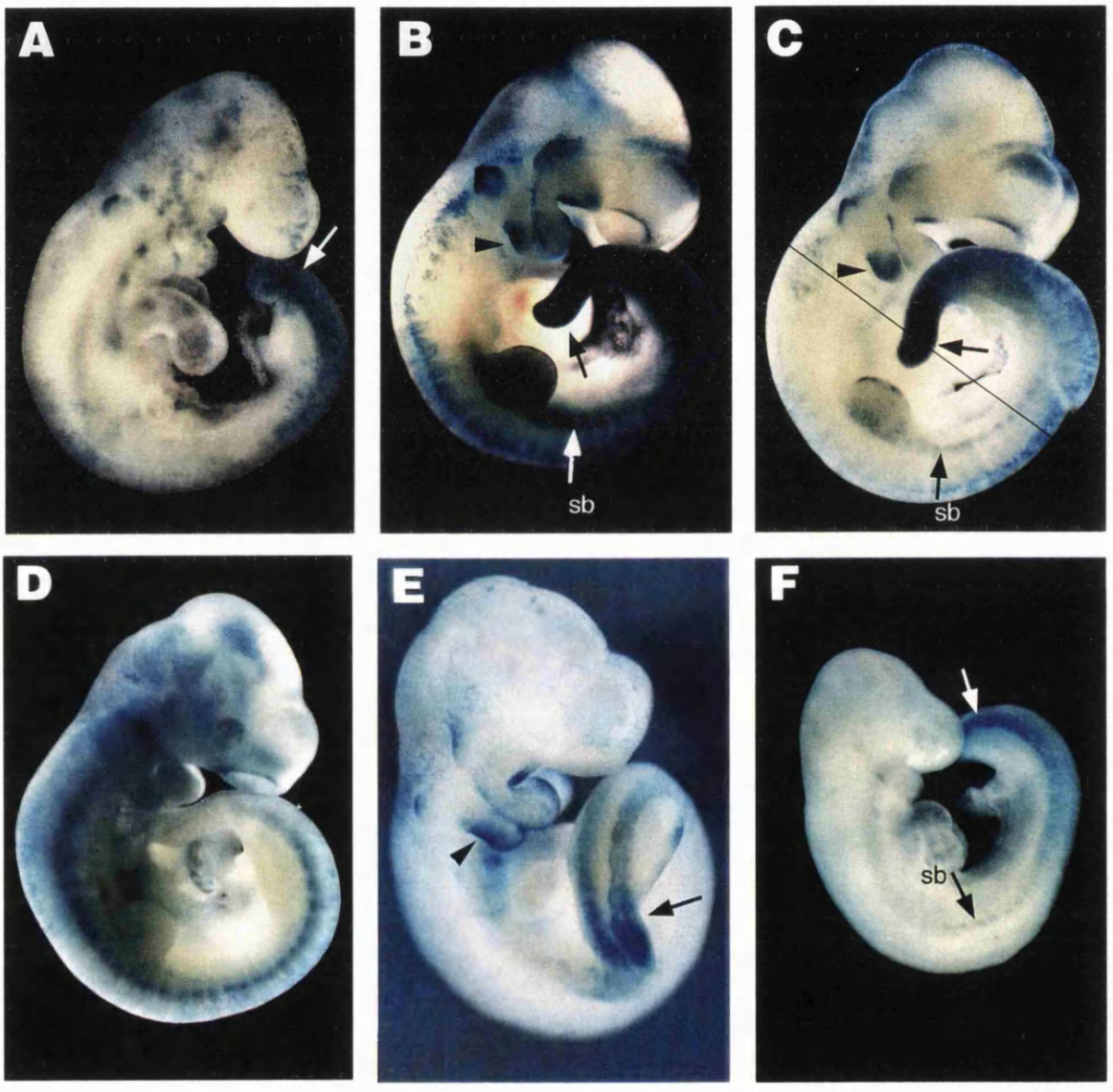
Table 4A:
Summary of the expression patterns obtained with the constructs indicated. **Exp/Tg** shows the number of animals expressing the transgene out of the total number of transgenic animals obtained (as detected by PCR). **TS** indicates expression in the tail and posterior somites, **TG** trigeminal ganglion, **DRG** dorsal root ganglion, **A** branchial arch expression. (+) shows whether or not (-) a construct produced the correct expression. (+/-) indicates that variable results were obtained.

Table 4:
The genomic structure of the *Fugu Myf-5* region is shown at the top. Exons are depicted as gray boxes, introns as red lines, and the 3' UTR is shown in light gray and boxed. Exon 3 of mouse *Myf-5* and the mouse 3' UTR are shown as hatched boxes, Exons of zebrafish *Myf-5* are in black. Vertical lines show the position of common restriction sites or primer binding sites amongst different constructs or in the genomic sequence. PCR-primers used to generate suitable fragments are indicated by labelled arrows. Each construct was made by cloning the indicated region of the *Myf-5* gene upstream of the minimal *hsp68* promoter driving the *LacZ* gene with an SV40 polyadenylation site. A + or - sign indicates whether correct expression was obtained.

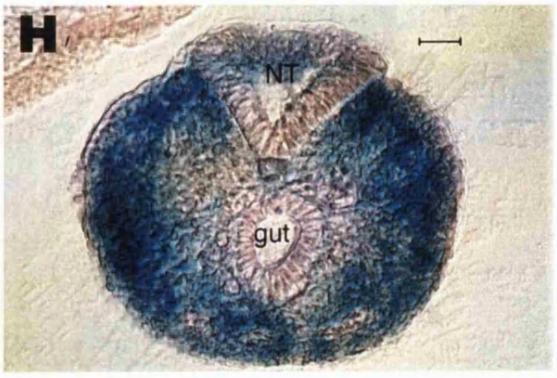
Fig. 28: Transgenic mouse embryos of ***Fugu*** **p12UTR** stained with X-Gal. Expression in the caudal somites and the ventral somitic bud (sb) resembled equivalent mouse constructs but intense staining throughout the presomitic mesoderm of the tail was unexpected (unlabelled arrows in 28A to F). Transverse sections through the tail (28H) showing staining throughout the entire presomitic mesoderm (excluding the neural tube (NT) and posterior diverticulum of the gut (gut). Fig. 28G shows a Transverse sections through the somites (plane of sectioning shown as line in 28C) with staining in the ventral dermomyotome (vDM) and in the dorsal root ganglia adjacent to the NT. In some embryos additional staining was also seen in the epidermis of the skin, the head mesoderm, the branchial arches (arrowheads 28B,C and E) and the trigeminal ganglion, all shared with the equivalent mouse constructs. Figure 28D shows an example of ectopic expression probably as an integration site dependent effect.

Fugu p12UTR


hsp68
LacZ
SV40 pA
utr
Intron1
Intron2


A
B
sb
C
sb
D
E
F
sb


G
vDM


H
NT
gut

staining in the dorsal root ganglia (also visible in section: just laterally from the dorsal neural tube in Fig. 28G) at progressively more rostral levels up to the level of the cervical somites similar to the mouse p12UTR construct. In some embryos additional staining was also seen in epidermis of the skin, the head mesoderm, the branchial arches (arrowheads in Fig. 28B, C and E) and the trigeminal ganglion similar to transgenic embryos carrying the equivalent mouse transgene. Two transgenic animals showed an anomalous pattern including mainly ectopic staining in the neural tube and epidermis of the skin, probably as an integration site dependent effect (one example is shown in Fig. 28D).

In analogy to the approach taken for the mouse *Myf-5* gene, the role of the *Fugu* introns was tested in construct ***Fugu* p12** (see table 4 and 4A). Ten transient transgenic animals were generated of which only one expressed the reporter gene in isolated cells and ectopic locations (data not shown). No other pattern of expression was observed for *Fugu* p12, suggesting that by themselves the *Fugu* introns were not able to direct the somitic expression pattern. Considering the absence of sequence conservation between *Fugu* and higher vertebrates, it might be possible to identify conserved elements between the zebrafish and *Fugu Myf-5* genes. To obtain the genomic sequence of the zebrafish gene, the *ZMyf-5* introns were amplified from genomic DNA by nested PCR using flanking exon primers (see Chapter 2, section 10.1.3. for details). The nucleotide sequence was determined as described. Surprisingly, the size of the zebrafish introns is significantly larger than that of any other species examined. The first intron of *ZMyf-5* is almost three times as large as the corresponding *Fugu* intron and more than twice as large as the human or mouse *Myf-5* intron (see table 5). A schematic representation of these differences is shown in Fig. 29. Pairwise sequence comparisons including the exons as well as the 5' and 3' UTRs of the zebrafish and *Fugu Myf-5* genes identified a single highly conserved block of 119bp near the 3' end of intron 1 (Fig. 30). Remarkably, this conserved region showed more than 85% nucleotide identity while none of the adjacent regions was significantly conserved. Analysis of the mouse and *Fugu Myf-5* introns showed that neither was sufficient to drive the proper somitic expression pattern in transgenic mice and that elements in the 3' UTR of mouse *Myf-5* needed to interact with the conserved intronic elements to activate the proper somitic expression pattern. Therefore, to test the function of the conserved element in the zebrafish intron, a hybrid construct (**pZFM**) comprising the zebrafish introns and the 3' UTR of mouse *Myf-5* upstream of hsp68-*LacZ* was used to generate transient transgenic mice (see table 4 and 4A and Fig. 31). A total of 11 transgenic animals was obtained of which four expressed the *LacZ* reporter gene. None of the transgenics showed a somitic expression pattern (Fig. 31A-C). Most of the expression observed was in the neural tube and dorsal root ganglia. Additional ectopic expression in the limbs was seen in two of the mice (Fig. 31A and B). To exclude the possibility that the absence of somitic

Myf-5 homologues: Size Comparison

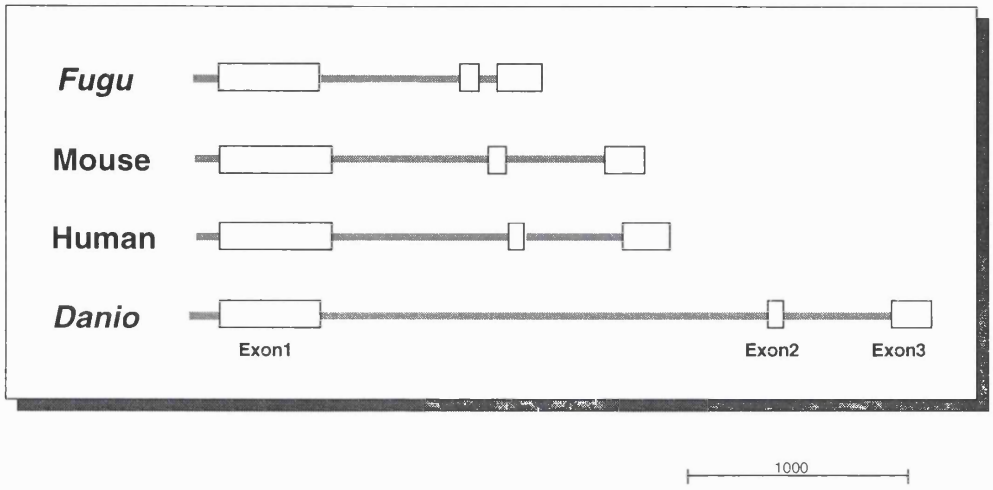


Fig. 29:
Schematic representation (drawn to scale) showing the relative size of the exons and introns of the *Myf-5* homologues from *Fugu*, mouse, human and zebrafish *(danio rerio)*. The exons are depicted as boxes, and the introns as grey lines. Surprisingly, the greatest size difference was observed between the two teleost species. The *Fugu* gene represents the smallest and the zebrafish gene the largest of the *Myf-5* homologues.

Myf-5	Mouse	Human	*Fugu*	*Danio*
Exon1	500	500	452	452
Exon2	75	76	72	75
Exon3	188	212	196	185
Intron1	708	793	634	1996
Intron2	435	427	85	472

Table 5:
Size of the exons and introns of *Myf-5* homologues in base pairs.

Dotplot: Zebrafish vs *Fugu Myf-5*

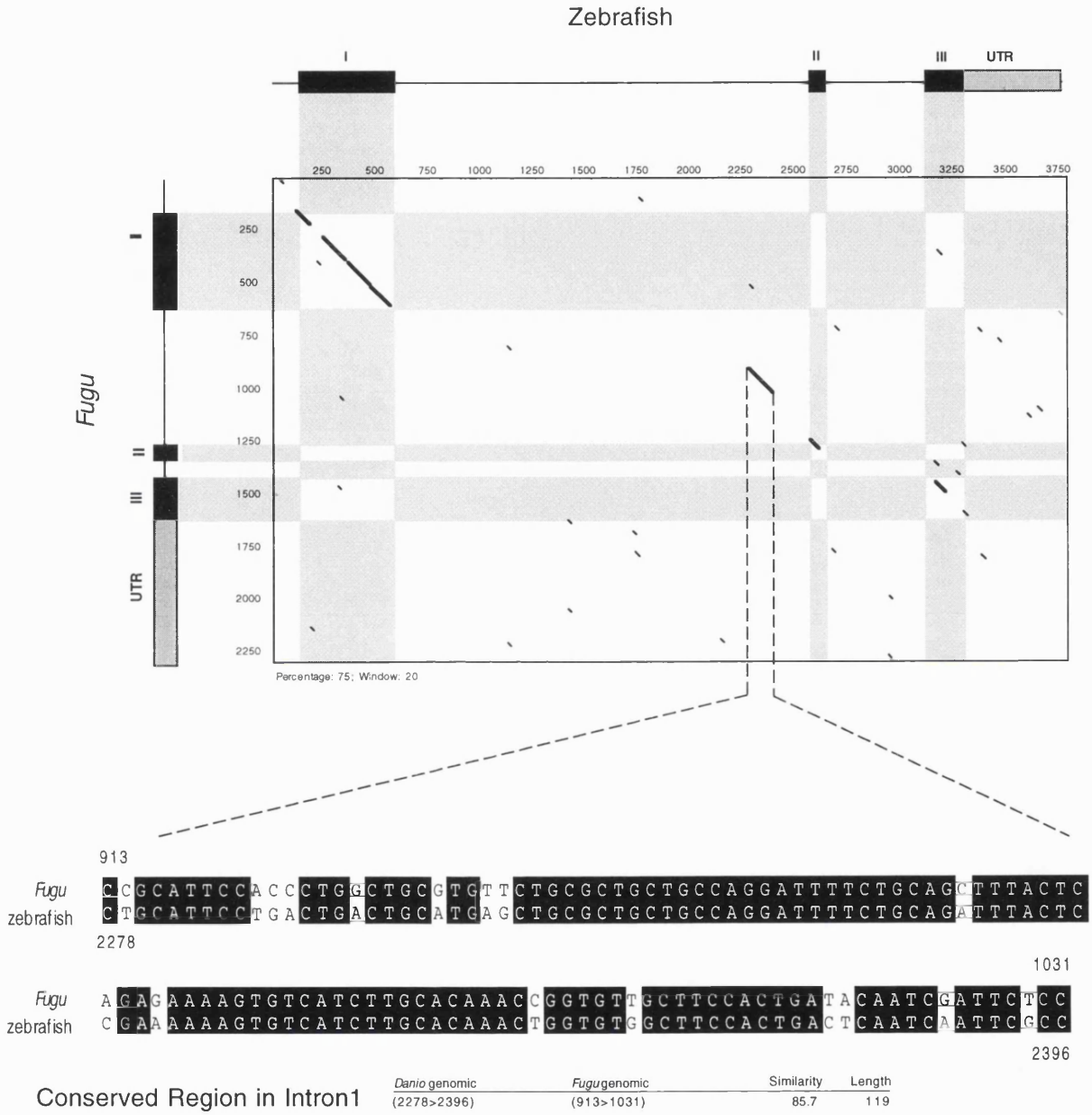


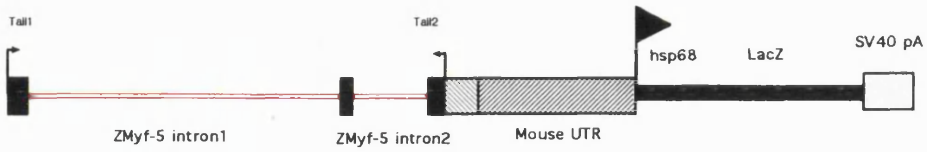
Figure 30

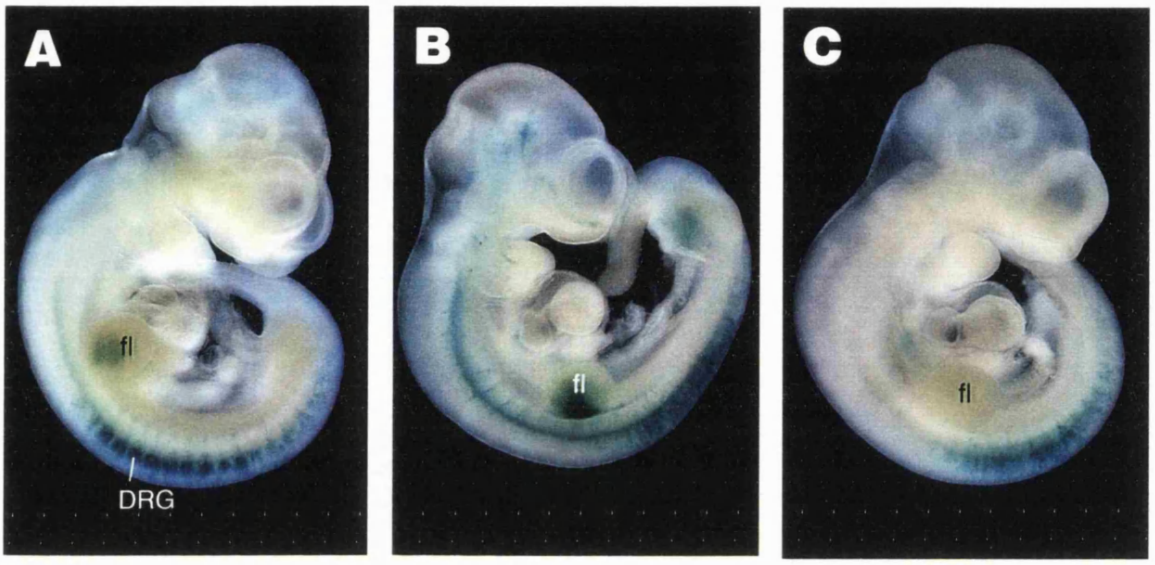
Two way comparison of the *Myf-5* genes from *Fugu* and zebrafish. Regions with at least 75% similarity over a 20bp window appear as lines on the diagonal. The exon - intron structure is superimposed onto the diagram to allow regions of similarity to be aligned with the genomic organisation. Apart from the coding regions a strikingly conserved block of more than 85% homology is found near the 3' end of intron 1 in zebrafish and *Fugu*. The nucleotide sequence of this block is shown below the diagram. Identical bases are shown on a black background.

Fig. 31: Transgenic mouse embryos of the hybrid construct **pZFM** comprising the zebrafish introns and the 3' UTR of mouse *Myf-5* . (31A-C) None of the transgenics showed a somitic expression pattern. Most of the expression observed was in the neural tube and dorsal root ganglia (DRG). Additional ectopic expression in the limbs was seen in two of the mice (31A and B). fl = forelimb

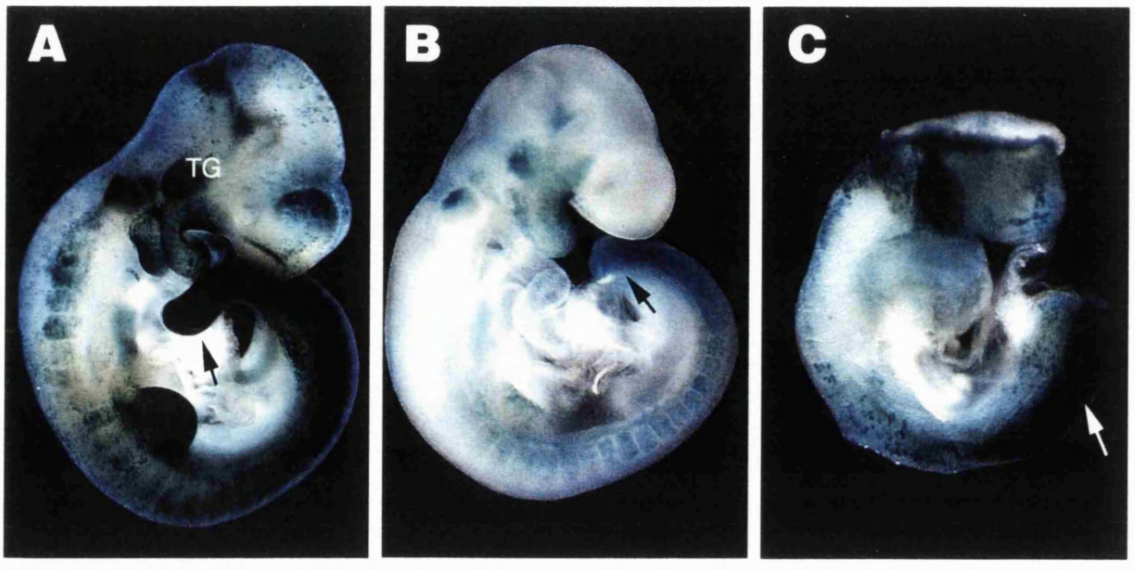
Fig. 32: Transgenic mouse embryos of ***Fugu* pUTR.** In this construct the reporter gene is driven only by the 3' UTR of the *Fugu Myf-5* gene. Figures 32A-C: All transgenics showed the typical expression in the tail (arrow) in the skin, the arches, the trigeminal ganglion (TG) and the limbs previously observed in *Fugu* p12UTR (Fig. 28). This would suggest that the 3' UTR of *Fugu Myf-5* is responsible for the pattern seen in both of the constructs. Therefore, the role of the conserved intron region between *Fugu* and zebrafish remains unclear.

pZFM





Fugu pUTR



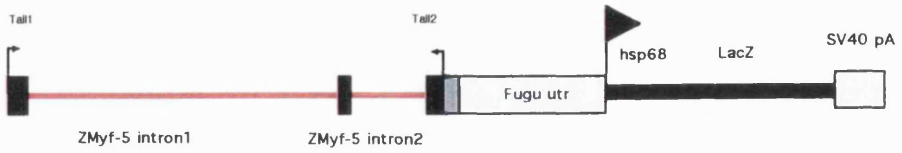
expression was due to incompatibility of the mouse UTR enhancer with the zebrafish introns, the mouse UTR was replaced by the *Fugu* UTR in the next construct **pZFF** (see table 4 and 4A and Fig. 33). Surprisingly, the pattern obtained with pZFF was very similar to that of *Fugu* p12UTR, including the typically intense staining in the tail (arrows), a rostrocaudal gradient of increasing *LacZ* expression in the somites and in the somitic bud as well as trigeminal ganglion and ectopic expression in the epidermis of the skin (compare with Fig. 28). The similarity in the pattern suggested that the *Fugu* UTR might be sufficient for the pattern observed in *Fugu* p12UTR and pZFF transgenic animals. To examine this possibility, a new reporter construct ***Fugu* pUTR** was injected and essentially the same pattern was obtained independently of the zebrafish or indeed the *Fugu* introns (Fig. 32A to C). All transgenics showed the typical expression in the tail (arrow), in the skin, the arches, the trigeminal ganglion (TG) and the limbs (compare Fig. 32A and B with Fig. 28 and Fig. 33). These results indicated clear differences in the regulatory organisation of the mouse and *Fugu Myf-5* genes suggesting that in contrast to the mouse *Myf-5* gene, the major regulatory elements of *Fugu Myf-5* may be located in the 3' UTR and that the *Fugu* introns may not be involved in the regulation of this aspect of the *Myf-5* pattern. In contrast, the mouse UTR itself was shown to direct the reporter gene only to the ventral neural tube, indicating that the conserved regions in the mouse introns must be responsible for the expression seen. However, since the pattern of the fish constructs is different from that of the equivalent mouse constructs, it is difficult to ascertain which, if any, of the expression domains obtained in these cross species experiments represent the true pattern and which is merely an artefact.

