

A MOLECULAR STUDY OF ONCOGENE
RELATED SEQUENCES IN XENOPUS

DAVID TANNAHILL

DOCTOR OF PHILOSOPHY
UNIVERSITY OF EDINBURGH

1987



Abstract

Cellular oncogenes form a class of evolutionarily conserved genes which are thought to be involved in the control of cell proliferation and differentiation and also may play a role during development. Organisms with an established embryology may offer valuable prospects in our understanding of cellular oncogene function. In this thesis I describe my attempts to study cellular oncogenes in *Xenopus* as a means to the above end.

By the use of highly labeled probes from a variety of birds, mammals and their viruses and low stringency hybridisation conditions I have shown the presence of *N-myc*, *c-myc*, *c-ha-ras*, *c-ki-ras*, *c-src*, *c-abl*, *c-erbB*, *c-fms*, *c-fos*, *c-sis* and *c-rel* related sequences in *Xenopus* genomic DNA. Similar experiments also show the presence of *N-myc*, *c-myc*, *c-ha-ras*, *c-ki-ras* and *c-fos* related sequences in *X.borealis* RNA.

The *c-myc* gene was chosen for further analysis with the isolation of homologous clones from a *X.borealis* previtellogenic ovary cDNA library which I had constructed. DNA sequence analysis and southern blotting experiments suggest that there are two distinct classes (A and B) of *c-myc* genes in *Xenopus* and in addition at least one of these classes may be polymorphic. Sequence analysis of the class A *c-myc* genes show they are highly homologous to the *c-myc* genes of other species.

The *Xenopus c-myc* gene(s) are expressed in a variety of tissues at a low level but oocytes contain a large store of maternal *c-myc* mRNA (about 5×10^6 copies per stage VI oocyte). After fertilisation the accumulated *c-myc* mRNA is degraded to less than 40 copies per cell by late blastula. New transcription of the *c-myc* gene(s) begins at late gastrula to maintain on average 30 - 130 copies of *c-myc* mRNA per cell throughout the rest of embryonic development. Preliminary evidence suggests that there is a localisation of *c-myc* expression during neural development. Attempts to generate an antiserum against the *Xenopus c-myc* protein are also described.

These results indicate *Xenopus* possesses a class of genes that are regarded as crucially important in other systems. In particular the *Xenopus c-myc* gene(s) are highly conserved throughout evolution and show an unusual regulation

during oogenesis and early development where high *c-myc* expression is uncoupled from cellular proliferation. In addition the *c-myc* gene(s) may be involved in neural differentiation.

Acknowledgements

I would like to thank my project supervisor, Dr. P.J. Ford, for his valuable help and encouragement throughout the period of my Phd study and for putting up with me so well. I also thank my advisor, Dr. D.J. Finnegan, for his guidance during my time in Edinburgh. I am extremely grateful to Nora Hunter and Darrell Sleep, former members of our group, for lots of help and advice. I am also grateful to Steve Jackson, Elaine McGlynn, Robert Saunders, Vince Cunliffe and many others (I apologise sincerely for any omissions) for stimulating discussions and for providing materials during financially difficult times. I thank Graham Brown and Alan McEwan for photography and Stephanie Hamilton for typing. Finally I really appreciate Lena Carlsson for love and attention during the last months of this thesis. This work was supported by the Science and Engineering Research Council of Great Britain.

Abbreviations

aa	amino acid
amp	ampicilin
APS	ammonium persulphate
ATP	adenosine-5'-triphosphate
bp	base pair
BSA	bovine serum albumin
butyl PBD	2-(4'-tert-butylphenyl) -5-(4''-diphenyl)-1,3,4-oxadiazole
⁰ C	degrees centigrade
cAMP	adenosine-3'-5'-cyclic phosphate
cDNA	complementary deoxyribonucleic acid
Ci	Curie(s)
cm	centimeter(s)
cpm	counts per minute
ctab	cetyltrimethylammonium bromide
CTP	cytidine-5'-triphosphate
(d)dATP	2'(3'-di)deoxyadenosine-5'- triphosphate
(d)dCTP	2'(3'-di)deoxycytidine-5'- triphosphate
(d)dGTP	2'(3'-di)deoxyguanosine-5'- triphosphate
(d)dTTP	2'(3'-di)deoxythymidine-5'- triphosphate
(d)dNTP	2'(3'-di)deoxynucleotide-5'- triphosphate
ddH ₂ O	double distilled H ₂ O
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
DTT	dithiothreitol
EDTA	diaminoethanetetra-acetic acid
g	gram(s)
GTP	guanosine-5'-triphosphate
>	greater than
Hepes	N-2-hydroxyethylpiperazine-N' -2-ethanesulphonic acid
hr	hour(s)
³ H	tritium (β emitting isotope of hydrogen)
kb	kilobase
kbp	kilobase pair
kd	kilodalton
klenow	large fragment of DNA polymerase I
krpm	kilorevolutions per minute
LTR	long terminal repeat
<	less than
M	molar
mA	milliampere(s)
mCi	millicurie(s)
mg	milligram(s)
ml	millilitre(s)

mm millimeter(s)
 mM millimolar
 mmol millimole(s)
 moi multiplicity of infection
 MOPS morpholinopropanesulfonic acid
 mRNA messenger ribonucleic acid
 MS22 ethyl-M-aminobenzoate
 mwt molecular weight
 MY million years
 NAD β -nicotinamide-adenine
 dinucleotide
 ng nanogram(s)
 nm nanometer(s)
 nmol nanomole(s)
 NP40 nonidet P40
 OD optical density
³²P β emitting isotope of phosphorus
 PEG polyethylene glycol
 % percentage
 pfu plaque forming units
 pg picogram(s)
 pH $-\log_{10}$ (hydrogen ion concentration)
 PMSF phenylmethylsulfonyl fluoride
 poly(A) polyadenylic acid
 poly(U) polyuridylic acid
 psi pounds per square inch
 RNA ribonucleic acid
 RNase ribonuclease
 rRNA ribosomal ribonucleic acid
 rpm revolutions per minute
 S Svedberg unit
³⁵S β emitting isotope of sulphur
 SAM S-adenosyl-L-methionine
 SDS sodium dodecyl sulphate
 str streptomycin
 tet tetracycline
 TEMED NNN'-tetra-methyl-1,2-diamino
 -ethane
 Tris tris(hydroxymethyl)-amino-methane
 Triton X-100 octylphenoxypolyethoxyethanol
 tRNA transfer ribonucleic acid
 U unit
 UTP uridine-5'-triphosphate
 uv ultraviolet
 μ Ci microcurie(s)
 μ g microgram(s)
 μ l microlitre(s)
 μ M micromolar
 μ mol micromole(s)
 V volt(s)
 v/v volume per volume
 w/v weight per volume

STANDARD AMINO ACID ABBREVIATIONS

A alanine
D aspartic acid
E glutamic acid
F phenylalanine
G glycine
H histidine
I isoleucine
K lysine
L leucine
M methionine
N asparagine
P proline
Q glutamine
R arginine
S serine
T threonine
V valine
W tryptophan
Y tyrosine

TABLE OF CONTENTS

Declaration	I
Abstract	II
Acknowledgements	IV
Abbreviations	V
Contents	VIII
Chapter 1 Introduction	1
1.1 Introduction	1
1.2 Discovery of Cellular Oncogenes	1
1.2.1 Proto-oncogenic progenitors of retroviral oncogenes	1
1.2.2 Functional tests identifying cellular oncogenes	2
1.2.3 Proviral insertions activating cellular oncogenes	3
1.2.4 Chromosomal translocations and cellular oncogenes	3
1.2.5 Cellular oncogenes in amplified DNA	4
1.2.6 Specific Cellular Oncogenes	4
1.3.1 Src	4
1.3.2 Erb B, fms and erb A	5
1.3.3 Sis	7
1.3.4 Ras	8
1.3.5 Fos	9
1.3.6 Other oncogenes	10
1.4 The C-myc Oncogene	11
1.4.1 Introduction	11
1.4.2 V-myc and c-myc	11
1.4.3 C-myc gene structure	12
1.4.4 The c-myc protein	14
1.4.5 C-myc is expressed in many normal cell types	16
1.4.6 C-myc uses two promoters	17
1.4.7 Post-transcriptional and transcriptional controls in c-myc expression	17
1.4.8 Regulation of c-myc expression in relation to the cell cycle	20
1.4.9 Regulation of c-myc expression in relation to differentiation	21
1.4.10 Activation of c-myc by retroviral insertion	22
1.4.11 Amplification of c-myc in tumour cell lines	23
1.4.12 Chromosomal translocations involving the c-myc locus	23
1.4.13 Studies with transfected myc genes and transgenic mice	24
1.4.14 Myc gene families: N-myc and L-myc	27
1.5 Gene Expression During Xenopus Development	29
1.5.1 Introduction	29
1.5.2 Gene expression in oogenesis	31
1.5.2.1 Yolk proteins	31
1.5.2.2 Ribosomes, rRNA and tRNA	32
1.5.2.3 mRNA	34
1.5.3 Gene expression during embryogenesis	36
1.5.3.1 RNA synthesis and midblastula transition	37
1.5.3.2 RNA populations during embryogenesis	38
1.5.4 Developmental decisions during embryogenesis	39
1.5.4.1 Homeobox containing genes	39
1.5.4.2 Localised factors	41
1.5.4.3 Expression at particular developmental stages	43
1.6 Objectives of this Thesis	44
Chapter 2 Materials and Methods	46
2.1 Solutions	46

2.2 Xenopus	49
2.2.1 Xenopus stocks	49
2.2.2 Egg and embryo production	49
2.2.2.1 Live matings	49
2.2.2.2 Artificial fertilisation	49
2.2.3 Oocytes and other tissues	49
2.3 Microbiological Strains and Media	50
2.3.1 Microbiological strains	50
2.3.2 Media	50
2.4 General Methods	53
2.4.1 Deproteinisation of nucleic acids —	53
2.4.2 Precipitation of nucleic acids —	53
2.4.3 Autoradiography	53
2.4.4 Deionised formamide	54
2.4.5 Restriction endonuclease digestion of DNA —	54
2.4.6 Quantitation of nucleic acids	54
2.4.7 Agarose gel electrophoresis	54
2.4.7.1 DNA gels	54
2.4.7.2 RNA gels	55
2.4.8 Gel photography	55
2.4.9 Ctab precipitation	55
2.5 Transformation of E.coli	56
2.6 Ligation of DNA	56
2.7 Extraction of DNA from Agarose Gels	56
2.7.1 Low melting point agarose gels	56
2.7.2 Electroelution	57
2.8 Elutip-d Column Purification of DNA	57
2.9 Plasmid DNA preparation	57
2.9.1 Mini-preparation of plasmid DNA	57
2.9.2 Maxi-preparation of plasmid DNA	58
2.10 Bacteriophage Lambda DNA Preparation	58
2.10.1 Plating of phages	58
2.10.2 Plate lysates	59
2.10.3 Liquid lysates	59
2.10.4 CsCl step gradients	59
2.10.5 Phage DNA extraction	59
2.11 Xenopus DNA Preparation	60
2.12 Xenopus RNA Preparation	60
2.12.1 Ovary RNA	60
2.12.2 Blood RNA	60
2.12.3 Other tissue RNA	61
2.12.4 Staged oocyte and embryo RNA	61
2.13 Poly(A)+ RNA Preparation	62
2.14 Poly(A) Content of RNA	62
2.15 Labeling of Nucleic Acids	62
2.15.1 End labeling and filling in of restriction sites	62
2.15.2 Kinasing of linkers	63
2.15.3 Kinasing of RNA	63
2.15.4 Nick translation	64
2.15.5 Oligolabeling	64
2.15.6 M13 single stranded probes	64
2.15.7 T7 RNA polymerase probes	65
2.16 Transfer of Nucleic Acids to Membrane Filters	66
2.16.1 Southern blots	66
2.16.2 Northern blots	66

2.16.3 Phage lifts	66
2.16.4 Dot blots	67
2.17 Hybridisation of Membrane Filters with Radioactive Probes	67
2.18 Preparation of a Previtellogenic Ovary cDNA library	67
2.18.1 First strand synthesis	68
2.18.2 Second strand synthesis	68
2.18.3 EcoRI methylation and end repair	68
2.18.4 EcoRI linker ligation and size fractionation	69
2.18.5 Cloning of cDNA into bacteriophage lambda NM1149	69
2.18.6 Screening of phage cDNA libraries	70
2.19 Dideoxynucleotide DNA Sequencing	70
2.19.1 General	70
2.19.2 Cloning into M13mp18 and M13mp19	72
2.19.3 M13 template preparation	72
2.19.4 Sequencing reactions	73
2.19.5 Polyacrylamide gel electrophoresis	73
2.20 Protein Methods	74
2.20.1 Laemmli polyacrylamide gels	74
2.20.2 Immunoblotting	75
2.20.3 Preparation of β galactosidase-fusion proteins	76
2.20.4 Preparation of <i>Xenopus</i> ovary proteins	76
2.20.5 Generation of rabbit anti-fusion antiserum	77
2.20.6 Affinity purification of antiserum	77
Chapter 3 Identification of Oncogene Related Sequences in <i>Xenopus</i>	79
3.1 Introduction	79
3.2 Detection of Oncogene Related Sequences in <i>Xenopus</i> DNA	80
3.3 Detection of Oncogene Related Sequences in <i>X.borealis</i> RNA	86
3.4 Discussion	90
Chapter 4 Isolation and Expression of C-myc Related Sequences in <i>X.borealis</i>	93
4.1 Introduction	93
4.2 Isolation of C-myc Related Sequences from <i>X.borealis</i> cDNA Libraries	94
4.2.1 Construction of a cDNA library	94
4.2.2 Isolation of c-myc related cDNA clones	94
4.3 Characterisation of C-myc Related cDNA clones	97
4.3.1 Northern blotting analysis	97
4.3.2 Southern blotting analysis	98
4.4 Expression of <i>X.borealis</i> C-myc Genes	100
4.4.1 Tissue specific expression	101
4.4.2 Level of expression in stage II oocytes	101
4.4.3 Expression during oogenesis and development	102
4.4.4 Expression in dissected embryos	104
4.5 Attempts to Analyse XbmycA Protein Expression	106
4.6 Discussion	107
Chapter 5 Sequence Analysis of C-myc Related Clones from <i>X.borealis</i>	112
5.1 Introduction	112
5.2 XbmycA1	112
5.3 XbmycA2 and XbmycA3	113
5.4 XbmycA and XlmycA	115
5.5 Features of XbmycA Nucleic Acid Sequences	116
5.5.1 Poly(A) addition site	116
5.5.2 Rapid degradation signals	117
5.5.3 Comparison with other c-myc sequences	117

5.6 Features of XbmycA Protein Sequences	120
5.6.1 Comparison with other c-myc protein sequences	120
5.6.2 Nuclear location signals	122
5.6.3 Rapid degradation signals	122
5.6.4 The myc boxes	123
5.7 Discussion	124
Chapter 6 Discussion	127
6.1 Oncogene Homologies in Xenopus	127
6.2 The X.borealis C-myc Genes	127
6.3 Expression of X.borealis C-myc Genes	128
6.4 The X.borealis C-myc Protein	131
6.5 Some Other Future Experiments	132
6.6 Recent Publications on the X.laevis C-myc Gene	133
6.7 Conclusions	135
References	136
Appendix 1	155
Appendix 2	156

CHAPTER 1

INTRODUCTION

1.1 Introduction

Although the causes of cancer are elusive it is the current belief in molecular oncology that a neoplastic cell develops from a normal progenitor cell as a result of mutations in a limited set of cellular genes. These mutant genes are known as active cellular oncogenes and their wild type counterparts as normal cellular oncogenes or proto-oncogenes (*c-onc's*). Normal cellular oncogenes are strongly implicated in the growth control of cells hence they may be involved in differentiation and development. Oncogene research is one of the most intensively studied areas in modern molecular biology and as a result this introduction will necessarily be selective. Many excellent reviews have been written which can be consulted for more detail (Bishop 1983, 1985, Deuschberg 1983, Muller and Verma 1984, Varmus 1984, Bishop and Varmus 1985, Hall 1986). *Xenopus* was used as a model organism to study cellular oncogenes in this thesis therefore the final part of this introduction will discuss gene expression in *Xenopus*.

1.2 Discovery of Cellular Oncogenes

In 1911 Rous discovered, in chickens, Rous sarcoma virus (RSV) – the first acutely transforming retrovirus (Rous 1911). Over 20 such viruses have now been isolated from a variety of birds and mammals (table 1.1) each carrying a distinct piece of genetic information responsible for the rapid induction of tumours and neoplastic transformation of cultured cells i.e. the viral oncogene (*v-onc*). In the early 1970's it was felt that cellular DNA might contain the potential to create viral or (active) cellular oncogenes by somatic mutation and/or reverse transcription (Huebner and Todaro 1969, Temin 1971). Now we know that cells from a variety of organisms harbour genetic loci homologous to retroviral oncogenes and that these loci are cellular not viral in origin. Also some cellular oncogenes with no viral counterparts have been isolated hence there are probably many genes important in the genesis of cancer.

1.2.1 Proto-oncogenic progenitors of retroviral oncogenes

The largest group of cellular oncogenes have been identified by the homology of their nucleotide sequences to retroviral oncogenes. The cellular

Retroviral oncogenes (*v-onc*'s)

Prototype virus(es)	<i>v-onc</i>	Species of origin	Genetic basis			
			Conditional mutants	Non-conditional mutants	In vitro recombinant *	Multiple transduction **
RSV	<i>src</i>	chicken	+	+	+	
Y73/ESV	<i>yes</i>	chicken	+			+
FuSV/PRCII	<i>fps</i>	chicken	+		+	+
UR2	<i>ros</i>	chicken	+			
MC29	<i>myc</i>	chicken		+	+	+
MH2	<i>myc</i> <i>mil</i>	chicken				+
AMV	<i>myb</i>	chicken	+			+
E26	<i>myb</i> <i>ets</i>	chicken				+
AEV	<i>erbA</i> <i>erbB</i>	chicken		+		+
SKV770	<i>ski</i>	chicken			+	
REV-T	<i>rel</i>	turkey		+	+	
Mo-MSV	<i>mos</i>	mouse	+	+	+	
Ab-MLV	<i>abl</i>	mouse		+	+	+
MSV-3611	<i>raf</i>	mouse				+
BALB-MSV	Ha- <i>ras</i>	mouse			+	+
Ha-MSV	Ha- <i>ras</i>	rat		+	+	+
RaSV	Ra- <i>ras</i>	rat				+
Ki-MSV	Ki- <i>ras</i>	rat	+		+	
FBJ-MSV	<i>fos</i>	mouse		+	+	+
FBR-MSV	<i>fos</i> <i>fox</i>	mouse				+
ST-Fe SV, GA-FeSV	<i>fes</i>	cat		+		+
SM-Fe SV	<i>fms</i>	cat				
GR-Fe SV	<i>fgr</i> <i>actin</i>	cat				+
PI-FeSV, SSV	<i>sis</i>	cat, woolly monkey			+	+

Abbreviations: RSV, Rous sarcoma virus; Y73, Yamaguchi 73 sarcoma virus; ESV, Esh sarcoma virus; FuSV, Fujinami sarcoma virus; MC29, myelocytomatosis-29 virus; MH2, Mill Hill-2 virus; AMV, avian myeloblastosis virus; AEV, avian erythroblastosis virus; REV-T, reticulo-endotheliosis virus, strain T; Mo-MSV, Moloney murine sarcoma virus; Ab-MLV, Abelson murine leukemia virus; Ha-MSV, Harvey murine sarcoma virus; RaSV, Rasheed rat sarcoma virus; Ki-MSV, Kirsten murine sarcoma virus; ST-, GA-, SM-, GR-, PI-FeSV, Snyder-Theilen, Gardner-Arnstein, Susan McDonough, Gardner-Rasheed, Parodi-Irgens feline sarcoma viruses; SSV, simian sarcoma virus.

+ means that the *v-onc* sequence has been found to have transforming activity when moved to a new context by in vitro manipulation of cloned DNA

+ means that the *onc* sequence has been encountered in two or more viruses that appear to have independently transduced the same *c-onc* from one or more host species

v-yes and *v-fgr* appear to be derived from the same cognate *c-onc*

v-fps and *v-fes* are derived from the same cognate *c-onc*

v-mil (also known as *v-mht*) and *v-raf* are derived from the same cognate *c-onc*

Ha-*ras* of BALB-MSV (also known as *v-bax*), Ha-*ras* of Ha-MSV, and Ra-*ras* of RaSV are all derived from the same cognate *c-onc*, called *c-Ha-ras*

Part of the mammalian β -actin sequence is found in the genome of GR-FeSV

Table 1.1 (From Varmus 1984)

oncogenes apparently have served as targets for genetic transduction by retroviruses. Mutations are introduced during the process rendering the cellular gene an active viral oncogene. The selective pressure for the evolution of oncogenic retroviruses remains unclear.

In 1976 it was established, by using cDNA specific for *v-src* of RSV, that closely related sequences existed in the genome of chickens (Stehelin *et al* 1976). Subsequently *src* related sequences were detected in a variety of vertebrates from fish to man (Spector, Varmus and Bishop 1978). With the advent of molecular cloning it has now been confirmed that all the viral oncogenes in table 1.1 have cellular counterparts (Bishop 1983, Varmus 1984, Muller and Verma 1984). In fact cellular oncogene related sequences have been isolated from *Drosophila* for *src*, *abl*, *erbB*, *ras* and *myb* (Hoffman-Falk *et al* 1983, Simon *et al* 1985, Hoffman *et al* 1983, Linveh *et al* 1985, Neumann-Silberberg *et al* 1984, Bishop *et al* 1985). Even yeast has *ras* related sequences (Defeo-Jones *et al* 1983, Powers *et al* 1984). Such high evolutionary conservation strongly implies that cellular oncogene products play an essential and basic role in normal cellular growth and development.

1.2.2 Functional tests identifying cellular oncogenes

An appropriate recipient cell line can be neoplastically transformed by the introduction of nuclear DNA from tumours or transformed cell lines (Cooper 1982, Land, Parada and Weinberg 1983). This perceives the cellular oncogene as an active oncogene and that the recipient cells acquire a dominant mutant allele. For example the lesion in the *ras* family of oncogenes is a single point mutation at amino acids 12, 13 or 61 (Hall 1984, Levinson 1986). Table 1.2 indicates the candidate oncogenes detected by this assay.

Two major limitations to this are that it examines oncogenic potential in a cell type and species that differ from the donor and that the recipient is already an established cell line. This may partly explain why 80% of spontaneous human tumours fail in this assay (Krontriss and Cooper 1981, Cooper 1982) and why many active genes identified belong to the *ras* family.

Primary cells have now been transformed using combinations of cloned oncogenes (Land, Parada and Weinberg 1983, Ruley 1983). Co-operation of a nuclear gene such as *myc*, adenovirus E1a, polyoma large T or p53 (with supposed immortalisation function) together with *ras*, adenovirus E1b or

Candidate oncogenes detected as transforming genes in the NIH/3T3 cell assay

Genes	Examples of cell types in which transforming activity is found ^a
<i>c-Ha-ras1</i>	Bladder carcinoma cell line (human) Urinary tract tumors (human) Lung carcinoma cell line (human) DMBA and NMU-induced mammary carcinomas (rat) DMBA/TPA-induced skin papillomas, benign and malignant (mice) Melanoma cell line (human) Mammary carcinosarcoma line (human) MC, BP, DEN and MNNG-transformed primary cells (guinea pig) Myeloid tumor cell line (mouse)
<i>c-Ki-ras2</i>	Lung carcinomas and cell lines (human) Colon carcinomas and cell lines (human) Pancreatic carcinoma cell line (human) Gall bladder carcinoma cell line (human) Rhabdomyosarcoma cell line (human) Ovarian carcinoma, primary (human) Acute lymphocytic leukemia line (human) MC-induced fibrosarcomas and MC-transformed fibroblasts (mouse) γ -irradiation induced thymoma (mouse) MC-induced thymic lymphoma and macrophage tumor lines (mouse) BP-induced fibrosarcoma (mouse)
<i>N-ras</i>	Neuroblastoma cell line (human) Burkitt lymphoma line (human) Fibrosarcoma and rhabdomyosarcoma cell lines (human) Promyelocytic leukemia line (human) T cell leukemia line (human) Melanoma cell lines (human) Teratocarcinoma line (human, late passage) Lung carcinoma line (human) Acute and chronic myeloblastic leukemia (human) Lung carcinoma cell line (mouse) Carcinogen-induced thymoma (mouse)
<i>c-mos</i>	Plasmacytoma cell lines (mouse)
<i>c-erbB-related</i>	ENU-related neuroblastoma cell lines (rat)
<i>Blym</i>	ALV-induced bursal lymphomas (chickens) Burkitt lymphoma lines (human)
<i>Tlym</i>	Intermediate T cell lymphoma lines (human) MLV-induced T cell lymphomas (mouse)
<i>tx-1^b</i>	Mammary carcinoma cell line (human) MMTV and DMBA-induced mammary carcinomas (mouse)
<i>tx-2^b</i>	PreB tumor cell lines (human) AbMLV-induced pre B tumor cell lines (mouse)
<i>tx-3^b</i>	Myeloma and plasmacytoma cell lines (human, mouse)
<i>tx-4^b</i>	Mature T cell lymphoma lines (human, mouse)
Unnamed	Pre B cell leukemia cell line (human)
Unnamed	MNNG-treated osteosarcoma cell line

^a Abbreviations: MC, methylcholanthrene; DMBA, dimethylbenzanthracene; TPA, tetradecanoyl-13-phorbol acetate; NMU, nitrosomethylurea; MNNG, N-methyl-N-nitro-N-nitroguanidine; BP, benz(a)pyrene; DEN, diethylnitrosamine; ENU, ethylnitrosourea.

Table 1.2 (From Varmus 1984)

polyoma middle T (with supposed transformation function) is sufficient to induce neoplastic transformation. It is unclear whether two oncogenes are inevitably required as very high levels of a mutant *ha-ras* allele alone can transform primary rodent cells (Spandidos and Wilkie 1984).

In vivo animal models using a variety of chemical carcinogens to generate tumours in a reproducible manner have identified active oncogenes (Table 1.3; Barbacid 1986). The *ras* family appear to predominate in this assay and table 1.4 describes the reproducibility of lesions in the *ha-ras* activated tumours. It seems these agents generate adducts with bases in the *ha-ras* DNA leading to mutation which occurs shortly after application of the carcinogen. Therefore oncogene activation is concomitant with the initiation of carcinogenesis.

1.2.3 Proviral insertions activating cellular oncogenes

Most retroviruses lack a viral oncogene but still induce tumours after a latent period. These tumours appear to contain mutant cellular oncogenes that are activated by proviral insertions (Nusse 1986). The insertion event seems to be of prime importance causing an increase in the level of expression of the target gene influenced by strong promoter or enhancer elements present in the long terminal repeat (LTR) of the provirus but sometimes the coding sequences also get altered. The first gene identified by this means was the chicken *c-myc* gene which had avian leukosis virus DNA integrated next to it (Hayward, Neel and Astrin 1981, and later). Table 1.5 lists the other oncogenes that have been identified and it is interesting to note many are previously unknown. Note that the provirus in all the resulting tumours act in a dominant way as a consequence of selection by investigators though there is evidence to suggest mutations in tumour cells are recessive. In the future it is possible insertional mutagenesis by retroviruses will help in the isolation of such sequences.

1.2.4 Chromosomal translocations and cellular oncogenes

Chromosomal translocations frequently occur in tumour cells. In solid tumours these changes often are heterogeneous and not well defined but in hematopoietic malignancies specific translocations are common (Rowley 1980, 1983). Interest has focused on whether any cellular oncogenes reside at the translocation breakpoints such that they undergo structural and functional alterations (Rowley 1983). Both the cellular *myc* and *abl* genes have been implicated as such oncogenes (*myc* will be discussed later). In chronic

Oncogenes in carcinogen-induced animal tumors

Species	Carcinogen	Tumor	Oncogene	Proportion of tested tumors containing transforming gene ^a (%)
Rat	NMU	Mammary carcinomas	H- <i>ras</i> -1	86 (80)
	DMBA	Mammary carcinomas	H- <i>ras</i> -1	23 (50)
	DMN	Renal carcinomas	K- <i>ras</i> -2	40 (25)
	ENU	Neuroblastomas	<i>neu</i>	100 (3)
	NMU	Schwannomas	<i>neu</i>	70 (10)
	MMS	Nasal carcinomas	?	100 (8)
Mouse	DMBA	Skin carcinomas	H- <i>ras</i> -1	90 (36)
	Several	Hepatomas	H- <i>ras</i> -1	100 (25)
	X-rays	Lymphomas	K- <i>ras</i> -2	57 (7)
	NMU	Lymphomas	N- <i>ras</i> -1	85 (6)
	3-MCA	Fibrosarcomas	K- <i>ras</i> -2	50 (4)

^aNumber in parentheses indicates number of tumors tested.

Table 1.3 (From Barbacid 1986)

Activation of H-ras-1 oncogenes by specific mutations in carcinogen-induced animal tumors

Species	Type of tumor	Carcinogen	Activating mutation	Incidence of activating mutation in tumors
Rat	Mammary carcinomas	NMU	G ³⁵ → A	61/61
	Mammary carcinomas	DMBA	A ¹⁸² /A ¹⁸³ → N	5/5
Mouse	Skin papillomas/carcinomas	DMBA	A ¹⁸² → T	33/34
Mouse	Skin carcinomas	DMBA	A ¹⁸² → T	3/3
	Skin papillomas/carcinomas	DBACR	A ¹⁸² → T	6/6
Mouse	Hepatocarcinomas	HO-AFF	C ¹⁸¹ → A	7/7
	Hepatocarcinomas	VC	A ¹⁸² → T	6/7

Table 1.4 (From Barbacid 1986)

Candidate oncogenes affected by insertion mutations

Gene	Insertion mutagen ^a	Tumor type	Species	Documented effect upon expression
<i>c-myc</i>	ALV, CSV, RPV	B cell lymphoma	chicken	+
	CSV	B cell lymphoma	chicken	
	RPV	B cell lymphoma	chicken	+
	MAV	B cell lymphoma	quail	
	Mo-MLV, MCF-MLV	T cell lymphoma	rat, mouse	+
	FeLV	T cell lymphomas	cat	
<i>c-erb B</i>	ALV	erythroblastosis	chicken	+
<i>c-Ha-ras</i>	MAV	nephroblastoma	chicken	+
<i>c-mos</i>	IAP	plasmacytomas	mouse	+
<i>c-myb</i>	Mo-MLV	plasmacytoid lymphosarcomas	mouse	
<i>int-1</i>	MMTV	mammary carcinoma	mouse	+
<i>int-2</i>	MMTV	mammary carcinoma	mouse	+
<i>Mlvi-1</i>	Mo-MLV	T cell lymphoma	rat	
<i>Mlvi-2</i>	Mo-MLV	T cell lymphoma	rat	
<i>Mlvi-3</i>	Mo-MLV	T cell lymphoma	rat	
<i>RMO-int-1</i>	Mo-MLV	T cell lymphoma	rat	
<i>pim-1</i>	MCF-MLV, MO-MLV	T cell lymphoma	mouse	+
<i>c-raf</i>	MLV-LTR ^b	transformed fibroblast	mouse	

^aAbbreviations: ALV, avian leukosis virus; CSV, chicken syncytial virus; RPV, ring-necked pheasant virus; MAV, myeloblastosis-associated virus; Mo-MLV, Moloney murine leukemia virus; FeLV, feline leukemia virus; IAP, intracisternal A particle; MMTV, mouse mammary tumor virus; MCF-MLV, mink cell focus-forming murine leukemia virus.

^bCloned DNA containing an MLV LTR was used as an insertion mutagen via transfection.

Table 1.5 (From Varmus 1984)

myelogenous leukemia the *abl* gene on chromosome 9 is translocated to a specific region of chromosome 22 called bcr (Heisterkamp *et al* 1983, Groffen *et al* 1984). Table 1.6 summarises the candidate oncogenes involved with chromosomal translocations.

1.2.5 Cellular oncogenes in amplified DNA

Many tumour cell lines manifest amplified cellular DNA detected cytologically as homogeneously staining regions (HSRs) in chromosomes or as small mini chromosomes called double minutes (DMs). It was hoped these regions might contain amplified cellular oncogenes that lead to unexpectedly high levels of the gene. Table 1.7 indicates the oncogenes detected in amplified DNA of tumour cells often such amplification does lead to the expression of the target gene but it should be noted that the amplified DNA spans a large region on either side of the gene and its effects on tumour formation are not clear. Most cases of amplified cellular oncogenes are sporadic but a consistent pattern of gene amplification has been found in human neuroblastomas of the *N-myc* gene and will be discussed later (Schwab 1985).

1.3 Specific Cellular Oncogenes

Many different oncogenes have now been uncovered and how they act can only be answered for a few in some detail. Cellular oncogenes appear to carry out tasks in many aspects of cell growth—some are growth factors, some are growth factor receptors, some may mediate intracellular signalling and some may be involved in the control of transcription. Here I only briefly outline some of the details concerning some oncogenes and readers are asked to consult the relevant reviews and references therein for further details. Section 1.4 discusses the *c-myc* oncogene in more detail as the majority of the work carried out in this thesis concerns this gene.

1.3.1 Src

Both *c-src* and *v-src* are 60kD phosphoproteins that possess tyrosine kinase activity (see review by Sefton 1985). *Src* is one member of a family of oncogene products with tyrosine kinase activity which includes *src*, *yes*, *abl*, *fgr*, *fps*, *ros*, *fms*, and *erbB*. A stretch of about 250 amino acids in *src* responsible for the kinase activity has been found to be conserved in the other

Translocated cellular oncogenes

Tumor	Known or candidate oncogenes	Chromosome	Partner chromosome	Locus on partner chromosome
Mouse plasmacytoma cell lines	<i>c-myc</i>	15(D2/3)	12(F1)	IgH
Rat immunocytoxa cell lines	<i>c-myc</i>	7	6	?
Human Burkitt lymphoma cell lines	<i>c-myc</i>	8(q24)	14(q32) 2(p13) 22(q11)	IgH IgK Igλ
Human chronic myelogenous leukemia cell lines	<i>c-abl</i>	9(q34→pter)	22(q11)	<i>bcr</i>
Human B cell leukemia and lymphoma	" <i>bcl-1</i> "(?)	11(q13)	14(q32)	IgH
Murine plasmacytoma	" <i>T_KNS-1</i> "(?)	10	6(C2)	C _K
Murine plasmacytomus	" <i>pvt</i> "(?)	15	6(C2)	C _K

Table 1.6 (From Varmus 1984)

Amplification of cellular oncogenes

Oncogene	Tumor	Source	Degree of amplification	DM/HSR
<i>c-myc</i>	Promyelocytic leukemia cell line, HL60 and primary tumor ^a	Human	20×	a
<i>c-myc</i>	APUDoma cell line, COLO320	Human	40×	+/+
<i>c-myc</i>	Small cell lung carcinoma cell lines (variants)	Human	5-30×	?
<i>c-myc</i>	CSV-induced bursal lymphoma	Chicken	5-10×	?
<i>N-myc</i>	Primary neuroblastomas (Stage III and IV) and neuroblastoma cell lines	Human	5-1000×	+/+
<i>N-myc</i>	Retinoblastoma cell line, Y79, and primary tumors	Human	10-200×	+/+
<i>N-myc</i>	Small cell lung carcinoma cell lines and tumor	Human	50×	+/?
<i>c-abl</i>	Chronic myeloid leukemia cell line, K562	Human	5-10×	?
<i>c-myb</i>	Colon carcinoma cell lines, COLO201/205	Human	10×	?
<i>c-myb</i>	Acute myeloid leukemia	Human	5-10×	?
<i>c-erbB</i>	Epidermoid carcinoma cell line, A431	Human	30×	?
<i>c-Ki-ras2</i>	Primary carcinomas of lung, colon, bladder and rectum	Human	4-20×	?
	Colon carcinoma cell line, SW480			?
<i>c-Ki-ras</i>	Adrenocortical carcinoma cell line, Y1	Mouse	30-60×	+/+
<i>N-ras</i>	Mammary carcinoma line, MCF7	Human	5-10×	?

^aDMs were observed in early passages of HL60 cells

Table 1.7 (From Varmus 1984)

tyrosine kinases. The *src* protein (and *yes*, *abl*, *fgr*, *fps*, *ros* also) have been found to be localised on the inner plasma membrane (Courtneidge *et al* 1980). Tyrosine phosphorylation appears to be a normal event in the regulation of cell division as growth factor receptors possessing tyrosine kinase activity activate this activity on binding the appropriate growth factor. Since *src* is located on the inner side of the plasma membrane it makes it unlikely that *src* is a growth factor receptor. The targets for phosphorylation by *src* are unclear but in RSV-transformed cells cytoskeletal, glycolytic and unknown proteins all show increased levels of phosphotyrosine though it is not clear if this plays a role in transformation. Phosphorylation of a 42kD protein is a common response in normal cells exposed to a variety of mitogens (Cooper *et al* 1984). *In vitro* *v-src* can phosphorylate phosphatidylinositol which breaks down to yield diacylglycerol and inositoltriphosphate (Sugimoto *et al* 1984). Activation of the serine specific protein kinase C results by increased Ca^{2+} levels mediated by inositoltriphosphate and by direct action of diacylglycerol. Protein kinase C is thought to play a central role in the various cellular responses to mitogenic stimulation including the phosphorylation of ribosomal protein S6 which seems to be a critical event in the mitogenic stimulation of quiescent cells. A cAMP dependent kinase can also activate protein kinase C in response to high cAMP levels generated by adenylate cyclase in response to growth factors receptors binding their agonist. It has been shown that *v-src* can phosphorylate cAMP dependent kinase (Graziani *et al* 1984) therefore there may be two distinct biochemical pathways by which *src* can play a role in mitogenic stimulation. Expression of *src* appears to be ubiquitous (Muller and Verma 1984) but high levels of *src* have been found in neural tissue during early development both in *Drosophila* and chickens (Simon *et al* 1985, Sorge *et al* 1985). *Src* appears to be expressed in non-proliferating but differentiating neural cells at high levels therefore maybe the role of *c-src* is not only as mitogenic signal but also a differentiation signal.

1.3.2 Erb B, fms and erb A

A connection between oncogenes and growth control was realised by the discovery that the *v-erb B* oncogene is almost certainly derived from the epidermal growth factor (EGF) receptor gene (Downward *et al* 1984, Ullrich *et al* 1984). The EGF receptor (*c-erb B*) gene product which is an integral transmembrane protein is one of the best characterised growth factor receptors (see Carpenter 1984, Cohen 1984 for review). EGF is a small

polypeptide hormone which can induce DNA synthesis and cell proliferation. On binding EGF, by the external domain of the protein, the receptor will autophosphorylate at a tyrosine residue in the cytoplasmic carboxyl terminal of the protein and also will phosphorylate other unidentified proteins. The receptor then clusters and subsequently becomes internalised leading to its eventual degradation. It is not clear how cellular proliferation is signaled by EGF binding to its receptor but it is interesting that the highly purified receptor has DNA nicking activity (Mroczkowski, Mosig and Cohen 1984) and that the receptor can cause increased guanine nucleotide binding and phosphorylation of *ras* gene products (Kamata and Feramisco 1984).

The *v-erb B* oncogene appears to be truncated version of the EGF receptor such that it has sustained a large deletion of the heavily glycosylated amino terminal domain which lies external to the cell and a small deletion of the cytoplasmic domain removing the site of autophosphorylation (see Hunter 1985, Hayman 1986 for reviews). This suggests *v-erb B* mimics the occupied EGF receptor, but it is difficult to demonstrate that *v-erb B* has tyrosine kinase activity *in vitro* and also avian erythroblastosis virus (AEV) transformed cells show only small increases in cellular phosphotyrosine levels. AEV also transforms erythroblasts which do not possess EGF receptors (fibroblasts do) therefore if *v-erb B* is mimicking a constitutively activated EGF receptor it is not a normal signal for these cells. This suggests *v-erb B* protein transforms cells by a pleiotropic mechanism indeed it seems different domains of the protein appear to influence the transformation of different cell types.

Since rather few growth factor receptors have been characterised then it is possible that other oncogene products with tyrosine (or serine) kinase activity could be growth factor receptors. The *v-fms* gene product has properties similar to growth factor receptors (Hunter 1985). It is an integral transmembrane protein with a heavily glycosylated amino terminal domain outside the cell. Also it has tyrosine kinase activity and can be clustered into coated pits. Recently it has been shown that the *c-fms* gene product is almost certainly the growth factor receptor for colony stimulating factor (CSF) 1 (Sherr *et al* 1985). The macrophage growth factor, CSF 1, stimulates hematopoietic precursor cells to form colonies containing mononuclear phagocytes. CSF 1 is required by cells to enter DNA synthesis and stimulates precursors of mononuclear phagocytes to proliferate and differentiate. Although *c-fms* transcripts can be detected in a wide variety of organs this might reflect the

presence of tissue macrophages since the CSF 1 receptor is apparently restricted to mononuclear phagocytes. Transcripts of *c-fms* have also been found in the developing mouse placenta and extra embryonal membranes (Muller *et al* 1983) but it is not clear if this is a characteristic of cells other than that of the mononuclear phagocyte lineage. It is possible *v-fms* transforms cells by a similar mechanism to *v-erb B*

Very recently it was discovered that the *c-erb A* gene of both human and chickens is a thyroid hormone receptor (Sap *et al* 1986, Weinberger *et al* 1986 and review by Green and Chambon 1986). Sequence analysis shows *c-erb A* is similar to steroid hormone receptors even though steroid and thyroid hormones are not structurally or biosynthetically related hence their receptors probably evolved from common ancestors. Steroid hormone receptors can activate transcription by specific binding of hormone-receptor complexes to enhancer elements of target genes. Thyroid hormone receptors are known to be localised in the nucleus and are believed to modulate gene expression in a similar manner to steroid hormone receptors. Sequence comparison shows the putative DNA binding domain is conserved between receptors. Thyroid hormones are known to have many effects on differentiation. The *v-erb A* gene cannot bind hormone and may be constitutively active. Therefore altered enhancer factors may be important in oncogenic transformation by interfering with target genes. Since both polypeptide and thyroid hormone receptors have turned out to be known oncogenes maybe steroid hormone receptors will also be found as oncogene products.

1.3.3 Sis

Another pathway at which oncogenic proteins could intercede is to mimic a growth factor which may stimulate cells in an autocrine manner. This potential was realised by the finding the *v-sis* gene from simian sarcoma virus (SSV) was highly related to the B chain of platelet derived growth factor (PDGF), (Waterfield *et al* 1983, Doolittle *et al* 1983 and reviews by Heldin and Westermark 1984, Hunter 1985). It has since been shown that *c-sis* does encode PDGF B chain. PDGF is a potent mitogen for connective tissue cells also it is secreted by many cell types and has many possible biological roles including wound healing, development, atherosclerosis and neoplasia (see Ross, Raines and Bowen-Pope 1986 for review).

PDGF probably forms a heterodimer of an A and B chain whereas *v-sis* only forms a homodimer. SSV transformed cells secrete a PDGF-like mitogen factor therefore it seems the transformed cell has escaped normal growth control by both secreting and responding to growth factors. Some non-SSV transformed cell lines also secrete a PDGF-like activity. The differences between *v-sis* and *c-sis* are probably unimportant since overexpression of *c-sis* can transform cells. Interestingly cytotrophoblastic cells from placenta which exhibit a pseudo-malignant phenotype express high levels of *c-sis*, produce PDGF and have PDGF receptors (Goustine *et al* 1985). Therefore autocrine stimulation may be important for growth of both transformed and normal cells. PDGF on binding to its receptor stimulates a tyrosine kinase activity which may then mediate mitogenic events through pathways similar to those of the EGF receptor. Shortly after PDGF binding *c-fos* then *c-myc* expression is also drastically increased (see later) hence many oncogene products may interact to produce the signals for mitosis.

1.3.4 Ras

The *ras* oncogenes appear to be involved in yet another level of growth control (see Hall 1984, Levinson 1986 for reviews). There are three types of *ras* genes: *ha-ras*, *ki-ras* which were discovered through homology to retroviral oncogenes and *N-ras* which was discovered to be the active oncogene in a neuroblastoma cell line. The great interest in the *ras* gene family began in 1982 when it was found that a single point mutation in the *ha-ras* gene appeared to be the active lesion in a human bladder carcinoma (Reddy *et al* 1982, Tabin *et al* 1982, Taparowsky *et al* 1982). Subsequently it has been found that a single point mutation at positions 12, 13, 59, 61 or 63 can generate an active *ras* oncogene. Indeed using site directed mutagenesis of the *ha-ras* gene shows that any amino acid except proline or glycine at position 12 generates an active oncogene; interestingly proline and glycine are known to be α -helix breaking residues. Similarly such somatic mutations of the *ras* oncogenes appear to be an important step in tumour development of chemical induced cancers. This step appears to be a general event in tumourogenesis but sometimes it occurs late hence pointing to multistep carcinogenesis. Therefore it appears qualitative changes are important in *ras* influenced cancers.

The mammalian *ras* proteins all are located on the inner side of the plasma membrane, bind guanine nucleotides, have GTPase activity and have molecular

weights of ~21kD. It appears that activating mutations inhibit GTPase activity. Recent interest focuses on the homology of *ras* proteins with G proteins. G proteins signal hormone binding to receptors by coupling to the generation of intracellular second messages and by doing so stimulate their intrinsic GTPase activity. Two *ras* gene homologs, RAS1 and RAS2, have been isolated from yeast (Powers *et al* 1984). Neither yeast gene is essential but one gene is required for the viability of haploid spores (Tatchell *et al* 1984). It has recently been shown that the yeast *ras* genes can function as GTP dependent regulators of adenylate cyclase (Toda *et al* 1985). Since a yeast *ras* gene can transform mammalian cells and a human *ras* gene can substitute for yeast *ras* genes this lead to the suggestion mammalian *ras* proteins might act in a similar fashion but apparently mammalian *ras* products have no effect on adenylate cyclase (Beckner *et al* 1986). Also microinjection of purified activated *ras* proteins into *Xenopus* oocytes leads to release from prophase arrest (i.e. maturation) through an apparently cAMP independent pathway whereas progesterone induced maturation leads to dramatic increase of cAMP levels (Birchmeier *et al* 1985). Therefore if *ras* proteins in higher eukaryotes are acting as coupling proteins for signal transduction from the cell surface to intracellular targets they must achieve this by some mechanism other than cAMP control. Very recently it has been shown that *N-ras* couples the bombesin receptor to phospholipase C leading to an increase in intracellular second messages (Wakelam *et al* 1986) hence other *ras* proteins may couple other receptors in a similar manner.

1.3.5 Fos (see Verma 1986 for review)

The *fos* gene was first identified in two retroviruses, FBJ murine sarcoma virus (MSV) and FBR-MSV and now both human and mouse *c-fos* genes have been isolated. Like *myc* the *fos* proteins are located in the nucleus. Both *c-fos* and *v-fos* proteins are similar but the last 48 carboxyl amino acids are different in *c-fos* due to a switch in reading frame. Interestingly although *v-fos* alone can transform cells *c-fos* must first be linked to a LTR and also have a deletion of 3' non-coding sequences such that constitutive expression can be maintained (Miller, Curran and Verma 1984), therefore *c-fos* might normally regulate its own synthesis mediated by 3' non-coding sequences.

The most remarkable feature of *c-fos* is its rapid induction on response to mitogen or differentiation specific agents. On differentiation of

myelomonocytic cells to mature macrophages (but not granulocytes) the *c-fos* gene is rapidly induced within a few minutes with maximal levels observed by 30 minutes of induction which decline to 20% of this level and are maintained for the next 10 days (Muller *et al* 1985) though *c-fos* protein is only detected for 120 minutes after induction. Exogenously introduced *c-fos* into F9 teratocarcinoma cells results in expression of *c-fos* mRNA and protein and which is accompanied by the appearance of morphologically altered cells displaying characteristics of differentiated cells (Muller and Wagner 1984, Muller *et al* 1985 b). *C-fos* is also expressed at high levels in fetal membranes and the placenta, in bone marrow, in early fetal liver and in differentiated macrophages (Muller *et al* 1983, Muller, Muller, Guilbert 1984). Therefore these studies strongly implicate *c-fos* being involved in cellular differentiation; note this is the opposite response of *c-myc* which is down-regulated upon differentiation (see later).

When quiescent fibroblasts are treated with serum or growth factors *c-fos* is rapidly induced within 2-3 minutes with maximal levels of induction (around 20 fold) occurring within 20 minutes and no transcripts can be detected by 240 minutes (Muller *et al* 1985). This extremely rapid appearance and disappearance of *c-fos* contrasts that of *c-myc* which is induced later and declines gradually (Muller *et al* 1985 and later). The transient transcriptional activation of *c-fos* upon serum stimulation requires a DNA element 300bp 5' to the mRNA cap site and has been shown to bind a protein(s) from HeLa cell nuclei specifically (Treisman 1986). Hence apparently *c-fos* is involved in two distinct pathways: one involving cellular differentiation and the other cellular proliferation. As to how the *c-fos* protein functions is unclear despite the many studies on the expression of the gene and protein but it is possible it is responsible for activation of other genes (including *c-myc* ?) in response to signals at the cell surface, more work will be required to elucidate its function.

1.3.6 Other oncogenes

In the previous sections I have highlighted some oncogenes about which a lot is known and which reflect current research in the oncogene field. For the many other oncogenes much less is understood about their behaviour and function. It is possible that many of them reflect variants of the oncogenes I have described above. For example maybe *abl* is another kind of growth factor receptor since it is located on the plasma membrane and exhibits tyrosine

kinase activity. Maybe the nuclear oncogenes such as *myb* and *ski* act in a similar manner to *myc* and *fos*. Other oncogenes such as *rel* do not appear to have any parallels in the known oncogenes. The number and variety of potential oncogenic sequences should keep molecular and cellular biologists occupied for a fair number of years to come. What is surprising is that in the face of such potential adversity a cell can manage to remain "normal" so perhaps looking at single oncogenes is a bit naive and a more unified approach will be more useful in understanding normal cell growth.

1.4 The C-myc Oncogene

1.4.1 Introduction

Most of the work described in this thesis concerns the isolation, expression and sequences of *c-myc* related genes from *X.borealis*. In this section I outline the current interest in the *c-myc* (and other *myc*) gene(s). Recently a major review on the regulation and expression of *c-myc* in normal and malignant cells has appeared (Kelly and Siebenlist 1986) and another major review is due to appear in the 1986 volume of the Annual Review of Genetics but I was unable to obtain the latter review before writing this thesis.

1.4.2 V-myc and c-myc

The *v-myc* gene was first identified in the acutely transforming defective avian myelocytomatosis virus MC29 (Sheiness, Fanshier and Bishop 1978, Mellon *et al* 1978, Roussel *et al* 1979). Subsequently the *v-myc* gene was found in three other independently isolated acutely transforming retroviruses (CMII, OK10 and MH2 (Roussel *et al* 1979, Hayman 1983). MH2 contains an additional oncogene referred to *v-mil* or *v-mht* (Jansen *et al* 1983, Kan *et al* 1983, Call *et al* 1983) and it is imagined each oncogene cooperates to alter the pathogenesis of the virus. The putative transforming gene products are hybrids composed of parts of the retroviral gene linked to the *v-myc* sequence, for example MC29 encodes a p110 Δgag -*myc* fusion protein, CMII a p90 Δgag -*myc* fusion protein and MH2 a p200 Δgag -*pol*-*myc* fusion protein (Abrams *et al* 1982, Hann *et al* 1983, Hayman 1983). Additionally OK10 and MH2 have proteins expressed from subgenomic mRNAs (p62 *v-myc* OK10 and p61/63 *v-myc* MH2). The *v-myc* proteins have been found localised in the nucleus (Abrams *et al* 1982, Donner, Greiser-Wilkie and Moelling 1982, Bunte *et al* 1982, Hann *et al* 1983) and can be shown to bind DNA *in vitro* (Donner, Greiser-Wilkie

and Moelling 1982, Bunte *et al* 1982).

Avian myelocytomatosis viruses induce a broad spectrum of malignancies *in vivo* including myelocytomas, endotheliomas, kidney and liver carcinomas and sarcomas. They also transform fibroblastic, myeloid and epitheloid cells *in vitro* (Graf and Beug 1978). Mutants have been isolated that have a reduced capacity to transform macrophages but still retain their ability to transform fibroblasts (Ramsay *et al* 1980). These mutants are useful in dissecting functionally important domains on the *myc* protein and it has been shown that small deletions in the 3' coding sequences are associated with the above phenotype but the protein still locates to the nucleus yet fails to bind DNA (Enrietto and Hayman 1985). Also it was shown that these mutations render the virus non-pathogenic in chickens. Hence there seems to be two domains in the *v-myc* protein: one responsible for transformation of macrophages and one for transformation of fibroblasts *in vitro*. Revertants of these mutants are similar to wild type MC29 *in vitro* but *in vivo* they have an altered pathogenic spectrum therefore it should be possible to analyse the specificity of tumour formation by the *v-myc* protein (Enrietto and Hayman 1985).

The cellular oncogene *c-myc* was first identified by its homology to the transforming gene (*v-myc*) of MC29 virus (Sheiness and Bishop 1979, Sheiness *et al* 1980, Robbins *et al* 1982). Nucleotide sequences of the *c-myc* gene are available from chickens, mice, human and trout (Papas *et al* 1984, Waston *et al* 1983, Shih *et al* 1984, Bernard *et al* 1983, Stanton *et al* 1984, Gazin *et al* 1984, Waston *et al* 1983 b, Colby *et al* 1983, Watt *et al* 1983, Van Beneden *et al* 1986). This sequence data for man, mouse and chicken has shown that the *c-myc* oncogene is composed of three exons with only the last two exons being translated into protein. The first exon suspected to play a role in the regulation of *c-myc* expression. The *v-myc* gene carried by the MC29 family of viruses is composed mainly of sequences derived from the last two exons of the chicken *c-myc* gene and presumably arose by recombination between avian leukosis virus and cellular *myc* sequences. The *v-myc* and *c-myc* coding sequences differ by only eight base pairs (Waston *et al* 1983).

1.4.3 C-myc gene structure

For human, mouse and chicken *c-myc* genes it has been shown that there are three exons and two introns (Papas *et al* 1984, Waston *et al* 1983, Shih *et*

al 1984, Bernard *et al* 1983, Stanton *et al* 1984, Gazin *et al* 1984, Waston *et al* 1983, Colby *et al* 1983, Watt *et al* 1983) and that trout *c-myc* gene has at least two exons (Van Beneden *et al* 1986). Exons 2 and 3 encode a protein product of about 439 amino acids. Immunological analysis using antisera generated to peptides homologous to parts of the predicted amino acid sequence as well as antisera generated to bacterial fusion proteins with portions of exon 2 and 3 have confirmed that exon 2 and 3 are translated in normal cells to give protein products of ~62kd and ~65kd (Alitalo *et al* 1983, Hann *et al* 1983, Persson *et al* 1984, Evan *et al* 1986). Exons 2 and 3 are very highly conserved between different *c-myc* genes for example the human and mouse *c-myc* amino acid sequences are ~93% homologous, the chicken and human *c-myc* amino acid sequences are ~68% homologous and the human and trout *c-myc* amino acid sequences are ~62% homologous. Exon 3 appears to be slightly more conserved than exon 2 perhaps reflecting a functional constraint on the exon 3 sequence such as DNA binding which has been suggested as a role for the exon 3 coding sequence (Persson and Leder 1984). *C-myc* protein coding regions have been conserved throughout evolution suggesting that it plays an important role in cellular growth, division and/or differentiation. The frequent alteration of *c-myc* gene structure and expression (see later sections) also reinforces the role of *c-myc* in growth control.

Exon 1 is about 600bp long and is ~70% conserved between humans and mice (Bernard *et al* 1983, Battey *et al* 1983, Gazin *et al* 1984). The mouse exon 1 sequence appears not to have any coding capacity but the human exon 1 sequence could give rise to a putative protein product of ~20kD (Gazin *et al* 1984). This is somewhat controversial since other published *c-myc* sequences have differences which would make them untranslatable. Support for the presence of an exon 1 translation product has recently been obtained from Gazin and co-workers who showed that antisera raised to synthetic peptides homologous to the putative protein coding sequence will detect a 32 kD protein which can dimerise to a 58 kD protein in human cells (Gazin *et al* 1986). It is possible there is true polymorphism in the first exon sequence but maybe the sequence divergence arose through somatic mutation in the tissues that was used to clone the exon 1 sequence or maybe this region is particularly subject to cloning artifacts. Sequencing germ-line exon 1 sequences should help clarify this point. Interestingly the mouse exon 1 sequence is untranslatable even though it is quite homologous to the human sequence this

suggests the coding capacity has been gained since the divergence of humans and mice but rudimentary conservation still exists. In the chicken, exon 1 is even more diverged; it is a lot shorter, bears no homology to human and mouse exon 1 sequences and cannot code for any protein (Linial and Groudine 1985, Shih *et al* 1984). This may either mean that exon 1 sequences are unimportant in the regulation of *c-myc* or that different ways of regulating the gene have evolved. Possibly human type exon 1 sequences are present in chickens but are encoded by a different gene. Exon 1 has been proposed to be important in the regulation of *c-myc* expression and this will be discussed in the appropriate sections. Note a coding function of exon 1 does not preclude any role in regulation of *c-myc* expression associated with exon 1 but it is intriguing that two independent polypeptides could be encoded by the *c-myc* gene and may reflect some kind of translational control.

1.4.4 The *c-myc* protein

To help analyse the *c-myc* protein expression antisera have been generated against synthetic peptides of the *c-myc* protein or against *c-myc* protein sequences synthesised in *E.coli* either as a fusion protein or not (Hann *et al* 1983, Ramsay, Evan and Bishop 1984, Evan *et al* 1986, Sullivan *et al* 1986, Watt *et al* 1985, Alitalo *et al* 1983, Persson *et al* 1984). At least two proteins are synthesised by the *c-myc* gene which run on SDS polyacrylamide gels with anomalous molecular weights of ~64kD and ~67kD (the calculated molecular weight from the DNA sequence is ~47kD). Bacterially synthesised *c-myc* protein also runs with an anomalous molecular weight (Watt *et al* 1985) hence this may be a property inherent to the *c-myc* protein itself and may reflect a high proline content. Phosphorylation is the only known post-translational modification of the *c-myc* and *v-myc* proteins (Hann *et al* 1983, Ramsay, Evan and Bishop 1984, Evan *et al* 1986). The *c-myc* proteins are found to have a very short half life of ~20-30 minutes (Ramsay, Evan and Bishop 1984, Evan *et al* 1986) which suggests that they are under tight cellular control and they can be rapidly modulated. This also correlates with the short half life of *c-myc* mRNA, described later, hence changes in transcription rate can be immediately translated into changes of protein level.