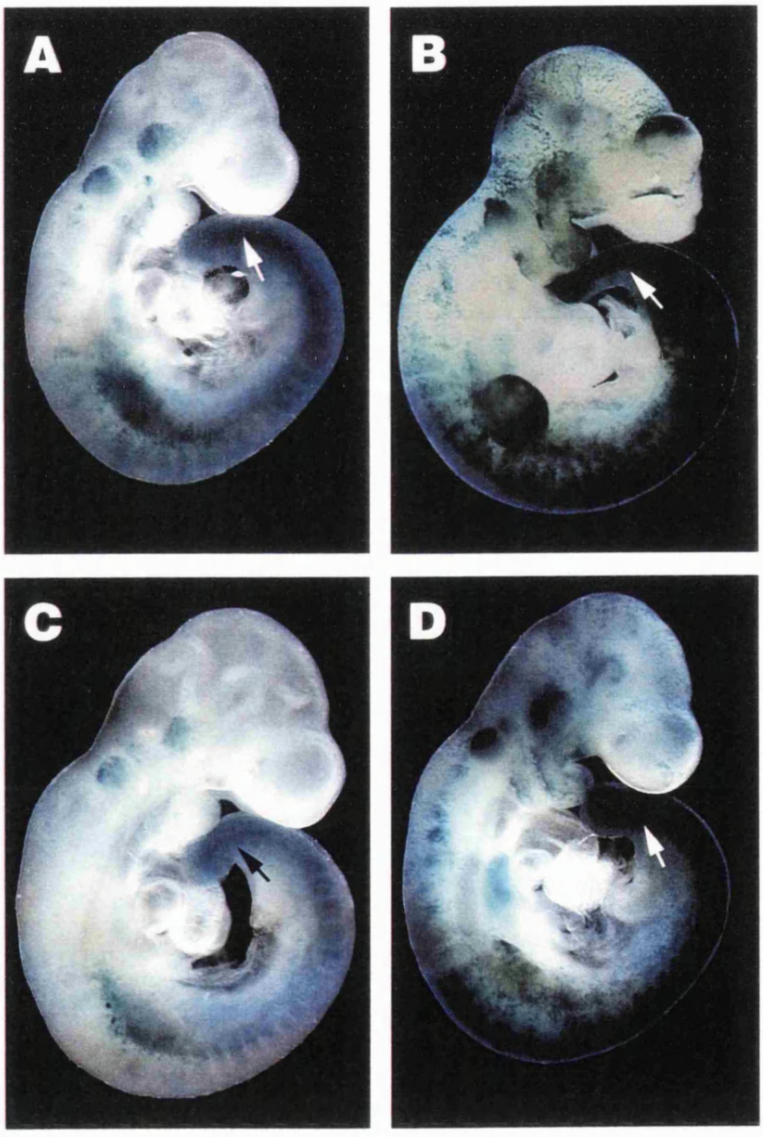
5. Summary

In summary, the transgenic analysis demonstrated that in the mouse expression of *Myf-5* in the ventral posterior dermomyotome of the somites is regulated by at least three distinct and necessary transcriptional elements. The pattern of expression involves separate conserved elements located near the 3' end of each of the mouse introns and additional elements probably involved in regulating the efficiency of expression are found in the 3' UTR of *Myf-5*. Unless the putative efficiency element is located directly at the *AseI* site in the UTR, it appears that either separate elements on both sides of the *Ase*I site are necessary or that the spacing between the elements in the introns and the element in the UTR is critical. In the *Fugu Myf-5* gene, regulatory elements are probably located in the 3' UTR although a conserved element of undefined function exists in the first intron of both *Fugu* and zebrafish *Myf-5*, indicating organisational and functional differences between teleost fish and mammals. Nevertheless, the *Fugu* gene can drive

Fig. 33: Transgenic mouse embryos of a hybrid construct **pZFF** comprising the zebrafish introns and the 3' UTR of *Fugu Myf-5* . The transgenics was consistently expressed in the caudal somites and the tail (arrow) . Additional ectopic expression in the limbs and dorsal root ganglia was observed as well as staining of the skin, the arches, the trigeminal ganglion similar to that of other constructs containing the *Fugu* UTR. Thus the 3' UTR of *Fugu Myf-5* appears to be responsible for the pattern seen and the role of the conserved region in the *ZMyf-5* introns remains unclear.

pZFF


Tail1
Tail2
hsp68
LacZ
SV40 pA
Fugu utr
ZMyf-5 intron1
ZMyf-5 intron2


A
B
C
D

expression of the *LacZ* reporter to the ventral dermomyotome of the somites. However, the expression pattern obtained with the elements differs significantly from the equivalent mouse construct in that it involves the presomitic mesoderm. This kind of pattern is remarkably similar to the native *Myf-5* expression pattern in zebrafish, suggesting that the *Fugu* gene can mimic aspects of its native pattern in transgenic mice. In zebrafish *Myf-5* is expressed in the adaxial as well as the lateral presomitic cells and in the somites. Since *MyoD* is not expressed in the lateral presomitic cells, it is possible that *MyoD* and *Myf-5* mark slow and fast muscle fibre precursors respectively.

Chapter 7
Discussion

1. Introduction

In vertebrates *MRF4* and *Myf-5* are neighbouring genes and several regulatory elements of *Myf-5* have been mapped to the intergenic region and there is evidence for such elements in the genes themselves (Braun *et al.*, 1994; Braun and Arnold 1995; Patapoutian *et al.*, 1995; Zhang *et al.*, 1995; Yoon *et al.*, 1997). Three complementary approaches have been used in this study to define more closely some of the elements controlling *Myf-5* expression specifically in the ventral posterior somite compartment. Sequence comparisons between *Myf-5* homologues of different species were carried out to find conserved sequence blocks. However, sequence conservation alone can not indicate functional relevance. Therefore, electrophoretic mobility shift assays were performed to test the ability of candidate regions to bind proteins from embryonic extracts *in vitro*. Finally, transgenic analysis was used to test putative control regions *in vivo*.

Work from our laboratory shows that in transgenic mice even our largest *LacZ* reporter construct (HMZ17, see Fig. 8B) containing the entire region between and including the *MRF4* and *Myf-5* genes, shows variations not seen by *in situ* hybridisation and fails to activate limb expression. Some of the differences observed may be due to missing distal regulatory elements, or influenced by the integration site and copy number of the transgene, the type and strength of the promoter and the relative persistence of β–galactosidase protein compared to the wild type protein or message. Also, a reconstituted set of putative regulatory elements in reporter constructs may not fully reproduce the wild type pattern because of their altered spatial distribution. However, these problems do not invalidate one or the other approach, rather they should be viewed as complementary to each other.

2. *Fugu* MRF4 *and* Myf-5 *Form a Syntenic Linkage Group*

The Japanese Pufferfish, *Fugu rubripes,* has an eight times smaller genome than the mouse with significantly smaller introns and intergenic distances and it should be particularly suited for studying regulatory elements (Brenner *et al.,* 1993). Therefore, a *Fugu rubripes* cosmid containing the homologues of the myogenic bHLH transcription factors *MRF4* and *Myf-5* was isolated and characterised. Although examples for conserved gene order exist in *Fugu*, it has been suggested that conservation of gene order in *Fugu* is confined to specialised regions such as the *Hox* cluster (Gilley *et al.,* 1997). The myogenic factors also fall into the group of developmentally regulated genes but unlike the *Hox* genes that are clustered even in invertebrates, multiple MRFs have so far only been found in vertebrate species and ascidians suggesting that they are not as ancient as *Hox* genes. Nevertheless, it was shown here that the homologues of *MRF4* and *Myf-5* of the pufferfish *Fugu rubripes* form a syntenic linkage group that has been maintained for 430 MYrs since teleosts started to evolve. This would indicate that conservation in *Fugu* is likely to extend to a much larger group of developmentally regulated genes. Adjacent to the *Myf-5* and *MRF4* genes sequences with high homology to the *C. elegans* transposable element TCB2 and the human oxysterol binding protein were located. The latter has been mapped to chromosome 19 in the mouse and chromosome 11 in human (GENBANK: accession AC003093). suggesting that this gene is not part of the *MRF4/Myf-5* linkage group in mouse or human, because *Myf-5* and *MRF4* map to mouse and human chromosomes 10 and 12, respectively. Thus the region of conserved synteny around the Myf-5 and *MRF4* genes appears to be relatively limited, suggesting that the duplication event associated with those genes was not chromosome-wide and supporting the notion that evolutionary pressures exist to maintain *MRF4* and *Myf-5* as a linked pair.

The genomic structure of the *Fugu MRF4* and*Myf-5* genes is similar to that in other species and the functional domains of the proteins are highly conserved. These include the recently described Cys-His rich region upstream of the basic domain that is thought to mediate changes in chromatin structure required by Myf-5 and MyoD for the activation of their target genes (Gerber *et al.,* 1997). Interestingly, a similar His-Cys-rich-bHLH region has even been found in the *MyoD* homologue of the ascidian *Ciona intestinalis* (Meedel *et al.,* 1997) which otherwise shares virtually no similarity with the other orthologues of the *MyoD* family, suggesting that evolutionarily even more distant species than *Fugu* have retained these structural motifs and probably operate in a similar fashion. In fact it has been shown that invertebrate myogenic factors from sea urchins (Venuti *et al.,* 1991) and nematodes (Krause *et al.,* 1992) can activate myogenesis in mammalian cells. This high degree of conservation would suggest that the mechanism by which these proteins act is extremely ancient.

3. *Fugu is Not a Good Model for* MRF4 *and* Myf-5

The Fugu MRF4/Myf-5 *region is larger than expected*

Although the *Fugu* MRF4 and *Myf-5* genes are separated by less than 3 kb compared to 7kb in the mouse, considering the more than seven times smaller *Fugu* genome, one might have expected the genes to be as close as 1kb. Surprisingly, despite the relatively modest size reduction of the region in *Fugu* there is no significant sequence conservation in the intergenic region or the introns of *MRF4* or *Myf-5*. It appears that in *Fugu* predominantly large genes are reduced in size (Baxendale *et al.*, 1995; Elgar *et al.*, 1996), whereas small genes like *MRF4* and *Myf-5* are less likely to be compressed, conceivably because of the limits imposed by the elements for processing and transcriptional regulation. Although the intergenic region in *Fugu* and the introns may only contain the essential regulatory elements without excess 'junk', the apparent lack of sequence conservation in comparison with other vertebrates including zebrafish would argue against that. From these results it is clear that *Fugu* is not a good genomic model for the study of regulatory elements of *MRF* and *Myf-5* because of the comparatively large size of this region in Fugu. More importantly however, the *Fugu* region is also poorly conserved, or contains very different regulatory elements compared with higher vertebrate species.

Noncoding Regions are Poorly Conserved in Myf-5 *- Evolutionary Considerations*

Comparisons of the *Fugu MRF4/Myf-5* region with homologues of mouse, bovine and human revealed little similarity outside of the coding regions. The lack of sequence conservation was unexpected because regulatory elements have been identified in this region in the mouse, and synteny between *MRF4* and *Myf-5* is conserved. However, these observations may be consistent with the concept that the origin of the myogenic factors lies early in vertebrate evolution and that the gene family as a whole must have diverged significantly in different vertebrate species (Atchley *et al.*, 1994; Holland *et al.*, 1994). Two lines of evidence support this view. Firstly, the absence of major differences amongst all four myogenic factors suggests that the precursor gene duplicated in short succession allowing each myogenic factor to evolve separately for a similar length of time. Secondly since the *MRF4* and *Myf-5* genes are already linked in *Fugu*, this must have happened early in vertebrate evolution. In agreement with this, the evolutionary position of the *Fugu* genes is at the base of the phylogenetic tree for *MRF4* and *Myf-5* (see Fig. 11B, D). The branching order for the *MRF4* gene corresponds to the known evolutionary branching patterns of the major vertebrate groups. Fish divergence occurs first, followed by amphibia, followed by birds and finally mammals (Benton, 1990). Interestingly, the *Myf-5* gene does not strictly adhere to this branching pattern. In *Myf-5*, teleost fish diverged first but then birds follow, before amphibia and finally mammals. In birds the roles of *Myf-5* and *MyoD* are reversed such that *Myf-5*

substitutes for the functions normally associated with *MyoD* and vice versa. In fact, chick *Myf-5* shares 88% amino acids with mouse *MyoD* but only 69% with mouse *Myf-5*. The high degree of pairwise amino acid similarity between *Myf-5* and *MyoD* and similarly between *MRF4* and *myogenin* is the result of their common evolutionary origin. Two distinct evolutionary models could account for the pairwise similarity of the four myogenic factors. First, the ancestral gene might have undergone two tandem duplications to give first two and then four genes. According to this model, *MyoD* and *myogenin*, were subsequently split up by a recombination or translocation event, while linkage between *MRF4* and *Myf-5* persisted. However, the lack of conserved flanking sequence lead Atchley *et al.* (1994) to suggest that probably three duplications occured, increasing the number of paralogous genes from one to two to three to four. They suggest that the ancestral gene was first duplicated, and then a second duplication of each individual gene took place as separate events on different chromosomes to yield the four myogenic factors.

Regardless of the evolutionary model, pairwise homology with *Myf-5* allowed *MyoD* in birds to substitute for the functions normally associated with *Myf-5* and *vice versa* (Pownall and Emerson, 1992) and the evolutionary pressures which allowed *Myf-5* in birds to substitute for *MyoD* might have permitted it to assume an evolutionary more distant position in the phylogenetic tree.

4. *Regions in the* Myf-5 *Introns Conserved Between Human and Mouse Regulate Ventral Somitic Expression.*

In the absence of sequence conservation between *Fugu* and mouse, other comparisons between more closely related vertebrates, particularly between mouse and human proved informative. Despite the low stringency associated with the evolutionarily small distance of such comparisons, more than 75% of the conserved sequence blocks identified in this way later proved to be essential for the ventral expression pattern of *Myf-5* when tested in transgenic mice, underlining the power of this strategy. Perhaps not surprisingly, comparisons between different species at the amino acid level correlated well with the level of conservation in the noncoding regions of *Myf-5*. It seems, therefore, that by comparing the amino acid sequence alone, a suitable model organism with the required evolutionary distance can be found before the genomic DNA sequence is determined for further analysis of regulatory elements in the noncoding regions. In both introns of mouse and human *Myf-5*, a remarkable bias in sequence conservation in the 3' half compared with the 5' half was apparent. While the 3' halves of both introns were highly conserved, showing more than 75% identity, the 5' halves shared less than 50% sequence identity, indicating that the second half of each of the introns contained functionally conserved elements (see Fig. 13A, B). Whether or not these conserved elements are functionally significant was investigated further both *in vitro* and *in vivo.*

5. Multiple Binding Sites Located in the Mouse Myf-5 Introns

To test if the conserved regions of the mouse *Myf-5* introns contain binding sites for transcription factors, in vitro binding assays were carried out. Surprisingly the electrophoretic mobility shift assays (EMSA) with embryonic and F9 EC cell extracts using subfragments of the introns of mouse *Myf-5* showed binding activities in both, the conserved 3' half as well as the nonconserved 5' half, suggesting that some nonspecific protein binding occurred under the *in vitro* conditions. Indeed, subsequent transgenic analysis demonstrated that the nonconserved regions play no part in the regulation of the somitic expression pattern of *Myf-5*. However, some of these activities appeared to be very efficiently competed (see Fig. 15E) raising the question as to what the function of these observed binding activities might be. One possibility is that these binding activities may be involved in other gene regulatory functions although the absence of evolutionary conservation of their binding sites does not support this view. An attempt to address this possibility was made by using *Myf-5* positive versus negative tissue extracts of mouse embryos (as judged by their *in situ* hybridisation pattern) in EMSA. However, no evidence of tissue specificity was found. Although it can not be excluded that some of the extracts may have suffered minor contamination from adjacent tissues, the relative differences should still be apparent. What this suggested is that most of the protein factors binding to the intron fragments might be ubiquitously distributed, or that minor contaminations of the embryonic protein extracts eliminated the intended tissue specificity. It is possible that the majority of protein factors are ubiquitous amongst the tissues tested and that the formation of specific complexes depends on minute concentration differences or a small fraction of 'somite- specific-cofactors' *in vivo*. The involvement of ubiquitous E2A gene products (E12, E47) and muscle specific MEF2 proteins in the regulation of the myogenic factors shows that interactions between differentially distributed transcription factors can synergistically activate bHLH gene expression. It is possible therefore, that tissue specific differences are too subtle to be observed in the *in vitro* assays used. Since such assays can not truly reflect the native binding conditions, it is also conceivable that the protein factors are unable to associate with their target sites or fail to form functional tertiary complexes and thus allow non-specific interactions to take place under non-optimised conditions. However, after the subsequent *in vivo* analysis of the *Myf-5* introns in transgenic mice, the remaining region in intron 1 is now only 110 bp in length and amenable to more detailed analysis in bandshift or footprinting studies, that should help to pinpoint transcription factor binding sites at nucleotide resolution.

Although so far the identity of most of the binding sites in the introns remains unknown, a potential NFY binding site matching the NFY consensus in 8 out of 9 bases, has been found near the 3' end of the second intron of mouse *Myf-5* (see Fragment G, Fig. 16G lanes 13 to 25). This was confirmed by competition with an NFY oligonucleotide

corresponding to the *MyoD* promoter region -628 to -602 (Zingg *et al.,* 1991). In the *Hox*b4 gene NFY is necessary for all of the somitic and lateral mesodermal expression and part of the neural expression pattern observed in transgenic mice. Since NFY sites have been found in the promoters of the *MyoD* and *myogenin* genes (Gilthorpe, unpublished; Gutman *et al.,* 1994) as well as in the *Myf-5* promoter (Summerbell, pers. comm.), NFY might play a general role in the regulation of myogenic factors. Since NFY appears to bind DNA as a heterotrimer, different tissue specific partners of NFY may be involved in regulating the activity of *Hox*b4 in the neural tube and the somites (Gilthorpe, pers. comm.) and it is possible that the binding activity observed in the *Myf-5* intron is related to NFY or that NFY binding to the *Myf-5* promoter involves a different combination of partners for binding. The observed change in consensus position 5 from T to C (both pyrimidines) is not as drastic as the reported mutation into a G which reduced NFY binding in the *Hox*b4 promoter to only 10-20% (Gutman *et al.,* 1994). Ultimately, mutational analysis of the NFY site in transgenic mice will provide information about its function in the regulation of *Myf-5*. It is interesting that the NFY site in the *Hox*b4 promoter is conserved between chicken and mouse, but the same is not true for the pufferfish, suggesting that in *Fugu* a different factor plays an equivalent role to that of NFY or alternatively, its regulatory elements are in a different location. In analogy to the *Hox*b4 gene, relocation of regulatory elements may have taken place in the *Fugu Myf-5* gene. Both possibilities might help to explain the apparent lack of sequence conservation between *Fugu* and higher vertebrates.

To obtain information about the *in vivo* role of the different regions of the *Myf-5* gene and particularly the conserved domains in each of the introns, a comprehensive analysis of the intron fragments and the 3' untranslated region was carried out using reporter gene assays in transgenic mice.

6. *Transgenic Analysis Reveals Complex Regulatory Mechanism for* Myf-5.

Myf-5 is initially expressed in cells derived from the dorsal medial portion of the dermomytome that give rise to the precursors of epaxial muscles, and *MyoD* is initially expressed in the cells derived from the ventro-lateral portion of the dermomyotome that give rise to the hypaxial muscles. On this basis it has been suggested that *Myf-5* and *MyoD* play distinct roles in the activation of the hypaxial and epaxial muscle programs (Rudnicki and Jaenisch, 1995; Smith *et al.,* 1994; Cossu *et al.,* 1996a; Kabler *et al.,* 1997). However, *Myf-5* is also clearly detectable in the ventral part of the somites by *in situ* hybridisation (Summerbell *et al.,* unpublished). Similarly, in explant experiments a subpopulation of the ventrolateral somite cells has been shown to express *Myf-5* (Cossu *et al.,* 1996a). While expression in the dorsal myotome is regulated by elements in the intergenic region between *MRF4* and *Myf-5*, the results presented here show that

distinct and evolutionarily conserved elements in intron 1 and intron 2 of *Myf-5* as well as the 3' UTR interact to control efficient expression of *Myf-5* in the posterior ventral margin of the dermomyotome. This raises the question as to what function Myf-5 could have in the latero-ventral somite compartment. Evidence that *Myf-5* could play a role in the activation of *MyoD* in most of the trunk comes from the recently observed (and previously not recognised) delay of *MyoD* expression in *Myf-5* null mutants (Tajbakhsh *et al.,* 1997). In these mutants *MyoD* expression is delayed by about 24 hours until *MyoD* becomes activated by Pax-3. Recently analysis of Splotch/ *Myf-5 LacZ* knock-in mice showed a dependence of *Myf-5* expression in the ventrolateral domain of the somites on the paired box transcription factor *Pax-3* (Tajbakhsh *et al.,* 1997). In the absence of *Pax-3* most if not all, hypaxial muscle is missing and this coincides with an absence of the ventral somitic bud in which *Myf-5* is also expressed, consistent with the notion that *Myf-5* is activated by *Pax-3* in these cells (Tajbakhsh *et al.,* 1997). In support of this view, *Myf-5/LacZ* knock-in mice show that the ventral posterior expression domain of *Pax-3* coincides with *Myf-5* expression (Natoli *et al.,* 1997; Tajbakhsh *et al.,* 1997) and appears to co-localise with the domains controlled by the regulatory regions defined in this study of the *Myf-5* gene. This would suggest that the elements characterised here play a part in regulating the *Pax-3* dependent population of *Myf-5* expressing cells in the ventral posterior dermomyotome. Analysis of the *Pax-3* expression pattern together with *LacZ* staining in the transgenics should confirm this proposal. Unlike *Pax-3* expression, *Myf-5* expression detected by *in situ* hybridisation in the ventral part of the somite is not confined to the posterior edge, which might indicate that the anterior and posterior halves of the somite are controlled by separate regulatory elements. Other examples from our laboratory show that the domains activated by isolated regulatory elements in the context of transgenic reporter constructs are not always in complete agreement with the *in situ* hybridisation patterns. In this regard, it is worth noting that our largest construct, HMZ17, is expressed in both the dermomyotome and the myotome of transgenic mice, whereas the *in situ* hybridisation pattern shows mainly myotomal expression, suggesting that additional negative control elements might be missing from the transgenic construct. It is possible that correct ventral expression of *Myf-5* depends on additional distal regulatory elements as does the expression of *Myf-5* in the limbs. It is presently unclear whether *Pax-3* activates *MyoD* directly or indirectly or how *Pax-3* might be involved in the ventrolateral expression of *Myf-5*, but the dissection of the regulatory elements controlling ventral *Myf-5* expression should help to address these questions.

7. Conserved Regulatory Elements in Both Introns and the UTR of Myf-5 are Required for Ventral Posterior Somite Expression in the Mouse.

Regulatory elements in the *Myf-5* gene have been shown to activate the *LacZ* reporter gene in the ventral posterior somite compartment (Yee *et al.*, unpublished). Here it was shown that a similar construct (p12UTR) - employing the heterologous heat shock promoter hsp68 in place of the previously used β-globin promoter or the endogenous *Myf-5* promoter - can reproduce faithfully that expression pattern (Summerbell, personal communication). An unusual but useful feature of the hsp68 promoter is that it frequently activates reporter gene expression in the ventral neural tube but in our experience this occurs only in the absence of other specific enhancer elements, thus providing a readout of transgene integration into a permissive site (our observation; Joyner *et al.*, 1987). In agreement with the notion that regulatory elements tend to be evolutionarily conserved, it was shown through successive deletion variants of p12UTR, that collectively the conserved regions in both introns and the 3' UTR of *Myf-5* are sufficient to drive expression of the *LacZ* reporter gene to the ventral posterior margin of the dermomyotome.

However, deletion of either intron or the UTR abolished ventral somitic expression (see constructs p1, p2, p12). Similar observations have also been made by Tajbakhsh *et al.*, 1996, showing that 5.5 kb of *Myf-5* upstream flanking sequence including exon 1 and most of intron 1 directed some skeletal muscle expression but not at a level that was quantitatively or qualitatively equivalent to the endogenous gene.

Sequence comparisons and bandshift analysis of the mouse *Myf-5* introns suggested that the candidate regulatory regions are located in the conserved 3' halves of both introns, corresponding to subfragments C and D for intron 1 and fragments F and G for intron 2. Successive deletion of the nonconserved 5' half of the introns (corresponding to fragments A, B and E, respectively) did not interfere with the ventral somitic expression pattern, confirming the assumption that the essential regulatory elements are evolutionarily conserved. The results clearly validate the approach used and show that sequence comparisons with appropriate models can be used efficiently to identify conserved regulatory elements in large genomic regions. More than 60% of nucleotide sequence were excluded from the *Myf-5* introns in this way, demonstrating that the conserved regions of both introns and the 3' UTR are sufficient to drive reporter gene expression to the correct anatomical domains.

8. Separate Elements Appear to Control Intensity and Spatial Distribution of Transgene Expression.

Although pairwise sequence comparisons between mouse and human were employed successfully in the identification of conserved functional regions in the *Myf-5* introns, a similar strategy was not possible to delimit the opposite end of the regulatory region in the 3' UTR. The degree of sequence homology in the 3' UTR was uniformly low despite the fact that the 3' UTR had been shown to be essential for the somitic expression pattern (see above). Therefore arbitrary deletions from the 3' end of the UTR were tested until the somitic expression pattern was lost.
Deletion of 467bp of sequence from the 3' UTR up to a *Hin*dIII site (pCFH) did not compromise the expression pattern in the ventral posterior domain of the thoracic somites. In contrast, digestion at the upstream *Ase*I site, which deleted nearly 700bp from the 3' UTR (pCFA) resulted in drastically reduced levels of expression. Nevertheless some residual *LacZ* expression was still found in the ventral posterior domain of the somites, indicating that elements influencing the level of transgene expression are located in the terminal 700bp fragment between the *Ase*I and *Hin*dIII site that had been deleted. This is supported by similar results (pCFAΔ) following deletion of the same 700bp from the 3' UTR of pCFHΔ , suggesting that indeed separate regulatory elements control the intensity and spatial distribution of transgene expression, the former located in the 700bp region upstream of the *Hin*dIII site in the 3' UTR, the latter most likely in the introns of the *Myf-5* gene. Surprisingly, when the terminal 700bp fragment of the 3' UTR was combined with the conserved intron regions it was not sufficient to drive the reporter gene (pΔAH) at wild type levels. Several reasons might account for this observation. First: the putative enhancer element in the 3' UTR may have been disrupted by *Ase*I digestion. Second: both the 5' and 3' halves of the untranslated region may be required for the expression of the transgene. However, these both seem unlikely because in sequence comparisons between the mouse and bovine the region around the AseI site is poorly conserved compared with the region upstream and downstream from the *Ase*I site. The most likely explanation is therefore that the spacing between the 3' UTR elements in the 700 bp region and the *Myf-5* introns is critical for the function of the 3' UTR element. However, none of these possibilities was explored further as the focus of additional experiments was directed at the conserved intron elements.

9. Fragments D, F and G, but not C, are Involved in Ventral Somitic Expression of Myf-5.

As mentioned above, the full ventral posterior somite pattern was retained in transgenic mice following successive deletion of the nonconserved intron fragments A, B and E from reporter constructs. Unexpectedly further deletion, this time of the conserved fragment C (pΔCPH) from pCFHΔ did not abolish the ventral expression of the reporter gene either. This meant that despite the evolutionary conservation of fragment C in mouse and human and the bandshift data, this region appeared not to be required for the somite pattern nor the other expression domains in the branchial arches and the trigeminal ganglion which were all unaffected. It is possible that fragment C may have other regulatory functions possibly unrelated to the *Myf-5* gene. Indeed, such long range effects of elements in the *MRF4* gene have been shown to affect expression of *Myf-5* in *cis* (Floss *et al.*, 1996; Yoon, *et al.*, 1997) and a *cis*-regulatory mechanism has also been implicated in the regulation of *MRF4* since one of the *Myf-5* mutations abolished myotomal *MRF4* expression, suggesting that this interaction may be reciprocal (Yoon *et al.*, 1997). Alternatively, fragment C may be involved in controlling aspects of temporal expression of the *Myf-5* gene, although no firm conclusions can be made without a more detailed analysis of multiple lines from this construct. However, similar deletions of fragment F (pΔFPH) and fragment G (pΔGPH) completely abolished the ventral posterior expression in the dermomyotome showing staining only in the ventral neural tube in both cases. Together with the data showing that elements of both introns and the 3' UTR are required for correct transgene expression, at least one of the necessary elements must be located in the only remaining part of intron 1 corresponding to fragment D at the 3' end. In addition, both fragments F and G of the second intron are likely to harbour essential regulatory elements. The proximity of fragments F and G might suggest that the binding activities in these fragments are overlapping or belong to a larger complex. The identification of a potential NFY binding site in fragment G has been described above, but the identities of the remaining binding activities are presently unknown, although more detailed bandshift analysis of fragment D indicated that its putative binding site(s) is located in the 110bp at the 5' end, and more detailed analysis *in vitro* and *in vivo* combined with mutation of the candidate target sites will help to resolve this issue.

10. A 'Branchial Arch Element' in the Myf-5 Gene?

Interestingly, aside from the somitic expression pattern, p12UTR and the majority of deletion constructs derived from it also showed reporter gene expression in the branchial arches which was unexpected since elements required for arch expression had previously been mapped to the intergenic region (our observations, unpublished; Patapoutian *et al.*, 1993). Since separate p12UTR lines and other deletion constructs

(including p2UTR, pCFUTRR, pCFH, pCFHΔ, pΔCAH and pΔCPH) showed the same arch expression, this is unlikely to be a position effect. *In situ* hybridisations on *Myf-5* by our laboratory showed that branchial arch expression is normally confined to the core of the second and third arch at about E9 and similar results have been obtained with both the intergenic arch element and in *Myf-5* LacZ knock-in mice (our observation, Patapoutian *et al.,* 1993; Tajbakhsh *et al.,* 1996). In contrast, although topologically correct, in this study the expression from the various deletion constructs mentioned above was delayed by about 24 hours and depending on the construct showed variable intensity, suggesting that elements in the *Myf-5* introns are unlikely to be responsible for the normal expression of *Myf-5* in the branchial arches. Sequence comparisons with the putative arch element from the intergenic region produced no significant matches that would explain the duplicate expression pattern. Interestingly *Pax-3*, which appears to be involved in the ventral posterior expression domain characterised in this study, is also expressed in the branchial arches (Natoli *et al.,* 1997). However, expression of *Myf-5* in the arches of Splotch mice that lack *Pax-3* indicates that the arch expression of *Myf-5* is independent of Pax-3 and of the elements required for the ventral posterior somitic expression (Tajbakhsh *et al.,* 1997). Conversely, it is not clear at present, if *Pax-3* expression in the arches is influenced by *Myf-5*.

11. *The 3' UTR of the Fugu* Myf-5 *Gene Drives Expression in the Somites and the Presomitic Mesoderm.*

Evidence from *MRF4* null mutations and transgenic analysis in the mouse indicates that regulatory elements controlling *Myf-5* are located in the intergenic region and in the *MRF4* and *Myf-5* genes themselves (Yoon *et al.,* 1997). Since synteny between both *Fugu* genes has been conserved over nearly the entire period of vertebrate evolution, it seems that regulatory mechanisms should also be conserved. However, sequence comparisons between *Fugu* and higher vertebrates may not necessarily identify regulatory elements since the binding sites may have adapted to the transcriptional machinery of *Fugu*. Furthermore, several observations suggest that the distribution of regulatory elements is different between *Fugu* and mouse. First, the size of introns in *Fugu* differs remarkably from the mouse, this applys particularly to the extremely short second intron of *Fugu Myf-5*. Second, pairwise sequence comparisons between the mouse and *Fugu* failed to identify conserved sequences in the introns or the UTR. Finally, transgenic analysis of the *Fugu Myf-5* gene showed that the 3' UTR alone is responsible for the somitic expression pattern observed in transgenic mice. Both p12UTR and pUTR directed expression to the ventral dermomyotome of the caudal somites similar to its mouse counterpart. However, the expression was not confined to the caudal edge of the somites and the *Fugu* transgene was expressed both in the dermomyotome and in the myotome of the caudal somites, whereas myotomal staining

in the mouse was only observed anterior from and including the thoracic somites. This may be a reflection of the fact that in teleost fish the myotome forms the major component of the somites and the process of somite maturation differs greatly from that of mammals. The most striking feature of these *Fugu* constructs, however, was the intense expression seen throughout the presomitic mesoderm of the tail in transgenic mice, a feature that is not shared with the equivalent mouse p12UTR construct. Although this may be the result of missing or non-functional negative control elements in the *Fugu Myf-5* gene, several lines of evidence argue against that. In amphibia *MyoD* and *Myf-5* are first expressed in the presomitic mesoderm where they initiate myogenic differentiation prior to somite formation (Hopwood *et al.,* 1992). It is therefore tempting to speculate that the presomitic expression seen with the *Fugu* transgene may reflect the native expression domain of the *Myf-5* gene in teleosts rather than that expected for the mouse, but does the pattern of *Myf-5* expression in teleosts support this view? Comparative studies conducted here in zebrafish, at least in part, support the idea. There are some striking similarities in the pattern of *ZMyf-5* expression in zebrafish and the *Fugu Myf-5* transgenics. In zebrafish *ZMyf-5* (but not *ZMyoD*) is expressed in the lateral presomitic cells of the unsegmented paraxial mesoderm reminiscent of the expression of the *Fugu Myf-5* gene in the presomitic mesoderm of transgenic mice. Like the *Fugu* transgene, *ZMyf-5* is observed in bands of cells in the caudal somites supporting the idea that the transgene, rather than mimicking the equivalent mouse construct, shows characteristics of its native expression pattern in fish. Such a scenario would imply that the transcription factors in the mouse can activate the *Myf-5* gene in the presomitic mesoderm, which in light of the limited sequence conservation seems unlikely. An alternative model would suggest that before amniotes evolved the myogenic factors were expressed in a presomitic phase possibly involved in both, myogenic and somitogenic functions and that their exclusion from the presomitic phase is a recent evolutionary event that requires active repression. Interestingly, for the *Pax-3* gene positive and negative elements have been identified that suppress expression in the caudal somites and the tail while activating *Pax-3* dorsally in the trunk and ventrally in the migratory myoblasts (Natoli *et al.,* 1997). If similar mechanisms operate on *Myf-5* or *MyoD* remains to be seen, although in the absence of further evidence it is equally possible that the signals upstream of *Myf-5* and *MyoD* might simply be delayed until after myogenesis. However, low levels of *Myf-5* transcripts have been detected in the presomitic mesoderm of E8.5 -10.5 mouse embryos using reverse transcription polymerase chain reaction (RT-PCR) (Rupp and Weintraub, 1991, Kopan and Weintraub, 1994) and in *LacZ-Myf-5* knock-in mice at E9 (Cossu *et al.,* 1996b) but not by *in situ* hybridisation (Ott *et al.,* 1991). This would argue in favour of a repressive mechanism since signals exist, at least to activate basal levels of *Myf-5* in amniotes prior to somitogenesis. Since amniotes develop in a relatively protected environment, the need

for rapid myogenic differentiation is much less important than for amphibia. This combined with the increasing complexity of somite anatomy and patterning might explain why in amniotes myogenic differentiation is delayed until after somite formation and perhaps actively repressed in the presomitic mesoderm. Conceivably, by delaying the expression of the myogenic factors until after somite formation it might be possible to cope with increasing complexity of regulatory functions perhaps associated with the establishment of distinct myogenic lineages arising from the medial and lateral somite compartments (Brand-Saberi *et al.,* 1996; Cossu *et al.,* 1996a; Pourquie et al., 1996) and increasing complexity of muscle specific structural genes required to make the complete set of skeletal muscle fibres. Although there is no firm evidence for a repressive mechanism, several candidate molecules could potentially fulfil such a role. In *Drosophila notch* activation inhibits differentiation of muscle cells by down regulating the only myogenic factor *nautilus (MyoD)* (Michelson *et al.,* 1990). The mouse homologue *mnotch1* when constitutively activated has been shown to suppress expression of *MyoD* and *Myf-5* in tissue culture (Kopan *et al.,* 1994). Since *mnotch* is expressed in the presomitic mesoderm of mouse embryos (Conlon *et al.,* 1995) it could be involved in suppressing myogenesis upon activation by homologues of the *delta* family in the segmental plate (Bettenhausen *et al.,* 1995). This fits well with suggestions that *Wnt* members from the neural tube and dorsal ectoderm could block the notch receptor and relieve repression to activate myogenesis as in *Drosophila* (Cossu *et al.,* 1996b; Ranganayakulu *et al.,* 1996). According to this model, presomitic expression of *Myf-5* in the lateral paraxial mesoderm of zebrafish embryos might most easily be explained by an absence of the regulatory sites for *notch* mediated repression in the fish gene. As in the mouse, notch is strongly expressed in the paraxial mesoderm of zebrafish and a putative *notch* mutant *white-tail (wit)* appeared to express *Myf-5* normally (Jiang *et al.,* 1996). However, there is no evidence of specificity in zebrafish for the anti- human Myf-5 antibody (Santa Cruz) used in these studies. Although myogenesis is normal in *wit* mutants, the formation of somite boundaries in *wit* mutants appears to be affected, suggesting that notch may be involved in aspects of somitogenesis in mice and teleosts. Since several notch receptors have already been found in the mouse (Williams *et al.,* 1995), the regulatory mechanisms may be more complicated and functional redundancy can not be excluded. In any case, failure to respond to repressive signals that might involve *notch* could account for the pattern observed with the *Fugu* gene in transgenic mice.

12. Conserved Element in Intron 1

Although the 3' UTR of the *Fugu Myf-5* gene appears to reproduce aspects of the endogenous *Myf-5* expression pattern in transgenic mice, there is surprisingly little sequence homology between the equivalent regions of the mouse and *Fugu* genes. Even more surprising, there was no significant sequence homology between the 3' UTR of zebrafish and *Fugu* either but a conserved region of 119bp with 85% homology near the 3' end of the first intron was identified. The remarkable conservation would suggest that some regulatory function is associated with the *Myf-5* introns in fish too. That function might be specific to teleost fish and could depend on transcription factors that are either not present or are too divergent in mice. However, in transgenic assays, the zebrafish introns combined with the *Fugu* UTR (pZFF) failed to contribute to the expression pattern which was indistinguishable from that of the *Fugu* UTR alone. When combined with the mouse UTR (pZFM), ectopic nonsomitic expression was obtained, indicating the conserved intron regions could not activate the reporter gene in the somites. Whether the fish introns are involved in the regulation of *Myf-5* in teleosts remains to be seen in a different context. It would therefore be interesting to compare the outcome of bandshift assays on this region using zebrafish and mouse embryonic extracts or examine the functional role of this region in transgenic zebrafish. Considering the broad distribution of regulatory elements involved in the somitic expression of *Myf-5* in the mouse, it is also possible that the conserved region in intron 1 of *Fugu Myf-5* requires additional elements outside of the region tested thus far, and that a larger construct that also includes the intron element might thus be able to reproduce a more comprehensive somitic pattern. To address the issue of incompatibility between the Fugu transgene and the transcriptional machinery in the mouse, it might be of interest to see how the *Fugu* or zebrafish gene would behave in more closely related amphibian embryos. A suitable GFP reporter construct is currently being tested in *Xenopus* embryos. Ultimate results can however only be reached in the context of the native environment in transgenic zebrafish and are unlikely to be obtained from amphibian models.

13. Myf-5 is Expressed in Lateral Presomitic Cells

To investigate the role of the myogenic factor Myf-5 in teleost fish, the zebrafish homologue of this gene was cloned and its expression pattern was studied by wholemount in situ hybridisation on zebrafish embryos. The expression pattern of *ZMyoD* and *ZMyf-5* in the somites and the presomitic mesoderm is consistent with a role in muscle fate specification in zebrafish. Both MRF are expressed in the medial muscle precursors adjacent to the notochord (adaxial cells), and in laterally extending stripes of the first somites and at the anterior margin of the presomitic mesoderm just

prior to somite formation (Weinberg *et al.,* 1996; and this study). Interestingly, *ZMyoD* and *ZMyf-5* show marked differences in their expression during the presomitic phase, suggesting that they serve distinct functions in myogenesis. Unlike *ZMyoD*, which is initially expressed in the adaxial cells flanking the notochord, *ZMyf-5* expression extends to nonadaxial lateral presomitic cells (LPS) (Devoto *et al.,* 1996). The pattern of *ZMyf-5* expression reported by Jiang *et al.* (1996) using anti-human Myf-5 antibodies (Santa Cruz) was similar to the pattern described here, however, there is no evidence for specificity of this antibody in zebrafish or mice.