The proteins specified by *c-myc* and *v-myc* are localised in the nucleus (Alitalo *et al* 1983, Hann *et al* 1983, Persson and Leder 1984, Eisenman *et al* 1985, Evan *et al* 1986). Indeed microinjection of bacterially synthesised *c-myc*

protein into cells efficiently localises it to the nucleus within 5 minutes after injection (Sullivan *et al* 1986). Immunofluorescence studies show that the *c-myc* proteins are located to discrete regions of the nucleus in a punctate fashion and are excluded from the nucleoli and nuclear membrane (Hann *et al* 1983, Sullivan *et al* 1986, Eisenman *et al* 1985). Both *c-myc* and *v-myc* proteins have been shown to bind to DNA with a relatively high but non-specific affinity (Donner *et al* 1982, Persson and Leder 1984, Eisenman *et al* 1985, Watt *et al* 1985). Initial studies suggested that the *c-myc* protein was firmly associated with the nuclear matrix (Eisenman *et al* 1985) but it now seems this is due to an artifact caused by insolubilisation of the protein when exposed to temperatures of 37°C during the isolation procedure. The *c-myc* protein appears now to be in the nuclear soluble fraction bound there by a very salt labile interaction (Evan *et al* 1986). These results may point to *c-myc* functioning as a transcriptional regulatory protein or in DNA replication. The *c-myc* protein has some structural similarity to the E1a transforming protein of adenovirus (Ralston and Bishop 1984) which is known to stimulate the transcription of other genes. Also a *c-myc* construct transfected into cells can activate the expression of an endogenously introduced heat shock protein gene upon expression of *c-myc* (Kingston, Baldwin and Sharp 1984). It still remains to be determined if *c-myc* has a trans-activating role which would help explain the effects of this gene in transformed cells but it is interesting that the E1a protein does not bind DNA directly, rather it activates the transcription of genes indirectly (Green, Triesman and Maniatis 1983). The significance of homology between E1a and *c-myc* proteins has recently been challenged (Ghrist and Ricciardi 1986). Recent excitement has been generated with the possibility that *c-myc* be involved in RNA splicing or metabolism. Of a variety of treatments only actinomycin D causes any effect on the immunolocalisation of *c-myc*, the punctate pattern becoming more dispersed (Sullivan *et al* 1986). If these results are validated then they would open up extremely interesting possibilities for cell growth and transformation.

Recently using two different antisera to *c-myc* protein antigens four distinct human *c-myc* proteins have been detected (Persson *et al* 1986). Two of these proteins (~64kD and ~67kD) are probably the proteins described above; the other two have molecular weights of ~65kD and ~68kD and are not phosphorylated. The 64, 67 and 68kD proteins have short half lives and may be precursors of a more stable 65kD protein which has a half life of ~8 hrs.

Immunofluorescence shows the 65 and 68kD proteins, in addition to being located in the nucleus in a speckled fashion as previously described, are also found in the nucleoli and nuclear membranes. The significance of these results remains to be determined since previous models have focused on the short half life of *c-myc* RNA and protein and there may be important implications understanding *c-myc* function.

1.4.5 C-myc is expressed in many normal cell types

The *c-myc* gene is expressed in a wide variety of cells and tissues as a mRNA of about 2.5–2.7kb (Gonda, Sheiness and Bishop 1982, Westin *et al* 1982, Stewart, Bellve and Leder 1984). In somatic cells the level of *c-myc* transcription tends to be low although there appears to be a correlation of the number of *c-myc* transcripts with the overall rate of cellular proliferation. For example the thymus has a high degree of cell division and has many *c-myc* transcripts whereas skeletal muscle with few dividing cells has very little *c-myc* RNA. An exception to this is that proliferating spermatogonia appear to express very little *c-myc* RNA as compared with normally dividing somatic cells which might be because these cells are committed to terminal differentiation (Stewart, Bellve and Leder 1984). Elevated *c-myc* RNA levels have been reported in hematopoietic tissues compared with other tissues such as brain (Gonda, Sheiness and Bishop 1982) and readily detectable levels of *c-myc* RNA are found in many human leukemia cell lines which are similar to immature hematopoietic cells (Westin *et al* 1982). In human placental development *c-myc* expression occurs primarily in the proliferative cytotrophoblasts and decreases upon differentiation into syncytiotrophoblasts (Pfeifer-Ohlsson *et al* 1984, Ohlsson and Pfeifer-Ohlsson 1986). Using *in situ* hybridisation it has been shown that *c-myc* expression is localised to only a subset of human embryonic cells such as brain, intestine, kidney, lung and skin suggesting cellular proliferation is not always associated with *c-myc* activity (Pfeifer-Ohlsson *et al* 1985). In murine development it has been shown for many tissues that *c-myc* RNA is expressed at the same level for newborn as in adult tissues but for brain *c-myc* expression is virtually undetectable by day 15 of postnatal development with hindbrain showing a more rapid decrease than forebrain (Zimmerman *et al* 1986). Recently it has been shown that *c-myc* RNA expression is quite high in Purkinje cells and granule precursor cells in the postnatal development of the murine cerebellum (Ruppert, Goldowitz and Wille 1986). From the expression patterns of *c-myc* described above it seems *c-myc*

activity is not simply a marker of proliferative activity but also reflects additional tissue-specific gene regulation especially during embryogenesis hence it is possible the *c-myc* product has many functions.

1.4.6 C-myc uses two promoters

The *c-myc* gene is transcribed from two promoters located about 150bp apart at the 5' end of exon 1 (Bernard *et al* 1983, Stewart, Bellve and Leder 1984, Battey *et al* 1983). The differential use of these promoters might be involved in processing or stability of *c-myc* RNA since the 5' ends of transcripts can be drawn as short stem loop structures (Battey *et al* 1983, Leder *et al* 1983) also this might influence the translational efficiency of each transcript. The ratio of promoter usage in normal tissues is variable for example in spleen and thymus the ratio of longer to shorter transcripts is 1:4 but in the preputial gland it is about 1:20 (Stewart, Bellve and Leder 1984). The relative promoter usage also changes during mouse cerebellar development. Before birth both promoters are used equally, at postnatal days 7-10 the ratio of longer to shorter transcripts is 1:4 and in mature cerebellar tissue transcription of *c-myc* is almost exclusively from the shorter promoter (Ruppert, Goldowitz and Wille 1986). In stimulated T and B cells induction of *c-myc* RNA can be up to 40 fold yet the promoter used does not change (Kelly *et al* 1983) also different tissues have individual promoter usage (Stewart, Bellve and Leder 1984). These observations suggest that each cell type analysed has a unique promoter usage of the *c-myc* gene. This may explain the change in promoter usage on mouse cerebellum development as the cell populations are changing from embryonic to mixed neonatal to adult cell populations.

1.4.7 Post-transcriptional and transcriptional controls in c-myc expression

In many cell lines containing chromosomal rearrangements of the *c-myc* gene where the breakpoint is within exon 1 or intron 1 high levels of truncated *c-myc* mRNA are produced whereas the unrearranged *c-myc* locus is transcriptionally silent which has led to a suggestion that a negative control element acting on exon one sequences regulates *c-myc* expression (Leder *et al* 1983). The inhibition of protein synthesis superinduces *c-myc* transcription which may suggest a labile protein either acting as a repressor or affecting *c-myc* mRNA stability regulates *c-myc* expression (Kelly *et al* 1983, Taub *et al*

1984).

It has been established that in many normal and tumour cell lines that *c-myc* mRNA has a half life of about 20 minutes (Dani *et al* 1984, Rabbitts *et al* 1985). In some cell lines there was no effect of protein synthesis inhibitors on *c-myc* mRNA stability but in others there was a dramatic stabilisation suggesting some kind of post-transcriptional controls (Dani *et al* 1984). G₀ arrested cells have low amounts of *c-myc* mRNA but on mitogenic stimulation *c-myc* mRNA levels increase dramatically (see later). One study has shown that this increase *c-myc* mRNA levels is not due to changes in the transcriptional rate of the *c-myc* gene but rather there is post-transcriptional regulation of *c-myc* expression at the level of mRNA degradation (Blanchard *et al* 1985). Another study using normal B cells suggests that the *c-myc* gene is not transcribed in quiescent cells (Smeland *et al* 1986). How these differences arise is not clear but may reflect changes in cells maintained in culture for a long time compared with fresh normal cells or may reflect epigenetic differences between cell types.

In many Burkitt's lymphoma and mouse plasmacytoma cell lines the truncated *c-myc* gene (lacking exon 1 sequences) gives rise to mRNA that is much more stable than normal *c-myc* mRNA (Piechaczyk *et al* 1985, Eick *et al* 1985, Rabbitts *et al* 1985 b, Piechaczyk *et al* 1986). Therefore it may be that exon 1 sequences impart instability on the *c-myc* mRNA but it has been found that exon 1 sequences cannot confer instability upon any other message (Piechaczyk *et al* 1985) and other *c-myc* mRNAs bearing exon 1 sequences are stable (Eick *et al* 1985, Bauer *et al* 1986). Probably the mRNA structure is important for its rapid degradation but the role of 3' sequences remains to be determined. Recently a short sequence located in the 3' non-coding sequences of colony stimulating factor 1 mRNA has been implicated in the rapid degradation of its mRNA (Shaw and Kamen 1986). Interestingly a short conserved sequence, AUUUA, has been found conserved in the 3' non-coding sequences of many genes whose mRNA is unstable including *c-myc*. The role of these sequences in *c-myc* mRNA stability should be open to experimental analysis.

Putative regulatory regions of the *c-myc* gene have been identified by the presence of 5 distinct DNase I hypersensitive sites in the human and murine *c-myc* genes (Siebenlist *et al* 1984, Fahrlander *et al* 1985). Two of these sites

are associated with each of the *c-myc* gene promoters and another two have sequences similar to the consensus for the binding of nuclear factor 1. One of these sites binds nuclear factor 1 *in vitro* (Siebenlist *et al* 1984). The function of nuclear factor 1 is unclear but it is essential for adenovirus replication *in vitro* and it may bind upstream of other promoters though it remains to be determined if it has a role in transcriptional activity (Dyran 1986). The remaining DNase I hypersensitive site is 1.8kb 5' of the transcriptional start sites and lies in a region of strong homology between the human and mouse *c-myc* genes (Fahrlander *et al* 1985). This may be the binding site for a repressor since this site is lost in many chromosomal translocations that deregulate the *c-myc* gene (Siebenlist *et al* 1984). The existence of a class of plasmacytomas which have chromosomal translocations lying ~400bp upstream of exon 1 has further defined the location of an upstream regulatory element (Yang *et al* 1985). By analysing the expression of *c-myc*-chloramphenicol acetyl transferase vectors transfected into cells it has been shown that a 760bp segment of the murine *c-myc* gene lying 428bp 5' from the first promoter can act as a negative transcriptional control element (Remmers, Yang and Marcu 1986). This element has been termed a dehaner since it has the opposite properties of a transcriptional enhancer. It is interesting that the DNase I hypersensitive sites which binds nuclear factor 1 is located in this element and that it lies in a region of good homology between the human and murine *c-myc* genes.

Recently it has been shown that intragenic pausing (or premature termination) has a profound effect on *c-myc* transcription (Bentley and Groudine 1986, Nepveu and Marcu 1986). There appears to be a significant block to the elongation of transcription located in the vicinity of the exon 1 - intron 1 boundary which and as judged by the use of protein synthesis inhibitors, appears to be mediated by the action of a labile protein acting in a negative fashion (Nepveu and Marcu 1986). These studies also showed the existence of a very minor promoter lying ~600bp upstream from the first promoter giving rise to a 3.1kb transcript (Bentley and Groudine 1986). On differentiation of HL60 cells the decreased transcription of *c-myc* appears to be due to an increased block in elongation at the exon 1 - intron 1 boundary (Bentley and Groudine 1986). Therefore the control of production of full length *c-myc* mRNA appears to be mediated through exon 1 sequences. These results disagree with those of Blanchard *et al*, 1985 which suggested that

transcription was constant through the gene and that other post-transcriptional mechanisms are important in *c-myc* regulation but this may be a cell type difference or a reflection on the fact a heterologous probe was used for this study. Interestingly there appears to significant antisense transcription of the *c-myc* gene (Bentley and Groudine 1986, Nepveu and Marcu 1986). The majority of antisense transcripts appear to begin about the 5' of exon 1 and these RNAs are very unstable. Their role in the regulation of *c-myc* transcription is unclear, maybe they are involved in the maturation of *c-myc* mRNA. How *c-myc* mRNA production is regulated appears to be a complex process involving both transcriptional and post-transcriptional controls and is likely to be finely balanced between the two mechanisms.

1.4.8 Regulation of *c-myc* expression in relation to the cell cycle

The regulation of *c-myc* expression during cellular proliferation has been investigated using systems that allow the manipulation of the cell cycle by growth inducing agents. Two separate signals are required for the entry of quiescent G_0 cells into DNA synthesis (S) phase. One control is an early priming event and is termed competence and the second acts at a later stage of G_1 . Both signals are either mediated by growth factors or polyclonal mitogens. Hence if cellular oncogenes are involved in growth control they might be expected to be involved in these events.

Treatment of growth arrested G_0 B lymphocytes with lipopolysaccharide, or T lymphocytes with concanavilin A or fibroblasts with platelet derived growth factor, serum or phorbol esters leads to a dramatic induction of *c-myc* mRNA levels of at least 20 fold within 1-2 hours after treatment occurring before any DNA synthesis (Kelly *et al* 1983, Campisi *et al* 1984). This suggests that *c-myc* expression is one of the earliest biochemical events that occur in response to mitogen binding therefore *c-myc* may play a role in the competence of a cell to progress through the G_0 - G_1 phase of the cell cycle. *C-myc* might be regarded as a cell cycle enabling signal whose presence is required either continuously or through a sensitive temporal window to allow a cell to progress through its cycle (Evan *et al* 1986). Since protein synthesis inhibitors superinduce *c-myc* expression it is suggested that a labile protein mediates these effects (Kelly *et al* 1983). Also it has been shown that, at least in part, that increased *c-myc* mRNA levels are due to transcriptional controls (Greenberg and Ziff 1984). It must be noted that increased *c-myc* mRNA levels

alone are not sufficient to stimulate cellular proliferation in the absence of late acting growth factors (Smeland *et al* 1985). Late acting growth factors, such as EGF, can induce *c-myc* expression in those cells that can show a proliferative response on binding the factor (Muller *et al* 1984). Therefore two post-receptor pathways, one mediated by protein kinase C (phorbol esters etc) and one mediated by EGF can induce *c-myc* activation (Ran *et al* 1986) and point to *c-myc* being an intracellular mediator of mitogen binding. In tumour cells there may be an inappropriate temporal expression of *c-myc* such that cells efficiently transit G₁. The high levels of *c-myc* mRNA in chemically transformed cells might suggest autocrine synthesis of growth factors leading to a continual activation of the *c-myc* gene (Campisi *et al* 1984).

Although there is a transient increase in *c-myc* expression by resting cells in response to growth factors it appears that in a variety of cell types the expression of both *c-myc* mRNA and protein are constant throughout the cell cycle in the continual presence of serum (Hann, Thompson and Eisenman 1985, Rabbitts *et al* 1985, Thompson *et al* 1985). These studies showed that both the turnover rates of *c-myc* mRNA and protein appear to be similar at each stage in the cell cycle as do modifications to the *c-myc* protein. Hence the increase in *c-myc* expression on stimulation of resting cells is only transient with high levels of mRNA and protein falling to a basal level which is maintained through the cell cycle. These results still imply *c-myc* may play a role in the G₀ to G₁ transition.

1.4.9 The regulation of c-myc expression in relation to differentiation

In a variety of cells that are induced to differentiate terminally there appears to be a downregulation of *c-myc* expression associated with cessation of proliferation upon differentiation. The promyelocytic leukemia cell line HL60 (Westin *et al* 1982), the mouse erythroleukemia cell line MEL (Lachman and Skoultchi 1984), the mouse myeloid leukemia cell line WEH1-3B (Gonda and Metcalf 1984), the human histiocytic lymphoma cell line U937 (Einat, Resnitzky and Kimchi 1985) and a mouse teratocarcinoma cell line (Campisi *et al* 1984) can by the action of a variety of agents be differentiated terminally *in vitro* and in each case there is concomitant decrease in *c-myc* expression. This decrease occurs within a few hours but for MEL cells there is a transient re-expression of *c-myc* about 15 hours later and apparently these changes occur before any commitment to differentiation (Lachman and Skoultchi 1984). Constitutive

expression of a transfected *c-myc* gene in MEL cells inhibits the ability of cells to terminally differentiate upon treatment with DMSO (Coppola and Cole 1986, Dmitrovsky *et al* 1986) but transient expression of *c-myc* before the second wave of endogenous *c-myc* expression causes cells to differentiate earlier (Lachman *et al* 1986). This suggests that *c-myc* has both positive and negative roles in the process of commitment of MEL cells. Low levels of *c-myc* mRNA may be due to the non-proliferative state of differentiated cells but recently it has been shown that in primary mouse keratinocytes *c-myc* expression remains high upon induction of differentiation with calcium but not with phorbol esters (Dotto *et al* 1986). Again post-transcriptional (Knight *et al* 1985, Bentley and Groudine 1986) or transcriptional regulation (Einat, Resnitzky and Kimchi 1985) might be involved in these processes but it is clear that the pattern of DNase I hypersensitive sites changes upon differentiation (Kelly and Siebenlist 1986, Bentley and Groudine 1986). Therefore it appears *c-myc* expression plays a role in both differentiation and proliferation of cells and will need to be analysed very carefully before assigning functions to the *c-myc* protein.

1.4.10 Activation of *c-myc* by retroviral insertion

A key discovery in oncogenesis was that in avian leukosis virus (ALV) induced B cell lymphomas a provirus was commonly integrated upstream of the *c-myc* gene (Neel *et al* 1981, Hayward, Neel and Astrin 1981). The ALV integration occurred mainly 5' of the second exon in the same transcriptional orientation leading to ~100 fold increase in *c-myc* expression (Hayward, Neel and Astrin 1981, Payne, Bishop and Varmus 1982). This is termed promoter insertion - activation due to readthrough of the powerful viral promoter leading to increased *c-myc* expression. In some cases the viral LTR is integrated upstream in the opposite transcription orientation or downstream of the *c-myc* gene suggesting increased *c-myc* expression is due to an enhancer element present in the viral LTR (Payne, Bishop and Varmus 1982). In some murine T cell lymphomas there appears to be an increase in *c-myc* expression due to integration by a retrovirus (Corcoran *et al* 1984, Selten *et al* 1984). Therefore the repeated isolation of retroviral insertions adjacent to the *c-myc* gene in T and B cell lymphomas strongly implicates *c-myc* in the genesis of tumour formation. Activation of *c-myc* is probably only one step in the development of these tumours and presumably other changes (in other cellular oncogenes?) are required. Since aberrant *c-myc* expression correlates with tumour

formation this suggests that *c-myc* plays an important role in normal cell growth.

1.4.11 Amplification of *c-myc* in tumour cell lines

The first human cell line reported to contain an amplified oncogene was the promyelocytic leukaemia line HL60 (Collins and Groudine 1982). It appears that increased expression of *c-myc* in this cell line is due to the amplification of the *c-myc* gene by about 20 fold. Other cell lines such as COLO 320 (Alitalo *et al* 1983) and SEWA (Schwab *et al* 1985) also show amplification of the *c-myc* gene with concomitant high levels of *c-myc* mRNA. It may be that *c-myc* amplification is a secondary event for increased tumourigenicity since amplification of *c-myc* is not often associated with tumour tissue from which the above cell types were derived and since the degree of amplification of *c-myc* in SEWA cells can be altered depending on whether they are grown *in vivo* or in *in vitro* (Schwab *et al* 1985). Also a greater degree of amplification of *c-myc* is found in highly malignant small cell lung carcinomas than less malignant variants (Little *et al* 1983). Again these studies point to a role of *c-myc* in tumourigenesis though in these cases the role may be much more a secondary event than in retroviral insertion but this may reflect cell or tumour type specificity.

1.4.12 Chromosomal translocations involving the *c-myc* locus

Many human Burkitt lymphomas (BLs) and murine plasmacytomas (PCs) are characterised by consistent reciprocal chromosomal translocations involving the *c-myc* locus and one of the three immunoglobulin (Ig) loci (see reviews by Perry 1983, Leder *et al* 1983, Varmus 1984, Kelly and Siebenlist 1986). A single detailed model to account for the consequences of these translocations is probably unrealistic due to the wide variety of translocations involved. Here I briefly outline some salient facts concerning *c-myc* and chromosomal translocations; for further details see the reviews cited above and references therein.

The most predominant translocation is the juxtaposition of the *c-myc* gene with the heavy chain Ig gene. In BLs the *c-myc* gene on chromosome 8 is translocated to the heavy chain gene on chromosome 14 and similarly in PCs the *c-myc* gene on chromosome 15 is translocated to the heavy chain Ig gene on chromosome 12. Normally the chromosome breakpoint in both BL and PCs

on the Ig carrying chromosome is located in the switch region. The breakpoint in the *c-myc* gene in BLs is normally 5' of the first exon but in PCs it can often include the first exon; note there is much variability in these locations. This generates a situation where both genes lie head to head in opposite transcriptional orientations hence the *c-myc* gene is expressed from its own or cryptic promoters. Since the reciprocal translocation event removes the Ig heavy chain enhancer then the transcription of *c-myc* is influenced by as of yet uncharacterised regulatory sequences in the 3' domain of the Ig heavy chain gene. Although BL cells have relatively high levels of *c-myc* mRNA they are apparently often not much greater than those in non-tumourigenic proliferating B cells therefore this points to other events involved in the genesis of tumour formation

The less well characterised variant translocations in BL and PCs involve the juxtaposition of *c-myc* to either of the light chain Ig loci. In these translocations the *c-myc* gene normally remains on its own chromosome and the Ig loci are translocated to it. For BLs this involves an 8:2 translocation for Kappa, 8:22 for lambda Ig light chain genes and in PCs a 15:6 translocation for Kappa Ig light chain genes resulting in the Ig genes located an indeterminate distance 3' to the *c-myc* gene. The *c-myc* mRNA levels appear to be no different in these tumours when compared with the other BL and PCs. Interestingly in many PCs bearing the 15:6 translocations a region called *pvt-1* is a common site for breakpoints on chromosome 15 and lies at least 94kb 3' of *c-myc*. *Pvt-1* is also a common site of proviral integration in retrovirally induced T cell lymphomas. Therefore *pvt 1* may represent a sequence that can act at a considerable distance from *c-myc*, since *pvt 1* is not transcribed and has no homology to other oncogenes, it may act by influencing the chromatin structure of the *c-myc* gene (Graham and Adams 1986).

In the BL and PCs only the *c-myc* gene that resides on the translocated chromosome is actively transcribed, the unrearranged *c-myc* gene remains transcriptionally silent. These results suggest that *c-myc* expression is negatively regulated by a process that can repress the normal allele but not the translocated allele. It is supposed that the *c-myc* protein or a protein synthesised in response to high levels of *c-myc* protein represses *c-myc* expression by acting on sequences 5' of exon 2 (i.e where most of the alterations occur in the translocated gene). Indeed it recently has been shown that a negative acting element with the opposite properties of a transcriptional

enhancer lies ~400bp 5' to *c-myc* first exon. (See earlier and Remmers, Yang and Marcu 1986).

Changes in the promoter usage of the *c-myc* gene may also occur in BL and PCs; maybe this alters translation of efficiency the mRNA but no consistent pattern of promoter usage is clear. Somatic mutation of the Ig locus is known to occur in normal B cell development and in some of the translocations described above mutations have been found in the translocated *c-myc* allele (Rabbitts, Hamlyn and Baer 1983). Few mutations occur in the coding sequences but many are observed in the first exon although they are inconsistent. These changes may alter putative repressor binding, mRNA stability or utilisation. Possibly deletion or mutations of first exon sequences are unimportant for *c-myc* regulation in these tumours since BLs exist which have translocated *c-myc* genes without any changes but these BLs may have alterations in chromatin structure leading to the same effect. Indeed the regulatory regions defined by DNase I hypersensitive sites appear to be altered in many different PCs and BLs. Some translocation events leave the Ig enhancer on the same chromosome as the *c-myc* gene hence it is likely the enhancer has positive effects on *c-myc* expression in these tumours. In summary it appears that the *c-myc* gene is highly regulated but in BL and PC cells a variety of mechanisms have given rise to a deregulated *c-myc* gene though it is not clear what role this deregulation has in tumour formation.

1.4.13 Studies with transfected myc genes and transgenic mice

DNA mediated gene transfer has been utilised to manipulate the *c-myc* and *v-myc* genes in order to analyse their transformation potential. Although viruses carrying *v-myc* cause morphological transformation of cells *in vitro* they are not immortalised and do not cause tumours in nude mice (Lautenberger *et al* 1981). In contrast transfection of heterologous promoter driven *c-myc* sequences into immortalised fibroblasts (Armelin *et al* 1984) or cloned *v-myc* DNA into secondary rat embryo fibroblasts does not produce morphological transformation (Land, Parda and Weinberg 1983), suggesting that other phenotypic changes are required for the full transformed phenotype in such transfected cells. It was found that co-transfection of a mutant *ha-ras* gene and an SV40 promoter driven *c-myc* gene into normal rat embryo fibroblasts could result in morphological transformation whilst either alone could not (Land, Parda and Weinberg 1983), suggesting groups of co-operating

oncogenes, one a nuclear gene supplying an immortalisation function (e.g. *myc*, E1a or SV40 large T) and the other a cytoplasmic gene supplying a transformation function (e.g. *ras* or polyoma middle T). An immortalisation function has been suggested for *myc* because transfected *myc* sequences alter the growth properties of cells. *C-myc* is induced by growth factors suggesting it is an intracellular mediator of mitogen binding. Balb/c3T3 cells transfected with *myc* sequences show an increased sensitivity to EGF and an increased probability of cell division in the absence of PDGF (Armelin *et al* 1984). Recently a model has been suggested in which the mechanism of co-operation between *myc* and *ras* oncogenes involves *ras* genes inducing growth factor production while *myc* genes increase the responsiveness of cells to these factors (Stern *et al* 1986). It should be noted that immortalised cell lines transfected with LTR promoter driven *c-myc* genes do result in tumourigenic cells even though they lack morphological transformation (Keath, Caimi and Cole 1984). Therefore constitutive *myc* expression leads to a variety of effects which probably relates to the genetic background of the cells but it could be indicative of a multi-functional *c-myc* protein.

Recently the transformation potential of *c-myc* has been analysed by the use of transgenic mice. When transgenic mice possess a *c-myc* gene under the control of the glucocorticoid inducible MMTV promoter they can develop monoclonal adenocarcinomas during one of their early pregnancies (Stewart *et al* 1984). F1 progeny of these mice inherit the predisposition to develop these tumours. Although the fusion gene in these mice can be expressed in a wide variety of tissues there is no obvious effect on early development. The active MMTV/*myc* gene does not appear to be sufficient for the development of tumours since uniform development of tumour masses involving the entire bilateral mammary gland would have resulted. Therefore further transforming events are required - possible mutation giving clonal antecedents of tumours and also the hormonal environment related to pregnancy. A further study on these mouse lines shows the deregulated MMTV/*myc* gene contributes to an increased variety of tumours including those of testicular, breast, lymphocyte, and mast cell origin (Leder *et al* 1986). Since tumours are not observed in every organ within the transgenic animal those which are tumourigenic begin to define the transformation spectrum of the *c-myc* gene which correlates with the pluripotent transformation ability of *c-myc* in other systems. In another study using transgenic mice different configurations of the *c-myc* gene linked

to enhancer sequences from Ig genes were studied (Adams *et al* 1985). Injection of *c-myc* genes linked to either heavy chain or light chain enhancers led to the induction of multifocal lymphomas in 90% of mice. No neoplastic disease was found when the normal *c-myc* gene or truncated *c-myc* gene was injected observations which argue against a negative acting role of exon 1 sequences in the normal regulation of *c-myc*. In BL and PCs the Ig loci must have a positive effect on *c-myc* expression. Again this study suggests dysregulated *c-myc* expression is insufficient for malignancy due to the clonality of tumours produced. Interestingly in these tumours the endogenous *c-myc* gene is not expressed whereas the transgene is expressed, which might point to an autoregulation model similar to BL and PCs but it is not clear if the tumours arose from cells in which the *c-myc* gene is not normally expressed. Also some of MMTV/*myc* carrying transgenic mice expressed both the endogenous gene and the transgene. It was suggested that maybe high levels of *c-myc* make cells that tend to proliferate rather than to generate differentiated progeny hence this will increase the proportion of less mature cells having a greater proliferative capacity. Presumably these cells are more susceptible to accidental events that would then create the malignant clone. Further studies on such transgenic mice should help define the role of *c-myc* in transformation and studies on the normal biochemistry of the protein should help in the design of future experiments with transgenic mice.