Most of the paraxial somitic mesoderm is thought to originate from the posterior tail bud (Kanki and Ho, 1997*). It was shown here that tailbud cells express both *ZMyoD* and *ZMyf-5*. It is interesting, however, that the axial tailbud cells do not express detectable levels of either myogenic factor. Axial posterior tailbud cells could represent uncommitted muscle precursors that are recruited to the paraxial mesoderm by migration to lateral paraxial regions as described by Kanki and Ho (1997) and initiate expression of *ZMyf-5* after they have escaped the repressive influence of the axial structures. The possibility that axial signals are repressing MRF expression is consistent with the observation that zebrafish *floating head (flh)* mutant embryos, which lack notochord, express *ZMyoD* in the axial midline during gastrulation (Halpern *et al.,* 1995*). The same signal may also be responsible for repression of *ZMyf-5* in the axial cells.

As the cells of the paraxial mesoderm migrate to the anterior margin of the segmental plate, both *ZMyf-5* and *ZMyoD* positive cells become incorporated into the newly formed somites. In both, zebrafish and amniotes, the first differentiating muscle cells are the most medial cells of the developing myotome (Felsenfeld *et al.,* 1991*). In zebrafish these are the *ZMyoD* expressing adaxial cells. 1-2 h after somite formation the adaxial cells elongate to span the length of the somite and then migrate radially through the somite to give rise to a superficial monolayer of cells that differentiates into slow muscle fibres. Expression of *ZMyoD* in the adaxial cells has recently been shown to be dependent on Shh signals from the notochord (Concordet *et al.,* 1996; Odenthal *et al.,* 1996; Weinberg *et al.,* 1996; Blagden *et al.,* 1997; Schier *et al.,* 1997*). After initiation of *ZMyoD* expression, subsequent exposure to Shh is not required for the adaxial cells to differentiate into slow muscle fibres suggesting that these cells are already committed to slow myoblast fate. At least one function of *ZMyoD* could therefore be to maintain cells that have been exposed to Shh in a committed state to form slow muscle until muscle differentiation is initiated. In their model, Blagden *et al.,* (1997) suggested that the decision whether to become a fast or slow myoblast is made concurrently with commitment to myoblast fate. This is in contrast with traditional models in which somitic

* see additional references on page 203

cells become committed first to become myoblasts and subsequently to fibre type specific subclasses.
Since fast muscle formation is independent of Shh expression (Blagden *et al.,* 1997) and *ZMyf-5* is expressed throughout the paraxial mesoderm, it is conceivable that *ZMyf-5* might play a role similar to Shh in the lateral presomitic cells and keep these cells committed to myoblast fate. The question remains however, how fast muscle differentiation is subsequently initiated.
Dye labelling experiments have shown, that unlike the adaxial cells, lateral presomitic cells do not migrate but develop *in situ* into deep muscle cells that give rise to fast muscle fibres (Devoto *et al.,* 1996). In contrast with *ZMyoD* which persists for some time after the onset of terminal differentiation of the myoblasts, the rapid decline of *ZMyf-5* in the newly formed somites makes it unlikely that *ZMyf-5* plays an analogous role to *ZMyoD* in fast muscle fibre differentiation. However, the decision whether to form a fast muscle fibre might be linked to the earlier phase of *ZMyf-5* expression or alternatively depend on additional extracellular signals. Interestingly expression of zebrafish myogenin appears to follow *ZMyoD* but is restricted to already formed somites (Weinberg *et al.,* 1996). *ZMyogenin* transcripts persist longer and are found in more anterior somites. *ZMyoD* and *ZMyogenin* may therefore have roles in early and late somite differentiation, respectively. Other genes like *snail1* (Hammerschmidt and Nuesslein-Vollhard, 1993*, Thisse *et al.,* 1993*) show a similar pattern of expression as *ZMyf-5* in the paraxial mesoderm and the developing somites and may be involved in the process of mesodermal segmentation (Thisse *et al.,* 1993). It will be important to study the interaction of *ZMyf-5* with these genes to gain a better understanding of these processes in vertebrate development.

14. Outlook

Using a combination of interspecies sequence comparisons, EMSA assays and transgenic analysis on the mouse *Myf-5* gene, I have identified separate regulatory regions involved in the activation of *Myf-5* in the ventral posterior somite compartment. Interestingly the regulatory regions that were identified in this study appear to have distinct functions. Elements within the 3' UTR of the *Myf-5* gene seem to act as '*enablers*', whereas regulatory elements in the introns of *Myf-5* direct expression to the ventral posterior somite compartment. The transcription factors involved are presently unknown, but by defining their target sites more closely it will be possible to reveal their identity. The regulatory region within the introns was mapped to separate conserved elements located in intron 1 between nts 461 to 645 (fragment D) and intron 2 between

* see additional referencs on page 203

nts 127 to 407 (fragments F and G) which are necessary for ventral posterior somitic expression in transgenic mice. Analysis of fragment D in EMSA identified the 84bp region at the 5' end as the most likely target site of transcription factors and additional EMSA analysis with overlapping oligonucleotide probes should delimit the binding sites further. Considering the relatively small size of fragment D, footprinting may be used to reveal the exact binding sites at nucleotide resolution.

Any putative binding site should be tested by mutation analysis. Mutation of the identified sites should abolish the binding activity *in vitro* and the expression in *in vivo* in transgenic mice. This would confirm the target site as an essential regulatory element. Because in the transgenic assays one might expected a negative result, a suitable assay system would be one where additional independent expression domains can be observed regardless of the introduced mutation. Mutation of the candidate region, in the context of HMZ17, for example, should leave the dorsal somitic expression domain and branchial arch expression unaffected, which could serve as a positive control.

If the analysis revealed a hitherto unknown transcription factor it would ultimately be necessary to embark on biochemical purification of that factor from embryonic protein extracts, determine its amino acid sequence and clone its gene. It would then be of interest to identify the expression pattern in the mouse, especially in relation to myogenesis, and perhaps to introduce null mutations into the gene.

A separate task is the characterisation of the enhancer elements in the 3' UTR between the *Ase*I and *Hin*dIII sites, involved in the regulation of the efficiency of expression of *Myf-5* . It will be of interest to examine more closely the role of spacing between the somite elements in the introns and the enhancer element in the *AseI/Hin*dIII fragment, by replacing the region extending from exon 3 to the *Ase*I site in the UTR by an independent inert DNA fragment of similar size in the context of pΔCPH.

Analysis of the *Fugu* and zebrafish *Myf-5* genes suggested that the distribution of the regulatory elements in the teleosts is different from that of the mouse. Transgenic analysis showed that in *Fugu* the somitic expression pattern is directed by the 3' UTR, while the function of the conserved region in intron 1 between *Fugu* and zebrafish is not clear. It is possible this region has important regulatory functions in teleost fish but is not functional in transgenic mice because of the large evolutionary distance between teleosts and higher vertebrates. In this regard it would be interesting to carry out EMSA analysis on this fragment with zebrafish embryonic extracts. Also, to address the issue of incompatibility between the *Fugu* transgene and the transcriptional machinery in the mouse, it might be of interest to see how the *Fugu* or zebrafish gene would behave in more closely related amphibian embryos. A suitable GFP reporter construct is currently being tested in *Xenopus* embryos. Ultimate results can however only be reached in the context of the native environment in transgenic zebrafish and are unlikely to be obtained from amphibian models.

Since conservation of synteny is observed in *Fugu* it is possible that other regulatory elements are shared between the mouse and *Fugu* genes. One possibility is that the conserved region depends on interactions with additional elements, for example from the intergenic region, that have so far not been tested, or that a different transgenic model is required to activate the conserved regulatory elements in the intron. It would therefore be interesting to examine if the intergenic region of *Fugu* could activate dorsal somitic expression as in the mouse, or if elements in the intergenic region could interact with elements in the 3' UTR and alter other aspects of the somitic expression pattern.
With the help of the *ZMyf-5* probe, it should be possible to study the expression pattern of the different myogenic factors in teleost fish using double labelling *in situ* hybridisations. It will also be of interest to examine the role of putative upstream signals like sonic hedgehog in the regulation of *ZMyf-5.* A collaboration to adress this question has been initiated.

Appendix

Oligonucleotides — Appendix I

Bandshift PCR primers for Mouse *Myf-5* intron fragments

Fragment A:	In1 for	CTGAGGGAACAGGTGGAGAAC
	Ar	TTTACTGCCATCTCCATGC
Fragment B:	Bf	GCAGTAAAAGCTTGGCTTGCC
	Br	GCCCTCTGAAAACAGTGTATC
Fragment C:	Cf	GATACACTGTTTTCAGAGGGC
	Cr	AGAGAAACCCCGCAGCAATGCG
Fragment D:	Df	GCATTGCTGCGGGGTTTCTCTC
	In1 rev	CTGTTCTTTCGGGACCAGACAG
Fragment E:	In2 for	GAAAGAACAGCAGCTTTGAC
	Er	TCTGGAGCACAAAACGGTCC
Fragment F:	Ff	CGTTTTGTGCTCCAGATTACC
	Fr	TAGGTCAGGTGATCACAG
Fragment G:	Gf	ATCACCTGGACCACCTATGC
	In2 rev	CAAGCTGGACACGGAGCTTTTATC

Bandshift Mouse *MyoD Hox*TF oligonucleotides

HOX-TF-for	TGCAGCAACCAGGGACTGGCGTGTGTC
HOX-TF-rev	GACACACGCCAGTCCCTGGTTGCTGCA

HLH probe amplification:

HLH For	AATCAACGAAGCCTTTGAG
HLH Rev	CGGCCACGTCGTCCAGGAC

Mouse *Myf-5* primer in 3' UTR used for transgenic construct generation:

UTR-R	CATGCTGTATAATTGCACCT

Fugu Myf-5 cDNA amplification:

nested forward set:

MC-A	AACGCCATCCAGTACATCGAG
MC-B	AGCAGGTGGAGAGCTACTACG

reverse primer:

MC-R	AGCCGGCGGTCTTATCGGTC

Fugu MRF4 cDNA amplification:

nested forward set:

HC-A	GCCTGCAAGATCTGCAAGC
HC-B	CCAGAGGCTGCCCAAGGT

reverse primer:

HC-R	CTGGAGGTCGCCGACGACT

Zebrafish *Myf-5* cDNA amplification:

nested forward set:

Zebra1	GGNMAYTGYYNMYNTGGGC
Zebra2	MYNTGGGCNTGYAATGCNTG

reverse primer:

Fugu Myf-5 Rev	CACCTGTTCCCGAGCAGCTCCTG

(Note: M=A or C, Y=C or T, N= A, C, G or T, degeneracy: Zebra1: 4096 Zebra2: 512, the underlined G in *Fugu* Myf-5 Rev is a mismatch of the *Fugu* and zebrafish sequence)

Zebrafish *Myf-5* intron amplification:

nested forward set:

ZM1	CAATCACGCCTTTGAGGCACTACG
ZM2	TAGCCAACGCCTCCCCAAGGTAGA

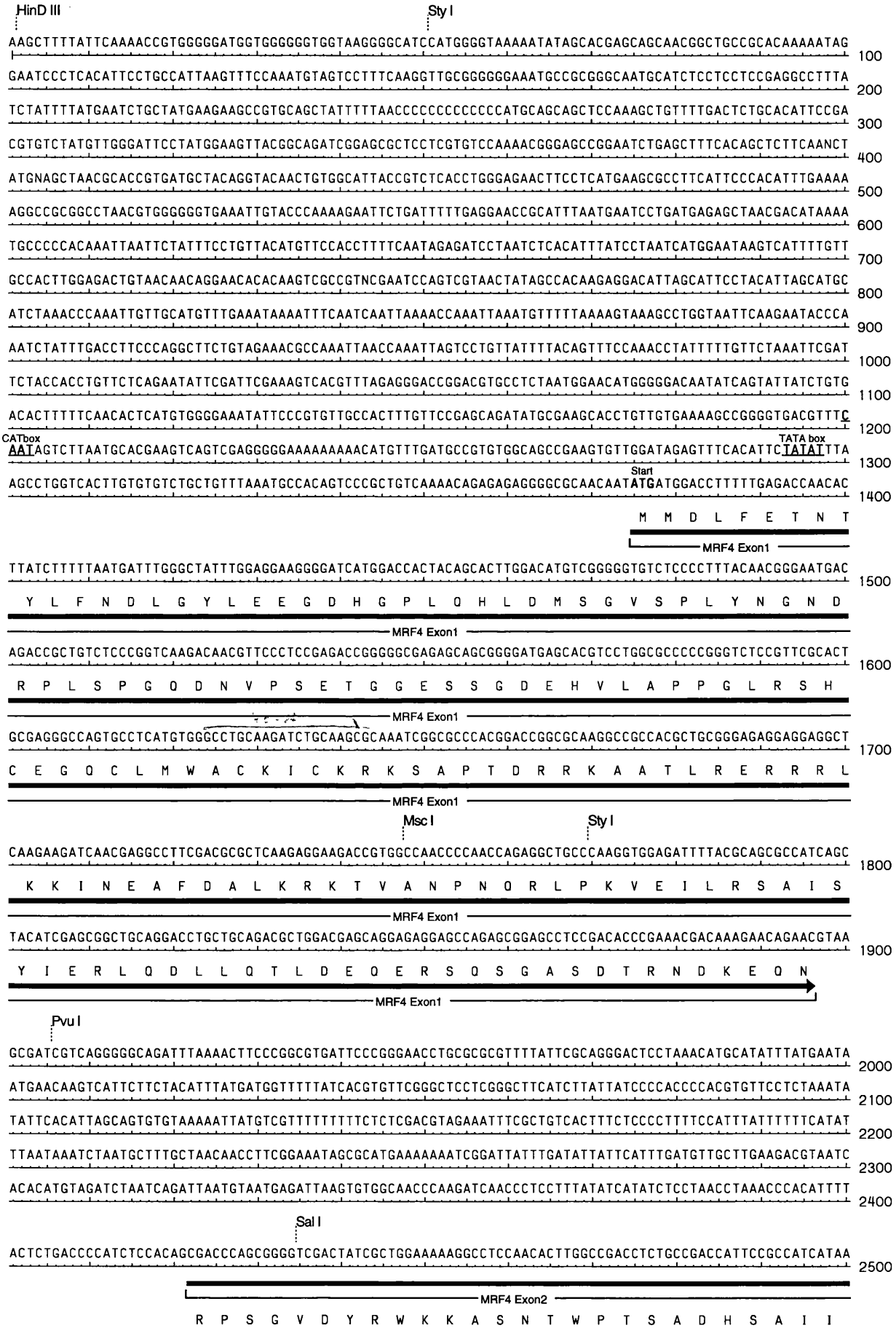
mested reverse set:

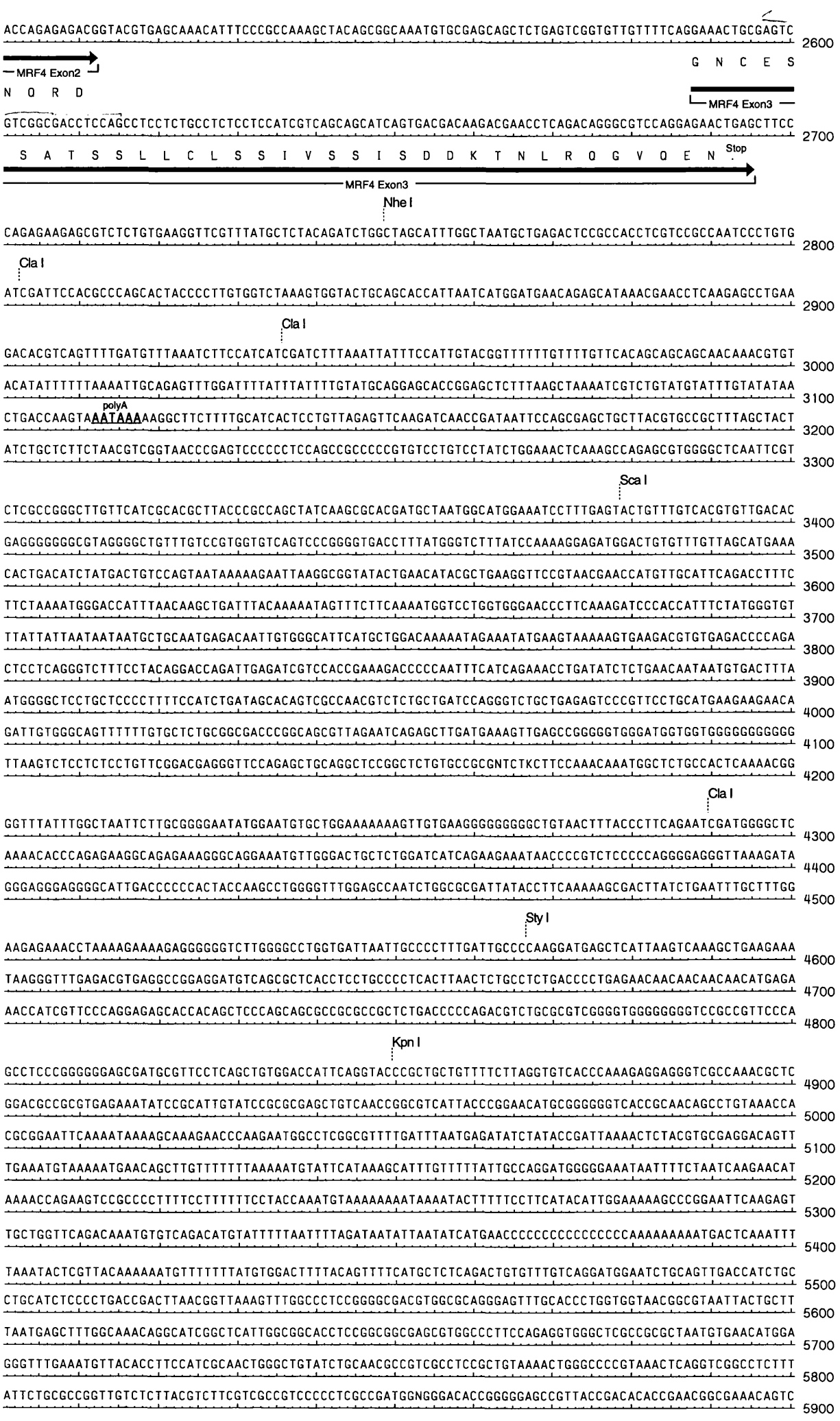
ZM1R	GTACAAGGAGTAAGGGCGTCCTAG
ZM2R	TCCTGTCAGACAGGTGCTACGACC

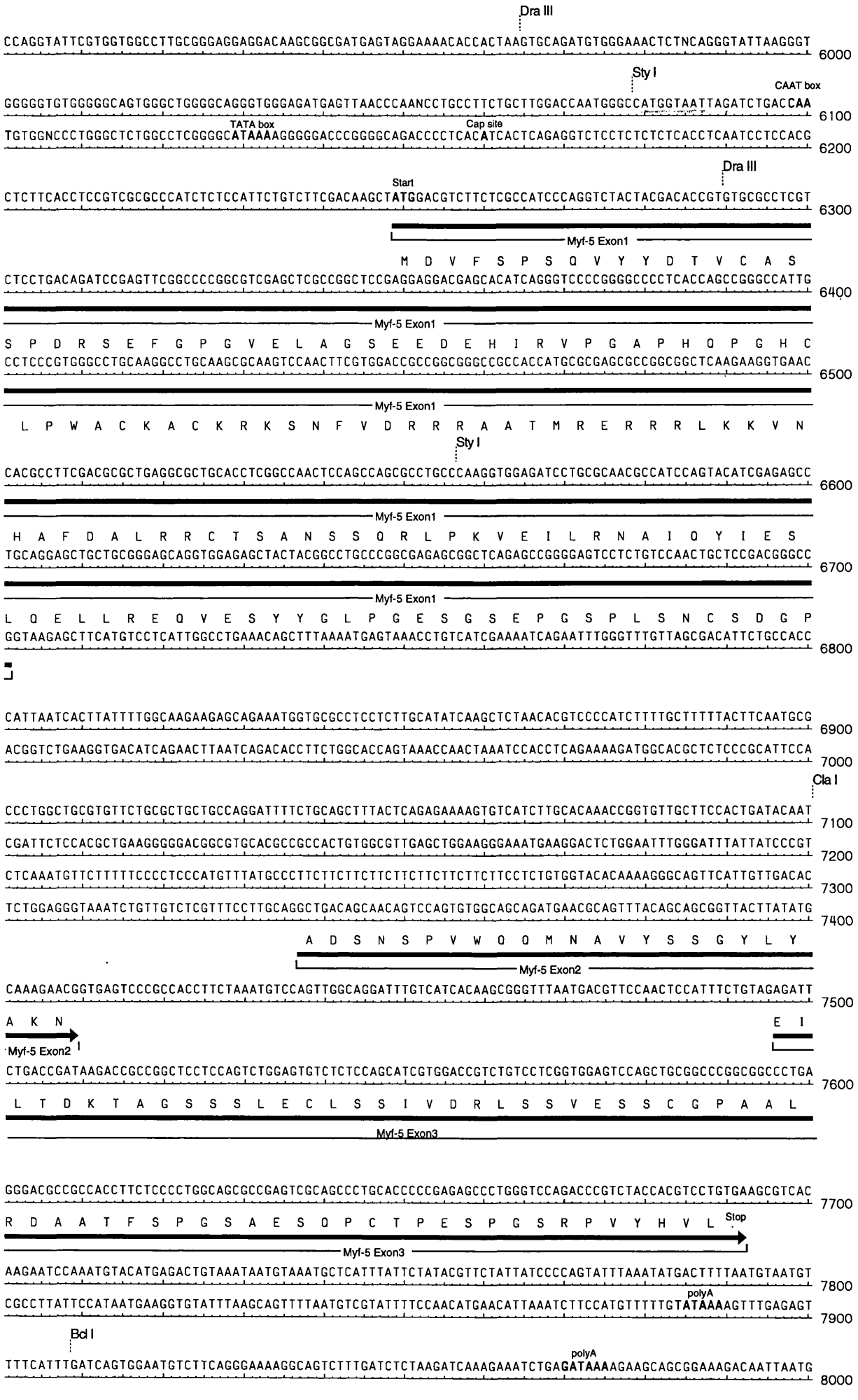
reamplificaton set:

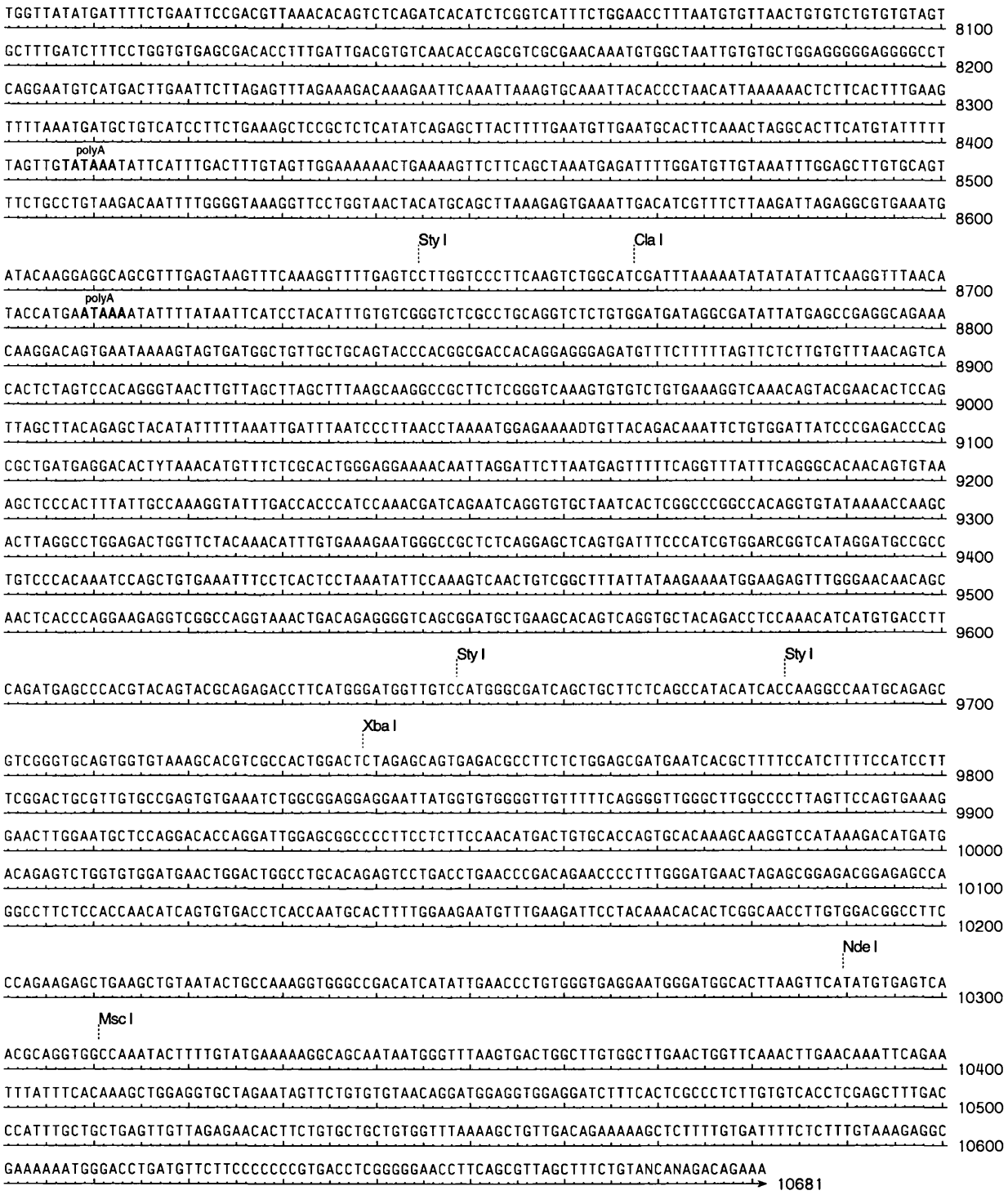
tail1	TGCGCATGCA/TAGCCAACGCCTCCCAAGGTAG
tail2	TCGCGATGCA/TAGGACAGTCTGTCCACGATGCTGG

(*Nsi*I restriction site is underlined and the cleavage site indicated by ' /')


HinD III
Sty I
100
200
300
400
500
600
700
800
900
1000
1100
1200
CATbox
TATA box
1300
Start
1400
M M D L F E T N T
MRF4 Exon1
1500
Y L F N D L G Y L E E G D H G P L Q H L D M S G V S P L Y N G N D
MRF4 Exon1
1600
R P L S P G Q D N V P S E T G G E S S G D E H V L A P P G L R S H
MRF4 Exon1
1700
C E G Q C L M W A C K I C K R K S A P T D R R K A A T L R E R R R L
MRF4 Exon1
Msc I
Sty I
1800
K K I N E A F D A L K R K T V A N P N Q R L P K V E I L R S A I S
MRF4 Exon1
1900
Y I E R L Q D L L Q T L D E Q E R S Q S G A S D T R N D K E Q N
MRF4 Exon1
Pvu I
2000
2100
2200
2300
2400
Sal I
2500
MRF4 Exon2
R P S G V D Y R W K K A S N T W P T S A D H S A I I

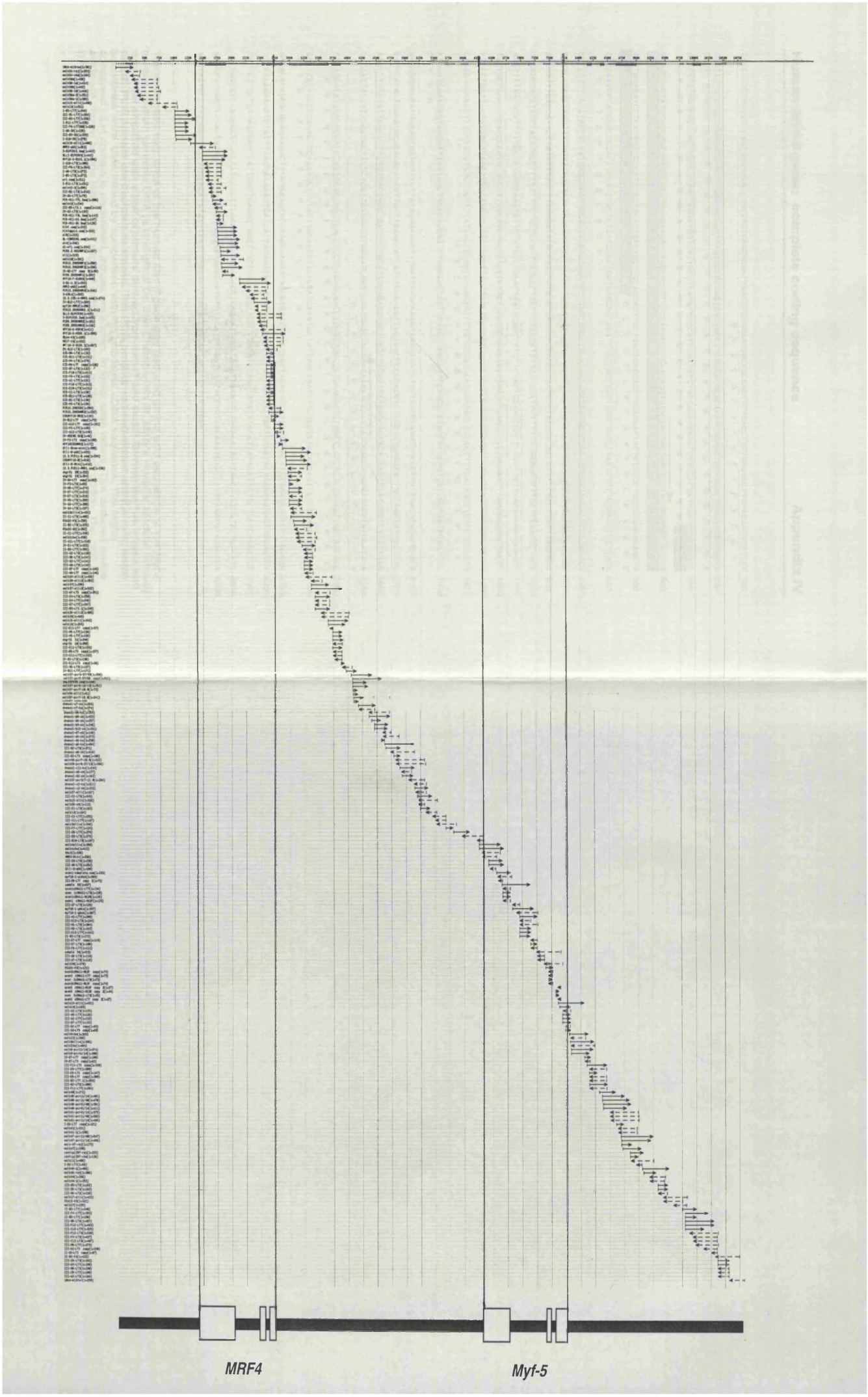

ACCAGAGAGACGGTACGTGAGCAAACATTTCCCGCCAAAGCTACAGCGGCAAATGTGCGAGCAGCTCTGAGTCGGTGTTGTTTTCAGGAAACTGCGAGTC
2600
MRF4 Exon2
G N C E S
N Q R D
MRF4 Exon3
GTCGGCGACCTCCAGCCTCCTCTGCCTCTCCTCCATCGTCAGCAGCATCAGTGACGACAAGACGAACCTCAGACAGGGCGTCCAGGAGAACTGAGCTTCC
2700
S A T S S L L C L S S I V S S I S D D K T N L R Q G V Q E N Stop
MRF4 Exon3
Nhe I
CAGAGAAGAGCGTCTCTGTGAAGGTTCGTTTATGCTCTACAGATCTGGCTAGCATTTGGCTAATGCTGAGACTCCGCCACCTCGTCCGCCAATCCCTGTG
2800
Cla I
ATCGATTCCACGCCCAGCACTACCCCTTGTGGTCTAAAGTGGTACTGCAGCACCATTAATCATGGATGAACAGAGCATAAACGAACCTCAAGAGCCTGAA
2900
Cla I
GACACGTCAGTTTTGATGTTTAAATCTTCCATCATCGATCTTTAAATTATTTCCATTGTACGGTTTTTTGTTTTGTTCACAGCAGCAGCAACAAACGTGT
3000
ACATATTTTTTAAAATTGCAGAGTTTGGATTTTATTTATTTTGTATGCAGGAGCACCGGAGCTCTTTAAGCTAAAATCGTCTGTATGTATTTGTATATAA
3100
polyA
CTGACCAAGTAAATAAAAAGGCTTCTTTTGCATCACTCCTGTTAGAGTTCAAGATCAACCGATAATTCCAGCGAGCTGCTTACGTGCCGCTTTAGCTACT
3200
ATCTGCTCTTCTAACGTCGGTAACCCGAGTCCCCCCTCCAGCCGCCCCCGTGTCCTGTCCTATCTGGAAACTCAAAGCCAGAGCGTGGGGCTCAATTCGT
3300
Sca I
CTCGCCGGGCTTGTTCATCGCACGCTTACCCGCCAGCTATCAAGCGCACGATGCTAATGGCATGGAAATCCTTTGAGTACTGTTTGTCACGTGTTGACAC
3400
GAGGGGGGGCGTAGGGGCTGTTTGTCCGTGGTGTCAGTCCCGGGGTGACCTTTATGGGTCTTTATCCAAAAGGAGATGGACTGTGTTTGTTAGCATGAAA
3500
CACTGACATCTATGACTGTCCAGTAATAAAAAGAATTAAGGCGGTATACTGAACATACGCTGAAGGTTCCGTAACGAACCATGTTGCATTCAGACCTTTC
3600
TTCTAAAATGGGACCATTTAACAAGCTGATTTACAAAAATAGTTTCTTCAAAATGGTCCTGGTGGGAACCCTTCAAAGATCCCACCATTTCTATGGGTGT
3700
TTATTATTAATAATAATGCTGCAATGAGACAATTGTGGGCATTCATGCTGGACAAAAATAGAAATATGAAGTAAAAAGTGAAGACGTGTGAGACCCCAGA
3800
CTCCTCAGGGTCTTTCCTACAGGACCAGATTGAGATCGTCCACCGAAAGACCCCCAATTTCATCAGAAACCTGATATCTCTGAACAATAATGTGACTTTA
3900
ATGGGGCTCCTGCTCCCCTTTTCCATCTGATAGCACAGTCGCCAACGTCTCTGCTGATCCAGGGTCTGCTGAGAGTCCCGTTCCTGCATGAAGAAGAACA
4000
GATTGTGGGCAGTTTTTTGTGCTCTGCGGCGACCCGGCAGCGTTAGAATCAGAGCTTGATGAAAGTTGAGCCGGGGGTGGGATGGTGGTGGGGGGGGGGG
4100
TTAAGTCTCCTCTCCTGTTCGGACGAGGGTTCCAGAGCTGCAGGCTCCGGCTCTGTGCCGCGNTCTKCTTCCAAACAAATGGCTCTGCCACTCAAAACGG
4200
Cla I
GGTTTATTTGGCTAATTCTTGCGGGGAATATGGAATGTGCTGGAAAAAAAGTTGTGAAGGGGGGGGGCTGTAACTTTACCCTTCAGAATCGATGGGGCTC
4300
AAAACACCCAGAGAAGGCAGAGAAAGGGCAGGAAATGTTGGGACTGCTCTGGATCATCAGAAGAAATAACCCCGTCTCCCCCAGGGGAGGGTTAAAGATA
4400
GGGAGGGAGGGGCATTGACCCCCCACTACCAAGCCTGGGGTTTGGAGCCAATCTGGCGCGATTATACCTTCAAAAAGCGACTTATCTGAATTTGCTTTGG
4500
Sty I
AAGAGAAACCTAAAAGAAAAGAGGGGGGTCTTGGGGCCTGGTGATTAATTGCCCCTTTGATTGCCCCAAGGATGAGCTCATTAAGTCAAAGCTGAAGAAA
4600
TAAGGGTTTGAGACGTGAGGCCGGAGGATGTCAGCGCTCACCTCCTGCCCCTCACTTAACTCTGCCTCTGACCCCTGAGAACAACAACAACAACATGAGA
4700
AACCATCGTTCCCAGGAGAGCACCACAGCTCCCAGCAGCGCCGCGCCGCTCTGACCCCCAGACGTCTGCGCGTCGGGGTGGGGGGGGTCCGCCGTTCCCA
4800
Kpn I
GCCTCCCGGGGGGAGCGATGCGTTCCTCAGCTGTGGACCATTCAGGTACCCGCTGCTGTTTTCTTAGGTGTCACCCAAAGAGGAGGGTCGCCAAACGCTC
4900
GGACGCCGCGTGAGAAATATCCGCATTGTATCCGCGCGAGCTGTCAACCGGCGTCATTACCCGGAACATGCGGGGGGTCACCGCAACAGCCTGTAAACCA
5000
CGCGGAATTCAAAATAAAAGCAAAGAACCCAAGAATGGCCTCGGCGTTTTGATTTAATGAGATATCTATACCGATTAAAACTCTACGTGCGAGGACAGTT
5100
TGAAATGTAAAAATGAACAGCTTGTTTTTTTTAAAAATGTATTCATAAAGCATTTGTTTTTATTGCCAGGATGGGGGAAATAATTTTCTAATCAAGAACAT
5200
AAAACCAGAAGTCCGCCCCTTTTCCTTTTTTCCTACCAAATGTAAAAAAAAATAAAATACTTTTTCCTTCATACATTGGAAAAAGCCCGGAATTCAAGAGT
5300
TGCTGGTTCAGACAAATGTGTCAGACATGTATTTTTAATTTTAGATAATATTAATATCATGAACCCCCCCCCCCCCCCAAAAAAAAATGACTCAAATTT
5400
TAAATACTCGTTACAAAAAATGTTTTTTTTATGTGGACTTTTACAGTTTTCATGCTCTCAGACTGTGTTTGTCAGGATGGAATCTGCAGTTGACCATCTGC
5500
CTGCATCTCCCCTGACCGACTTAACGGTTAAAGTTTGGCCCTCCGGGGCGACGTGGCGCAGGGAGTTTGCACCCTGGTGGTAACGGCGTAATTACTGCTT
5600
TAATGAGCTTTGGCAAACAGGCATCGGCTCATTGGCGGCACCTCCGGCGGCGAGCGTGGCCCTTCCAGAGGTGGGCTCGCCGCGCTAATGTGAACATGGA
5700
GGGTTTGAAATGTTACACCTTCCATCGCAACTGGGCTGTATCTGCAACGCCGTCGCCTCCGCTGTAAAACTGGGCCCCGTAAACTCAGGTCGGCCTCTTT
5800
ATTCTGCGCCGGTTGTCTCTTACGTCTTCGTCGCCGTCCCCCTCGCCGATGGNGGGACACCGGGGGAGCCGTTACCGACACACCGAACGGCGAAACAGTC
5900


Dra III
Sty I
CAAT box
TATA box
Cap site
Dra III
Start
Myf-5 Exon1
Myf-5 Exon1
Myf-5 Exon1
Sty I
Myf-5 Exon1
Myf-5 Exon1
Cla I
Myf-5 Exon2
Myf-5 Exon2
Myf-5 Exon3
Stop
Myf-5 Exon3
polyA
polyA
Bcl I



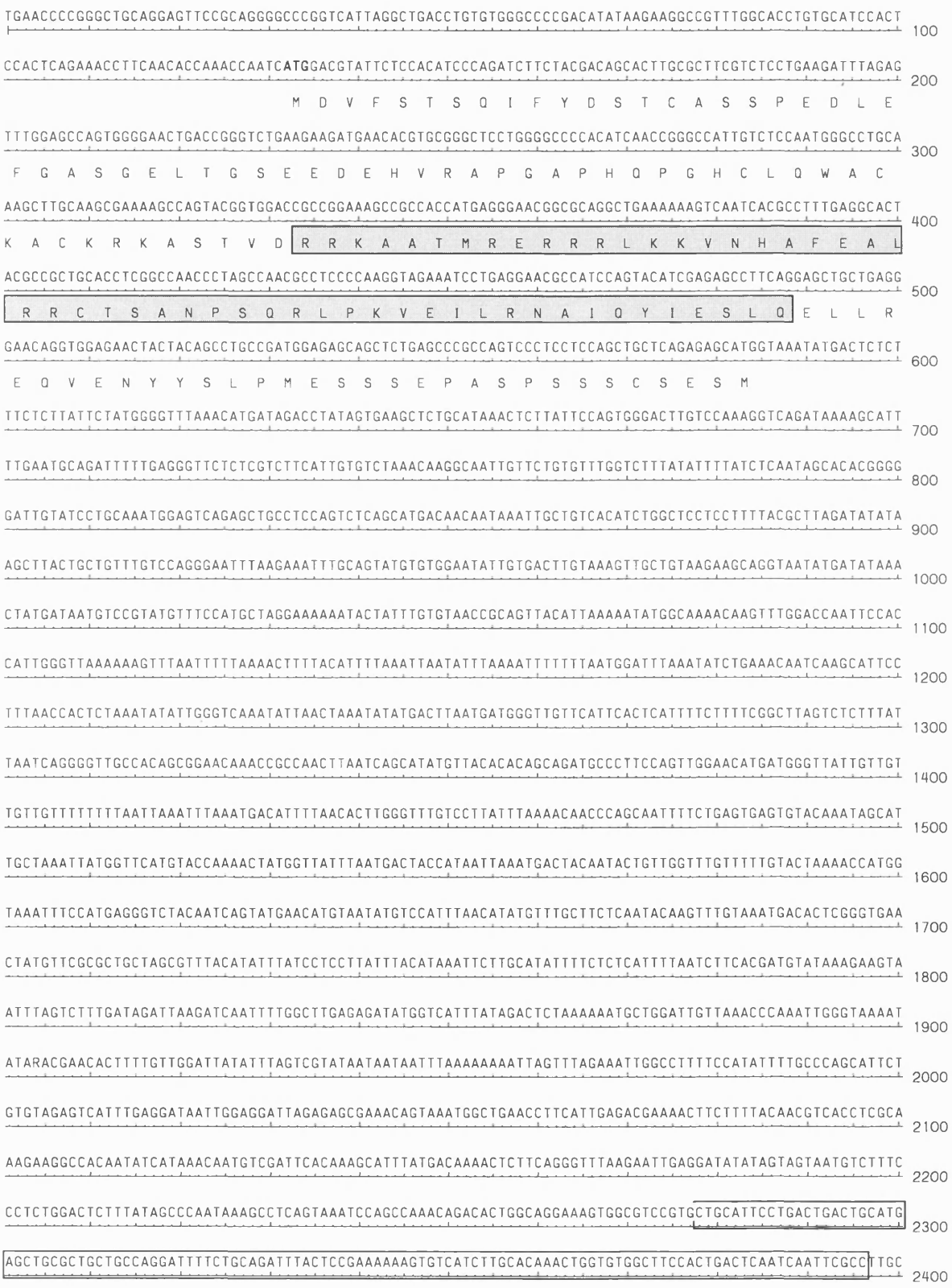
A 10681bp *Hin*dIII subclone of a *Fugu* cosmid containing the *MRF4* and *Myf-5* genes and flanking genomic region was isolated and the nucleotide sequence determined. The genomic structure was derived from cDNA sequence analysis. Exons are underlined and labelled, and the deduced amino acids are shown as one letter code, aligned with the second nucleotide of each codon. Selected restriction sites are indicated above the sequence. Putative transcription start sites (Start), TATA boxes, CAT boxes and polyadenylation sites (polyA) as predicted by GRAIL and labelled as indicated.