1.4.14 Myc gene families: N-myc and L-myc

Studies of tumour cell lines with gene amplifications has led to the identification of *c-myc* related genes (Schwab 1985). Initially it was shown by using a *v-myc* probe that amplified DNA from a neuroblastoma cell line contained a novel oncogene, termed *N-myc*, with limited homology to *c-myc* (Schwab *et al* 1983). The strongest similarity lies in two major domains separated by a heterologous sequence. These so called *myc*-boxes are extremely conserved: box A has 19 out 24 amino acids identical between *N-myc* and *c-myc* and box B has 13 out of 16 amino acids identical. The complete human *N-myc* nucleotide sequence shows that like *c-myc* the *N-myc* gene is composed of three exons both having large untranslated first exons which bear no sequence homology to each other (Kohl *et al* 1986). The overall homology between *N-myc* and *c-myc* is about 70% with the strongest homology in exon 3. The *N-myc* gene maps to human chromosome 2 (Kohl *et al* 1983). Anti-peptide antiserum to the *myc* boxes detect a *N-myc* protein of

~66kD which has a short half life and is located in the nucleus (Evan *et al* 1986). Therefore *N-myc* appears to be a member of a *myc* gene family and has properties very similar to that of *c-myc* and may reflect similar functions.

Most neuroblastoma cell lines have a 10-700 fold amplification of the *N-myc* gene leading to increased *N-myc* expression (Schwab *et al* 1983), Kohl *et al* 1983, Schwab 1985). In a study of tumour patients it was found that amplification of *N-myc* only occurred in the later, more metastatic neuroblastomas (Brodeur *et al* 1984). 50% of advanced neuroblastomas had an amplification of the *N-myc* gene whereas no amplification of *N-myc* was detected in the earlier stages. Similarly *N-myc* amplification has been detected in cell lines derived from highly malignant small cell lung carcinomas (Nau *et al* 1984) and in retinoblastoma cell lines, though a high level of *N-myc* expression is independent of amplification in retinoblastoma cell lines (Lee, Murphee and Benedict 1984). *N-myc* amplification and increased expression may not be the only factor in the formation of the above tumours since 50% of the metastatic neuroblastomas do not have such amplification. Indeed transfection of a *N-myc* construct driven by a viral enhancer into rat embryo cells results in established cell lines that are non-tumorigenic in animals (Schwab 1985, Schwab, Varmus and Bishop 1985) but in some cases co-transformation of *N-myc* with the mutated *ha-ras* gene can give rise to fully transformed cells which are tumorigenic in animals (Schwab 1985). This situation is very reminiscent of that described for the *c-myc* gene. Neuroblastomas and retinoblastomas are both of neural origin while small cell lung carcinomas are of endodermal origin they share common membrane proteins and express neuron specific enolase (Nau *et al* 1984) hence *N-myc* expression has only been found in tumours that have neural characteristics. *In situ* hybridisation suggests that high levels of *N-myc* expression are limited to primary neuroblasts (Schwab *et al* 1984). Retinoic acid induced differentiation of a human neuroblastoma cell line results in decreased *N-myc* expression (Thiele, Reynolds and Israel 1985) as does differentiation of murine embryonal carcinoma cell lines which may reflect expression in normal early embryonic embryos (Jakobovits *et al* 1985). *N-myc* is only expressed in a limited number of adult tissues and is expressed at high levels in the murine mid-gestation embryo which decreases as the embryo approaches term (Jakobovits *et al* 1985). It may be *N-myc* expression is high in inner cell mass cells at around implantation but upon differentiation into endodermal cells expression

decreases. High levels of *N-myc* expression have been also found to be restricted in the developing mouse to hindbrain, forebrain and kidney with mRNA levels decreasing drastically between day 15 of prenatal development to day 15 of postnatal development (Zimmerman *et al* 1986). Hence it appears the expression of *N-myc* is much more limited than *c-myc* and the *N-myc* gene product may play a role in multiple (but related) differentiation pathways.

A third *myc* related gene, termed *L-myc* was identified in the amplified DNA of small cell lung carcinoma cell lines by virtue of its homology to *N-myc* (Nau *et al* 1985). In humans there is a restriction polymorphism associated with this gene and either allele becomes amplified in the small cell lung carcinoma cell lines. Using an *L-myc* specific probe a 2.2kb *L-myc* transcript is detected in cell lines containing an amplified *L-myc* gene. In any small cell lung carcinoma only one of the three *myc* genes is expressed which may reflect overlapping functionality of the gene products. Limited sequence analysis shows that the *L-myc* gene contains highly conserved *myc* boxes. The *L-myc* gene maps to human chromosome 1. Apparently anti-peptide antisera recognise a 64kD *L-myc* protein product (Evan *et al* 1986). During mouse development *L-myc* showed a very restricted expression pattern that was similar to that of *N-myc* (Zimmerman *et al* 1986). Therefore it seems there is a family of *myc* genes all of which play a part in the formation of some tumours. It could be they all have similar functions which have been tailored to a particular need, indeed restricted developmental expression and expression in certain tumour cells suggests such a role and it may be there as yet undiscovered *myc* related genes. Understanding the relationship of these genes in normal cell growth and tumour formation should be a fruitful area in the next few years.

1.5 Gene Expression During *Xenopus* Development

1.5.1 Introduction

In this thesis the study of cellular oncogene function was carried out using *Xenopus* as a model system. *Xenopus* is an ideal organism for studying the molecular biology of development for a variety of reasons. Large numbers of oocytes and embryos are easily obtained and occur in well defined developmental stages (Dumont 1972, Neuwkoop and Faber 1956). Development is extremely rapid, a fertilised egg reaches gastrulation in about 15 hr and by the time a *Xenopus* embryo has reached 80,000 cells the mouse embryo has

only just completed its first division. *Xenopus* embryos have a well known histology and established methods exist for the micromanipulation of the embryos. Factors governing the maturation of the oocyte are easily studied *in vitro* (e.g. Wasserman *et al* 1984). Indeed the *Xenopus* oocyte has been used by many workers as a living test tube in the design of their experiments. As a vertebrate *Xenopus* provides a valuable model system for studying vertebrate developmental biology.

A great deal is known about gene organisation and expression in *Xenopus* (Dawid, Kay and Sargent 1983, Dawid *et al* 1985, Dawid and Sargent 1986) although it must be remembered that other organisms (e.g. *Drosophila*) have been studied extensively at the molecular level. *Xenopus* oocytes and embryos are very large hence the microinjection of molecules is simple. Microinjection of cloned DNA into fertilised eggs provides a powerful tool for studying mechanisms that underlie tissue specific gene expression since the cloned DNA will be inherited by daughter blastomeres that enter separate differentiation pathways. Also the microinjection of other molecules such as antibodies to try and block protein function (e.g. Scheer 1986) and antisense RNA to try and block mRNA function (Drummond and Colman 1986, Melton and Rebagliati 1986) are rapidly becoming possible. Many different genes have been studied during *Xenopus* development, for example histones (Woodland 1980), keratins (Dawid and Sargent 1986), actins (Gurdon *et al* 1985 a), ribosomal proteins (Pierandrei-Amaldi *et al* 1982), small nuclear RNAs (Mattaj *et al* 1985), 5S RNAs (Brown 1984), homeoboxes (Carrasco *et al* 1984) and many other genes have all been studied during oocyte and embryo growth. *Xenopus* oocytes and eggs have also provided experimental systems for the study of RNA localisation (Rebagliati *et al* 1985), meiosis (Miake-Lye and Kirschner 1985), nuclear assembly (Forbes *et al* 1983), transport of proteins to the nucleus (DeRobertis *et al* 1983) and cell cycle events (Kirschner, Newport and Gerhart 1985) as well as other processes. Techniques such as *in situ* hybridisation to embryos using antibodies and antisense RNA are currently under development and should provide valuable information on the temporal and spacial controls during embryonic gene expression. *Xenopus*, as a system, has provided a few firsts in modern molecular biology such as the isolation of the first eukaryotic gene (Birnstiel *et al* 1968) and the cloning of the first eukaryotic transcription factor (Ginsberg, King and Roeder 1984) to name but a few.

Although *Xenopus* has a good grounding in molecular biology it is not a

completely ideal organism to study developmental biology. The most important limitation is that *Xenopus* lacks standard genetics which severely hinders the study of gene activity and is better understood in *Drosophila* or mice. Transgenic *Xenopus* will be difficult to obtain because animals take 1–2 years to reach sexual maturity and since microinjected cloned DNA appears not to integrate into the host genome. Studying the biosynthesis of macromolecules is also difficult since *Xenopus* embryos are impermeable to radioactive precursors. Also blocking translation of mRNA with antisense RNA embryos is not yet completely satisfactory (Drummond and Colman 1986, Melton and Rebagliati 1986) though this is apparently not true in oocytes; for example antisense RNA to ribosomal protein L1 will block translation of L1 mRNA in oocytes (Wormington 1986).

1.5.2 Gene expression in oogenesis

The growth of *Xenopus* oocytes takes at least 6 months during which time the cells grows to a diameter of about 1.5mm and accumulates a stockpile of many substances for utilisation during early development (see Davidson 1976, for review). Oogenesis in *Xenopus* has been subdivided into six major stages, I to VI, based on morphological criteria such as oocyte diameter, pigmentation and the stage of meiotic prophase reached by the nucleus (Dumont 1972). Only fully grown stage VI oocytes are capable of responding to the steroid hormone progesterone and resuming meiosis from prophase arrest (e.g. Wasserman *et al* 1984). This culminates in the dissolution of the nuclear membrane, condensation of the chromosomes and reductive meiotic divisions prepare the egg for fertilization. The first polar body (one set of condensed chromosomes) is extruded at this point and meiosis becomes arrested at metaphase. On fertilisation of the egg by a sperm the second meiotic division is completed and the second polar body is extruded (Balinsky 1970). Further development of the egg thus ensues.

1.5.2.1 Yolk proteins

Quantitatively the largest accumulation of material by the growing oocyte is yolk protein and can be as much as 80% of the dry weight of an oocyte. Yolk proteins are synthesised in the female liver, under the hormonal control of oestrogen, as high molecular weight precursors known as vitellogenins (see reviews by Wahli *et al* 1981, Wahli and Ryffel 1985). Vitellogenins are secreted

into the bloodstream and then taken up by the growing oocyte by receptor-mediated endocytosis. These precursor vitellogenins are cleaved to give phosphoproteins and lipoproteins which are assembled to form crystalline yolk platelets. Yolk proteins serve as a nutrient store during embryogenesis until the embryo reaches an independent feeding stage.

1.5.2.2 Ribosomes, rRNA and tRNA

Xenopus oocytes accumulate large amounts of the components of protein synthesising machinery during oogenesis. 5S RNA and tRNA are made very early in oogenesis representing about 70% of the total RNA in previtellogenic oocytes (Ford 1971, Mairy and Denis 1972). Enough 5S RNA is made during previtellogenesis to allow a hundred fold increase in ribosome content without need for further synthesis. At early stages of oogenesis tRNA and 5S RNA are stored in 42S ribonucleoprotein complexes (Ford 1971, Denis and Mairy 1972) and 5S RNA in 7S ribonucleoprotein complexes (Picard and Wegnez 1979). In fully grown oocytes 42S ribonucleoprotein complexes are undetectable (Dixon and Ford 1982) and much of the 5S RNA becomes recruited into ribosomes. Accumulation of 18S and 28S rRNA starts at the onset of vitellogenesis and continues throughout oogenesis such that in stage VI oocytes 18S and 28S rRNA represent ~90% of the total RNA (Ford 1971, Mairy and Denis 1972). Ribosomal protein synthesis also begins at the onset of vitellogenesis and their accumulation parallels that of 18S and 28S rRNA (Dixon and Ford 1982). Stage VI oocytes are thought to have accumulated their full ribosome complement although continued synthesis and turnover can occur (Leonard and La Marca 1975).

The differential accumulation of 18S and 28S rRNA compared with 5S RNA and tRNA is reflected in their gene organisation. There are about 200 genes for each tRNA, about 20,000 genes for oocyte specific 5S RNA and about 400 genes for somatic specific 5S RNA. The oocyte specific 5S RNA differs by 6 nucleotides from somatic 5S RNA (Ford and Southern 1973) and only oocyte specific 5S RNA is accumulated in oocytes although somatic 5S RNA is synthesised but not accumulated (Denis and Wegnez 1977). The 5S RNA and tRNA synthesised from these genes during previtellogenesis is stable for 1-2 years (Ford, Mathieson and Rosbash 1977). The 18S and 28S rRNA genes are organised into repeating units containing 18S and 28S DNA interspersed with spacer DNA and are present in about 450 copies per haploid genome. During

very early oogenesis these genes are selectively amplified to give many extrachromosomal nucleoli (see Lewin 1980 for review). This amplification gives a total number of ribosomal genes of about 10^6 which are utilised at the onset of oogenesis. The extra rDNA is not replicated during cleavage and either is diluted by nuclear replication or becomes degraded. Hence the oocyte uses two different mechanisms to accumulate the resources of 5S RNA and tRNA, 18S and 28S rRNA used during early development. 18S and 28S rRNA are synthesised from genes which are selectively amplified over a relatively short period whereas 5S RNA and tRNA are made on a fixed number of genes over a much more protracted period of time.

The transcription of 5S genes is developmentally regulated and is reasonably well understood. Transcription of 5S genes requires three factors (TFIIIA, B and C) in addition to RNA polymerase III (Segall, Matsui and Roeder 1980). TFIIIA is specifically required for the initiation of 5S gene transcription and is a protein of ~38kD which binds to a 50bp internal control region of 5S genes (Bogenhagen, Sakonju and Brown 1980, Engelke *et al* 1980). This complex is further stabilised by the sequential binding of factors IIIC and IIIB (see Brown 1984 for review). TFIIIA also interacts with 5S RNA to form 7S ribonucleoprotein complexes which serve as storage particles during oogenesis (Pelham and Brown 1980), therefore it is possible an autoregulatory mechanism exists for 5S gene expression during oogenesis. The activation of all 5S genes during early oogenesis is ensured by the very high levels of TFIIIA protein ($\sim 10^{12}$ molecules/cell). The decrease in 5S gene transcription during oogenesis can be accounted by the decreased expression of the TFIIIA gene (Ginsberg, King and Roeder 1984) and the recruitment of TFIIIA into 7S ribonucleoprotein particles in the face of the accumulating 5S RNA. In somatic cells 95% of the 5S RNA arises from somatic type genes and in cell free extracts from oocytes or somatic cells cloned somatic 5S genes are transcribed more efficiently than cloned oocyte specific genes (Pelham, Wormington and Brown 1981). Oocyte 5S RNA genes appear only to be transcribed when there is a large excess of TFIIIA. In somatic cell chromatin repressed oocyte 5S genes lack transcription complexes and this repression appears to be mediated by histone H1 (Schlissel and Brown 1984). During embryogenesis there is transient expression of both kinds of 5S gene at midblastula due to TFIIIA inherited by the egg but as cell division continues and the concentration of TFIIIA drops only the somatic 5S genes become transcribed (Wormington and

Brown 1983). Injection of TFIIIA protein into cleaving embryos greatly increases the transcription of endogenous oocyte 5S genes whereas injection of mixture of cloned 5S genes lead to the preferential transcription of somatic 5S genes (Brown and Schlissel 1985). Hence it seems the preference of somatic over oocyte 5S gene transcription in somatic cells is due to the differences in binding constants of TFIIIA for the somatic and oocyte 5S genes and to the concentration of TFIIIA protein in the cell.

1.5.2.3 mRNA

A particularly interesting group of materials accumulated by oocytes is the relatively large amount of poly(A)⁺ RNA. The total amount of poly(A)⁺ RNA has reached its final abundance in previtellogenic oocytes (Rosbash and Ford 1974, Cabada *et al* 1977, Dolecki and Smith 1979). This RNA population comprises about 20,000 different sequences and apparently lacks a high abundance class which is present in tadpole or adult tissues (Perlman and Rosbash 1978, Rosbash 1981). Of these sequences 5% are about 15 times more abundant than the remaining 95%; the majority of sequences being present at ~0.005% of the poly(A)⁺ population. These results have been extended to show that several individual poly(A)⁺ species, as measured by cloned probes, cease to accumulate after stage II of oogenesis (Golden, Schafer and Rosbash 1980). The poly(A)⁺ RNA synthesised in previtellogenic oocytes is stable over a 18 month period (Ford, Mathieson and Rosbash 1977) hence poly(A)⁺ RNA is accumulated before the onset of maximum extension of lampbrush chromosomes. Although RNA synthesis continues throughout oogenesis most is ribosomal but some poly(A)⁺ RNA is synthesised. Since the poly(A)⁺ RNA made early in oogenesis does not turnover then this suggests the poly(A)⁺ RNA made later turns over differentially. The significance of this is unclear but it is interesting that the poly(A)⁺ RNA made earlier has a larger poly(A)⁺ tail than that made later (Cabada *et al* 1977).

What is the nature and function of the poly(A)⁺ in oocytes? A simple interpretation is that it is mRNA of which most is stored during oogenesis for subsequent translation in the course of oocyte maturation and early embryogenesis since only a small fraction of the RNA is found in polysomes (Woodland 1974). Undoubtedly some of the oocyte mRNAs are translated in the oocyte to synthesise many stored products and to support its own metabolism e.g. ribosomal proteins (Pierandrei-Amaldi *et al* 1982). Most of the

poly(A)⁺ RNA is never translated in the oocyte and will only be used during embryogenesis many weeks after its synthesis. This storage of mRNA in *Xenopus* oocytes is probably one of the clearest examples of the translational regulation of gene expression known.

There are a number of ways that this translational regulation could arise. Fully grown oocytes could have a limited translational capacity due to some factor lacking in the translational machinery. Apparently this is partly true since the injection of heterologous mRNAs only results in protein synthesis at the expense of proteins translated from endogenous mRNAs (Laskey *et al* 1977, Richter and Smith 1981). Hence the oocyte has only the capacity to translate the mRNA actually loaded on polysomes even though this may be less than 5% of the total poly(A)⁺ RNA. There is no direct evidence in support of this due to the absence of the isolation of the limiting components.

Another hypothesis is that proteins bind to the poly(A)⁺ RNA and mask its translation prior to maturation and embryogenesis. In previtellogenic oocytes the majority of the poly(A)⁺ RNA is located in the cytoplasm and up to 70% is polysome associated (Darnbrough and Ford 1976) but less than 10% is polysome associated in full grown oocytes (Rosbash and Ford 1974). The remainder is complexed with protein sedimenting heterogeneously over the 30S to 120S range. Specific poly(A)⁺ RNA binding proteins have been identified in oocytes and cannot be detected in liver or reticulocytes (Darnbrough and Ford 1981, Dixon and Ford 1982). These proteins are independently regulated during oogenesis and their accumulation follows that of poly(A)⁺ RNA. Using a cell free translation assay it has been shown that the poly(A)⁺ RNA in ribonucleoprotein complexes can be translated as efficiently as free poly(A)⁺ RNA and that their coding capacities are similar (Darnbrough and Ford 1976). A contrasting result has been obtained by Smith and co-workers who found that when globin mRNA is mixed with RNA binding proteins from stage I and II oocytes then injected into oocytes that the globin mRNA is inhibited for translation (Smith, Richter and Taylor 1984). Therefore the role of masking proteins suppressing the translation of maternal mRNA remains unclear.

An additional possibility is that the poly(A)⁺ RNA molecules in oocytes possess structural features which inhibit their own translation. Up to 70% of the oocyte poly(A)⁺ RNA displays an interspersed sequence organisation in which regions transcribed from single copy and repetitive sequences are

covalently associated whereas only 15% of tadpole poly(A)⁺ RNA displays this organisation (Anderson *et al* 1982). Recently it has been shown that poly(A)⁺ RNAs which lack a renaturable repeat are translated efficiently in both *in vitro* and *in vivo* assays whereas interspersed poly(A)⁺ RNAs are translationally inactive (Richter *et al* 1984). There may be a number of reasons why the interspersed poly(A)⁺ RNAs which lack a renaturable repeat are translated efficiently in both *in vitro* and *in vivo* assays whereas interspersed poly(A)⁺ RNAs are translationally inactive (Richter *et al* 1984). There may be a number of reasons why the interspersed poly(A)⁺ RNAs are translationally inactive such as the separation of ribosome binding sites from start codon, the translation start signals could be preceded by repeated elements which contain stop codons, they could contain unprocessed intervening sequences which inhibit translation or they could provide binding sites for oocyte proteins that prevent translation. This might suggest a precursor role for much of the oocyte poly(A)⁺ RNA but the finding that many individual RNAs have the same size in oocytes and tadpoles tends to argue against this (Golden, Schaffer and Rosbash 1980, Dworkin *et al* 1984).

Therefore it is most unclear how the translational regulation of the poly(A)⁺ RNA arises but it is likely to involve the co-ordinate control of all the mechanisms described above. A calculation described in Richter *et al* 1984 might suggest that the mass of translatable maternal mRNA present in oocytes excluding the interspersed poly(A)⁺ RNA might be sufficient to support protein synthesis until the midblastula stage when new RNA synthesis begins (see below) but it must be borne in mind there are many uncertainties in this calculation. Also it remains to be determined whether the poly(A)⁺ interspersed RNAs are processed in later development (maybe in particular tissues or at different developmental stages) to render them translatable. It is worth noting no one has yet reported the presence of an interspersed repeated sequence within the body of a coding region in cDNA clones derived from oocyte libraries.

1.5.3 Gene expression during embryogenesis

The previous section described that the fully grown *Xenopus* oocyte contains a large store of poly(A)⁺ RNA some of which is mRNA. This maternal mRNA is used to support protein synthesis during the first hours of development before the onset of embryonic transcription which begins around

midblastula (see below). About 2% of ribosomes are found in polysomes in stage VI oocytes then during the first 5 hours of development this figure increases to 20% which then remains constant for the next 24 hours and then ribosome recruitment increases to a somatic level of 70% by the feeding tadpole stage (Woodland 1974). Because there is little accumulation of new mRNA during the first hours of development (see below) then some of the maternal poly(A)⁺ RNA must be recruited into polysomes and hence does represent stored RNA. Indeed it has been shown that the 2-4 fold increase in protein synthesis during oocyte maturation is due to the recruitment of mRNA onto polysomes (Richter, Wasserman and Smith 1982). Also using cloned cDNA probes it has been shown that many sequences that are non-polysomal in the stage VI oocyte become mobilised onto polysomes by the 16 cell stage although some (5/18) are mobilised poorly (Dworkin, Shrutkowski and Dworkin-Rastl 1985). This study also suggested there is a class of moderately expressed maternal mRNA in the oocyte which decreases in titre during development but is loaded onto polysomes and is perhaps involved basic physiological processes during early development or related to pattern formation or differentiation.

1.5.3.1 RNA synthesis and the midblastula transition

The first 12 cleavage divisions proceed synchronously and take the single cell egg to the 4000 cell midblastula in about 7 hours. No RNA synthesis is detectable in the embryo until the midblastula stage (Brown Littna 1966a, Brown and Littna 1966 b, Shiokawa *et al*/ 1981, Newport and Kirschner 1982 a). The specific time this occurs is termed the midblastula transition since many other changes occur in the embryo at this time (Newport and Kirschner 1982 b) cell division becoming asynchronous, cells becoming motile and the cell cycle slowing down. During the first 12 cleavages the cell cycle is ascribed to an endogenous free running oscillator present in each blastomere and operating independent of the nucleus (Kirschner, Newport and Gerhart 1985). It is not clear what activates the midblastula transition but it may be due to nuclear to cytoplasmic volume or recruitment of maternal factors (Kirschner, Newport and Gerhart 1985).

It appears that many different classes of RNA begin to accumulate at or shortly after the midblastula transition. Amongst the most rapid are the small nuclear RNAs, 5S RNA and tRNAs (Newport and Kirschner 1982, Brown and

Littna 1966 b). Ribosomal RNAs also begin to accumulate after the midblastula transition (Shiokawa, Misumi and Yamana 1981) though this experiment used dissociated embryos which may affect synthetic behaviour. Poly(A)⁺ RNAs undoubtedly begin to accumulate soon after the midblastula transition (Shiokawa *et al* 1981). It has been calculated that about 40% of the poly(A)⁺ RNA in gastrula polysomes and in whole embryos is newly synthesised. Since the majority of poly(A)⁺ RNA does not vary (see below) between the egg and gastrula embryo then this suggests that the embryo resynthesises RNAs that were already present.

1.5.3.2 RNA populations during embryogenesis

Using cDNA:RNA hybridisation it has been concluded that oocyte and tadpole poly(A)⁺ RNA populations are very similar except that the oocyte population lacks a high abundance class present in tadpole RNA (Perlman and Rosbash 1978). Only a small fraction of poly(A)⁺ RNAs are newly synthesised during development. This has been extended using *in vitro* protein translation assays and is in agreement with previous results (Rosbash 1981). It has also been shown that oocyte poly(A)⁺ RNA is larger than tadpole poly(A)⁺ RNA (Rosbash 1981). Perhaps this decrease in size is partly due to the synthesis of shorter molecules and partly due to the loss of interspersed repeat elements described earlier. Analysis of RNA populations in embryos has also been studied by using labeled poly(A)⁺ RNA from different developmental stages to screen several hundred cloned sequences from gastrula and tadpole cDNA libraries (Dworkin and Dawid 1980 a, Dworkin and Dawid 1980 b, Dworkin *et al* 1981). These studies confirmed the previous results and suggested that oocyte, egg and gastrula poly(A)⁺ RNA populations are similar, changes only become apparent from the tailbud stage onwards with some RNA species increasing in concentration from undetectable levels which also leads to the presence, in tadpoles, of the highly abundant poly(A)⁺ RNA species. Also this showed in early embryos that mitochondrial RNAs were the most abundant poly(A)⁺ RNAs. A study of individual cloned cDNA probes showed that the accumulation of prominent tadpole poly(A)⁺ RNAs that are rare in eggs starts to increase in titre at the beginning of neurulation suggesting in the late gastrula there is a major transition in gene expression which accompanies cellular differentiation and morphogenesis (Dworkin *et al* 1984). Abundant RNAs which appeared in polysomes during later development were found to be newly synthesised RNA (Dworkin and Hershey 1981). Hence this result and the results

described earlier implies that initially during development polysomes are assembled using stored maternal mRNA and then later in development polysome assembly involves newly made RNAs some of which may be qualitatively different from the RNAs in the egg.

Since total poly(A)⁺ RNA levels remain constant during gastrulation but newly synthesised mRNA begins to accumulate from blastula onwards the poly(A)⁺ RNA in the embryo must turnover. It has been estimated that 40% of the mass of gastrula poly(A)⁺ RNA is newly synthesised (Shiokawa *et al* 1981). It has been shown that stored histone mRNA turns over during early development such that by gastrulation it is entirely replaced by newly synthesised mRNA (Woodland 1980). For other endogenous mRNAs turn over rates are not clear but it is interesting to note that injected mouse globin mRNA is stable through embryogenesis (Gurdon *et al* 1974). Therefore turn over rates may be a property of the individual mRNAs.

1.5.4 Developmental decisions during embryogenesis

The information described above is at a general level and refers to population classes of RNA molecules. Much insight has been brought into the expression of individual genes by the application of recombinant DNA technology and much is understood about the developmental regulation of known gene products such as histone, actin and tubulin. A number of approaches have been adopted to try and identify genes or factors that play critical roles in developmental decisions. In other words the initial information that differentiates populations of blastomeres must come from somewhere.

1.5.4.1 Homeobox containing genes

One approach to isolate genes that are developmentally important is by virtue of their homology to genes that are important in pattern formation in other systems. The homeotic and segmentation genes in *Drosophila* are important in specifying body segment identity, number and polarity and it is thought these genes function by selecting between different developmental pathways. The homeobox is a protein coding domain that is conserved in many of the homeotic and segmentation genes (see Gehring 1985 for review). It is most striking that this conserved sequence resides on proteins that are involved in important features of *Drosophila* embryogenesis. The function of these proteins remains to be determined but the homology of homeoboxes

with yeast mating type regulatory proteins and bacterial DNA binding proteins has led to the suggestion that regulatory proteins are directed to binding sites on DNA by the homeobox domain (Shepherd *et al* 1984, Laughon and Scott 1984, DiNardo *et al* 1986). This has been supported by the fact homeobox proteins are located in the nucleus (White and Wilcox 1984, Beachy *et al* 1985, DiNardo *et al* 1986) and since the engrailed gene product can bind DNA (Desplain, Theis and O'Farrel 1985). There is genetic evidence that homeobox proteins can regulate the transcription of other homeotic genes (Struhl and White 1985); indeed a hierarchy of when these gene products are required by the *Drosophila* embryo has been recently described (MacDonald, Ingham and Struhl 1986). That the homeobox plays an important role in developmental events has been reinforced by the discovery that the sequences are conserved in frogs (Carrasco *et al* 1984, Muller, Carrasco and DeRobertis 1984, Harvey, Tabin and Melton 1986), mice (McGinnis *et al* 1984, Rabin *et al* 1985, Hart *et al* 1985) and humans (Levine, Robin and Tijan 1985, Hauser *et al* 1985) all of which display a segmented body plan. Most homeobox containing genes appear to be developmentally regulated or expressed in a cell type specific manner hence this predicts a role for homeoboxes in development. The isolation of a homeobox gene in a non-segmented organism, the sea urchin, which is also developmentally expressed (Dolecki *et al* 1986) suggests homeobox genes are not simply involved in the establishment of body plan *per se* but rather they might be the regulatory genes that choose between different developmental programs.

The fact homeobox genes appear to be developmentally important has led to the isolation of a number of homeobox containing genes from *Xenopus* (Carrasco *et al* 1984, Muller, Carrasco and De Robertis 1984, Harvey, Tabin and Melton 1986). The homeobox containing domain is very highly conserved (greater than 65% when compared to *Drosophila* sequences). Interestingly two homeoboxes were isolated on the same genomic clone (Xhox-1). The sequences outside the Xhox-1B homeobox are similar to the human Hu-1 and mouse Mu-1 homeobox containing genes hence it will be of great interest to identify the function of this gene. Both Xhox-1A and B transcripts can first be detected in gastrula stage embryos and are expressed through to the swimming tadpole stage. The Xhox-1A transcript is present at very low levels in unfertilised eggs whereas the MM3 homeobox transcript is expressed maternally at a very high level. During cleavage the MM3 transcript disappears

rapidly but it is re-expressed from gastrulation onwards with an additional transcript. So the MM3 gene product might determine the fate of cells during the initial stages of development. The AC-1 homeobox is expressed from the beginning of neurulation through to the swimming tadpole. The AC-1 probe also detects two other transcripts which are only expressed transiently during neurulation. Injection of Xhox-1A RNA into *Xenopus* oocytes shows that the protein encoded by this gene is located in the nucleus. Deletion of 13 amino acids from the homeobox domains as well as 32 carboxyl-terminal amino acids of the Xhox-1A protein produces a protein which no longer localises to the nucleus. Therefore the *Xenopus* homeobox genes have features which might say they are important in developmental decisions but much more work will be required in order to elucidate their function.

1.5.4.2 Localised factors

It has been a long standing hypothesis in developmental biology that embryonic cells acquire specific developmental fates by the inheritance of maternal factors which are differentially distributed among cleavage cells. The localisation of maternal factors in the egg which are inherited by specific blastomeres provides an attractive explanation of how cell fate (gene expression) is determined. The nature of these determinants is unknown but it is often suggested that maternal mRNA could fulfill such a function, indeed it is implicated in insect development (Anderson and Nusslein-Volhard 1984). If maternal mRNA is such a cytoplasmic determinant then some of the mRNAs should be regionally localised within the egg. cDNA:RNA hybridisation suggests that small number (3-5%) of poly(A)⁺ RNA sequences are enriched in the vegetal third of *Xenopus* eggs and embryos (Carpenter and Klein 1982). Also concentrations of both total and poly(A)⁺ RNA are known to change along the animal-vegetal axis of oocytes and embryos (Capco and Jeffery 1982, Capco and Jackle 1982). A group of abundant soluble proteins have been found to be localised along the animal-vegetal axis but their relationship to localised RNAs is unknown.

Two recent studies have shown the existence of localised RNAs in frog eggs. In the first study poly(A)⁺ RNA isolated from different regions of oocytes was translated in *vitro* and analysed by two-dimensional gel electrophoresis (King and Barklis 1985). A comparison of the resulting protein products showed that at least 17 specific translatable mRNA are regionally localised in

the oocyte. The second study used a differential screening procedure to isolate cDNA clones for localised mRNAs (Rebagliati *et al* 1985). This showed most maternal mRNAs are uniformly distributed but a rare class of localised mRNAs existed. Three of these localise to the animal pole (An1,2,3) and one to the vegetal pole (Vg1) and they are all locally inherited in cleaving blastomeres. An1 and Vg1 are only synthesised in oogenesis and during development this mRNA disappears by gastrula. An2 and An3 are similarly expressed in oogenesis but their mRNAs are re-expressed in late gastrula and neurula respectively. Thus their expression is consistent with a role in cell determination. Each of these mRNAs appear to be located on polysomes during development and hence probably encode proteins (Weeks *et al* 1985). There are many ways how these mRNAs could affect the developmental fates of embryonic cells. A simple idea would be their protein products might turn on genes in one region of the embryo. Further work should elucidate the roles of localised mRNA in cell determination.

Another way to analyse how cells acquire their developmental fates is to use genes that are known to be activated in restricted cell types during embryogenesis. The cardiac actin gene is transcriptionally activated at the end of gastrulation only in those cells that will subsequently form muscle (Mohun *et al* 1985, Gurdon *et al* 1985 a). Ligation experiments with fertilised but uncleaved eggs showed that all the materials required for the eventual activation of the cardiac actin genes are localised in a subequatorial region of the egg (Gurdon *et al* 1985 a, Gurdon *et al* 1985 b). It is suggested that some muscle cells of an embryo might be formed by the activity of cytoplasmic determinants localised in the uncleaved egg and partitioned by cleavage to the third tier of a 32 cell embryo. It is possible to argue that inductive processes rather than cytoplasmic determinants are involved in producing muscle cells from the tier-3 cells. It has been shown that vegetal cells can induce animal cells to activate the cardiac actin genes (Gurdon *et al* 1985 a, Gurdon *et al* 1985 c). The induction signal does not require cell contacts but the ability of cells to induce and respond is strictly limited in time (Gurdon *et al* 1985 a, Gurdon *et al* 1985c) also cell division is not required for the signal (Gurdon and Fairman 1986). Recently it has been demonstrated that upstream sequences are required for the temporal and tissue specific expression of the cardiac actin gene as shown by the microinjection of chimeric plasmids containing cardiac actin promoter - bacterial chloramphenicol acetyl transferase gene constructs

into fertilised eggs (Mohun, Garrett and Gurdon 1986). Therefore it appears that the developmental regulation of cardiac actin expression (i.e. muscle cell differentiation) is influenced by many parameters and hence it could be difficult to obtain an overall picture of how muscle cells are determined.

1.5.4.3 Expression at particular developmental stages

Although the RNA populations of eggs and gastrula embryos are similar some qualitatively new mRNAs are synthesised immediately after the midblastula transition. These genes might encode products that are important in the formation of germ layers and tissue anlagen and might also be involved in gastrulation processes. There are a few ways that this class of gene sequence can be isolated. The first approach is to use labeled poly(A)⁺ RNA from eggs and gastrula embryos to differentially screen a gastrula cDNA library but only relatively abundant sequences will be isolated by this method. One such clone, called GS17, was isolated by this procedure (Krieg and Melton 1985). GS17 is one of the first genes to be expressed by the embryonic genome exactly at the midblastula transition and is only expressed transiently during gastrulation. It is interesting to note, as judged by microinjection of recombinant plasmids, that only about 2kb at the 5' end of the GS17 gene are required for both the correct switch on and off of transcription during development. No function has yet been assigned to the GS17 gene product but its promoter will certainly be useful in studying the function of other genes during development.

The second approach is to use cDNA libraries that have been enriched in the sequences of interest. By employing a subtractive cDNA cloning procedure a library containing sequences that are differentially expressed in gastrula (called DG genes) has been obtained (Sargent and Dawid 1983). Each DG gene is activated in a specific temporal manner during early development and none are expressed in adult (Jamrich, Sargent and Dawid 1985). Sequence analysis of many of the clones has revealed two families (Jonas, Sargent and Dawid 1985, Dawid and Sargent 1986). One family represents cytokeratins which form the intermediate filament system of epidermal and epithelial cells. These genes appear to be embryonic forms of adult cytokeratins and in the gastrula their mRNA is located exclusively in the ectoderm. The second family represents sequences that accumulate rapidly after the midblastula transition and the decay during neurulation. One sequence, DG 42, has no sequence homology to

any other known gene and its mRNA is exclusively located in the endoderm of the gastrula.

Both classes of genes have been used as molecular markers to analyse inductive events during early differentiation (Jamrich, Sargent and Dawid 1985, Dawid and Sargent 1986). Using exogastrula embryos it has been shown that the correct shape of the embryo is not important for the expression of the DG genes or cardiac actin. Suppression of mesoderm formation by UV treatment (which is a purely positional effect) or the interruption of cell communication by the use of calcium and magnesium free medium and dissociation of cells does not inhibit the expression of the DG genes but does cardiac actin. The keratin genes are switched off but not other DG genes and cardiac actin genes when the cells are in close proximity yet have lost tight adhesivity. All these results suggest that the regulatory substances which cause the activation of the DG genes could be maternally derived but the activation of cardiac actin (and other muscle specific genes) requires substances to be formed in early development. Since expression of ectodermal and endodermal markers is cell autonomous the endoderm/ectoderm dichotomy is determined by the polarity of components in the egg. The mesoderm originates from an interaction between ectodermal and endodermal cells which is independent of calcium mediated tight adhesion and expression of mesodermal markers is suppressed when cell interactions during the blastula stage are inhibited. Elucidation of the nature of these cell interactions and how the differentiated state arises will be a very exciting area in future studies on molecular biology of development.

1.6 Objectives of this Thesis

Initially I wanted to make use of the observation that cellular oncogenes are highly conserved in evolution to encompass the genus *Xenopus*. Since cellular oncogenes are strongly implicated in regulation of normal cell growth, division and differentiation and during development it should be possible to exploit the potential of *Xenopus* to analyse cellular oncogenes during oogenesis, very early embryogenesis and cellular differentiation.

Interest would focus on one particular oncogene (*c-myc*) with the isolation on homologous sequences from *Xenopus* gene libraries helping in the analysis. Sequence analysis of such clones would then establish the *Xenopus* sequences relationship with other sequences and may point to important functional

domains in the protein. Clues to the function of these genes might be found by analysing their expression during oogenesis and development. Also parallels with other systems might become apparent and unique features peculiar to aspects of *Xenopus* development might be found. Antisera to putative gene products could be generated to try and study aspects of protein expression. Also clues to the evolution of the *Xenopus* genus might be gained. In summary I would hope to have begun an initial characterisation of a gene that plays an important role in development - the choice of this gene having been influenced by its apparently important role in normal and tumour cell growth in other systems.

CHAPTER 2

MATERIALS AND METHODS

2.1 Solutions

Generally solutions were made sterile and RNase/DNase free where possible by autoclaving (15psi/15 minutes) or by treating with 0.01%(v/v) diethyl pyrocarbonate overnight at 65°C. Stocks solutions were made in sterile, baked glassware and were also filtered through a 0.45µm cellulose nitrate filter to remove particulate matter. Solutions not detailed in the text are described below.

<u>TE</u>	10mM Tris-Cl, 1mM EDTA pH 7.4 - 8.0
<u>0.4M EDTA</u>	0.4M diaminoethanetetra-acetic acid pH 8.0
<u>1x DeBoers</u>	0.11M NaCl, 1.3mM KCl, 0.44mM CaCl pH 7.2 with NaHCO ₃
<u>Barth X (Barth and Barth 1954)</u>	<u>solution A:</u> 51.50g NaCl, 0.75g KCl, 2.04g MgSO ₄ ·7H ₂ O 0.62g Ca(NO ₃) ₂ ·H ₂ O, 0.60g CaCl ₂ ·2H ₂ O made to 1 litre divided into 100ml <u>solution B:</u> 2g NaHCO ₃ made to 500ml and divided into 50ml aliquots <u>solution C:</u> 24.22g Tris, 20mg phenol red pH to 7.6 with HCl made to 500ml and divided into 50ml aliquots 1x Barth X is made by mixing 1 aliquot of solution A, B and C then made to 1 litre
<u>MS222</u>	10 - 20%(w/v) ethyl-M aminobenzoate in ddH ₂ O
<u>10x high salt restriction buffer</u>	1M NaCl, 0.5M Tris-Cl pH 7.5, 0.1M MgCl ₂ , 10mM DTT
<u>10x medium salt restriction buffer</u>	0.5M NaCl, 0.1M Tris-Cl pH 7.5, 0.1M MgCl ₂ , 10mM DTT
<u>10x low salt restriction buffer</u>	0.1M Tris-Cl pH 7.5, 0.1M MgCl ₂ , 10mM DTT
<u>10x universal restriction buffer</u>	0.33M Tris-acetate pH 7.9, 0.1M Mg-acetate, 0.66M K-acetate, 10mM DTT, 1mg/ml gelatin

<u>10x TAE gel buffer</u>	0.4M Tris-acetate pH 8.3, 20mM EDTA
<u>10x TBE gel buffer</u>	0.89M Tris-borate-0.89M boric acid, 10mM EDTA
<u>DNA gel sample buffer</u>	50%(v/v) glycerol, 0.25%(w/v) bromophenol blue, 0.25%(w/v) xylene cyanol FF in ddH ₂ O
<u>10x MOPS gel buffer</u>	0.2M Na-MOPS pH 7.0, 50mM sodium acetate, 10mM EDTA
<u>RNA sample buffer</u>	50% formamide, 2.2M formaldehyde, 1x MOPS gel buffer
<u>10x T4 DNA ligase buffer</u>	0.66M Tris-Cl pH 7.6, 50mM MgCl ₂ , 50mM DTT
<u>STET</u>	8%(w/v) sucrose, 0.5%(v/v) Triton X-100, 50mM EDTA pH 8.0, 10mM Tris-Cl pH 8.0
<u>Kirby salts</u> <u>(modified from Kirby 1965)</u>	0.1M Tris-Cl pH 9.0, 10mM EDTA, 0.5%(w/v) SDS, 1.0%(w/v) tris-iso-propylnaphtalene sulphonic acid, 6.0%(w/v) 4-aminosalicylic acid, 1.0%(w/v) NaCl, 6.0%(w/v) phenol
<u>10x Linker-kinase buffer</u>	0.7M Tris-Cl pH 7.6, 0.1M MgCl ₂ , 50mM DTT
<u>30% Acrylamide stock</u>	30%(w/v) acrylamide, 1%(w/v) N-N'-methylene bis-acrylamide, deionised with MB-1 (BDH) stored in dark at 4°C
<u>TEMED</u>	NNN'N'-tetra-methyl-1,2-diamino-ethane (BDH)
<u>APS</u>	10 or 25%(w/v) ammonium persulphate in ddH ₂ O
<u>DNase I mix</u>	100ng/ml DNase I, 50mM Tris-Cl pH 8.0, 5mM MgCl ₂
<u>Pol I mix-C</u>	50mM KPO ₄ pH 7.4, 7mM MgCl ₂ , 1mM DTT, 0.33mM dATP 0.33mM dGTP, 0.33mM dTTP
<u>OLB</u>	solution O: 0.125M MgCl ₂ , 1.25M Tris-Cl pH 8.0 solution A: 0.95ml solution O, 18µl 2-mercaptoethanol, 25µl 20mM dATP, 25µl 20mM dTTP, 25µl 20mM dGTP solution B: 2M Hepes pH 6.6

	solution C: Hexadeoxyribonucleotides (Pharmacia) suspended in TE at 90 OD U/ml
	OLB is made up by mixing solution A, B, and C in the ratio 2:5:3 and stored at -20°C
<u>10x transcription buffer</u>	0.4M Tris-Cl pH 7.9, 0.1M NaCl, 0.1M DTT, 0.06M MgCl ₂ 0.02mM spermidine.
<u>10x transcription rNTPs</u>	5mM rATP, 5mM rGTP, 5mM rTTP, 120µM UTP
<u>20x SSPE</u>	3.6M NaCl, 0.2M NaH ₂ PO ₄ pH 7.4, 0.02M EDTA
<u>50 x Denhardt</u>	1%(w/v) Ficoll, 1%(w/v) polyvinyl pyrrolidone, 1%(w/v) BSA (pentax fraction V)
<u>Denatured herring sperm DNA</u>	Is made up as a 5mg/ml stock and sonicated till the viscosity is reduced. It is also phenol extracted as in 2.4.1
<u>cDNA mix-C</u> <u>(modified from Efstratidis</u> <u>et al 1976)</u>	50µl 1M Tris-Cl pH 8.3 (measured at 37°C), 40µl DTT 30µl 0.2M MgCl ₂ , 20µl 3M NaCl, 50µl 20mM dATP, 50µl 20mM dGTP, 20mM dTTP, 100µg/ml p(dT) ₁₂₋₁₈ 460µl ddH ₂ O. Stored at -20°C
<u>2x Second strand salts</u>	40mM Tris-Cl pH 7.5, 10mM MgCl ₂ , 20mM ammonium sulphate, 200mM KCl
<u>Packaging buffer A</u>	20mM Tris-Cl pH 8.0, 3mM MgCl ₂ , 0.05%(v/v) 2-mercaptoethanol, 1mM EDTA
<u>Packaging buffer M1</u>	110µl ddH ₂ O, 6µl 0.5M Tris-Cl pH 7.5, 300µl 0.05M spermidine, 0.1M putrescine 9µl 1M MgCl ₂
<u>5x Laemmli gel buffer</u>	144g glycine, 30g Tris, 5g SDS made to 1 litre
<u>2x Protein sample buffer</u>	0.125M Tris-Cl pH 6.8, 20%(w/v) glycerol, 5%(v/v) 2-mercaptoethanol, 2%(w/v) SDS, 0.01%(w/v) bromophenol blue.
<u>Protein extraction buffer</u>	0.25M sucrose, 50mM Tris-Cl pH 7.5, 50mM KCl 10mM MgCl ₂ , 1mM DTT with 150µg/ml PMSF added just before use
<u>10x PBS</u>	0.1M Sodium phosphate pH 7.5, 1.3M NaCl
<u>X-gal (BCIG)</u>	20mg/ml 5-bromo-4-chloro-3-indoyl-β galactoside in dimethyl formamide
<u>IPTG</u>	20mg/ml Isopropyl-β-D-thiogalactoside in ddH ₂ O

2.2 *Xenopus*

2.2.1 *Xenopus* stocks

Xenopus borealis, *Xenopus laevis* and *Xenopus tropicalis* frogs were kept in large aquaria and maintained in a constant environment at 22°C. Animals were raised from our own matings and were fed on pellets supplied by *Xenopus* Ltd. Aquaria were cleaned weekly.

2.2.2 Egg and embryo production

2.2.2.1 Live matings

Mature male and female frogs are injected with 150–300 units of human chorionic gonadotrophin (Pregnyl, Organon) and a couple allowed to mate in small aquaria. Eggs are laid 8–12 hr later. Fertilised eggs are harvested and allowed to develop to the required stage (Nieuwkoop and Faber 1956). Stages 1–6 are cleavage, stages 7–9 are blastula, stages 10–13 are gastrula, stages 14–20 are neurula and stages 23–24 are tailbud. embryos are dejellied in 2%(w/v) cysteine–hydrochloride pH 8.0 before use.

2.2.2.2 Artificial fertilisation

Females are injected as in 2.2.2.1 and eggs collected 8–12 hr later in to 1x DeBoers by gentle squeezing of the animal. Sperm are obtained from males by macerating testes in 1x DeBoers (Wolf and Hedrick 1971). Eggs and sperm are mixed in a small volume of 0.05x DeBoers and fertilisation monitored by rotation of the eggs. Developing eggs are transferred to 1x DeBoers. Upon reading blastula embryos are transferred to 0.05x DeBoers to allow gastrulation to proceed correctly.

2.2.3 Oocytes and other tissues

Frogs are anaesthetised with a lethal dose of MS222. Ovaries and other tissues are dissected, washed in Barth X and used immediately or stored at -70°C. Previtellogenic ovaries are isolated from immature frogs. Blood is obtained by heart puncture; injection of a little Barth X containing heparin helps stop coagulation. Anemic frogs are made by the injection, on three consecutive days, of 6–10mg of phenylhydrazine in Barth X into the dorsal lymph sac. These animals are bled 10 days after the final injection. Embryos

are dissected under a microscope using fine tungsten needles.

Oocytes are obtained from adult ovaries by treatment of 1mg/ml collagenase in Barth X plus 20mM EDTA. After gentle stirring dissociated oocytes are washed in Barth X. These oocytes are essentially free of surrounding follicle cells and are staged as described by Dumont (Dumont 1972). Table 2.1 summarises these stages.

2.3 Microbiological Strains and Media

2.3.1 Microbiological strains

All bacterial strains, plasmid vectors and bacteriophage vectors used in this study are described in table 2.2. Oncogene and other plasmids are described in Table 2.3. Bacterial stocks are stored at -70°C in freezing media and are also maintained on the appropriate plates. Lambda bacteriophage are stored in phage buffer with a few drops of CHCl_3 at 4°C . M13 bacteriophage are stored as DNA in sequencing TE at -20°C or in M13 phage buffer.

2.3.2 Media

Media is sterilized by autoclaving. Antibiotics, vitamins and sugars are made up in sterile water than filter sterilised. Where appropriate antibiotics are added to plates and media : ampicillin to $50\mu\text{g/ml}$, streptomycin to $200\mu\text{g/ml}$, tetracycline to $50\mu\text{g/ml}$ and chloramphenicol to $100\mu\text{g/ml}$.

<u>L broth</u>	10g Difco bacto Tryptone, 5g Difco Bacto yeast extract, 5g NaCl per litre, pH 7.2
<u>L agar</u>	As L broth plus 15g Difco agar per litre
<u>BBL agar</u>	10g Baltimore Biological Laboratories trypticase, 5g NaCl, 10g Difco agar per litre, pH unaltered
<u>BBL top agar</u>	As BBL agar but only 6.5g Difco agar per litre
<u>Spitzen minimal salts</u>	10g $(\text{NH}_4)_2\text{SO}_4$, 20g K_2HPO_4 , 30g KH_2PO_4 , 5g tri-sodium citrate, 1g MgSO_4 per litre
<u>Minimal medium</u>	80ml spitzen minimal salts, 4mls 20% (w/v) glucose 0.2ml 1mg/ml vitamin B1 per 400ml
<u>Minimal agar</u>	As minimal medium plus 6g Difco Bacto agar
<u>Phage buffer</u>	3g KH_2PO_4 , 7g Na_2HPO_4 , 5g NaCl, 10ml 0.1M MgSO_4 ,

Xenopus oocyte stages

		Stage	Size (μ)	General appearance	Chromosomes
Pre-vitellogenic		I	50 to 300	Transparent cytoplasm; nucleus and mitochondrial mass clearly visible	Zygotene, pachytene, and very early diplotene
Vitellogenic	Early	II	300 to 450	Cytoplasm translucent early, white and opaque later; nucleus and mitochondrial mass visible in early Stage II	Diplotene and early lampbrush
	Mid	III	450 to 600	Pigment forms — light brown early, uniformly blackish-brown late; lightly pigmented at area of attachment to ovary	Mid-diplotene; maximum lampbrush
		IV	600 to 1000	Animal and vegetal hemispheres become differentiated	Late diplotene; lampbrush and chromosome frame retract
	Late	V	1000 to 1200	Hemispheres clearly delineated at equator; animal hemisphere appears light brown	Condensed; massed in center of nucleus
Post-vitellogenic	VI	1200 to 1300	Unpigmented equatorial band		

Table 2.1 (From Dumont 1972)

	10ml 0.1M CaCl ₂ , 1 ml 1%(w/v) gelatin per litre
<u>M13 Phage buffer</u>	20mM Tris-Cl pH 7.5, 20mM NaCl, 1mM EDTA
<u>Freezing medium</u> <u>Gergen <i>et al</i> 1979)</u>	6.3g K ₂ HPO ₄ , 1.8g KH ₂ PO ₄ , 0.45g Na citrate, 0.09g MgSO ₄ .7H ₂ O, 0.9g (NH ₄) ₂ SO ₄ , 44ml glycerol to 1 litre L broth.

Hosts	Genotype	Reference
C600	F ⁻ ,thi1,thr1,leuB6,lacY1 tonA2I,supE44,λ ⁻	Appleyard 1954
HB101	F ⁻ ,hsdS20(r ⁻ _B ,m ⁻ _B),recA13 ara14,proA2,lacY1,galk2, rpsL20(Sm ^r),xyl5,mtl1,supE44,λ ⁻	Boyer 1969
RR1	As HB101 except recA ⁺	Bolivor 1977
JM83	ara,Δ(lac-proAB),rpsL(strA), φ80,lacZΔM15	Yanisch-Perron 1983
NM522	hsdΔ(M ⁻ S ⁻ R ⁻),Δlac,Δpro,supE, thi,/F'proA ⁺ B ⁺ ,lacI ^q ,lacZΔM15, traD36	Gough 1983
pop2136	MM294 (F ⁻ endA,thi,hsdR)with malT, Prcl857, mal PQ. This was constructed by Dr.Raibaud using pOM40 to transfer Pr,CI857 onto C600 (by Mal ⁻ Tet ^r selection) then by PI cotransduction using the marker araB onto MM294 (Personal communication)	Raibaud 1984
NM514	lyc7, hsdM ⁺ S ⁺ R ⁻ hfl	Murray 1983
NM259	supE, supF, hsd M ⁺ S ⁺ R ⁻ ,met trp3	Murray 1977
Vectors	Genotype	Source/Reference
pAT153	Amp ^r , Tet ^r (cloning vector)	Twigg 1980
pBR322	Amp ^r , Tet ^r (cloning vector)	Sutcliffe 1978
puc18/19	Amp ^r (lacZ colour selection, cloning vector)	Yanisch-Perron 1983
pGEM1/2	Amp ^r , (T7/SP6 transcription)	Promega Biotech (Melton 1984)
pEX1/2/3	Amp ^r (β galactosidase-fusion)	Stanley 1984



	protein expression vector)	
λNM1149	Imm ⁴³⁴ b(538) (cloning vector)	Murray 1983
M13mp18/19	(lacZ colour selection, sequencing vectors)	Yanisch-Perron 1983

Table 2.2 Bacterial hosts, plasmid vectors and bacteriophage vectors

Oncogene	Size Fragment	Source/Reference
<i>v-abl</i>	7.8kb <i>EcoRI</i>	N.Hastie(Srinivasan 1981)
<i>v-erbA/B</i>	2.5kb <i>PvuII</i>	A.Hall(Vennstrom 1982)
<i>v-erb B</i>	0.5kb <i>BamHI</i>	subcloned from <i>v-erb A/B</i>
<i>v-fos</i>	1.1kb <i>Pst I</i>	R.Muller(Curran 1982)
<i>v-fms</i>	1.4kb <i>Pst I</i>	A.Hall(Donner 1982)
chicken <i>myc</i>	11.8kb <i>BamHI</i>	N.Hastie(Watson 1983)
chk-2- <i>myc</i>	0.5kb <i>PstII Sall</i>	subcloned from chicken <i>myc</i>
human <i>myc</i>	1.7kb <i>EcoRI/ClaI</i>	N.Wilkie(Dalla-Favera)
<i>N-myc</i> (human)	1.0kb <i>EcoRI/Bam HI</i>	P.Middleton(Schwab 1983)
<i>v-ras Harvey</i>	2.3kb <i>BamHI/ EcoRI</i>	N.Hastie(Ellis 1981)
<i>v-ras Kirsten</i>	3.2kb <i>BamHI/ EcoRI</i>	N.Hastie(Ellis 1981)
<i>v-rel</i>	1.0kb <i>EcoRI</i>	A.Hall(Chen 1981)
<i>v-sis</i>	3.1kb <i>EcoRI</i>	A.Hall(Czernilopsky 1983)
Others	Size Fragment	Source/Reference
<i>X.borealis</i> cytoskeletal actin(p35A)	3.6kb <i>EcoRI</i>	H.Woodland(unpublished)
<i>X.borealis</i> histone H4(pX1H4W1)	0.4kb <i>BamHI</i>	H.Woodland(Turner 1982)
<i>X.laervis</i> mitochondria genome(pXIm32)	17kb <i>BamHI</i>	I.Dawid(Rastl 1979)

Table 2.3 Oncogene and other plasmids

2.4 General Methods

2.4.1 Deproteinisation of nucleic acids

Phenol was first redistilled under N₂, adjusted to 0.1%(w/v) 8-hydroxyquinoline and saturated with TE. Purified phenol is stored at -20°C until required.

Nucleic acids are extracted as follows. 5µg of *E.coli* tRNA is added as carrier if required then the aqueous phase is adjusted to 0.4M NaCl, 20mM EDTA. An equal volume of phenol is added, incubated at 37°C for 5 minutes and occasionally vortexed to maintain the emulsion. After cooling on ice an equal volume of CHCl₃/isoamyl alcohol (24:1) is added and incubated for 5 minutes vortexing occasionally. The phases are separated under centrifugation and the aqueous phase removed. The organic phase is back-extracted with an equal volume of TE and the aqueous phases combined then re-extracted with an equal volume of CHCl₃/isoamyl alcohol (24:1) on ice. After centrifugation the aqueous phases are precipitated as required.

2.4.2 Precipitation of nucleic acids

The aqueous phase is adjusted to 0.3M sodium acetate (pH 5.2) or 0.4M NaCl then 2 volumes (for DNA) or 2.5 volumes (for RNA) of ethanol or 1 volume of isopropanol (for DNA) is added and incubated at -20°C for at least 1hr. For small quantities of material (<5µg/ml) carrier tRNA may be added to 5-10µg/ml and the precipitation left overnight. The precipitate is recovered under centrifugation at 4°C in a Sorvall centrifuge (10K rpm, 20 minutes, HB4) or microfuge (12Krpm, 10 minutes). The pellet is washed in 70% ethanol, air dried and redissolved in ddH₂O or TE.

DNA can also be precipitated by the addition of an equal volume of 4M ammonium acetate and 2 volumes of ethanol or 1 volume of isopropanol. After incubating on dry/ice ethanol for 15 minutes the DNA is recovered as above. Two such precipitations remove >99% of unincorporated nucleotides from reactions (Okayama and Berg 1982).

2.4.3 Autoradiography

Autoradiography was performed using Cronex 4 X-ray film and cassettes. For ³²P film were preflashed and exposed at -70°C using intensifying screens

(Laskey and Mills 1977). Films were developed in a Agfa X1 automatic film processor.

2.4.4 Deionised formamide

100ml formamide was deionised with 10g BioRad analytical grade mixed bed resin AG 501-X8(D), 20-50 mesh by stirring at room temperature for 30 minutes. Formamide was then filtered through Whatman No.1 filter paper and stored at -20°C .