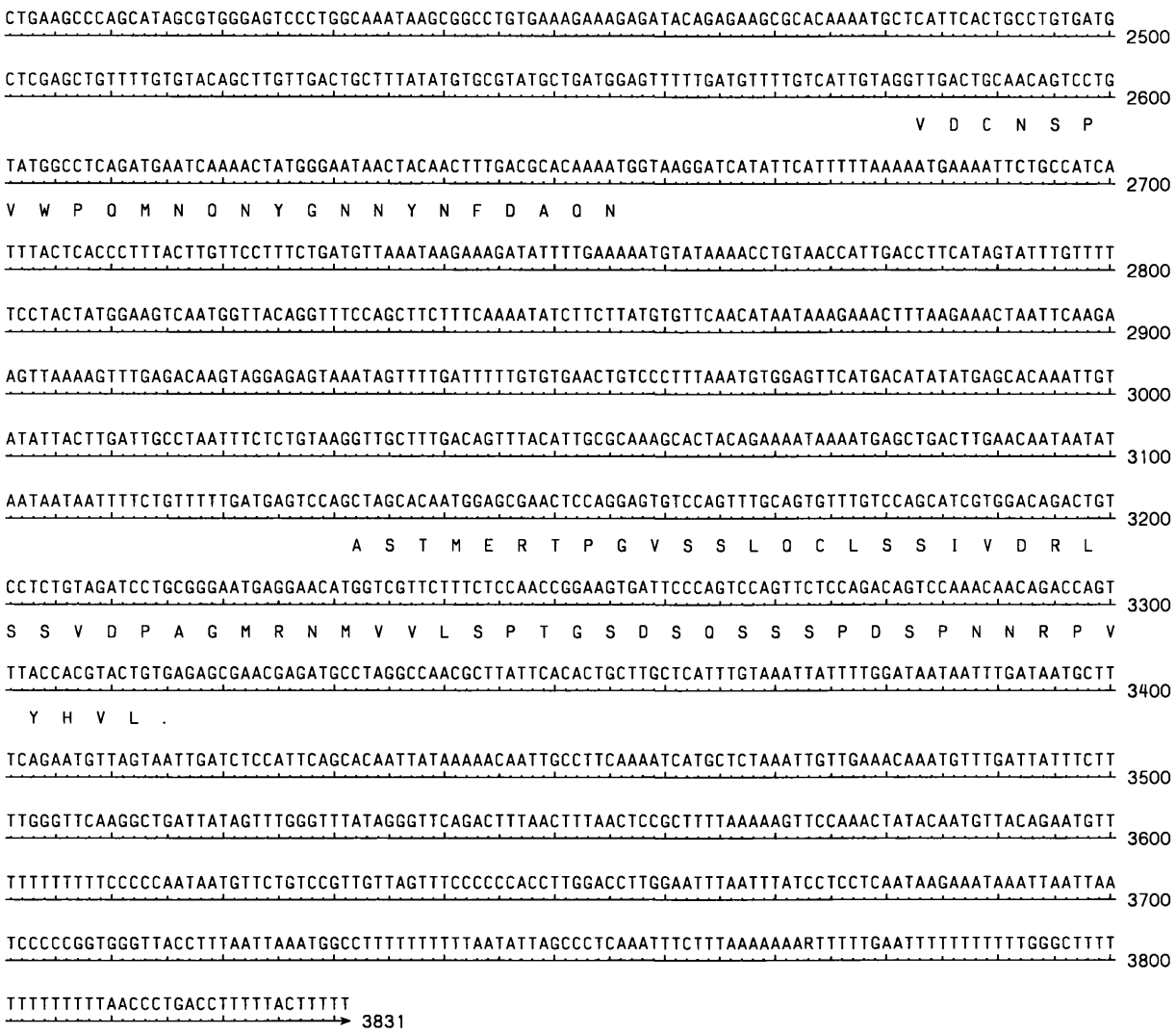
A 10681bp *Hin*dIII subclone of a *Fugu* cosmid containing the *MRF4* and *Myf-5* genes was isolated and the nucleotide sequence determined by shotgun sequencing using a variation of the dideoxynucleotide chain termination method (Sanger *et al.*, 1977). A total of approximately 415 individual sequence reads (arrows) was assembled, initially into 38 contigs. Gaps between the contigs were closed by 48 primer walks, until all contigs merged. The average coverage was 7.4 fold. The schematic shows the name, position, length and direction of the individual sequence reads. The positions of the *MRF4* and *Myf-5* genes and their general genomic structure is illustrated underneath the sequence reads. Exons are depicted as open boxes, and introns or intergenic sequences are represented by a solid bar. The genomic organisation was initially deduced by performing comparisons with the NCBI database. Matches of the *Fugu* sequence obtained with entries in the database were aligned along the contig to position the genes. The exact exon-intron structure was subsequently determined by sequencing *Fugu Myf-5* and *MRF4* muscle cDNA. The organisation of the *Fugu* genes is very similar to that of the mouse *MRF4* and *Myf-5* genes and the other myogenic regulatory genes. Both have three exons separated by two introns. The first, and largest, exon in both genes is the most highly conserved, owing to the characteristic basic-helix loop-helix (bHLH) domain shared by all of the myogenic factors. The intergenic distance between *MRF-4* and *Myf-5* is much larger than might be expected for the small genome of the pufferfish and there is no significant sequence conservation outside of the coding region between *Fugu* and higher vertebrates.


MRF4
Myf-5

```
CCTCTCGCTGCCGTCCAGGTGCACCGCCTGCCTCTCAGCAGGATGGACGTGATGGATGGCTGCCAGTTCTCACCTTCTGAGTACTTCTACGACGGCTCCT 100
                                                  M  D  G  C  Q  F  S  P  S  E  Y  F  Y  D  G  S

GCATACCGTCCCCCGAGGGTGAATTTGGGGACGAGTTTGTGCCGCGAGTGGCTGCCTTCGGAGCGCACAAAGCAGAGCTGCAGGGCTCAGATGAGGACGA 200
C  I  P  S  P  E  G  E  F  G  D  E  F  V  P  R  V  A  A  F  G  A  H  K  A  E  L  Q  G  S  D  E  D  E

GCACGTGCGAGCGCCTACCGGCCACCACCAGGCTGGTCACTGCCTCATGTGGGCCTGCAAAGCCTGCAAGAGGAAGTCCACCACCATGGATCGGCGGAAG 300
  H  V  R  A  P  T  G  H  H  Q  A  G  H  C  L  M  W  A  C  K  A  C  K  R  K  S  T  T  M  D  R  R  K

GCAGCCACTATGCGCGAGCGGAGGCGCCTGAAGAAGGTCAACCAGGCTTTCGAAACCCTCAAGAGGTGTACCACGACCAACCCCAACCAGAGGCTGCCCA 400
  A  A  T  M  R  E  R  R  R  L  K  K  V  N  Q  A  F  E  T  L  K  R  C  T  T  T  N  P  N  Q  R  L  P

AGGTGGAGATCCTCAGGAATGCCATCCGCTACATCGAGAGCCTGCAGGAGTTGCTGAGAGAGCAGGTGGAGAACTACTATAGCCTGCCGGGACAGAGCTG 500
K  V  E  I  L  R  N  A  I  R  Y  I  E  S  L  Q  E  L  L  R  E  Q  V  E  N  Y  Y  S  L  P  G  Q  S  C

CTCGGAGCCCACCAGCCCCACCTCCAACTGCTCTGATGGCATGGTAAGCAATAGATCTGGTACCTGCTAGGCTACCCTAATCTTTTCTAAAGTCCTTACA 600
  S  E  P  T  S  P  T  S  N  C  S  D  G

CCTCATTTAACCTGGTGTGGTGGGGAGAGTGGGGTGGAGGCAGATGCTGAGTTGCTTTAAAAGAAGAGAGGGGTCCACATTTAGAAAGACTCCCCAAACC 700

GCTGCTGAACAAGATTTTAGTTTAACTTTCTAGCAGGTTCTAGGTGTACACTGTAATCGAGTGTTGACATGGAAGATGGGTGGCTGTGAATGATCACTCA 800

GATGTTTTCTCCATTCCTGAATTTATTTTCAAAATATGCCATCTGTGGATCATGCCCTACGCTAATATCTAAAGGCACCGTTTCTAACTTAATGAGGAAA 900

TGGAAAGAAATACCCACACGGCCCCAGTTCCTGCTCCAATGAGGCCTGGCTGAAAGATGTTGATGCATTCTTTTTAGAGGGCGTTTGCTCCAAGGCTGCC 1000

AGGTTTTAATGTGTTTTTGCCCTGGGAAAGTGTTCTTTCCCTGAATTAGTGTGGCTTTCTTCTACTCCAATCCATTTTGCATGGTTAACCCAATGCACAT 1100

TGCTGCTGAATTCCACCCCCTCTTCCCTTTGCTGCTGCTCTCCTCTTCTTCAAGCACAGAGATTGACCTCAGTGCCCTGGGAATTTGGAGAGGGCTACCC 1200

CTTCCTAAATCAAGGCAGTGAAGGTGACTGACAGTGTTCGGTTACAGAGCTGGTGGGCAAGCACAGCCTCACCTTTGGTCAGAACATCTTTTGCCAAAAC 1300

CTGAAAACAAACTTTGTTGTGTGTCTTGTATTATAGCCCGAATGTAACAGTCCTGTCTGGTCCAGAAAGAGCAGTACTTTTGACAGCATCTACTGTCCTG 1400
                                  P  E  C  N  S  P  V  W  S  R  K  S  S  T  F  D  S  I  Y  C  P

ATGTATCAAATGGTAAGAATTGATAACTTCACAGGAGTTTAAAGACCAGTTCAACCTAACAATTCAGCCTATAACATTCTGTTCTTGCTGATAGTATTGG 1500
D  V  S  N

GGAAGGGAGAATGGAAGTGATGGTTCTTATAGGGAGGCTTTGGTAAAGCAAAATAAACACATCTTCTGCTCCAAATCCCCCTAGCAGACACGCACGCACA 1600

CATGCATACACACATGCACACACAATGTTGCTTGAAATATTATCAGGGGGCTTCCCCACTCCCCACGTCTACCCCTCAGGAATTGCCAGATATTTGTTGC 1700

AAATTTCTATGTTAGGCTTTCTGTGACCACCTGACCTCTGGGTGTCAGAGGAGCTGACCTACAATTTAAGGAGCAACATAAGCAAATCTGTCTATCTTGG 1800

GCTAATTATTTTTTAATGCTTTTCTCCTTGTATCCTTAGTATCAAATGTATATGCCACAGATAAAAACTCCTTATCCAGCTTGGATTGCTTATCCATTTA 1900
                                              V  Y  A  T  D  K  N  S  L  S  S  L  D  C  L  S  I  Y

TTTAGGATATGGATGGGACATAGTGGACCGGATCACCTCCTCAGAGCAACCTGGGTTGCCTCTCCAGGATCTGGCTTCTCTCTCTCCAGTTGCCAGCACC 2000
  L  G  Y  G  W  D  I  V  D  R  I  T  S  S  E  Q  P  G  L  P  L  Q  D  L  A  S  L  S  P  V  A  S  T

GATTCACAGCCTCGAACTCCAGGGGCTTCTAGTTCCAGGCTTATCTATCATGTGCTATGAACTAATTTTCTGGTCTATATGACTTCTTCCAGGAGGGCCT 2100
D  S  Q  P  R  T  P  G  A  S  S  S  R  L  I  Y  H  V  L  .

AATACACAGGACGAAGAAGGCTTCAAAAAGTCCCAAACCAAGACAACATGTACATAAAGATTTCTTTTCAGTTGTAAATTTGTAAAGATTACCTTGCCAC 2200

TTTATAAGAAAGTGTATTTAACTAAAAAGTCATCATTGCAAATAATACTTTCTTCTTCTTTATTATTCTTTGCTTAGATATTAATACATAGTTCCAGTAA 2300

TACTATTTCTGATAGGGGGCCATTGATTGAGGGTAGCTTGTTCGAATGCTTAACTTATATATACATATATATATTATAAATATTGCTCATCAAAATGT 2400

CTCTGGTGTTTAGAGCTTTATTTTTTTCTTTAAAACATTAAAACAGCTGAGAATCAGTTAAATGGAATTTTAAATATATTTAACTATTTCTTTTCTCTTT 2500

AATCCTTTAGTTATATTGTATTAAATAAAAATATAATACTGCCTAATGTATATATTTTGATCTTTTCTTGTAAGAAATGTATCTTTTAAATGTAAGCACA 2600

AAATAGTACTTTGTGGATCATTTCAAGATATAAGAAATTTTGGAAATTCCACCATAAATAAAATTTTT 2668
```

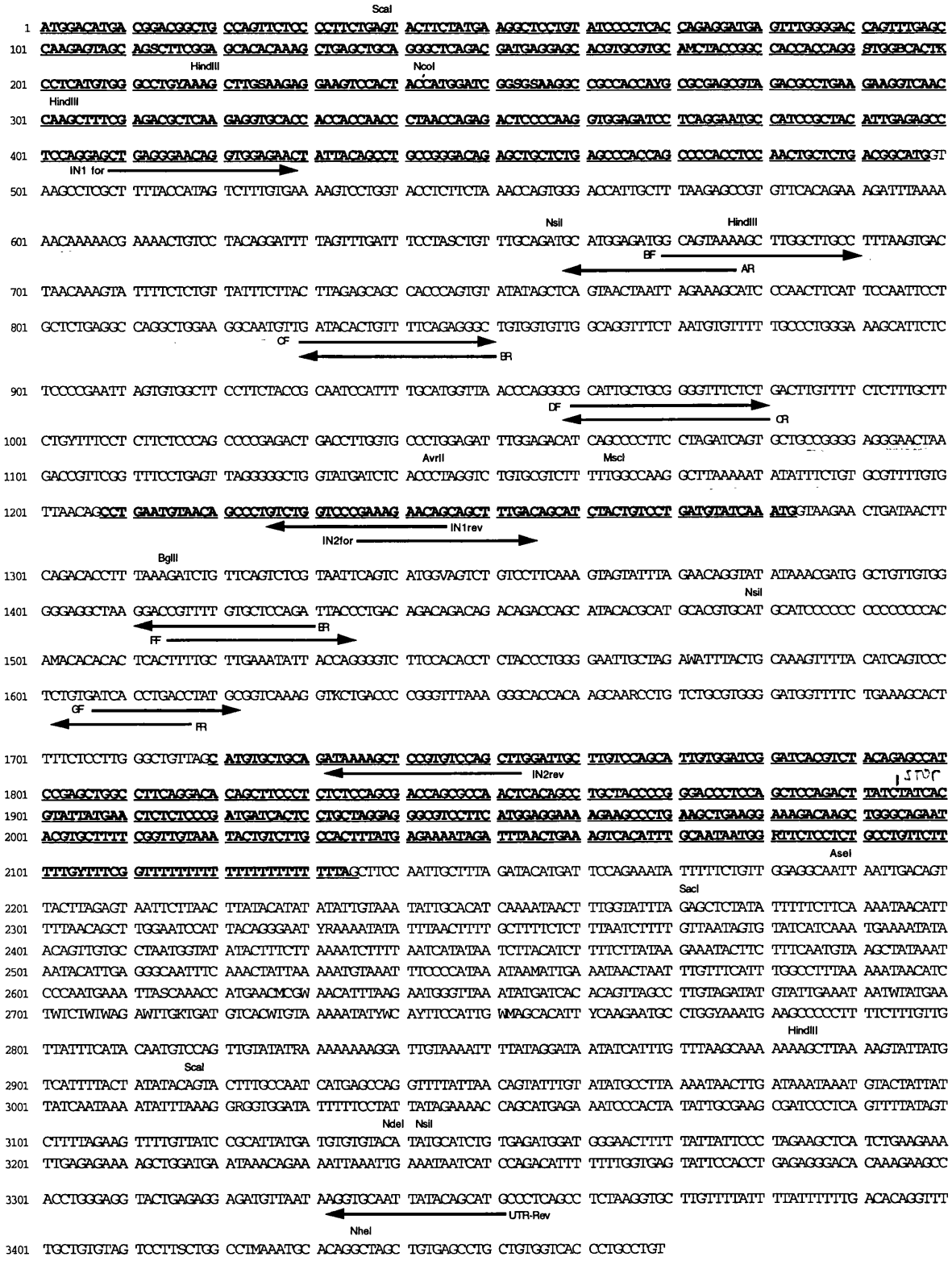
The nucleotide sequence of the human *Myf-5* gene and deduced amino acid sequence of the *Myf-5* protein. Human genomic DNA (a kind gift of Dr. Sue Chamberlain St. Mary's Hospital Medical School) was PCR amplified with a pair of primers derived from exon1 and exon3 flanking the two introns and exon2. PCR primers are depicted as arrows. The PCR product was sequenced and the exon intron boundaries were determined by comparison with the published cDNA sequence. The highly conserved HLH domain is underlined. Introns are shown in lower case letters. The deduced amino acids are shown as one letter code, aligned with the second nucleotide of each codon. The stop codon is marked as a dot.


TGAACCCCGGGCTGCAGGAGTTCCGCAGGGGCCCGGTCATTAGGCTGACCTGTGTGGGCCCCGACATATAAGAAGGCCGTTTGGCACCTGTGCATCCACT
100
CCACTCAGAAACCTTCAACACCAAACCAATCATGGACGTATTCTCCACATCCCAGATCTTCTACGACAGCACTTGCGCTTCGTCTCCTGAAGATTTAGAG
200
M D V F S T S Q I F Y D S T C A S S P E D L E
TTTGGAGCCAGTGGGGAACTGACCGGGTCTGAAGAAGATGAACACGTGCGGGCTCCTGGGGCCCCACATCAACCGGGCCATTGTCTCCAATGGGCCTGCA
300
F G A S G E L T G S E E D E H V R A P G A P H Q P G H C L Q W A C
AAGCTTGCAAGCGAAAAGCCAGTACGGTGGACCGCCGGAAAGCCGCCACCATGAGGGAACGGCGCAGGCTGAAAAAAGTCAATCACGCCTTTGAGGCACT
400
K A C K R K A S T V D R R K A A T M R E R R R L K K V N H A F E A L
ACGCCGCTGCACCTCGGCCAACCCTAGCCAACGCCTCCCCAAGGTAGAAATCCTGAGGAACGCCATCCAGTACATCGAGAGCCTTCAGGAGCTGCTGAGG
500
R R C T S A N P S Q R L P K V E I L R N A I Q Y I E S L Q E L L R
GAACAGGTGGAGAACTACTACAGCCTGCCGATGGAGAGCAGCTCTGAGCCCGCCAGTCCCTCCTCCAGCTGCTCAGAGAGCATGGTAAATATGACTCTCT
600
E Q V E N Y Y S L P M E S S S E P A S P S S S C S E S M
TTCTCTTATTCTATGGGGTTTAAACATGATAGACCTATAGTGAAGCTCTGCATAAACTCTTATTCCAGTGGGACTTGTCCAAAGGTCAGATAAAAGCATT
700
TTGAATGCAGATTTTTGAGGGTTCTCTCGTCTTCATTGTGTCTAAACAAGGCAATTGTTCTGTGTTTGGTCTTTATATTTTATCTCAATAGCACACGGGG
800
GATTGTATCCTGCAAATGGAGTCAGAGCTGCCTCCAGTCTCAGCATGACAACAATAAATTGCTGTCACATCTGGCTCCTCCTTTTACGCTTAGATATATA
900
AGCTTACTGCTGTTTGTCCAGGGAATTTAAGAAATTTGCAGTATGTGTGGAATATTGTGACTTGTAAAGTTGCTGTAAGAAGCAGGTAATATGATATAAA
1000
CTATGATAATGTCCGTATGTTTCCATGCTAGGAAAAAATACTATTTGTGTAACCGCAGTTACATTAAAAATATGGCAAAACAAGTTTGGACCAATTCCAC
1100
CATTGGGTTAAAAAAGTTTAATTTTTAAAACTTTTACATTTTAAATTAATATTTAAAATTTTTTTAATGGATTTAAATATCTGAAACAATCAAGCATTCC
1200
TTTAACCACTCTAAATATATTGGGTCAAATATTAACTAAATATATGACTTAATGATGGGTTGTTCATTCACTCATTTTCTTTTCGGCTTAGTCTCTTTAT
1300
TAATCAGGGGTTGCCACAGCGGAACAAACCGCCAACTTAATCAGCATATGTTACACACAGCAGATGCCCTTCCAGTTGGAACATGATGGGTTATTGTTGT
1400
TGTTGTTTTTTTTAATTAAATTTAAATGACATTTTAACACTTGGGTTTGTCCTTATTTAAAACAACCCAGCAATTTTCTGAGTGAGTGTACAAATAGCAT
1500
TGCTAAATTATGGTTCATGTACCAAAACTATGGTTATTTAATGACTACCATAATTAAATGACTACAATACTGTTGGTTTGTTTTTGTACTAAAACCATGG
1600
TAAATTTCCATGAGGGTCTACAATCAGTATGAACATGTAATATGTCCATTTAACATATGTTTGCTTCTCAATACAAGTTTGTAAATGACACTCGGGTGAA
1700
CTATGTTCGCGCTGCTAGCGTTTACATATTTATCCTCCTTATTTACATAAATTCTTGCATATTTTCTCTCATTTTAATCTTCACGATGTATAAAGAAGTA
1800
ATTTAGTCTTTGATAGATTAAGATCAATTTTGGCTTGAGAGATATGGTCATTTATAGACTCTAAAAAATGCTGGATTGTTAAACCCAAATTGGGTAAAAT
1900
ATARACGAACACTTTTGTTGGATTATATTTAGTCGTATAATAATAATTTAAAAAAAATTAGTTTAGAAATTGGCCTTTTCCATATTTTGCCCAGCATTCT
2000
GTGTAGAGTCATTTGAGGATAATTGGAGGATTAGAGAGCGAAACAGTAAATGGCTGAACCTTCATTGAGACGAAAACTTCTTTTACAACGTCACCTCGCA
2100
AAGAAGGCCACAATATCATAAACAATGTCGATTCACAAAGCATTTATGACAAAACTCTTCAGGGTTTAAGAATTGAGGATATATAGTAGTAATGTCTTTC
2200
CCTCTGGACTCTTTATAGCCCAATAAAGCCTCAGTAAATCCAGCCAAACAGACACTGGCAGGAAAGTGGCGTCCGTGCTGCATTCCTGACTGACTGCATG
2300
AGCTGCGCTGCTGCCAGGATTTTCTGCAGATTTACTCCGAAAAAAGTGTCATCTTGCACAAACTGGTGTGGCTTCCACTGACTCAATCAATTCGCCTTGC
2400



A cDNA library of post-somitogenesis zebrafish constructed by Robert Riggleman and Karthryn Helde in Lambda ZAP II (Stratagene), distributed by David Jonah Grunwald (Dept. of Human Genetics, University of Utah, USA), was screened with a PCR probe derived from the hlh domain of zebrafish *Myf-5* using a combination of degenerate primers and *Fugu* primers (see Chapter 2, section 8 for details). A phage clone was isolated; the insert was PCR amplified and sequenced. The intron sequence was subsequently determined by analysis of PCR products obtained from genomic zebrafish DNA (as described in Chapter 2). The complete sequence of 3831bp comprises the entire coding region and most of the 5' and 3' untranslated regions. The deduced amino acids are shown as one letter code, aligned with the second nucleotide of each codon.