2.4.5 Restriction endonuclease digestion of DNA

In general DNA was digested with a 2-4 fold excess of enzyme for at least 2hr in conditions recommended by the manufacturer. Genomic DNA was digested for at least 6hr with a 4 fold excess of enzyme. Alternative reaction conditions used high salt, medium salt or low salt restriction buffers (Maniatis *et al* 1982) or universal buffer (O'Farrel *et al* 1980). Multiple enzyme digests were carried out in the same buffer where possible, otherwise after completion of one digest the salt conditions are altered for the next or the sample phenol extracted and precipitated before the next reaction.

2.4.6 Quantitation of nucleic acids

The optical density at 260nm was used to quantify nucleic acids. An OD_{260} of 1.0 is $50\mu\text{g/ml}$ for DNA or $40\mu\text{g/ml}$ for RNA. Nucleic acid was free of contaminating protein if $\text{OD}_{260}/\text{OD}_{280}$ was 1.8-2.0. The background value at OD_{320} was subtracted from all readings. A rough estimate of nucleic acid concentration can be obtained by ethidium bromide staining of agarose gels with known quantities of material.

2.4.7 Agarose gel electrophoresis

2.4.7.1 DNA gels

DNA was size fractionated by submerged agarose gel electrophoresis. Molten Pharmacia agarose (0.3-2.0%w/v in 1x TAE gel buffer) was cooled to 60°C and poured into horizontal gel formers. Set gels were submerged in homemade electrophoretic tanks containing 1x TAE gel buffer. 1/5th volume of DNA gel sample buffer was added to the DNA sample and then applied to the sample well. Electrophoresis was performed at various voltages (5-100 V/cm)

and times (0.5–18 hr) at room temperature until the required separation was achieved. Restriction digests of bacteriophage lambda cl857 or pAT153 were used as molecular weight markers. Gels were stained with 5µg/ml ethidium bromide in ddH₂O for 0.5hr and destained in ddH₂O for 2hr at room temperature. Low melting point agarose (BRL) gels were poured and run at 4°C, 7.5V/cm in 1x TAE gel buffer containing 0.5 µg/ml ethidium bromide.

2.4.7.2 RNA gels (Lehrach et al 1977)

RNA was similarly size fractionated by electrophoresis through 1.2% agarose (Pharmacia) gels. Molten agarose was cooled to 60°C and adjusted to 1x MOPS gel buffer, 2.2M formaldehyde before pouring into gel formers. Dried RNA samples were redissolved in 20µl RNA sample buffer and heated to 65°C for 5 minutes. After cooling on ice 2µl of 50% glycerol, 0.2%(w/v) bromophenol blue was added and the sample applied to the gel well. Electrophoresis was at 5V/cm for 18hr in 1x MOPS gel buffer at room temperature. Gels were stained for 20 minutes with 5µg/ml ethidium bromide in 0.1M ammonium acetate at room temperature and destained for 2hr in 0.1M ammonium acetate. *Xenopus borealis* 18S and 28S ribosomal RNA was used as markers.

2.4.8 Gel photography

Nucleic acid was visualised in stained gels using an ultraviolet transilluminator (254nm). Photographs were taken through a red (A1) filter onto Ilford HP5 film (f4.5, 20s). Films were developed in Ilford microphen (5 min), stopped in 3%(v/v) acetic acid (30s) and fixed in Ilford hypem (5 min) at room temperature. Films were washed well in water, dried and the relative mobilities of the sample measured directly from the negative (distance proportional to log₁₀ mwt).

2.4.9 Ctab precipitation

The incorporation of radiolabeled nucleotides into nucleic acids was monitored by ctab precipitation. A sample is added to 1ml of 1mg/ml yeast tRNA in 0.5M sodium acetate pH 4.6 and 0.5ml 4% (w/v) ctab is added, mixed and left at room temperature for 5 minutes. The precipitate is collected by vacuum filtration onto Whatman GF/C filters prewetted in 5% (w/v) sodium pyrophosphate. The precipitates are then washed 1ml 5%(w/v) sodium

pyrophosphate and 2ml ddH₂O. The filters are dried at 80°C put into scintillation vials, covered in scintillant (8g/litre butyl-PBD in toluene) and counted.

2.5 Transformation of E.coli (Mandel and Higa 1970)

A single colony of *E.coli* is picked into 5ml L broth and grown up overnight at 37°C. 1ml of this overnight culture is used to inoculate a 100ml culture and is grown till OD₆₅₀=0.2. Cells are chilled on ice then spun down. The cell pellet is gently resuspended in 50ml of cold 100mM CaCl₂ and kept on ice for 15 minutes. Cells are respun and the pellet resuspended in 10ml of cold 100mM CaCl₂ and kept on ice at least 1hr before use.

0.2ml of competent cells are added to the DNA sample (<100ng) in glass tubes and left on ice for 1hr. The tubes are heat shocked for 2 minutes at 42°C and 1ml L broth added and incubated at 37°C for 1hr. 2.5ml of molten BBL top agar at 45°C is added and poured onto appropriate plates, inverted and incubated at the correct temperature. If lacZ colour selection is used 30µl X-gal and 30µl of IPTG is added before pouring. This method gives 10⁵-10⁷ transformants/µg DNA.

2.6 Ligation of DNA

The vector and insert DNA are cut to completion with the appropriate restriction endonucleases but 5 minutes before completion 1 unit of calf intestinal phosphatase (Boehringer) is added. Samples are phenol extracted, precipitated and dissolved in ddH₂O. Ligation conditions are chosen such that the vector to insert DNA concentrations are at a 2:1 molar ratio. Typically 100ng vector and 20ng insert DNA are ligated in 15µl reactions containing 1x T4 DNA ligase buffer, 1mM ATP and 1 unit T4 DNA ligase at 16°C for 18hr. The ligation mixes are transformed into *E.coli* or *in vitro* packaged into lambda.

2.7 Extraction of DNA from Agarose Gels

2.7.1 Low melting point agarose gel (Weislander 1979)

Low melting point agarose gels are run as in section 2.4.7.1 and the required DNA fragment is excised cleanly from the gel. 5 volumes of 0.1M NaCl, 10mM Tris-Cl pH8.0, 1mM EDTA are added and heated to 65°C to melt the gel. The gel is phenol extracted (no CHCl₃) once at 65°C then the aqueous

phase is phenol extracted as in section 2.4.1 and precipitated with addition of 5µg carrier *E.coli* tRNA.

2.7.2 Electroelution

Pharmacia agarose gels are run as in section 2.4.7.1 and a small trough is cut in front of the desired fragment. The trough is then lined with dialysis tubing and the fragment electrophoresed into the trough. The current is reversed for 30 seconds and the DNA is sucked off with a pasteur pipette. DNA is then phenol extracted as in section 2.4.1 and precipitated with addition of 5µg carrier *E.coli* tRNA.

2.8 Elutip-d Column Purification of DNA

For some purposes (e.g. nick translation of insert DNA) it is found that further purification of DNA is required. Elutip-d columns (Schleicher and Schull) are used exactly as described by the manufacturer. Briefly columns are primed by washing in high salt buffer (1.0M NaCl, 20mM Tris-Cl pH 7.5, 1mM EDTA) then low salt buffer (0.2M NaCl, 20mM Tris pH 7.5, 1mM EDTA). DNA, in low salt buffer, is applied to the column (which has a 0.45µm cellulose acetate filter attached to remove particulate matter). The DNA is then eluted in high salt buffer (no filter attached) and the sample precipitated.

2.9 Plasmid DNA Preparation

2.9.1 Mini-preparation of plasmid DNA (Holmes and Quigley 1981)

This method can yield 5–10µg of plasmid DNA. It is convenient to prepare multiples of twelve as required. Single bacterial colonies are used to inoculate 2ml cultures in L broth with appropriate antibiotics and grown overnight at the correct temperature. Cultures are transferred to microfuge tubes and the cells collected by centrifugation for 30 seconds. Supernatants are removed and 90µl of cold STET and 10µl of 10mg/ml lysozyme (made fresh in STET) are added to the pellet and mixed thoroughly. The tubes are boiled for 40 seconds and the cellular debris is immediately pelleted by centrifugation for 10 minutes. Supernatants are removed, deproteinised (2.4.1) and isopropanol precipitated. The DNA pellet is redissolved in 30µl ddH₂O and 5–10µl are used for restriction analysis including RNase A at 50µg/ml to destroy contaminating RNA.

2.9.2 Maxi-preparation of plasmid DNA (modified from Maniatis et al 1984)

A single bacterial colony was used to inoculate a 2ml culture in L broth (with antibiotics) and grown overnight (normally 37°C). A fresh 25ml culture was grown for 3hr using 1ml of the overnight culture. The 25ml culture was used to start a 500ml culture and after 2hr 75mg chloramphenicol was added and the culture grown overnight. Cells were harvested by centrifugation in a Sorvall GSA rotor (4Krpm, 5 minutes, 4°C) and resuspended in 5ml sucrose buffer (25%(w/v) sucrose, 50mM Tris-Cl pH 8.1, 40mM EDTA) on ice. 1ml 10mg/ml lysozyme (in sucrose buffer) and 1ml 0.5 M EDTA added and incubated for 15 minutes on ice. Lysis was achieved by the addition of 13 ml Triton buffer (0.1%(w/v) Triton X-100, 60mM EDTA, 50mM Tris-Cl pH 8.0) and incubated on ice for a further 10 minutes. Chromosomal DNA and cellular debris was removed by centrifugation in a Sorvall SS34 rotor (17Krpm, 1hr, 4°C). The supernatant was recovered, deproteinised (2.4.1) and ethanol precipitated.

Nucleic acids were recovered by centrifugation and resuspended in TE containing 100µg/ml RNase A, 5U/ml RNase T1 and incubated at 37°C for 30 minutes followed by deproteinisation and ethanol precipitation. Plasmid DNA was recovered and redissolved in 3ml TE, 3g of caesium chloride and 0.15ml 5mg/ml ethidium bromide in TE were added. Exactly 6.15g were removed to a vertical rotor centrifuge tube and plasmid DNA was banded under centrifugation in an OTD50 Sorvall ultracentrifuge (45Krpm, 18°C, TV-865) for 18hr. The supercoiled plasmid DNA band was recovered via a syringe needle pushed through the centrifuge tube wall. Plasmid DNA was extracted four times with butan-2-ol (saturated in 4 NaCl, TE) to remove ethidium bromide. The plasmid DNA was diluted to 2ml with TE, precipitated with 2ml isopropanol, recovered and precipitated twice more from ammonium acetate. Plasmid DNA was stored at -20°C in ddH₂O.

2.10 Bacteriophage Lambda DNA Preparation

2.10.1 Plating of phages

An overnight culture of NM514 (recombinant phage) or NM259 (non recombinant phage) was grown in L broth supplemented with 0.4% (w/v) maltose, and harvested by centrifugation then resuspended in 1/2 volume of 10mM MgSO₄. Appropriate dilutions of phage are mixed with 0.2ml of plating

cells and incubated on ice for 20 minutes then at 37°C for 20 minutes. 3ml of molten top agar supplemented with 10mM MgSO₄ is added and poured onto 9cm BBL agar plates. Set plates are inverted and incubated at 37° for 6–12hr.

2.10.2 Plate lysates

A well isolated phage plaque was picked into 1ml phage buffer. 50µl phage were incubated with 100µl plating cells and plated onto freshly poured L plates. Phage are grown (non-inverted) for 12hr at 37°C then 4ml L broth supplemented with 10mM MgSO₄ was added. The top layer and liquid were removed, vortexed briefly and cells and agar removed under centrifugation. The supernatant was removed, titred and stored at 4°C with the addition of a few drops of CHCl₃.

2.10.3 Liquid lysates

A 100ml culture of NM259 or NM514 was grown at 37°C to OD₆₅₀=0.5 in L broth supplemented with 10mM MgSO₄. Cultures were infected with 25µl of plate lysate (moi=0.1) and incubated till OD₆₅₀ reached a minimum (5–7 hr). 4g NaCl, 100µg RNase A and 100µg DNase I were added and incubated at room temperature per 1hr. Phages were left to precipitate overnight at 4°C by the addition of 25ml of 50%(w/v) PEG-6000. The phage particles were recovered by centrifugation at 8Krpm in a Sorvall GSA rotor. The phage particles are gently resuspended in 5ml phage buffer.

2.10.4 CsCl step gradients

The phage particles were further purified on a CsCl step gradient. Three 2ml CsCl steps (in phage buffer) were used with densities of 1.3, 1.5 and 1.7g/ml. The phage sample was carefully layered onto the step gradient in 14ml MSE centrifuge tubes and phage banded at 35Krpm at 20°C for 2hrs in a 6 x 14ml Ti swing out rotor. The bluish phage band was collected through the side of the tube with a syringe. The phage suspension was then dialysed twice for 1hr at room temperature against 1 litre of 10mM MgCl₂, 10mM NaCl, 50mM Tris-Cl pH 8.0.

2.10.5 Phage DNA extraction

The dialysed phage suspension was diluted to 2ml with TE and 100µl 0.4M EDTA, 100µl 10%(w/v) SDS and 10µl 20mg/ml proteinase K added and

incubated at 65°C for 1hr. The phage DNA was then deproteinised and precipitated with ethanol. The DNA pellet was left to redissolve in 500µl ddH₂O overnight at 4°C before storing at -20°C.

2.11 *Xenopus* DNA Preparation (Blin and Stafford 1976)

Xenopus erythrocyte cells are pelleted and resuspended in 1 volume of 0.1M NaCl. Cells are pelleted and resuspended in 10 volumes of lysis buffer (0.5M EDTA, 0.5%(w/v) sarkosyl, 100µg/ml proteinase K) then incubated at 50°C with gentle swirling for 3hr. The DNA is gently extracted three times with one volume of phenol at 50°C then extracted with two volumes of CHCl₃/isoamyl alcohol (24:1) on ice. DNA is then dialysed extensively against TE then treated with RNaseA (to 100µg/ml) at 37°C for 3hrs. Proteinase K is added to 50µg/ml and incubated for 1hr at 37°C before extracting three times with phenol and twice with CHCl₃/isoamyl alcohol (24:1). The DNA is precipitated and the pellet left to redissolve in ddH₂O at 4°C overnight before storing at -20°C.

2.12 *Xenopus* RNA Preparation

2.12.1 Ovary RNA (Darnbrough and Ford 1976)

Dissected ovary tissue is homogenised in Kirby salts (10ml/g tissue) until no large tissue fragments could be observed. An equal volume of phenol is added, vortexed and incubated at 37°C for 15 minutes. The emulsion is spun in a Sorvall HB4 rotor (10Krpm, 4°C, 15 minutes). The upper phase is recovered and 1/10 volume 3M NaCl and one volume phenol is added and shaken then incubated at 37°C for 15 minutes. After respinning the aqueous phase is extracted with two volumes of CHCl₃/isoamyl alcohol (24:1) on ice. Following centrifugation the aqueous phase is precipitated twice from ethanol and the RNA is stored as an ethanol precipitate at -70°C.

2.12.2 Blood RNA

Isolated blood cells are washed twice in Barth X and resuspended in five cell volumes of lysis buffer (0.1M NaCl, 5mM MgCl₂, 20mM Tris-Cl pH 7.5). Lysis is achieved by the addition of another five cell volumes of lysis buffer containing 1/40th volume of 10% (v/v) NP40. Nuclei are removed by centrifugation in a Sorvall HB4 rotor (10Krpm, 10 minutes, 4°C). The supernatant is removed adjusted to 20mM EDTA, 0.5% SDS and added to ten

cell volumes of 0.1M NaCl, 0.2% SDS in TE. The RNA is deproteinised as in section 2.4.1, ethanol precipitated and stored as for ovary RNA.

2.12.3 Other tissue RNA (Feramisco et al 1982)

Tissue fragments are frozen in liquid N₂ and then powdered in dry ice before homogenisation in guanidinium mix (4M guanidinium isothiocyanate, 10mM EDTA 50mM Tris-Cl pH 7.6, 10% (w/v) sarkosyl, 1% (v/v) 2-mercaptoethanol). The mixture is brought to 60°C and is passed through a 19 gauge needle until the viscosity of the solution is reduced. An equal volume of phenol is added and the solution again passed through the needle. After cooling the mixture on ice, 0.1 volume of 0.1M sodium acetate pH 5.2, 10mM Tris-Cl pH 7.4, 1mM EDTA and an equal volume of CHCl₃/isoamyl alcohol (24:1) is added then mixed well. The aqueous phase is recovered by centrifugation in a Sorvall HB4 rotor (10Krpm, 10 minutes 4°C) and is re-extracted once with an equal volume of phenol/CHCl₃/isoamyl alcohol (25:24:1) and twice with an equal volume of CHCl₃/isoamyl alcohol (24:1). Two volumes of ethanol are added to the supernatant and after storing at -20°C for 2hr the RNA is recovered by centrifugation (10Krpm, 20 minutes, 4°C, HB4) and the pellet redissolved to the original volume in 0.1M Tris-Cl pH 7.4, 50mM NaCl, 10mM EDTA, 0.2% (w/v) SDS. Proteinase K is added to 200µg/ml and the mixture incubated at 37°C for 2hr before heating to 60°C. 0.5 volumes of phenol are added and mixed at 60°C for 10 minutes then cooled on ice and 0.5 volumes of CHCl₃/isoamyl alcohol (24:1) is added and mixed. The aqueous phase is recovered by centrifugation and the extraction repeated once more. The aqueous phase is extracted twice with CHCl₃/isoamyl alcohol (24:1) and the RNA is ethanol precipitated and stored as for ovary RNA.

2.12.4 Staged oocyte and embryo RNA

Total RNA is prepared from staged oocytes and embryos by a modification of the previous method. Ten oocytes or embryos are vortexed hard in 100µl of guanidinium mix, then 100µl phenol and 100µl 0.1M sodium acetate pH 5.2, 10mM Tris-Cl pH 7.4, 1mM EDTA are added and vortexing continued until the mixture is homogenous. The mixture is spun in a microfuge for 5 minutes and the supernatant removed and re-extracted with 200µl phenol/CHCl₃/isoamyl alcohol (25:24:1). The aqueous phase is precipitated with ethanol and the pellet redissolved in 200µl of ddH₂O. To remove contaminating polysaccharides 200µl

8M LiCl is added, left overnight at -20°C and the RNA recovered by spinning in a microfuge for 10 minutes at 4°C . The RNA pellet is redissolved in $20\mu\text{l}$ ddH₂O and stored at -70°C .

2.13 Poly (A)⁺ RNA Preparation (Aviv and Leder 1972)

Poly (A)⁺ RNA is selected from total RNA by affinity chromatography on oligo (dT)-cellulose columns. The column is equilibrated in sterile loading buffer (20mM Tris-Cl pH 7.6, 0.5M NaCl, 1mM EDTA, 0.1% (w/v) SDS). RNA in ddH₂O is heated to 65°C for 5 minutes, cooled on ice and an equal volume of 2x loading buffer is added then the RNA is applied to the column. The flow through is collected, heated to 65°C , cooled and reapplied to the column. Bound RNA is washed with 10 column volumes of loading buffer then eluted with 2-3 column volumes of sterile elution buffer (10mM Tris-Cl pH 7.5, 1mM EDTA, 0.05% (w/v) SDS). The column is re-equilibrated in loading buffer and the bound RNA put through a second cycle of chromatography after adjusting the bound RNA to loading buffer conditions. The poly (A)⁺ RNA is ethanol precipitated and then stored as for ovary RNA. Oligo (dT)-cellulose columns are regenerated by washing in 0.1M NaOH, ddH₂O then loading buffer checking the pH of the column effluent is less than 8.0

2.14 Poly(A) Content of RNA (Rosbash and Ford 1974)

To measure the poly(A) content of RNA a standard ³H poly(U) hybridisation assay was performed. Dilutions of RNA (in $100\mu\text{l}$ ddH₂O) were added to $400\mu\text{l}$ annealing buffer (50mM Tris-Cl pH 7.0, 0.3M NaCl, 1mM EDTA) containing excess ³H poly(U) ($2-5\mu\text{l}$ of ³H poly(U) specific activity 2×10^6 cpm/ μg). Annealing was allowed to take place at 45°C for 30 minutes then the annealed RNA was cooled on ice for 5 minutes and incubated with $10\mu\text{g}$ RNase A for 20 minutes. Ribonuclease resistant material was ctab precipitated, counted and the poly(A) content calculated as in Appendix 1.

2.15 Labeling of Nucleic Acids

2.15.1 End labeling and filling in of restriction sites (Maniatis et al 1982)

For end labelling of restriction sites the following simple method is employed. After 2hr of restriction digestion of DNA $10\mu\text{Ci}$ α ³²P dNTP (complementary to the protruding 5' termini), 2 U klenow are added directly to

the reaction and incubated at room temperature for 30 minutes. The end labelled DNA is precipitated twice from ammonium acetate/ethanol to remove unincorporated nucleotides.

Filling in of recessed 3' termini is carried out as above except 10 μ Ci of the appropriate 3 H dNTPs are used and after 30 minutes incubation all four dNTPs are added to 0.25mM then incubated for a further 30 minutes. Successful filling in is monitored by ctab precipitation. Such filled in fragments can be used in blunt end ligations.

2.15.2 Kinasing of Linkers (Maniatis et al 1982)

Commercially available linkers, were resuspended at 0.5 μ g/ μ l in TE and stored at -20° C. 4 μ l linkers, 20 μ Ci α^{32} P ATP, 1 μ l 10x linker kinase buffer, 10U T4 polynucleotide kinase and ddH₂O (to 10 μ l) were incubated at 37° C for 15 minutes. Then 1 μ l 10x linker kinase buffer, 1 μ l 10mM ATP, 10U T4 polynucleotide kinase and ddH₂O (to 20 μ l) were added and incubated at 37° C for a further 30 minutes before storing at -20° C.

The kinased linkers were checked they could be ligated as follows. 5 μ l kinased linkers were self ligated as in section 2.6 but at 4° C for 4hr then heated to 65° C for 10 minutes to inactivate the ligase, 5 μ l of ligated linkers were removed and digested with 10U of the appropriate restriction enzyme as in section 2.4.5. Finally samples of the kinased, ligated and restricted linkers were electrophoresed through a 10% polyacrylamide gel prepared as follows. 16.7ml 30% acrylamide stock, 2.5ml 10x TBE, 100 μ l 25%(w/v) APS, 100 μ l TEMED and ddH₂O (to 50ml) were mixed and allowed to polymerise in a vertical gel apparatus. 2 μ l 50% glycerol, 0.2% (w/v) bromophenol blue were added to the samples before loading onto the gel. The gel was run in 0.5x TBE until the dye had run half way. The gel is then dried down and exposed to X-ray film.

2.15.3 Kinasing of RNA (Donis-Keller et al 1977)

RNA is first treated as follows to increase the number of 5'ends available to accept a labeled phosphate. 2.5 μ g RNA is taken up in 50 μ l 0.1M NaOH and left on ice for 1hr before neutralising with 5 μ l 1M Tris-Cl pH 8.0, 5 μ l 1M HCl. The RNA is ethanol precipitated, redissolved in 25 μ l 10mM Tris-Cl pH 7.4 1mM spermidine, 0.1mM EDTA and heated to 50° C for 3 minutes before chilling on ice. 10 μ l RNA (1 μ g) is added to 15 μ l 2x RNA kinase buffer (100mM Tris-Cl pH

9.0, 20mM MgCl₂, 10mM DTT 5µl α ³²P ATP (10mCi/ml 3000Ci/mmol and 4U T4 polynucleotide kinase. The reaction is incubated for 1 hr at 37°C, a further 4U T4 polynucleotide kinase added and incubated for one more hour (monitoring incorporation by ctab precipitation). The labeled RNA is ethanol precipitated, redissolved in 50µl ddH₂O and unincorporated nucleotides removed by passing over a 15cm x 0.4cm Sephadex G-75 gel filtration column in ddH₂O. The peak fractions are pooled and used as a hybridisation probe, heating to 65°C for 3 minutes and cooling on ice before use.

2.15.4 Nick translation (Rigby et al 1977)

10µg plasmid DNA (in 10µl ddH₂O) plus 50µl DNase I mix are incubated at 37°C for 20 minutes then phenol extracted as in section 2.4.1. The supernatant is ethanol precipitated and the DNA redissolved at 0.1µg/µl in ddH₂O. For DNA fragments the optimum DNase I treatment must be determined empirically. Successful nicking can be checked by running on agarose gel or by a trial labeling reaction.

Nicked DNA is labeled as follows: 50µl Pol I mix-C, 5µl α ³²P dCTP (10mCi/ml, 3000Ci/mmol), 200ng DNA and 2U *E.coli* DNA polymerase I are mixed and incubated at 16°C for 90 minutes. The reaction is monitored by ctab precipitation and if necessary the incubation continued with the addition of more DNA polymerase. Unincorporated nucleotides are removed either by gel filtration chromatography over Sephadex G-75 (see 2.15.3) or by two precipitations from ammonium acetate/ethanol with the addition of 5µg *E.coli* tRNA. The labeled DNA is redissolved in 100µl ddH₂O, boiled for 10 minutes then cooled on ice before use.

2.15.5 Oligolabeling (Feinberg and Vogelstein 1983, 1984)

Highly labelled insert probes with a specific activity of 5x10⁸ – 2x10⁹ cpm/µg can be generated by this procedure. 5-10µg plasmid or phage DNA is cleaved with restriction endonucleases and the fragments separated on one lane of a 0.8-1.5% low melting point agarose gel. The desired fragment is located by uv visualisation of the stained gel and excised cleanly. 3ml ddH₂O/g is added to the gel fragment and then boiled for 7 minutes to denature the DNA. 50µl aliquots are then stored at -20°C boiling for 3 minutes before use.

The reaction is carried out at room temperature by adding in the following

order: ddH₂O (to 50μl), 10μl OLB, 2μl 10mg/ml BSA (pentax fraction V), 30μl DNA 2–5μl α ³²P dCTP (10mCi/ml, 3000Ci/mmol) and klenow (40U home made stock).

After 5–6hr the reaction is normally completed with 70–100% of the label getting incorporated. The reaction is stopped by heating to 65°C for 5 minutes. An equal volume of ddH₂O is added and the labeled DNA precipitated twice from ammonium acetate/ethanol (using 5μg carrier *E.coli* tRNA) to remove unincorporated nucleotides. The DNA is redissolved in 100μl ddH₂O and boiled for 10 minutes then chilled on ice before use.

2.15.6 M13 single stranded probes (Hu and Messing 1982)

Single stranded M13 probes are generated by using the M13 universal probe primer pHM235 (NEB, 5'-GAAATTGTTATCC-3'). 5μl template DNA (see section 2.19.3), 1μl 10x medium salt buffer, 1μl primer and 3μl ddH₂O are drawn into a capillary. The ends are sealed and the sample boiled for 3 minutes then allowed to cool slowly to room temperature. The annealed primer/template are added to 3μl ddH₂O, 3μl α ³²P α dCTP (10mCi/ml, 3000Ci/mmol), 3μl 0.5mM dNTP-C (dATP, dTTP, dGTP), 1μl klenow (5 units) and the reaction incubated at room temperature for 90 minutes then terminated by addition of 2μl 0.4M EDTA. Unincorporated nucleotides are removed as for nick translation. The sample is not denatured before using.

2.15.7 T7 RNA polymerase probes (Melton et al 1984, Krieg and Melton 1986)

DNA cloned into the pGEM 1/2 vectors (Promega Biotech) is suitable for transcribing T7 or SP6 RNA polymerase transcripts. 5μg of plasmid DNA is linearised with the appropriate restriction enzyme then phenol extracted (2.4.1) and redissolved in ddH₂O at 0.25μg/μl. At room temperature 2μl 10x transcription buffer, 2μl 10x transcription rNTPs, 1μl RNasin (20 units, Amersham), 2μl linearised plasmid DNA, 3μl α ³²P UTP (10mCi/ml, 410Ci/mmol) or 30μCi ³H UTP (13.6Ci/mmol dried down and redissolved in 3μl ddH₂O), 1μl T7 RNA polymerase (10U, Boehringer), 9μl ddH₂O are mixed and then incubated at 37°C for 1hr. 1μl of RNase free DNase I (0.5μg) is added and incubated for 15 minutes at 37°C then the reaction phenol extracted and the RNA ethanol precipitated. This procedure generates RNA of high specific activity (>2x10⁸cpm/μg) or large amounts of ³H labelled RNA (>1μg).

2.16 Transfer of Nucleic Acids to Membrane Filters

2.16.1 Southern blots

This procedure is essentially as described by Southern (1975), Wahl et al (1979) and Maniatis et al (1982). Genomic DNA gels are first treated before transfer. The gel is soaked in 0.25 M HCl for 15 minutes to depurinate the DNA, then in 0.5M NaOH, 1.5M NaCl for 1 hr to denature the DNA and then in 1M Tris-Cl pH 8.0, 1.5M NaCl for 1hr to neutralize the gel. DNA is transferred to nitrocellulose (Schleicher and Schull) by capillary action for 18 hr using 20x SSPE as transfer buffer in a blotting sandwich as previously described. The blot is dismantled and the nitrocellulose filter air dried and then baked at 80°C for 2hr before use. Gels other than genomic gels are transferred similarly except that the gels are neutralised in 1M ammonium acetate, 0.02M NaOH and the gel transferred without a reservoir of transfer buffer. This can be modified to bidirectionally transfer the gel to two filters simultaneously.

2.16.2 Northern blots (Thomas 1983)

RNA gels are transferred by capillary action in a similar manner to that of southern blots. After running the gel the marker track is cut off and stained (it is important not to stain RNA gels before transfer). The gel is rinsed briefly in ddH₂O then the transfer set up using a nylon membrane (Hybond-N, Amersham) instead of nitrocellulose. After 18hr the blot is dismantled, the filter air dried, exposed to a uv transilluminator (254 nm, RNA side down) for 5 minutes and then baked at 80°C for 1hr before use.

2.16.3 Phage lifts (Benton and Davis 1977)

DNA from phage plaques can be transferred to nitrocellulose filters by the following procedure. Phage are plated onto dry BBL bottom plates using 0.7% agarose in 10mM MgSO₄ instead of BBL top agar and allowed to grow for 6-12hr. after cooling at 4°C for 1hr a single sheet of nitrocellulose, cut to the dimensions of the plate, is laid on the agarose surface for 1 minute marking with needle holes for subsequent orientation. The nitrocellulose was removed and laid, plaque side up, on blotting paper soaked in 0.5M NaOH, 1.5M NaCl for 5 minutes, then immersed in 0.5M Tris-Cl pH 8.0, 1.5M NaCl for 5 minutes, then 2x SSPE for 5 minutes and finally air dried and baked at 80°C for 2hr. Multiple filters can be prepared from each plate in this manner.

2.16.4 Dot blots

Sometimes in testing hybridisation conditions it is useful to directly spot nucleic acid onto nitrocellulose filters presoaked in 20x SSPE. RNA is first heated to 65°C for 5 minutes and cooled on ice before applying in a small volume (<1.5µl). DNA is first denatured by heating to 60°C in 0.1M NaOH for 1hr then neutralising with 0.1M HCl before spotting on to the filter. Filters are air dried then baked at 80°C before use.

2.17 Hybridisation of membrane Filters with Radioactive Probes

Filters were first prehybridised in either high or low stringency conditions shaking for 4–12hr (0.2ml/cm²) at 37°C. Hybridisation was then carried out for 18–20hr in the same solution (~0.1ml/cm²) but with the addition of denatured radiolabeled probe. High stringency conditions were 50% formamide, 5x SSPE, 5x Denhardtts, 0.1% (w/v) SDS and 100µg/ml denatured herring sperm DNA. Low stringency conditions were similar but using 35–40% formamide and 6x SSPE. Poly (U) (Sigma) was added to 5µg/ml if the probe contained a poly (A) tail. Also if a nylon filter was used then 0.5% (w/v) SDS was used.

Filters were then washed four times at room temperature in 2x SSPE, 0.1% (w/v) SDS (10 minutes each wash). The filters were then washed for 1hr at the required final stringency. This stringency depended on the nature of the probe. Homologous probes had a final stringency of 0.2x SSPE, 0.1% (w/v) SDS at 65°C whereas for heterologous probes the stringency varied from 2x SSPE, 0.1% (w/v) SDS at 37°C to 0.5x SSPE, 0.1% (w/v) SDS at 50°C. If bad background was a problem it was found that treatment of the filter with proteinase K at 25µg/ml at 37°C (or 65°C for bad background) for 1hr would solve the problem. Filters were then air dried and exposed to preflashed X-ray film. Filters were reused by adding them to boiling ddH₂O which was allowed to cool to room temperature.

2.18 Preparation of a Previtellogenic Ovary cDNA Library

The general methods for this technique have recently been described by Huynh *et al* (1985) and Watson and Jackson (1985). I have modified the methods to my own requirements as described below.

2.18.1 First strand synthesis

10µg poly(A)⁺ RNA (in 10µl ddH₂O) is denatured at 65°C for 5 minutes and cooled rapidly on ice. 100µl cDNA mix-C, 2µl RNasin (Amersham), 6µl 20mM dCTP, 20µCi α ³²P dCTP (3000Ci/mmol) or 20µCi ³H dCTP (26.5Ci/mmol), 2µl AMV reverse transcriptase (BioRad, 40 units) are added, incubated at 42°C for 2hr then 20U reverse transcriptase added and the incubation continued for a further 2hr. The reaction is then phenol extracted (2.4.1) and precipitated twice from ammonium acetate/ethanol. It is calculated that 30–50% of the RNA input mass was converted to cDNA (see Appendix 2). The size distribution is checked by running ³²P labelled cDNA on an alkaline agarose gel (McDonnell 1977). This is a normal DNA type gel but is poured in 50mM NaCl, 1mM EDTA and run in 30mM NaOH, 1mM EDTA. The gel is dried down and exposed to X-ray film. This shows the average size of the cDNA is about 2.0kb (see figure 4.1).

2.18.2 Second strand synthesis (Gubler and Hoffman 1983)

Optimum second strand synthesis is achieved by processing <200ng of cDNA/RNA hybrid in 100µl reactions as follows: 200ng cDNA/RNA hybrid, 50µl 2x second strand salts, 5.0µl 3.0mM NAD (Sigma), 10µl 1mM dNTPs, 1µl 5µg/µl BSA (BRL), 0.8µl RNase H (IU, Pharmacia), 1µl *E.coli* DNA ligase (0.4µg, Pharmacia) 20µCi α ³²P dCTP (3000Ci/mmol) and ddH₂O to 100µl are mixed and kept on ice for 4 minutes. 0.4µl *E.Coli* DNA polymerase I (159U/µl, home made stock) is added and incubated at 12°C for 1hr then at 23°C for 2hr. A further 50U DNA polymerase I is added and the reaction incubated 23°C for 2hr more before phenol extracting and precipitating from ammonium acetate/ethanol. It is calculated 90–100% of input hybrid is converted to double stranded cDNA (Appendix 2). The size distribution is checked by running a sample on a DNA agarose gel, drying down and exposing to X-ray film. This shows the average size of the double stranded cDNA is at least 2.0kb (figure 4.1).

2.18.3 EcoRI methylation and end repair

1–2µg of double stranded cDNA are methylated using *EcoRI* methylase (NEB). 106.5µl cDNA in ddH₂O, 30µl 5x methylase buffer, 12µl 5µg/µl BSA (BRL), 0.4µl 32mM SAM (in 5mM H₂SO₄, 10% ethanol, NEB). 2µl *EcoRI* methylase (40U) are mixed and incubated at 37°C for 2hr. The reaction is heated at 65°C for 10 minutes then phenol extracted and precipitated. The cDNA can then be tested

to see if it is protected to *EcoRI* digestion.

The methylated cDNA is end repaired as follows: the cDNA is redissolved in 21 μ l ddH₂O then 5 μ l 1mM dNTPs, 3 μ l 10x T4 DNA ligase buffer and 1 μ l Klenow (5U) are added and incubated at room temperature for 30 minutes. Following phenol extraction the cDNA is precipitated from ethanol.

2.18.4 *EcoRI* linker ligation and size fractionation

1 μ g methylated, end repaired cDNA is redissolved in 10 μ l ddH₂O and added to 5 μ l phosphorylated *EcoRI* linkers (0.5 μ g, prepared as in 2.15.2), 2 μ l 10x T4 DNA ligase buffer, 1 μ l 20mM ATP, 2 μ l T4 DNA ligase (6U, Pharmacia) and incubated at 15°C for 2 days. After terminating the reaction at 65°C for 10 minutes the cDNA is precipitated once from ammonium acetate/ethanol and digested with 240U *EcoRI* (120U/ μ l, Boehringer) for 2hr at 37°C in a 50 μ l reaction. The reaction is then phenol extracted, precipitated from ammonium acetate/ethanol and redissolved in 20 μ l ddH₂O.

Removal of unincorporated linkers and size fractionation of the cDNA is carried out by gel filtration chromatography essentially as described by Huynh *et al* (1985). A 1ml bio-gel A 50m, 100–200 mesh (BioRad) column is poured in a serological pipette (250 x 3mm). The column is run in 0.1M NaCl, 10mM Tris–Cl pH 7.4, 1mM EDTA and is first calibrated by running end labelled DNA markers through the column collecting 4 drop (100 μ l) fractions. After extensive washing of the column 2 μ l 50% glycerol, 0.2% (w/v) bromophenol blue is added to the cDNA and applied to the column. Fractions containing cDNA greater than 500–700bp (Figure 4.1) are pooled and stored at –20°C for future cloning. Size fractionation can be checked by running a sample on a DNA gel, drying down and exposing to X-ray film.

2.18.5 Cloning of cDNA into bacteriophage lambda NM1149

A 50 μ g *EcoRI* restricted stock of NM1149 DNA is prepared such that the titre of restricted DNA drops by at least 10³ on *in vitro* packaging and also that no recombinant phages are generated by self ligation of the vector (checked by plating on NM514). 20ng cDNA is co-precipitated with 1 μ g restricted NM1149 DNA and redissolved in 9 μ l ddH₂O. 1.2 μ l 10x T4 DNA ligase buffer, 0.6 μ l 20mM rATP, 1.2 μ l T4 DNA ligase (4U) are added and incubated at 15°C for 18hr.

The ligated lambda DNA is packaged *in vitro* essentially as described by Hohn and Murray (1977). The sonicated extract and the freeze thaw lysate were gifts from Dr. N.Hunter. DNA is packaged as follows: 7µl packaging buffer A is mixed with lambda DNA (up to 0.5µg in volume of 3µl), 1µl packaging buffer M1, 6µl sonicated extract and 10µl freeze thaw lysate then incubated at 20°C for 1hr. Maximum packaging efficiency ($5 \times 10^7 - 1 \times 10^8$ pfu/µg) is obtained by adding 10µl freeze thaw lysate, 1µl packaging buffer M1 and incubating for a further 2hr. The packaging is terminated by adding 0.5ml phage buffer and a drop of CHCl₃. The packaged DNA is stored at 4°C.

The percentage of recombinant phages is determined by plating dilutions on NM514 (only recombinant clear plaques grow) and NM259 (both recombinant clear plaques and non-recombinant turbid plaques grow). Using these procedures a library of approximately 5×10^4 recombinant phages was actually produced (with 1% total recombinants) which was used successfully.

2.18.6 Screening of phage cDNA libraries

Approximately 1.5×10^4 phages are used to infect 0.8ml NM514 plating cells (see 2.10.1) and plated onto warm 12x12cm BBL bottom agar plates (using 8ml of 0.7% (w/v) agarose in 10mM MgSO₄). The plaques are left to grow until the plaques are just touching each other (7-10hr). Plaque lifts are performed as in 2.16.3 and the filters hybridised as in 2.17. Positive plaques are picked with the blunt end of a pasteur pipette into 1ml phage buffer and then are rescreened until a positive signal corresponds to an individual well isolated plaque.

After use the library is recovered by overlaying each plate with 25ml L broth plus 10mM MgSO₄ and leaving overnight at 4°C. The liquid is taken off, debris spun out and the supernatant stored at 4°C over 2ml CHCl₃. The amplified library has to be titred before use, typically it has 6.5×10^5 pfu/µl.

2.19 Dideoxynucleotide DNA Sequencing

2.19.1 General

DNA sequencing is performed by the dideoxynucleotide chain termination method as described by Sanger *et al* (1977), Sanger and Coulson (1978), Biggen (1983) and Messing (1983). For convenience all sequencing stock solutions are listed below.

<u>BCIG(X-gal)</u>	20mg/ml 5-bromo-4-chloro-3-indoyl- β galactoside in dimethylformamide
<u>IPTG</u>	20mg/ml Isopropyl- β -D-thiogalactoside in ddH ₂ O
<u>M13 phage buffer</u>	20mM Tris-Cl pH 7.5, 20mM NaCl, 1mM EDTA
<u>TM buffer</u>	100mM Tris-Cl pH 8.5, 50mM MgCl ₂
<u>Sequencing TE</u>	10mM Tris-Cl pH 8.5, 0.1mM EDTA
<u>M13 universal 17mer sequencing primer</u>	5'-GTAAAACGACGGCCAGT-3' 2.5 μ g lyophilite(NEB) resuspended in 1000 μ l sequencing TE
<u>Oligonucleotide sequencing primer</u>	5'-TGGGAACGTGACACCGC-3' a gift from M.Bownes, at 2.5ng/ μ l in ddH ₂ O
<u>³⁵S-dATP</u>	deoxyadenosine 5' α ³⁵ S thiotriphosphate 8.0mCi/ml, 410 Ci/mmol (Amersham)
<u>dNTP stocks</u>	10mM of each dNTP (Sigma) in sequencing TE
<u>ddNTP stocks</u>	10mM of each ddNTP (Boehringer) in sequencing TE
<u>ddNTP termination mixes</u>	The composition of each ddNTP termination mix is given below, all values are in microlitres.

Termination mix	ddTTP	ddCTP	ddGTP	ddATP
0.5mM dTTP	25	500	500	500
0.5mM dCTP	500	25	500	500
0.5mM dGTP	500	500	25	500
10mM ddTTP	40	-	-	-
10mM ddCTP	-	6	-	-
10mM ddGTP	-	-	16	-
10mM ddATP	-	-	-	0.6
Sequencing TE	935	969	959	500

<u>Sequencing chase</u>	0.25mM dGTP, 0.25mM dCTP, 0.25mM, dATP, 0.25mM dTTP in ddH ₂ O
<u>Formamide dyes</u>	10ml deionised formamide, 0.01g xylene cyanol FF, 0.01g bromophenol blue, 0.25 ml 0.4M EDTA

<u>10x TBE</u>	0.9M Tris-borate pH 8.3, 25mM EDTA
<u>40% Acrylamide</u>	380g Acrylamide, 20g bis-acrylamide made to 1 litre with ddH ₂ O, deionised with 20g MB-1 (BDH) and filtered
<u>0.5x TBE gel mix</u>	150ml 40% acrylamide, 50ml 10xTBE, 460g urea made to 1 litre with ddH ₂ O and filtered
<u>2.5x TBE gel mix</u>	150ml 40% acrylamide, 250ml 10xTBE, 460g Urea, 50g Sucrose, 50g bromophenol blue made to 1 litre with ddH ₂ O and filtered
<u>Gel fix</u>	250ml glacial acetic acid, 250ml methanol made to 2.5 litre with ddH ₂ O.

2.19.2 Cloning into M13mp18 and M13mp19

An insert DNA preparation of the fragment to be sequenced is first prepared as in 2.7.1. To sequence each fragment entirely it is necessary to clone sub-fragments of less than 350bp into the M13 vectors. This is achieved by restriction of the fragment with a battery of 4bp and 6bp recognition restriction enzymes then cloning into the appropriate site in the polylinker of the vector. Ligation of the fragment to vector is carried out as in 2.6 and transformation of NM522 is carried out using lacZ colour selection as described in 2.5 but with the addition of 50µl of an overnight culture of NM522 to each plate and no addition of L broth and incubation before plating. Recombinant white plaques are picked into 0.5ml M13 phage buffer or used to produce templates immediately.

2.19.3 M13 template preparation

1ml of a fresh NM522 overnight culture is diluted into 100ml L broth and 2ml aliquoted into 1.5x15cm tubes. White M13 plaques are picked into the tubes and grown with vigorous shaking at 37°C for 5-6hr. 1.5ml are transferred to microfuge tubes and spun for 5 minutes. 1.2ml of the supernatants are transferred to tubes containing 150µl 2.5M NaCl, 20% (w/v) PEG-6000, mixed and stored at 4°C for 30 minutes. The precipitated phages are spun for 10 minutes and the supernatants removed. Tubes are then respun and the remaining supernatants removed. The phage pellets are resuspended in 100µl of 0.1M NaCl, sequencing TE and 50µl phenol added, vortexed extensively then spun. The supernatants are removed and precipitated with

sodium acetate/ethanol. DNA pellets are recovered by centrifugation, washed in 70% ethanol, air dried and redissolved in 30µl sequencing TE.

2.19.4 Sequencing reactions

Primer mix/template: 1µl M13 sequencing primer or 1µl oligonucleotide sequencing primer, 1µl TM buffer, 3µl ddH₂O

Annealing mix/template: 5µl primer mix, 5µl template DNA

Enzyme mix/template: 0.62µl ³⁵S thio dATP, 1µl 80mM Tris-Cl pH 8.5, 1µl 80mM DTT, 5.1µl ddH₂O, 0.31µl Klenow (5U/µl Boehringer), made up on ice just before use.

Templates are annealed in annealing mix at 60°C for 1hr then allowed to room temperature for 15 minutes. 2µl of the annealed template is added to 4 wells of a round bottomed siliconised microtitre plate. 2µl of ddTTP termination mix is added to one well, similarly 2µl of the remaining ddNTP termination mixes are added to each of the other wells. 2µl of enzyme mix is added to each well and the plate tapped gently to mix the components then incubated at room temperature for 25 minutes. Finally 2µl of sequencing chase is added to each well and incubated at room temperature for a further 20 minutes before freezing at -20°C. Before loading onto a gel the samples are first denatured by adding 2µl of formamide dyes, floating the plate on a 80°C water bath for 10 minutes and cooling on ice.

2.19.5 Polyacrylamide gel electrophoresis

Two alternative gel systems are used to analyse the products of the sequencing reactions. The gradient gel system resolves bases 0-250 from the sequencing primer binding site whereas the straight gel can resolve bases 150-350 or more depending on the run time.

The gradient gel is prepared as follows: 70µl 25%(w/v) APS, 70µl TEMED and 35ml 0.5x TBE gel mix are mixed. 14µl 25%(w/v) APS, 14µl TEMED, and 7ml 2.5x TBE gel mix are mixed. 4ml of the polymerising 0.5x TBE gel mix are taken up in a 10ml pipette. a crude gradient is formed by introducing 3-4 air bubbles, then pouring it quickly into a glass gel former (40 x 20 x 0.035cm). the rest of the polymerising 0.5x TBE gel mix is added, a well former is inserted and the gel allowed to polymerise. Similarly a straight gel is poured as above

but it is made only with 0.5x TBE gel mix. (i.e. it does not have a gradient)

The gels are then assembled into an electrophoresis kit with 1x TBE gel buffer in both upper and lower gel tanks. If sharktooth combs (BRL) are used they are inserted now. The wells are flushed clear of urea and half the denatured sequencing reactions are loaded onto the gels in the order TCGA. Gradient gels are run at 25 watts constant power until the bromophenol blue has just ran off the bottom (2.5–3hr) and straight gels are run at 25 watts constant power for 6–9hr. The gels are dismantled such that they remain attached to one plate (non-siliconised) then transferred to a tray of gel fix for 15 minutes. Gels are transferred to prewetted blotting paper, covered in Saran Wrap and dried under vacuum at 80°C on a gel drier. Dried gels are placed directly in contact with X-ray film for 1–3 days before developing. Sequence information is read off the autoradiograph and compiled by version 2.0 of the programs of Staden (Staden 1980) and analysed by version 4.0 of the programs of the University of Wisconsin Genetic Computer Group (Devereux *et al* 1984).

2.20 Protein Methods

2.20.1 Laemmli polyacrylamide gels (Laemmli 1970)

A stock solution of 30% (w/v) acrylamide, 1% (w/v) N, N'-methylene-bis-acrylamide is deionised and stored at 4°C. Gels are cast between 25 x 20cm glass plates separated by 1mm perspex spacers. The separating gel (see below) is poured, overlaid with ddH₂O and allowed to polymerise for 2hr. The ddH₂O is poured off and the stacking gel poured then a comb inserted. After allowing to polymerise for 1hr the comb is removed and assembled into a vertical gel kit. 1x Laemmli buffer is used in top and bottom tanks. An equal volume of 2x protein sample buffer is added to the sample, boiled for 5 minutes then loaded onto the gel and run at 7mA for 18hr or at 20mA for 30 minutes then 50mA till completed. The gel is then either stained or immunoblotted (2.20.2). To stain gels they are first fixed in 50% (v/v) methanol, 10% (v/v) acetic acid then stained in 50% (v/v) methanol, 10% (v/v) acetic acid plus 0.1% (w/v) Coomassie brilliant blue R250. Gels are then destained in several changes of 5% (v/v) methanol, 7% (v/v) acetic acid. Markers for protein gel electrophoresis are obtained from BRL and are 200, 97.4, 68, 43, 25.7, 18.4, and 14.3KD.

Polyacrylamide gels (volume in ml):

	Separating		Stacking
	7.5%	10%	
0.75M Tris-Cl pH 8.8	15.0	15.0	-
0.625M Tris-Cl pH 6.8	-	-	2.0
30% Acrylamide Stock	7.5	10.0	1.67
ddH ₂ O	7.5	5.0	6.33
10% SDS	0.3	0.3	0.1
TEMED	0.015	0.015	0.005
10% APS	0.15	0.15	0.05

2.20.2 Immunoblotting

This method for transferring proteins from polyacrylamide gels to nitrocellulose is a modification of Towbin *et al* (1979). The gel is run as above then soaked for 30 minutes in transfer buffer (20% methanol, 25mM Tris, 192mM glycine pH 8.3). A gel sandwich is constructed in a BioRad transfer cassette as follows: all components are first prewetted in transfer buffer then layered on to the cassette in this order, removing bubbles between layers; a Scotch brite pad, blotting paper cut to same size as the gel, the gel, nitrocellulose (0.45µm Schleicher and Schull) cut 10% large than the gel, more blotting paper and finally another Scotch brite pad. The cassette is closed and then added to a BioRad electrotransfer kit filled with 3 litres of transfer buffer such that the nitrocellulose faces the anode. Proteins are then transferred at 250mA overnight or at 600mA for 4hr. After transfer the nitrocellulose sheet can be stained with 0.5% Ponceau S (Sigma) in 3% TCA and washed in ddH₂O to visualise the transferred proteins.

Nitrocellulose strips are blocked in MTS (5% (w/v) skimmed milk 50mM Tris-Cl pH 7.8, 150mM NaCl) for 2hr at room temperature then probed with the appropriate dilution of rabbit antiserum (typically 1:200) in MTS for 2-4hr at room temperature or overnight at 4°C with gentle shaking. The strips are washed 4 times, 10 minutes each wash, in TS (50mM Tris-Cl pH 7.8, 150mM NaCl). To visualise the antibody-antigen interactions the filter is next incubated

with goat anti-rabbit immunoglobulins (1gG fraction) conjugated with horseradish peroxidase (Sigma) at a 1:500 dilution in MTS for 2hr at room temperature. The filters are then washed again extensively in TS and developed with 10mM Imidazole pH 7.4, 0.25mg/ml dianisidine hydrochloride (Sigma), 0.37% (v/v) hydrogen peroxide (BDH) until the bands are visible. Filters are washed in plenty ddH₂O and stored away from the light until photographed.

2.20.3 Preparation of β galactosidase-fusion proteins

DNA fragments containing protein coding sequences are cloned into the appropriate restriction site of one of the pEX series of vectors (Stanley and Luzio 1984) such that the coding sequence to be expressed is in the correct reading frame fused onto the 3' end of β galactosidase. Plasmids are then transformed (2.5) into the *E.coli* strain pop 2136 and grown at 30°C on plates containing ampicillin. Mini-preparations of plasmid DNA (2.9.1) are done and analysed by restriction digestion. Cells containing the appropriate plasmid are used to make proteins as follows: Colonies are grown up at 30°C overnight in L broth plus ampicillin then a 1:100 dilution is made and grown up for 6hr at 30°C. This culture is then transferred to a 42°C water bath and shaken vigorously for 90 minutes. 1.5ml of the culture is spun down in a microfuge tube; resuspended in 75 μ l PBS and 75 μ l 2x protein sample buffer is added then vortexed hard. After boiling for 5 minutes the cell debris is spun down and the supernatant removed and stored at -70°C. 5-10 μ l are then used for one track of a protein gel, boiling for 3 minutes before loading. This procedure is simply modified for cells without plasmids (no antibiotics), no induction of fusion proteins (no heat shock) or large preparation of proteins (increase volumes appropriately).

2.20.4 Preparation of *Xenopus* ovary proteins

This method is modified from Darnbrough and Ford (1981). A piece of ovary tissue containing about 1000-5000 oocytes is homogenised in 1ml of protein extraction buffer on ice. the homogenate is spun in a Sorvall HB4 rotor for 20 minutes (10Krpm, 4°C) then the supernatant is removed, avoiding the fat layer and stored at -70°C. Before use an equal volume of protein sample buffer is added then boiled for 5 minutes. About 2-5 oocyte equivalents are loaded on one track of a gel.

2.20.5 Generation of rabbit anti-fusion antiserum

A preparative gel of the required fusion protein was prepared by running 0.75–1.5ml of the induced *E.coli* cell extract on the whole width of a protein gel. The fusion protein was localised by staining two small strips from the gel and the protein band was cut out, then homogenised in 3–5 volumes of PBS. One third of this homogenate (about 100µg protein) was used for one rabbit on each day of injection.

Three young half lop rabbits (Redhill Farm, Surrey) were subcutaneously injected at 6 sites on the back as follows: Before any injection about 20ml blood was taken from the ear as a preimmune control. The first injection was with (1:1) homogenate/complete freunds adjuvant (Difco) emulsion. The second injection was with (1:1) homogenate/incomplete freunds adjuvant (Difco) emulsion 3 weeks later. The subsequent injections were with homogenate alone at approximately 2 week intervals. 20ml ear bleeds were taken after the 4th and 5th injection. All rabbits were bled from the heart 16 weeks after the first injection.

The blood collected from the rabbits was allowed to clot overnight at 4°C then the serum was removed, clarified by centrifugation in a Sorvall HB4 (10Krpm, 5 minutes, 4°C) and stored at -20°C or purified further by ammonium sulphate precipitation as follows: 1/2 volume of saturated NH₄SO₄ pH 7.8 was added drop by drop to the serum and allowed to stir at room temperature for 3hr. the precipitate is collected by centrifugation (HB4, 10Krpm, 30 minutes, room temperature) and then redissolved in 1/2 volume saline. The purified serum is then dialysed extensively against borate buffered saline at 4°C. Cryoglobulins were spun out (HB4, 10Krpm, 30 minutes, 4°C) and the supernatant removed then stored at -20°C.

2.20.6 Affinity purification of antiserum

Affinity purification of antiserum against antigen can be done by a simple modification to the immunoblotting technique (2.20.2) as follows. The antigen is run on 1/3 width of a protein gel, transferred and stained with Ponceau S. The band containing (or thought to contain) the antigen is cut out, blocked and probed with 1:20 dilution of antiserum for 2hr at room temperature. After extensive washing in TS the bound antibody is eluted from the nitrocellulose with 1ml 0.2M glycine-HCl pH 2.2 for 5 minutes at 4°C. 150µl of 2M Tris base is

added to neutralize the glycine. An equal volume of 2x TS is added, the supernatant removed and stored at -20°C .

CHAPTER 3

IDENTIFICATION OF ONCOGENE RELATED SEQUENCES IN XENOPUS

3.1 Introduction

As discussed in Chapter 1 cellular oncogenes have been detected in a wide variety of organisms from yeast to man. An analysis of cellular oncogenes in different organisms (exploiting the special potential of each) may lead to clues about their function in normal cells which cannot be obtained in other ways and hence may help us to understand their roles in carcinogenesis. This kind of analysis has been dramatically illustrated by the biochemical and genetic analysis of the yeast *ras* gene products which appear to be G proteins that act as GTP dependent positive regulators of adenylate cyclase (Toda *et al* 1985). Also the *Drosophila src* gene product may play a role in the early development of neural tissue and smooth muscle (Simon *et al* 1985). Although studying cellular oncogene functions in different organisms sometimes gives misleading results (for example adenylate cyclase is unaffected by *ras* proteins in mammalian systems, Beckner, Hattori and Shih 1985) the general mechanisms of how these proteins act and how they interact with other proteins is still a valid scientific approach in the dissection of how normal cells behave.

Initially I wanted to see if it was possible to detect cellular oncogene related sequences in *Xenopus* DNA using a variety of oncogene probes from viruses and vertebrates. As a vertebrate with an established embryology, *Xenopus* is an ideal organism to study cellular oncogene function during development. Large numbers of oocytes and embryos are readily available which are amenable to micromanipulation. Gene expression during *Xenopus* development is reasonably well understood (Chapter 1). Modern techniques such as *in situ* hybridisation to oocytes and embryos and the blocking of translation of specific mRNAs with antisense RNA in *Xenopus* are rapidly becoming commonplace (Melton 1985, Melton and Rebagliati 1986, Colman and Drummond 1986). Switching on of genes at specific developmental timepoints to assess their function will also soon be possible (Krieg and Melton 1985).

Closely related proteins that are encoded by small gene families are sometimes differentially expressed. For example the best characterised gene family is the globin family whose members are expressed at different times

during the development of an individual. Apparently these gene families have arisen by duplication of an ancestral gene. Studies on DNA content, number of chromosomes and immunological distances between albumins of different species of the genus *Xenopus* suggest a whole genome duplication has occurred in *X.laevis* and *X.borealis* about 30 million years ago (Bisbee *et al* 1977, Thiebaud and Fischberg 1977). *X.laevis* and *X.borealis* have 36 chromosomes whereas *X.tropicalis*, with 20 chromosomes and half the DNA content of *X.laevis* and *X.borealis*, probably is a modern representative of an original diploid ancestral form. Clues to the evolution of cellular oncogenes might therefore become apparent by studying various species of *Xenopus*.

Cellular oncogenes are known to be expressed in a variety of normal tissues (Chapter 1, Muller and Verma 1984) therefore it might be possible to detect expression of these sequences in *Xenopus* RNA using heterologous probes. If so this may provide further clues for cellular oncogene function and a starting point for the examination of cellular oncogene expression during development. In this thesis all RNA studies were performed using *X.borealis*. *X.borealis* was used instead of the more commonly used *X.laevis* for a number of reasons. Oncogene research is highly competitive hence it is likely that *X.laevis* would be chosen by others for study. So by using *X.borealis*, I would not directly repeat anyone else's work and also provide the basis for a comparative species analysis which may be informative and useful. In our hands *X.borealis* are easier to rear, can be kept at higher densities and the tadpoles are much larger than *X.laevis*. Eventually experiments could be designed using interspecies hybrids to follow the expression of specific molecules in the developing egg. In retrospect it may have been more sensible to use *X.tropicalis* since analysis can be difficult in *X.laevis* and *X.borealis* due to the increased gene number.