The organisation of the zebrafish *Myf-5* gene *(ZMyf-5)* is similar to that of other vertebrate species and other myogenic regulatory genes. *ZMyf-5* has three exons separated by two introns. The first, and largest, exon is the most highly conserved, owing to the characteristic basic-helix loop-helix (bHLH) domain (boxed in grey) and this motif is shared by all of the myogenic factors. Interestingly, the length of the first intron of *ZMyf-5* is significanly larger than that of any other species examined. In fact, the *ZMyf-5* intron is larger than the entire *Fugu Myf-5* gene. This means that the two teleost homologues have the greatest difference in size. Despite the size difference, sequence comparisons between the *Myf-5* genes of *Fugu* and zebrafish showed a highly conserved region of 119bp near the end of the first intron (boxed). The functional signifcance of this conserved region is presently not known.



The nucleotide sequence of the mouse *Myf-5* gene was compiled from the published cDNA sequence (Buonanno *et al.*, 1992) and genomic sequence obtained by Summerbell and Halai from our laboratory. The Exons are bold and underlined. Primers are indicated by arrows and the position of restriction sites is indicated.

Bibliography

Aissani B. and Bernardi G. (1991). CpG islands: features and distribution in the genomes of vertebrates. *Gene.* **106** 173-83.

Altschul S.F., Gish W., Miller W., Myers E.W. and Lipman D.J. (1990). Basic local alignment search tool. *J Mol Biol.* **215** 403-10.

Antequera F. and Bird A. (1993). Number of CpG islands and genes in human and mouse. *Proc Natl Acad Sci U S A.* **90** 11995-9.

Aoyama H. and Asamoto K. (1988). Determination of somite cells: independence of cell differentiation and morphogenesis. *Development.* **104** 15-28.

Aparicio S., Hawker K., Cottage A., Mikawa Y., Zuo L., Venkatesh B., Chen E., Krumlauf R. and Brenner S. (1997). Organization of the *Fugu rubripes Hox* clusters: evidence for continuing evolution of vertebrate *Hox* complexes. *Nat Genet.* **16** 79-83.

Aparicio S., Morrison A., Gould A., Gilthorpe J., Chaudhuri C., Rigby P., Krumlauf R. and Brenner S. (1995). Detecting conserved regulatory elements with the model genome of the Japanese puffer fish, *Fugu rubripes. Proc Natl Acad Sci U S A.* **92.** 1684-8.

Araki I., Terazawa K. and Satoh N. (1996). Duplication of an *amphioxus* myogenic bHLH gene is independent of vertebrate myogenic bHLH gene duplication. *Gene.* **171** 231-236.

Atchley W.R., Fitch W.M. and Bronner-Fraser M. (1994). Molecular evolution of the *MyoD* family of transcription factors. *Proc Natl Acad Sci U S A.* **91** 11522-6.

Barth JL, Worrell RA, Crawford JM, Morris J and Ivarie R. (1993). Isolation, sequence, and characterization of the bovine myogenic factor-encoding gene *myf-5. Gene.* **127.** 185-91.

Baxendale S., Abdulla S., Elgar G., Buck D., Berks M., Micklem G., Durbin R., Bates G., Brenner S. and Beck S. (1995). Comparative sequence analysis of the human and pufferfish Huntington's disease genes *Nat Genet.* **10** 67-76.

Beckers J., Gerard M. and Duboule D. (1996). Transgenic analysis of a potential *Hoxd-11* limb regulatory element present in tetrapods and fish. *Dev Biol.* **180** 543-53.

Beddington R.S.P. (1981). An autoradiographic analysis of the potency of embryonic ectoderm in the 8th day postimplantation mouse embryo. *J Embryol Exp Morphol.* **64** 87-104.

Benton M.J. (1990). Phylogeny of the major tetrapod groups: morphological data and divergence dates. *J Mol Evol.* **30** 409-24.

Bettenhausen B., Hrabe de Angelis M., Simon D., Guenet J.L. and Gossler A. (1995). Transient and restricted expression during mouse embryogenesis of *Dll1*, a murine gene closely related to *Drosophila Delta. Development.* **121** 2407-18.

Blagden C.S., Currie P.D., Ingham P.W. and Hughes S.M. (1997). Notochord induction of zebrafish slow muscle mediated by Sonic hedgehog. *Genes Dev.* **11** 2163-75.

Bober E., Brand-Saberi B., Ebensperger C., Wilting J., Balling R., Paterson B.M., Arnold H.H. and Christ B. (1994). Initial steps of myogenesis in somites are independent of influence from axial structures. *Development.* **120** 3073-82.

Bober E., Lyons G.E., Braun T., Cossu G., Buckingham M. and Arnold H.H. (1991). The muscle regulatory gene, *Myf-6*, has a biphasic pattern of expression during early mouse development. *J Cell Biol.* **113** 1255-65.

Bouvet J. (1976). Enveloping layer and periderm of the trout embryo *(Salmo trutta fario L.). Cell Tissue Res.* **170** 367-82.

Brand-Saberi B., Ebensperger C., Wilting J., Balling R. and Christ B. (1993). The ventralizing effect of the notochord on somite differentiation in chick embryos. *Anat Embryol (Berl).* **188** 239-45.

Brand-Saberi B., Muller T.S., Wilting J., Christ B. and Birchmeier C. (1996). Scatter factor/hepatocyte growth factor (SF/HGF) induces emigration of myogenic cells at interlimb level in vivo. *Dev Biol.* **179** 303-8.

Braun T. and Arnold H.H. (1995). Inactivation of *Myf-6* and *Myf-5* genes in mice leads to alterations in skeletal muscle development. *EMBO J.* **14** 1176-86.

Braun T. and Arnold H.H. (1996). *Myf-5* and *MyoD* genes are activated in distinct mesenchymal stem cells and determine different skeletal muscle cell lineages. *EMBO J.* **15** 310-18.

Braun T., Bober E., Buschhausen-Denker G., Kohtz S., Grzeschik K.H., Arnold H.H. and Kohtz. (1989b). Differential expression of myogenic determination genes in muscle cells: possible autoactivation by the *Myf* gene products *EMBO J.* **8** 3617-25.

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** 821-31.

Braun T., Buschhausen-Denker G., Bober E., Tannich E. and Arnold H.H. (1989a). A novel human muscle factor related to but distinct from MyoD1 induces myogenic conversion in 10T1/2 fibroblasts. *EMBO J.* **8** 701-9.

Braun T., Rudnicki M.A., Arnold H.H. and Jaenisch R. (1992). Targeted inactivation of the muscle regulatory gene *Myf-5* results in abnormal rib development and perinatal death. *Cell.* **71** 369-82.

Brennan T.J. and Olson E.N. (1990). Myogenin resides in the nucleus and acquires high affinity for a conserved enhancer element on heterodimerization. *Genes Dev.* **4** 582-95.

Brenner S., Elgar G., Sandford R., Macrae A., Venkatesh B. and Aparicio S. (1993). Characterization of the pufferfish *(Fugu)* genome as a compact model vertebrate genome. *Nature.* **366** 265-8.

Buonanno A, Apone L, Morasso MI, Beers R, Brenner HR, Eftimie R. (1992). The *MyoD* family of myogenic factors is regulated by electrical activity: isolation and characterization of a mouse *Myf-5* cDNA. *Nucleic Acids Res.* **20**. 539-44.

Buckingham M. (1992). Making muscle in mammals. *Trends Genet.* **8** 144-8.

Buckingham M. (1994). Molecular biology of muscle development. *Cell.* **78** 15-21.

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** 2627-40.

Chen X., Raab G., Deutsch U., Zhang J., Ezzell R.M. and Klagsbrun M. (1995). Induction of heparin-binding EGF-like growth factor expression during myogenesis. Activation of the gene by MyoD and localization of the transmembrane form of the protein on the myotube surface. *J Biol Chem.* **270** 18285-94.

Cheng T.C., Wallace M.C., Merlie J.P., Olson E.N. (1993). Separable regulatory elements governing myogenin transcription in mouse embryogenesis. *Science.* **261** 215-218

Choi J., Costa M.L., Mermelstein C.S., Chagas C., Holtzer S. and Holtzer H. (1990). MyoD converts primary dermal fibroblasts, chondroblasts, smooth muscle, and retinal pigmented epithelial cells into striated mononucleated myoblasts and multinucleated myotubes. *Proc Natl Acad Sci U S A.* **87** 7988-92.

Christ B., Brand-Saberi B., Grim M. and Wilting J. (1992). Local signalling in dermomyotomal cell type specification. *Anat Embryol (Berl).* **186** 505-10.

Church G.M. and Gilbert W. (1984). Genomic sequencing. *Proc Natl Acad Sci U S A.* **81** 1991-5.

Concordet J.P., Lewis K.E., Moore J.W., Goodrich L.V., Johnson R.L., Scott M.P. and Ingham P.W. (1996). Spatial regulation of a zebrafish patched homologue reflects the roles of sonic hedgehog and protein kinase A in neural tube and somite patterning. *Development.* **122** 2835-46.

Conlon R.A., Reaume A.G. and Rossant J. (1995). *Notch1* is required for the coordinate segmentation of somites. *Development.* **121** 1533-45.

Cossu G., Kelly R., Tajbakhsh S., Di Donna S., Vivarelli E. and Buckingham M. (1996a). Activation of different myogenic pathways: *myf-5* is induced by the neural tube and *MyoD* by the dorsal ectoderm in mouse paraxial mesoderm. *Development.* **122** 429-37.

Cossu G., Tajbakhsh S. and Buckingham M. (1996b). How is myogenesis initiated in the embryo? *Trends Genet.* **12** 218-23.

Couso J.P. and Martinez Arias A. (1994). *Notch* is required for *wingless* signaling in the epidermis of Drosophila. *Cell.* **79** 259-72.

Crnogorac-Jurcevic T., Brown J.R., Lehrach H. and Schalkwyk L.C. (1997). *Tetraodon fluviatilis*, a new puffer fish model for genome studies. *Genomics.* **41** 177-84.

Currie P.D. and Ingham P.W. (1996). Induction of a specific muscle cell type by a hedgehog-like protein in zebrafish. *Nature.* **382** 452-5.

Daston G., Lamar E., Olivier M. and Goulding M. (1996). *Pax-3* is necessary for migration but not differentiation of limb muscle precursors in the mouse. *Development.* **122** 1017-27.

Davis R.L. and Weintraub H. (1992). Acquisition of myogenic specificity by replacement of three amino acid residues from MyoD into E12. *Science.* **256** 1027-30.

Davis R.L., Weintraub H. and Lassar A.B. (1987). Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell.* **51** 987-1000.

Denetclaw W.F. Jr, Christ B. and Ordahl C.P. (1997). Location and growth of epaxial myotome precursor cells. *Development.* **124** 1601-10.

Devoto S.H., Melancon E., Eisen J.S. and Westerfield M. (1996). Identification of separate slow and fast muscle precursor cells in vivo, prior to somite formation. *Development.* **122** 3371-80.

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 program. *Genes Dev.* **3** 628-40.

Edmondson D.G., Cheng T.C., Cserjesi P., Chakraborty T. and Olson E.N. (1992). Analysis of the myogenin promoter reveals an indirect pathway for positive autoregulation mediated by the muscle-specific enhancer factor MEF-2. *Mol Cell Biol.* **12** 3665-77.

Elgar G. (1996). Quality not quantity: the pufferfish genome. *Hum Mol Genet.* **5** 1437-42.

Elgar G., Sandford R., Aparicio S., Macrae A., Venkatesh B. and Brenner S. (1996). Small is beautiful: comparative genomics with the pufferfish *(Fugu rubripes). Trends Genet.* **12** 145-50.

Fan C.M. and Tessier-Lavigne M. (1994). Patterning of mammalian somites by surface ectoderm and notochord: evidence for sclerotome induction by a hedgehog homolog. *Cell.* **79** 1175-86.

Fan C.M., Porter J.A., Chiang C., Chang D.T., Beachy P.A. and Tessier-Lavigne M. (1995). Long-range sclerotome induction by sonic hedgehog: direct role of the amino-terminal cleavage product and modulation by the cyclic AMP signaling pathway. *Cell.* **81** 457-65.

Feldman B., Poueymirou W., Papaioannou V.E., DeChiara T.M. and Goldfarb M. (1995). Requirement of FGF-4 for postimplantation mouse development. *Science.* **267** 246-249.

Fields C., Adams M.D., White O. and Venter J.C. (1994). How many genes in the human genome? *Nat Genet.* **7** 345-6.

Floss T., Arnold H.H. and Braun T. (1996). *Myf-5(m1)/Myf-6(m1)* compound heterozygous mouse mutants down-regulate *Myf-5* expression and exert rib defects: evidence for long-range *cis* effects on *Myf-5* transcription. *Dev Biol.* **174** 140-7.

Frank D. and Harland R.M. (1991). Transient expression of *XMyoD* in non-somitic mesoderm of *Xenopus* gastrulae. *Development.* **113** 1387-93.

Franz T., Kothary R., Surani M.A., Halata Z. and Grim M. (1993). The *Splotch* mutation interferes with muscle development in the limbs. *Anat Embryol (Berl).* **187** 153-60.

Fujisawa-Sehara, A., Nabeshima, Y., Komiya, T., Uetsuki, T., Asakura, A. and Nabeshima,Y.-I.(1992). Differential trans-activation of muscle-specific regulatory elements including the mysosin light chain box by chicken *MyoD*, *myogenin*, and *MRF4. J.Biol.Chem.* **267** 10031-38.

Gardiner-Garden M. and Frommer M. (1987). CpG islands in vertebrate genomes. *J Mol Biol.* **196** 261-82.

Gardner R.L. (1982). Investigation of cell lineage and differentiation in the extraembryonic endoderm of the mouse embryo. *J Embryol Exp Morphol.* **68** 175-98.

Gardner R.L. and Rossant J. (1979). Investigation of the fate of 4-5 day post-coitum mouse inner cell mass cells by blastocyst injection. *J Embryol Exp Morphol.* **52** 141-52.

George-Weinstein M., Gerhart J.V., Foti G.J. and Lash J.W. (1994). Maturation of myogenic and chondrogenic cells in the presomitic mesoderm of the chick embryo. *Exp Cell Res.* **211** 263-74.

Gerber A.N., Klesert T.R., Bergstrom D.A. and Tapscott S.J. (1997). Two domains of MyoD mediate transcriptional activation of genes in repressive chromatin: a mechanism for lineage determination in myogenesis. *Genes Dev.* **11** 436-50.

Gilley J., Armes N. and Fried M. (1997). *Fugu* genome is not a good mammalian model. *Nature.* **387** 140.

Gilthorpe J.D. and Rigby P.W.J (in press). 'Reporter Genes fore the Study of Transcriptional Regulation in Transgenic Mouse Embryos'. In Molecular Embryology: Methods and Protocols. Sharpe and Mason, Eds. (Humana Press, Totowa, N.J.)

Gossett L.A., Kelvin D.J., Sternberg E.A. and Olson E.N. (1989). A new myocyte-specific enhancer-binding factor that recognizes a conserved element associated with multiple muscle-specific genes. *Mol Cell Biol.* **9** 5022-33.

Goulding M., Lumsden A. and Paquette A.J. (1994). Regulation of *Pax-3* expression in the dermomyotome and its role in muscle development. *Development.* **120** 957-71.

Grass S., Arnold H.H. and Braun T. (1996). Alterations in somite patterning of *Myf-5*-deficient mice: a possible role for FGF-4 and FGF-6. *Development.* **122** 141-50.

Gutman A., Gilthorpe J. and Rigby P.W.J. (1994). Multiple positive and negative regulatory elements in the promoter of the mouse homeobox gene *Hoxb-4*. *Mol Cell Biol.* **14** 8143-54.

Hanneman E. and Westerfield M. (1989). Early expression of acetylcholinesterase activity in functionally distinct neurons of the zebrafish. *J Comp Neurol.* **284** 350-61.

Harvey R.P. (1992). MyoD protein expression in *Xenopus* embryos closely follows a mesoderm induction-dependent amplification of *MyoD* transcription and is synchronous across the future somite axis. *Mech Dev.* **37** 141-9.

Hasty P., Bradley A., Morris J.H., Edmondson D.G., Venuti J.M., Olson E.N. and Klein W.H. (1993). Muscle deficiency and neonatal death in mice with a targeted mutation in the *myogenin* gene *Nature.* **364** 501-6.

Hatta K. and Takahashi Y. (1996). Secondary axis induction by heterospecific organizers in zebrafish. *Dev Dyn.* **205** 183-95.

Hatta K., Bremiller R., Westerfield M. and Kimmel C.B. (1991). Diversity of expression of *engrailed*-like antigens in zebrafish. *Development.* **112** 821-32.

Haub O. and Goldfarb M. (1991). Expression of the fibroblast growth factor-5 gene in the mouse embryo. *Development.* **112** 397-406.

Hebert J.M., Tosenquist T., Gotz J. and Martin G.R. (1994). FGF-5 as a regulator of the hair growth cycle: Evidence from targeted spontaneous mutations. *Cell.* **78** 1017-1025.

Hinegardner R. and Rosen E. (1972). Cellular DNA content and evolution of teleostean fishes. *Am. Naturalist.* **106** 621-644.

Hinterberger T.J., Mays J.L. and Konieczny S.F. (1992). Structure and myofiber-specific expression of the rat muscle regulatory gene *MRF4*. *Gene.* **117** 201-7.

Hinterberger T.J., Sassoon D.A., Rhodes S.J. and Konieczny S.F. (1991). Expression of the muscle regulatory factor *MRF4* during somite and skeletal myofiber development. *Dev Biol.* **147** 144-56.

Hogan B., Beddington R., Costantini F. and Lacy E. (1994). *Manipulating the Mouse Embryo - A Laboratory Manual.* 2nd Edition, Cold Spring Harbour Laboratory Press.

Holland P.W., Garcia-Fernandez J., Williams N.A. and Sidow A. (1994). Gene duplications and the origins of vertebrate development. *Dev Suppl.* **120** 125-33.

Hooft van Huijsduijnen R, Li XY, Black D, Matthes H, Benoist C, Mathis D. (1990). Co-evolution from yeast to mouse: cDNA cloning of the two NF-Y CP-1/CBF) subunits. *EMBO J.* **9** 3119-3127.

Hoppler S., Brown J.D. and Moon R.T. (1996). Expression of a dominant-negative Wnt blocks induction of *MyoD* in *Xenopus* embryos. *Genes Dev.* **10** 2805-17.

Hopwood N.D., Pluck A., Gurdon J.B. and Dilworth S.M. (1992). Expression of *XMyoD* protein in early *Xenopus laevis* embryos. *Development.* **114** 31-8.

Hughes S.M., Taylor J.M., Tapscott S.J., Gurley C.M., Carter W.J. and Peterson C.A. (1993). Selective accumulation of *MyoD* and *myogenin* mRNAs in fast and slow adult skeletal muscle is controlled by innervation and hormones. *Development.* **118** 1137-47.

Jennings, C.G. (1992). Expression of the myogenic gene MRF4 during *Xenopus* development. *Dev Biol* . **151** 319-32.

Jiang Y.J., Brand M., Heisenberg C.P., Beuchle D., Furutani-Seiki M., Kelsh R.N., Warga R.M., Granato M., Haffter P., Hammerschmidt M., Kane D.A., Mullins M.C., Odenthal J., van Eeden F.J. and Nusslein-Volhard C. (1996). Mutations affecting neurogenesis and brain morphology in the zebrafish, *Danio rerio.* *Development.* **123** 205-16.

Johnson R.L., Laufer E., Riddle R.D. and Tabin C. (1994). Ectopic expression of *Sonic hedgehog* alters dorsal-ventral patterning of somites. *Cell.* **79** 1165-73.

Joyner A.L. and Martin G.R. (1987). *En-1* and *En-2,* two mouse genes with sequence homology to the *Drosophila engrailed* gene: expression during embryogenesis [published erratum appears in Genes Dev 1987 Jul;1(5):521] *Genes Dev.* **1** 29-38.

Kablar B., Krastel K., Ying C., Asakura A., Tapscott S., Rudnicki M. (1997). *MyoD* and *Myf-5* differentially regulate the development of limb versus trunk skeletal muscle. *Dev.* **124** 4729-4738

Kimmel C.B. and Warga R.M. (1987). Cell lineages generating axial muscle in the zebrafish embryo. *Nature.* **327** 234-7.

Kimmel C.B., Warga R.M. and Schilling T.F. (1990). Origin and organization of the zebrafish fate map. *Development.* **108** 581-94.

Konieczny S.F. and Emerson C.P. Jr (1984). 5-Azacytidine induction of stable mesodermal stem cell lineages from 10T1/2 cells: evidence for regulatory genes controlling determination. *Cell.* **38** 791-800.

Kopan R., Nye J.S. and Weintraub H. (1994). The intracellular domain of mouse Notch: a constitutively activated repressor of myogenesis directed at the basic helix-loop-helix region of MyoD. *Development.* **120** 2385-96.

Krause M., Fire A., White-Harrison S., Weintraub H. and Tapscott S. (1992). Functional conservation of nematode and vertebrate myogenic regulatory factors. *J Cell Sci Suppl.* **16** 111-5.

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 to the muscle creatine kinase enhancer. *Cell.* **58** 823-31.

Lassar A.B., Paterson B.M. and Weintraub H. (1986). Transfection of a DNA locus that mediates the conversion of 10T1/2 fibroblasts to myoblasts. *Cell.* **47** 649-56.

Lee J.J., Ekker S.C., von Kessler D.P., Porter J.A., Sun B.I. and Beachy P.A. (1994). Autoproteolysis in hedgehog protein biogenesis *Science.* **266** 1528-37.

Lewin B. (1994). Chromatin and gene expression: constant questions, but changing answers. *Cell.* **79** 397-406.

Li L., Zhou J., James G., Heller-Harrison R., Czech M.P. and Olson E.N. (1992). FGF inactivates myogenic helix-loop-helix proteins through phosphorylation of a conserved protein kinase C site in their DNA- binding domains. *Cell.* **71** 1181-94.

Lundin L.G. (1993). Evolution of the vertebrate genome as reflected in paralogous chromosomal regions in man and the house mouse. *Genomics.* **16** 1-19.

Mansour S.L., Goddard J.M. and Capecchi M.R. (1993). Mice homozygous for a targeted disruption of the proto-oncogene *int-2* have developmental defects in the tail and inner ear. *Development.* **117** 13-28.

Marigo V. and Tabin C.J. (1996). Regulation of *patched* by sonic hedgehog in the developing neural tube. *Proc Natl Acad Sci U S A.* **93** 9346-51.

Maroto M., Reshef R., Munsterberg A.E., Koester S., Goulding M. and Lassar A.B. (1997). Ectopic *Pax-3* activates *MyoD* and *Myf-5* expression in embryonic mesoderm and neural tissue. *Cell.* **89** 139-48.

Marshall H., Studer M., Popperl H., Aparicio S., Kuroiwa A., Brenner S. and Krumlauf R. (1994). A conserved retinoic acid response element required for early expression of the homeobox gene *Hoxb-1*. *Nature.* **370** 567-71.

Marti E., Takada R., Bumcrot D.A., Sasaki H. and McMahon A.P. (1995). Distribution of Sonic hedgehog peptides in the developing chick and mouse embryo. *Development.* **121** 2537-47.

Martindale M.Q., Meier S. and Jacobson A.G. (1987). Mesodermal metamerism in the teleost, *Oryzias latipes* (the medaka). *J Morphol.* **193** 241-52.

Meedel T.H., Farmer S.C. and Lee J.J. (1997). The single *MyoD* family gene of *Ciona intestinalis* encodes two differentially expressed proteins: implications for the evolution of chordate muscle gene regulation. *Development.* **124** 1711-21.

Megeney L.A., Kablar B., Garrett K., Anderson J.E. and Rudnicki M.A. (1996). *MyoD* is required for myogenic stem cell function in adult skeletal muscle. *Genes Dev.* **10** 1173-83.

Michelson A.M., Abmayr S.M., Bate M., Arias A.M. and Maniatis T. (1990). Expression of a *MyoD* family member prefigures muscle pattern in *Drosophila* embryos. *Genes Dev.* **4** 2086-97.

Miner J.H. and Wold B. (1990). *Herculin*, a fourth member of the *MyoD* family of myogenic regulatory genes. *Proc Natl Acad Sci U S A.* **87** 1089-93.

Miyaki K., Tabeta O. and Kayano. H. (1995). Karyotypes in six species of Pufferfishes, Genus *Takifugu (Tetraodontiformes). Fish. Sci.* **61** 594-598.

Morin-Kensicki E.M. and Eisen J.S. (1997). Sclerotome development and peripheral nervous system segmentation in embryonic zebrafish. *Development.* **124** 159-67.

Muensterberg A.E., Kitajewski J., Bumcrot D.A., McMahon A.P. and Lassar A.B. (1995). Combinatorial signaling by *Sonic hedgehog* and *Wnt* family members induces myogenic bHLH gene expression in the somite. *Genes Dev.* **9** 2911-22.

Murre C., McCaw P.S. and Baltimore D. (1989). A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. *Cell.* **56** 777-83.

Myer A., Wagner D.S., Vivian J.L., Olson E.N. and Klein W.H. (1997). Wild-type myoblasts rescue the ability of *myogenin*-null myoblasts to fuse in vivo. *Dev Biol.* **185** 127-38.

Nabeshima Y., Hanaoka K., Hayasaka M., Esumi E., Li S., Nonaka I. and Nabeshima Y. (1993). *Myogenin* gene disruption results in perinatal lethality because of severe muscle defect *Nature.* **364** 532-5.

Natoli T.A., Ellsworth M.K., Wu C., Gross K.W. and Pruitt S.C. (1997). Positive and negative DNA sequence elements are required to establish the pattern of *Pax3* expression. *Development.* **124** 617-26.

Nicolas R.H. and Goodwin G.H. (1993). 'Purification and cloning of transcription factors'. In Transcription Factors - A Practical Approach. 1st Edition, Ed. Latchman D.S. (The Practical Approach Series, IRL Press, Oxford). 85-6.

Niswander L. and Martin G.R. (1992). *Fgf-4* expression during gastrulation, myogenesis, limb and tooth development in the mouse. *Development.* **114** 755-68.

Nüsslein-Volhard C. (1994). Of flies and fishes. *Science.* **266** 572-4.

Odenthal J., Haffter P., Vogelsang E., Brand M., van Eeden F.J., Furutani-Seiki M., Granato M., Hammerschmidt M., Heisenberg C.P., Jiang Y.J., Kane D.A., Kelsh R.N., Mullins M.C., Warga R.M., Allende M.L., Weinberg E.S. and Nüsslein-Volhard C. (1996). Mutations affecting the formation of the notochord in the zebrafish, *Danio rerio. Development.* **123** 103-15.

Olson E.N., Arnold H.H., Rigby P.W.J.and Wold B.J. (1996). Know your neighbors: three phenotypes in null mutants of the myogenic bHLH gene *MRF4. Cell.* **85** 1-4.

Ordahl C.P. and Le Douarin N.M. (1992). Two myogenic lineages within the developing somite. *Development.* **114** 339-53.

Ott M.O., Bober E., Lyons G., Arnold H. and Buckingham M. (1991). Early expression of the myogenic regulatory gene, *myf-5*, in precursor cells of skeletal muscle in the mouse embryo. *Development.* **111** 1097-107.

Packard D.S. Jr (1978). Chick somite determination: the role of factors in young somites and the segmental plate. *J Exp Zool.* **203** 295-306.

Patapoutian A., Miner J.H., Lyons G.E. and Wold B. (1993). Isolated sequences from the linked *Myf-5* and *MRF4* genes drive distinct patterns of muscle-specific expression in transgenic mice. *Development.* **118** 61-9.

Patapoutian A., Yoon J.K., Miner J.H., Wang S., Stark K. and Wold B. (1995). Disruption of the mouse *MRF4* gene identifies multiple waves of myogenesis in the myotome. *Development.* **121** 3347-58.

Pette D. and Vrbova G. (1985). Neural control of phenotypic expression in mammalian muscle fibers. *Muscle Nerve.* **8** 676-89.

Porter J.A., Young K.E. and Beachy P.A. (1996). Cholesterol modification of hedgehog signaling proteins in animal development [published erratum appears in Science 1996 Dec 6;274(5293):1597] *Science.* **274** 255-9.

Pourquie O., Coltey M., Breant C. and Le Douarin N.M. (1995). Control of somite patterning by signals from the lateral plate. *Proc Natl Acad Sci U S A.* **92** 3219-23.

Pourquie O., Coltey M., Teillet M.A., Ordahl C. and Le Douarin N.M. (1993). Control of dorsoventral patterning of somitic derivatives by notochord and floor plate. *Proc Natl Acad Sci U S A.* **90** 5242-6.

Pourquie O., Fan C.M., Coltey M., Hirsinger E., Watanabe Y., Breant C., Francis-West P., Brickell P., Tessier-Lavigne M. and Le Douarin N.M. (1996). Lateral and axial signals involved in avian somite patterning: a role for *BMP4*. *Cell.* **84** 461-71.

Power M.A. and Tam P.P. (1993). Onset of gastrulation, morphogenesis and somitogenesis in mouse embryos displaying compensatory growth. *Anat Embryol (Berl).* **187** 493-504.

Pownall M.E. and Emerson C.P. Jr (1992). Sequential activation of three myogenic regulatory genes during somite morphogenesis in quail embryos. *Dev Biol.* **151** 67-79.

Ranganayakulu G., Schulz R.A. and Olson E.N. (1996). Wingless signaling induces *nautilus* expression in the ventral mesoderm of the *Drosophila* embryo. *Dev Biol.* **176** 143-8.

Rawls A., Morris J.H., Rudnicki M., Braun T., Arnold H.H., Klein W.H. and Olson E.N. (1995). Myogenin's functions do not overlap with those of MyoD or Myf-5 during mouse embryogenesis *Dev Biol.* **172** 37-50.

Rhodes S.J. and Konieczny S.F. (1989). Identification of *MRF4*: a new member of the muscle regulatory factor gene family. *Genes Dev.* **3** 2050-61.

Roelink H., Porter J.A., Chiang C., Tanabe Y., Chang D.T., Beachy P.A. and Jessell T.M. (1995). Floor plate and motor neuron induction by different concentrations of the amino-terminal cleavage product of sonic hedgehog autoproteolysis. *Cell.* **81** 445-55.

Rong P.M., Teillet M.A., Ziller C. and Le Douarin N.M. (1992). The neural tube/notochord complex is necessary for vertebral but not limb and body wall striated muscle differentiation. *Development.* **115** 657-72.

Rosenthal A., Coutelle O. and Craxton M. (1993). Large-scale production of DNA sequencing templates by microtitre format PCR. *Nucleic Acids Res.* **21** 173-4.

Rossant J., Zirngibl R., Cado D., Shago M. and Giguere V. (1991). Expression of a retinoic acid response element-*hsplacZ* transgene defines specific domains of transcriptional activity during mouse embryogenesis. *Genes Dev.* **5** 1333-44.

Rudnicki M.A., Braun T., Hinuma S. and Jaenisch R. (1992). Inactivation of *MyoD* in mice leads to up-regulation of the myogenic HLH gene *Myf-5* and results in apparently normal muscle development. *Cell.* **71** 383-90.

Rudnicki M.A., Schnegelsberg P.N., Stead R.H., Braun T., Arnold H.H. and Jaenisch R. (1993). *MyoD* or *Myf-5* is required for the formation of skeletal muscle. *Cell.* **75** 1351-9.

Rudnicki, M.A., Jaenisch, R. (1995). The MyoD family of transcription factors and skeletal myogenesis.*Bioessays.* **17** 203-209

Rupp R.A. and Weintraub H. (1991). Ubiquitous *MyoD* transcription at the midblastula transition precedes induction-dependent *MyoD* expression in presumptive mesoderm of *X. laevis.* *Cell.* **65** 927-37.

Saccone S., De Sario A., Wiegant J., Raap A.K., Della Valle G. and Bernardi G. (1993). Correlations between isochores and chromosomal bands in the human genome. *Proc Natl Acad Sci U S A.* **90** 11929-33.

Saitoh O, Fujisawa-Sehara A, Nabeshima Y, Periasamy M. (1993) Expression of myogenic factors in denervated chicken breast muscle: isolation of the chicken *Myf-5* gene. *Nucleic Acids Res.* **21** 2503-9.

Salminen A., Braun T., Buchberger A., Jurs S., Winter B. and Arnold H.H. (1991). Transcription of the muscle regulatory gene *Myf4* is regulated by serum components, peptide growth factors and signaling pathways involving G proteins. *J Cell Biol.* **115** 905-17.

Sambrook J., Fritsch E. and Maniatis T. (1989). *Molecular Cloning - A Laboratory Manual.* 2nd Edition. Cold Spring Harbour Laboratory Press.

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 *MyoD1* during mouse embryogenesis. *Nature.* **341** 303-7.

Schwarz J.J., Chakraborty T., Martin J., Zhou J.M. and Olson E.N. (1992). The basic region of myogenin cooperates with two transcription activation domains to induce muscle-specific transcription. *Mol Cell Biol.* **12** 266-75.

Selleck M.A. and Stern C.D. (1991). Fate mapping and cell lineage analysis of Hensen's node in the chick embryo. *Development.* **112** 615-26.

Smith C.K. 2nd, Janney M.J. and Allen R.E. (1994). Temporal expression of myogenic regulatory genes during activation, proliferation, and differentiation of rat skeletal muscle satellite cells. *J Cell Physiol.* **159** 379-85.

Snow M.H. (1981). Growth and its control in early mammalian development. *Br Med Bull.* **37** 221-6.

Southern E.M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol.* **98** 503-517.

Spoerle R., Gunther T., Struwe M. and Schughart K. (1996). Severe defects in the formation of epaxial musculature in *open brain (opb)* mutant mouse embryos. *Development.* **122** 79-86.

Sulik K., Dehart D.B., Iangaki T., Carson J.L., Vrablic T., Gesteland K. and Schoenwolf G.C. (1994). Morphogenesis of the murine node and notochordal plate. *Dev Dyn.* **201** 260-78.

Swiatek P.J., Lindsell C.E., del Amo F.F., Weinmaster G. and Gridley T. (1994). *Notch1* is essential for postimplantation development in mice. *Genes Dev.* **8** 707-19.

Tajbakhsh S. and Buckingham M.E. (1994). Mouse limb muscle is determined in the absence of the earliest myogenic factor *myf-5*. *Proc Natl Acad Sci U S A.* **91** 747-51.

Tajbakhsh S., Bober E., Babinet C., Pournin S., Arnold H. and Buckingham M. (1996). Gene targeting the *myf-5* locus with *nlacZ* reveals expression of this myogenic factor in mature skeletal muscle fibres as well as early embryonic muscle. *Dev Dyn.* **206** 291-300.

Tajbakhsh S., Rocancourt D., Cossu G. and Buckingham M. (1997). Redefining the genetic hierarchies controlling skeletal myogenesis: Pax- 3 and Myf-5 act upstream of MyoD. *Cell.* **89** 127-38.

Tam P.P. (1981). The control of somitogenesis in mouse embryos. *J Embryol Exp Morphol.* **65** 103-28.

Tam P.P. (1986). A study of the pattern of prospective somites in the presomitic mesoderm of mouse embryos. *J Embryol Exp Morphol.* **92** 269-85.

Tam P.P. and Trainor P.A. (1994). Specification and segmentation of the paraxial mesoderm. *Anat Embryol (Berl).* **189** 275-305.

Trinkaus J.P. (1988). Directional cell movement during early development of the teleost *Blennius pholis:* I. Formation of epithelial cell clusters and their pattern and mechanism of movement. *J Exp Zool.* **245** 157-86.

Trower M.K., Orton S.M., Purvis I.J., Sanseau P., Riley J., Christodoulou C., Burt D., See C.G., Elgar G., Sherrington R., Rogaev E.I., St. George-Hyslop P., Brenner S. and Dykes C.W. (1996). Conservation of synteny between the genome of the pufferfish *(Fugu rubripes)* and the region on human chromosome 14 (14q24.3) associated with familial Alzheimer disease (AD3 locus) *Proc Natl Acad Sci U S A.* **93** 1366-9.

Überbacher E.C. and Mural R.J. (1991). Locating protein-coding regions in human DNA sequences by a multiple sensor-neural network approach. *Proc Natl Acad Sci U S A.* **88** 11261-5.

Venuti J.M., Goldberg L., Chakraborty T., Olson E.N. and Klein W.H. (1991). A myogenic factor from sea urchin embryos capable of programming muscle differentiation in mammalian cells. *Proc Natl Acad Sci U S A.* **88** 6219-23.

Wahl G.M., Stern M. and Stark G.R. (1979). Efficient transfer of large DNA fragments from agarose gels to diazobenzyloxymethyl-paper and rapid hybridization by using dextran sulfate. *Proc Natl Acad Sci U S A* **76** 3683-3687

Waterman R.E. (1969). Development of the lateral musculature in the teleost, *Brachydanio rerio:* a fine structural study. *Am J Anat.* **125** 457-93.

Weinberg E.S., Allende M.L., Kelly C.S., Abdelhamid A., Murakami T., Andermann P., Doerre O.G., Grunwald D.J. and Riggleman B. (1996). Developmental regulation of zebrafish *MyoD* in wild-type, *no tail* and *spadetail* embryos. *Development.* **122** 271-80.

Weintraub H., Dwarki V.J., Verma I., Davis R., Hollenberg S., Snider L., Lassar A. and Tapscott S.J. (1991). Muscle-specific transcriptional activation by *MyoD*. *Genes Dev.* **5** 1377-86.

Weintraub H., Tapscott S.J., Davis R.L., Thayer M.J., Adam M.A., Lassar A.B. and Miller A.D. (1989). Activation of muscle-specific genes in pigment, nerve, fat, liver, and fibroblast cell lines by forced expression of *MyoD*. *Proc Natl Acad Sci U S A.* **86** 5434-8.

Whiting J., Marshall H., Cook M., Krumlauf R., Rigby P.W.J., Stott D. and Allemann R.K. (1991). Multiple spatially specific enhancers are required to reconstruct the pattern of *Hox-2.6* gene expression. *Genes Dev.* **5** 2048-59.

Williams B.A. and Ordahl C.P. (1994). *Pax-3* expression in segmental mesoderm marks early stages in myogenic cell specification. *Development.* **120** 785-96.

Williams R., Lendahl U. and Lardelli M. (1995). Complementary and combinatorial patterns of *Notch* gene family expression during early mouse development. *Mech Dev.* **53** 357-68.

Wright W.E., Sassoon D.A. and Lin V.K. (1989). Myogenin, a factor regulating myogenesis, has a domain homologous to MyoD. *Cell.* **56** 607-17.

Yamaguchi T.P., Harpal K., Henkemeyer M. and Rossant J. (1994). *fgfr-1* is required for embryonic growth and mesodermal patterning during mouse gastrulation. *Genes Dev.* **8** 3032-44.

Yee S.P. and Rigby P.W.J. (1993). The regulation of myogenin gene expression during the embryonic development of the mouse. *Genes Dev.* **7** 1277-89.

Yoon J.K., Olson E.N., Arnold H.H. and Wold B.J. (1997). Different *MRF4* knockout alleles differentially disrupt *myf-5* expression: *cis*-regulatory interactions at the *MRF4/Myf-5* locus *Dev Biol.* **188** 349-62.

Zhang W., Behringer R.R. and Olson E.N. (1995). Inactivation of the myogenic bHLH gene *MRF4* results in up-regulation of *myogenin* and rib anomalies. *Genes Dev.* **9** 1388-99.

Zingg J.M., Alva G.P. and Jost J.P. (1991). Characterisation of a genomic clone covering the structural mouse *MyoD1* gene and its promoter region. *Nucleic Acids Res.* **19** 6433-9.

Additional References

Felsenfeld A.L., Curry M. and Kimmel C.B. (1991). The *fub-1* mutation blocks initial myofibril formation in zebrafish muscle pioneer cells. *Dev Biol.* **148** 23-30.

Halpern M.E., Thisse C., Ho R.K., Thisse B., Riggleman B., Trevarrow B., Weinberg E.S., Postlethwait J.H. and Kimmel C.B. (1995). Cell-autonomous shift from axial to paraxial mesodermal development in zebrafish *floating head* mutants. *Development.* **121** 4257-4264.

Hammerschmidt M. and Nüsslein-Volhard C., (1993). The expression of a zebrafish gene homologous to Drosophila *snail* suggests a conserved function in invertebrate and vertebrate gastrulation. *Development.* **119** 1107-1118.

Kanki J.P. and Ho R.K. (1997). The development of the posterior body in zebrafish. *Development.* **124** 881-893.

Schier A.F., Neuhauss S.C., Helde K.A., Talbot W.S. and Driever W. (1997). The *one-eyed pinhead* gene functions in mesoderm and endoderm formation in zebrafish and interacts with *no tail. Development.* **124** 327-342.

Thisse C., Thisse B., Schilling T.F. and Postlethwait J.H. (1993). Structure of the zebrafish *snail1* gene and its expression in wild-type, *spadetail* and *no tail* mutant embryos. *Development.* **119** 1203-1215.