3.2 Detection of Oncogene Related Sequences in Xenopus DNA

A variety of oncogene probes from mammals, birds and their viruses (table 2.3) were obtained and used to probe southern blots of genomic DNA from *X.borealis*, *X.laevis* and *X.tropicalis* (also human and chicken DNA as positive controls when appropriate) to see if any related sequences could be detected. Initially probes were made from complete plasmid DNA, labeled by nick translation (Chapter 2). Routinely 50% of the label (50 μ Ci was used for each reaction) would be incorporated giving probes of specific activities between

$\sim 10^7 - 10^8$ cpm/ μ g. Since the vector DNA sequences can represent up to 90% of the labeled DNA the DNA fragment of interest has only incorporated $\sim 5\%$ of the label used in the reaction. Using these probes under standard high stringency hybridisation conditions (50% formamide, 5x SSPE, 37°C) and moderately high washing conditions (2-1x SSPE at 65°C) no homologies were detected in *Xenopus* DNA.

X.laevis has a haploid DNA content of 3.1×10^6 kbp which is similar to that of humans (Lewin 1980). To detect gene sequences that are present in roughly single copy it is therefore essential that the probes are labeled to as high a specific activity as possible since the molar ratio of each gene is very low compared with that of *Drosophila* (~ 20 x less) or yeast (~ 140 x less). Loading high amounts of DNA helps to increase sensitivity by increasing the mass of each DNA fragment to be detected. Sensitivity is also increased by long exposures of the X-ray film. Therefore to try and increase the specific activity of the fragment of interest insert DNA preparations were made as described in Chapter 2 and used in standard nick translation reactions. It was found, even after elutip-d column chromatography, that nick translation of insert DNA was extremely unreliable. Anywhere between 5-50% of the label would be incorporated representing specific activities of $\sim 10^7 - 2.0 \times 10^8$ cpm/ μ g. When these probes were used under standard high stringency hybridisation conditions and moderately high washing conditions any homologies in *Xenopus* DNA were detected inconsistently although bands were frequently detected in the positive control chicken DNA.

Oligolabeling was found to be the most consistent and reliable technique for producing probes of high specific activity (see Chapter 2). Routinely greater than 80% of the label would become incorporated into the probe giving specific activities of the fragment of $\sim 5 \times 10^8 - 1.5 \times 10^9$ cpm/ μ g (20-50ng of fragment DNA was used for each reaction). Even with oligolabeling detection of oncogene related sequences in *Xenopus* DNA under standard high stringency conditions and moderately high washing conditions was inconsistent using heterologous probes. When homologous probes were used on *Xenopus* DNA bands of homology were detected with relative ease, band intensities were strong, there were few non-specific bands and the exposures required were reasonably short (see following chapter for examples).

The heterologous oncogene probes probably have a high degree of

sequence divergence from the *Xenopus* sequences hence hybridisation rates and the stability of partially matched duplexes would be effected. Many parameters govern the formation and stability of hybrid duplexes in filter hybridisation (see Anderson and Young 1985 for discussion) but in general the use of high salt, reduced formamide concentrations and low temperatures aids the detection of diverged sequences. It must be borne in mind, however, that each probe used was unique - mismatching between duplexes formed may be localised leaving large regions of good pairing or it may be dispersed leaving the overall duplex unstable, the base compositions and probe lengths were different also the specific activities varied by roughly 2-3 fold. Routinely I used probes made by oligolabeling (specific activities of 5×10^8 - 1.5×10^9 cpm/ μ g DNA) and hybridisation conditions of 35-40% formamide, 6x SSPE at 37°C for at least 18hr (no advantage is gained by leaving the hybridisation longer due to renaturation of the probe) to try and detect oncogene related sequences in *Xenopus* DNA. Washing of filters was done differentially to try and preserve the *bona fide* hybrids over the inevitable non-specific hybrids. This was done by washing initially at low temperatures and high salt (37°C, 1x SSPE) and then washing at higher temperature and/or lower salt (e.g. 50°C, 1x SSPE then 50°C, 0.5x SSPE) exposing the filters between each wash. When this strategy was adopted I could consistently find oncogene related sequences in the genomic DNA of *Xenopus* as illustrated in figures 3.1, 3.2 and 3.3.

Figure 3.1(a) shows the result when human *N-myc* is used as a probe. The high degree of divergence between the probe and the *Xenopus* sequences is illustrated because only 5 μ g of human DNA (which is homologous for this probe) gives rise to a very intense expected ~2.1kb *EcoRI* band (Schwab *et al* 1983) compared with the weak bands in 15 μ g of *Xenopus* DNA.

If the whole genome duplication theory is correct for *Xenopus* genome evolution then in the simplest case one gene in *X.tropicalis* would become duplicated in *X.borealis* and *X.laevis*. Both these genes being on different chromosomes may then evolve independently giving rise to two detectably separate genes which may eventually acquire slightly different functions. The practical consequences for southern hybridisations is that when equal amounts of DNA are loaded there should be two bands of equal intensity in *X.borealis* and *X.laevis* and one band of twice the intensity in *X.tropicalis* DNA since effectively twice the number of genome equivalents of *X.tropicalis* DNA have been used. This also assumes that restriction enzyme sites do not split the

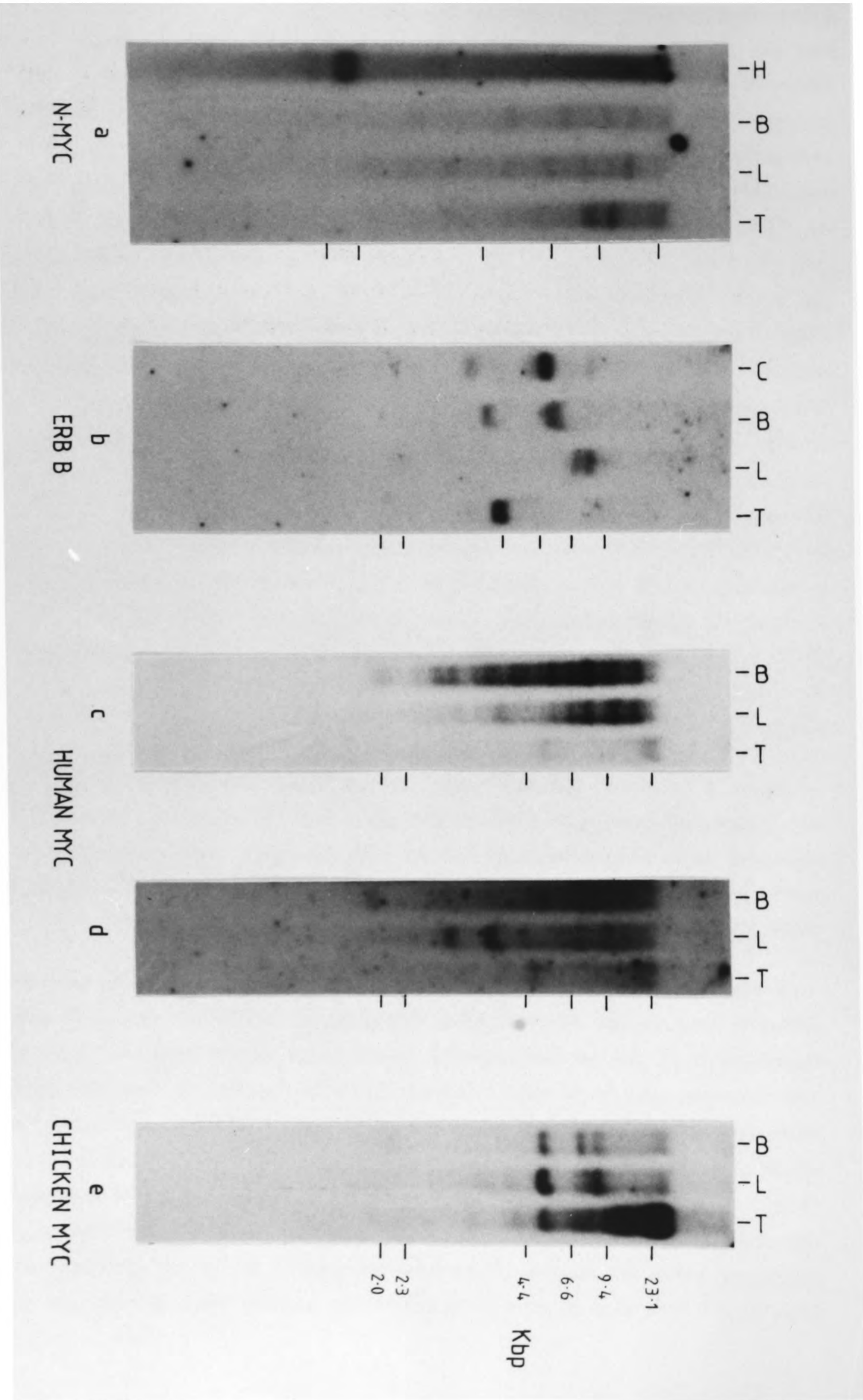


FIGURE 3.1

Genomic southern blots with heterologous oncogene probes. H = 5 μ g human DNA, C = 5 μ g chicken DNA, B = 15 μ g *X.borealis* DNA, L = 15 μ g *X.laevis* DNA, T = 15 μ g *X.tropicalis* DNA. All DNAs were digested with *EcorI*. Size markers were *Hind3* digested λ DNA. All probes had specific activities of $\sim 7 \times 10^8 - 1.5 \times 10^9$ cpm/ μ g. (a) *N-myc* probe, stringency 2x SSPE, 65 $^{\circ}$ C, 3 day exposure; (b) *v-erb B* probe, stringency 2x SSPE, 65 $^{\circ}$ C, 8 day exposure; (c) human *c-myc* third exon probe, stringency 0.5x SSPE, 50 $^{\circ}$ C, 1 day exposure; (d) filter from (c) washed at 0.5x SSPE, 65 $^{\circ}$ C, 5 day exposure; (e) chicken *c-myc* second exon probe, stringency 0.25x SSPE, 60 $^{\circ}$ C, 3 day exposure.

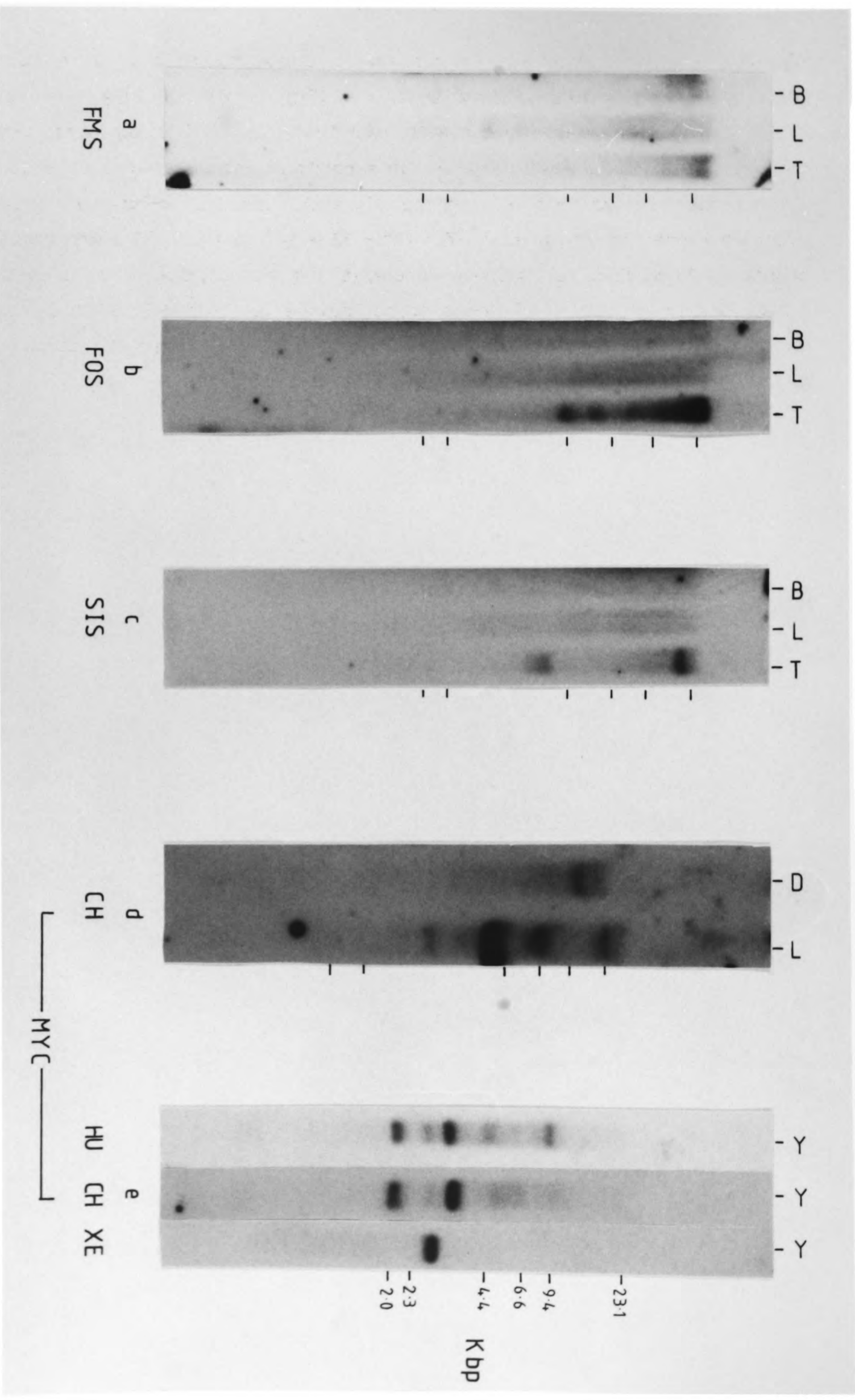


FIGURE 3.3

Genomic southern blots with heterologous oncogene probes. B = 15 μ g *X.borealis* DNA, L = 15 μ g *X.laevis* DNA, T = 15 μ g *X.tropicalis* DNA, D = 5 μ g *Drosophila* DNA, Y = 5 μ g yeast DNA. All DNAs were *Eco*RI digested. Size markers were *Hind*III digested λ DNA. Ch = chicken *c-myc* second exon probe, Hu = human *c-myc* third exon probe, Xe = *X.borealis* 3' *c-myc* probe. All probes had specific activities of $\sim 7 \times 10^8 - 1.5 \times 10^9$ cpm/ μ g. (a) *v-fms* probe, stringency 1x SSPE, 50 $^{\circ}$ C, 21 day exposure; (b) *v-fos* probe, stringency 1x SSPE, 50 $^{\circ}$ C, 21 day exposure; (c) *v-sis* probe, stringency 1x SSPE, 50 $^{\circ}$ C, 7 day exposure; (d) *c-myc* probe, stringency 0.5x SSPE, 65 $^{\circ}$ C, 12 day exposure; (e) *c-myc* probes, stringency 1x SSPE, 25 $^{\circ}$ C, 1 day exposure.

region of homology spanned by the probe and that both alleles at each locus are identical.

When a multigene family is involved the analysis becomes more complicated especially when heterologous probes are used. This would appear to be the case for *N-myc* although there are fewer more intense bands in *X.tropicalis* DNA. The presence of many bands in *Xenopus* DNA may reflect a multigene family but some might be non-specific hybridisation, although *myc* has been described as a small multigene family. Also the presence of introns may complicate analysis of this pattern.

Figure 3.1(b) shows the result using chicken *v-erb B* as a probe. Remember it is difficult to compare the results between different filters for the reasons described previously. Again 5 μ g of chicken DNA was used as a positive homologous control compared with 15 μ g of *Xenopus* DNA hence it is easier to detect the chicken sequence even though the probe is actually from a chicken virus. There are two *c-erb B* alleles (α and β) in chickens (Raines *et al* 1985) and they can have either or both alleles. The ~6.6kb *EcoRI* band I detect is in agreement with a ~6.4kb *EcoRI* band from the β allele. It would appear the *X.tropicalis c-erb B* gene is single copy and has evolved in a simple doubling fashion in *X.borealis* and *X.laevis* as described above.

Figures 3.1(c) and (d) shows the results when a human *c-myc* probe containing the third exon is used as a probe. The effect of differential washing is shown here : figure 3.1(c) is a lower stringency wash compared with the same filter washed at higher stringency in figure 3.1(d), which is exposed for longer. Notice the presence of many non-specific bands (especially in *X.laevis*) that disappear on washing at higher stringency. A common problem is a background smear in the DNA which obscures bands and appears to have occurred in the *X.borealis* track.

Figure 3.1(e) illustrates the hybridisation of a chicken *c-myc* probe which contains the second exon. The hybridisation is clear with stronger intensity in *X.tropicalis* DNA. This pattern of bands is different from figure 3.1(d), a result expected assuming the presence of *EcoRI* sites in an intron and or in the coding sequences. The chicken *c-myc* sequence may be more closely related to *Xenopus* DNA than the human *c-myc* sequence due to the clearer hybridisation but each species probe used was a different exon. It is difficult

to interpret these results with the evolution of the *Xenopus* genus but the general impression is that the pattern of bands is simpler in *X.tropicalis* and the bands are stronger supporting the genome doubling notion. The multiple bands again hint at the presence of *myc* being a multigene family.

Figures 3.2(a) and (b) shows the hybridisation of *v-ha-ras* and *v-ki-ras* to *Xenopus* DNA. Under these low stringency hybridisation conditions both probes pick up a similar pattern of bands with some differences whereas under high stringency in human and rat DNA they pick up separate bands (Hall 1984). This suggests the homology between the two probes is high enough to cross hybridise to related sequences in *Xenopus* DNA. The presence of so many bands can be partly explained by the fact that in other systems *ras* is a multigene family (e.g. human DNA has at least 5 members, 2 of which are pseudogenes, Hall 1984). Also since the probes used span the complete coding domain of *ras* and since human *ras* genes can have up to 4 introns, assuming the same to be true in *Xenopus* this might give rise up to 4 bands for each gene in the family. Each species of *Xenopus* shows a different band pattern implying a different genomic organisation of the *ras* genes.

Hybridisation patterns obtained when *v-src* is used as a probe are shown in figures 3.2(c). *Src* related sequences have already been cloned from *X.laevis* (Steele 1985). The ~10.0kb and ~2.3kbp *EcoRI* fragments I detect on southern analysis agrees well with the restriction map of a λ genomic clone and is good evidence that the conditions adopted, which are similar to Steele's, are detecting *bona fide* oncogene related sequences in *Xenopus* DNA. Steele suggests on the basis of cDNA clones and southern analysis, that *X.laevis* has two *src* genes. My results suggest this is probably also true for *X.borealis*, and *X.tropicalis* probably has just one gene although there a number of weak bands in addition to the single strong 23kbp band.

Figure 3.2(d) shows the pattern of hybridisation obtained when a *v-abl* probe that spans the tyrosine kinase domain (a 1.7kbp *PvuII* fragment) is used. A common progenitor sequence to both *src* and *abl* (cDash) has been isolated from *Drosophila* DNA as well as a *src* specific sequence (cDsrc2, Hoffman-Falk *et al* 1983, Hoffman *et al* 1983) suggesting *src* and *abl* have evolved from the same ancestral gene. The hybridisation pattern of *v-src* and *v-abl* to *Xenopus* DNA is quite different implying cDash and cDsrc2 have diverged to give rise to distinct *src* and *abl* genes in the lineage leading to vertebrates. Note that the

~2.3kb band common to both the *src* and *abl* filters in *X.tropicalis* may be a progenitor sequence but this is unlikely since the *X.borealis* and *X.laevis* patterns are so different. Further southern blot analysis is required to determine the relationship of *src* and *abl* genes in *X.tropicalis* DNA.

Weak hybridisation in all three species of *Xenopus* to a *rel* probe is shown in figure 3.3(e). When *v-fms*, *v-fos* and *v-sis* are used as probes only very weak hybridisation is detected (figure 3.3 a,b,c). Definite weak bands can be detected in all the *Xenopus* DNAs with each of the probes used. In general *X.tropicalis* DNA has stronger bands than *X.laevis* or *X.borealis* which probably reflects the genome doubling event, however I cannot be sure these homologies are real because they are so weak. Also I am unaware of the detection of these sequences in any organism other than a mammal.

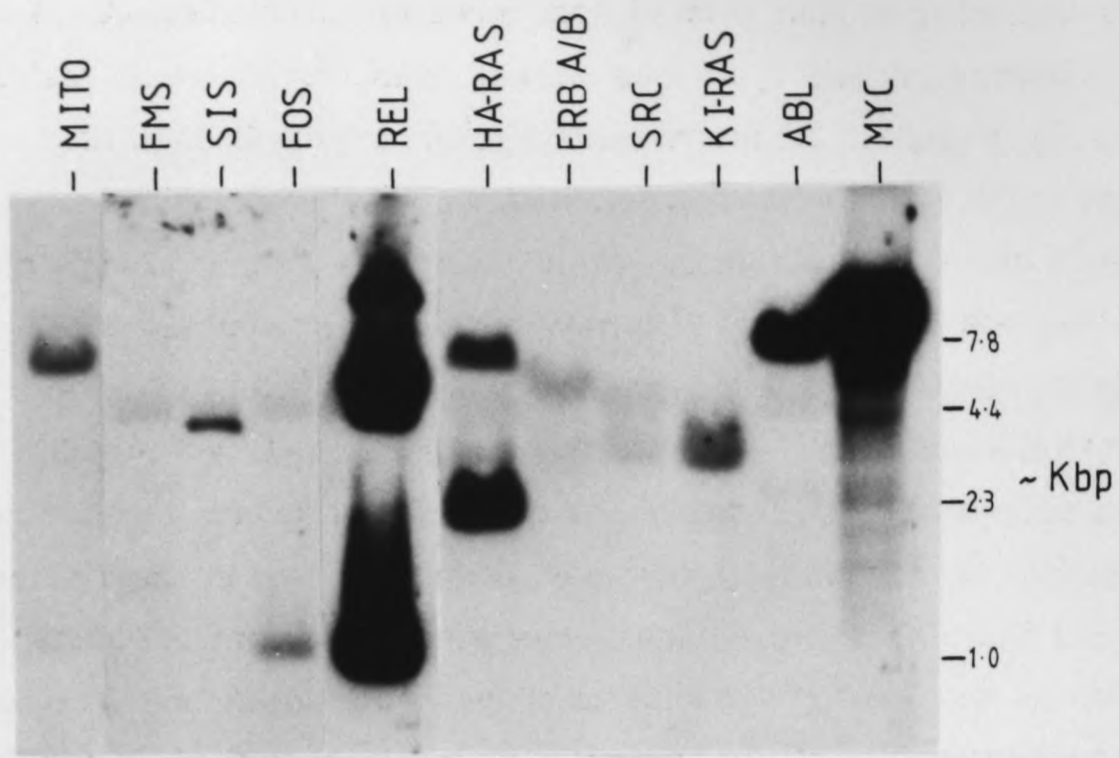
Chapter 1 describes why the *myc* genes are an interesting class of oncogene to study. Therefore as background I wanted to see if I could detect *c-myc* related genes in *Drosophila* and yeast DNA. Since the genomes of these organisms are much smaller it is possible to load more genome equivalents per track, which may aid detection compared with *Xenopus*. Figure 3.3(d) shows the hybridisation of a human *c-myc* third exon probe to *Drosophila* DNA showing a weak band (*Xenopus* DNA is used as a control). Figure 3.3(e) shows the hybridisation of a human *c-myc* third exon probe, a chicken second exon *c-myc* probe and *Xenopus* 3' *c-myc* probe (Xbmyc A1 see later) to yeast DNA. Second and third exon probes detect similar bands while the *Xenopus* probe detects a single band strongly which is also detected by the other probes but is weaker. It is unclear if these truly represent detection of *c-myc* homologies or are just spurious hybridisations. Note that a *Drosophila* *c-myc* related sequence has been reported (Madhavan *et al* 1985). The bands of hybridisation in yeast DNA appear to be close to repeated sequences that generate bands on *EcoRI* digestion and can be seen on an ethidium stained gel. Therefore these hybridisations may be fortuitous. It is encouraging that the *Xenopus* probe, which contains 300bp of 3' coding sequence, detects strongly a band that is moderately strong with the human third exon probe and weaker with the chicken second exon probe. It would be premature to interpret these results any further.

3.3 Detection of Oncogene Related Sequences in *X.borealis* RNA

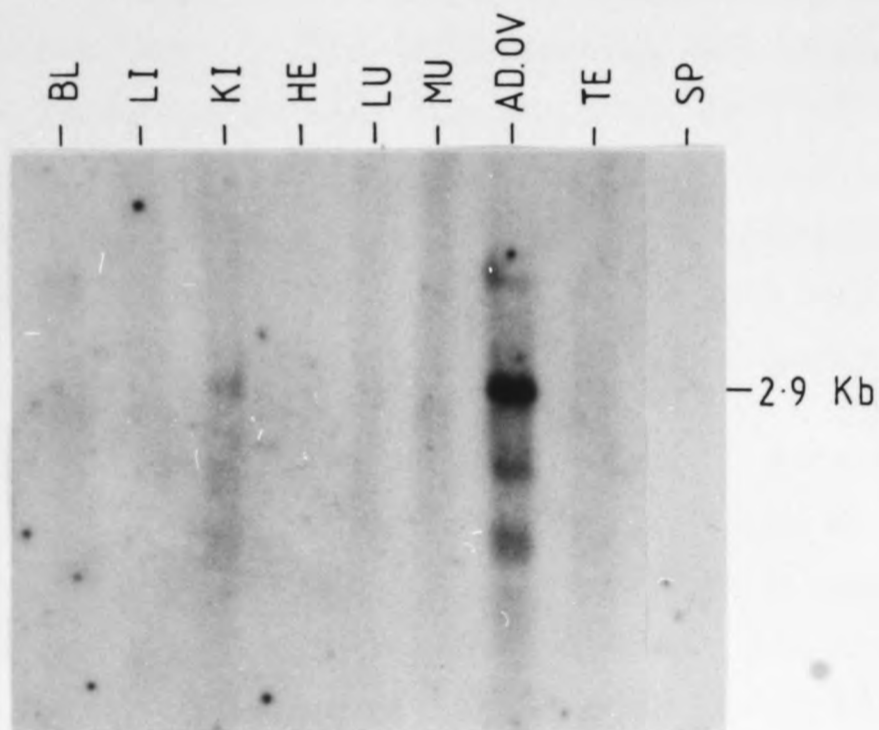
Having detected the presence of oncogene related sequences in *Xenopus* DNA I wanted to see if I could extend these observations to see if I could detect the presence of any transcripts in *X.borealis* RNA using heterologous probes. If I could find the presence of transcripts it would provide more evidence that the homologies I see in *Xenopus* DNA are in fact real. The detection of any transcripts using heterologous probes depends on parameters similar to those described for southern blots and therefore similar hybridisation conditions were used. A major difference though, is that the detection of transcripts depends very much on the level of gene expression in the tissues of interest. A combination of factors may lead to an apparently negative but incorrect result especially if the probe used is highly diverged and the mRNA is not highly expressed. So it is difficult to quantitate gene expression using heterologous oncogene probes. These results may be only used as a guide to examine which tissues are expressing particular sequences.

Initially to determine rapidly what if any oncogene sequences were expressed in adult ovary RNA, southern blots of a variety of oncogene plasmids digested with appropriate restriction endonucleases were probed with ³²P kinased labeled poly(A)⁺ RNA. This technique will only detect reasonably abundant mRNAs (estimated to be >0.06% of the total, Dworkin and Dawid 1980), also there may be fortuitous hybridisation because heterologous probes are used, hence this experiment only gives a guide to which oncogenes might be expressed allowing further work to proceed more rapidly.

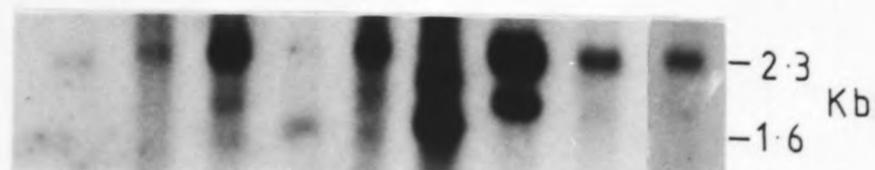
Figure 3.4(a) shows the result of such an experiment. A *HindIII* digested *X.laevis* mitochondrial plasmid was used as a positive control – the strong band at ~6.0kbp corresponds to the fragment containing the mitochondrial ribosomal RNA genes (Rastl and Dawid 1979). Since non poly(A)⁺ RNAs contaminate poly(A)⁺ selected RNAs by up to 50%, as judged by ³H poly(U) hybridisation estimates of mRNA purity, it is not surprising that mitochondrial rRNA can be detected in this RNA preparation. However it is unclear why abundantly expressed polyadenylated mitochondrial RNAs are not detected. The non-specific hybridisation can be estimated by the intensity of the bands corresponding to vector sequences; this seems to be quite low (most vector sequences are at ~4.4kb on this gel) although slight partial digestion has given quite high background levels in the *erb A/B* and *rel* tracks. The strong 4.4kb



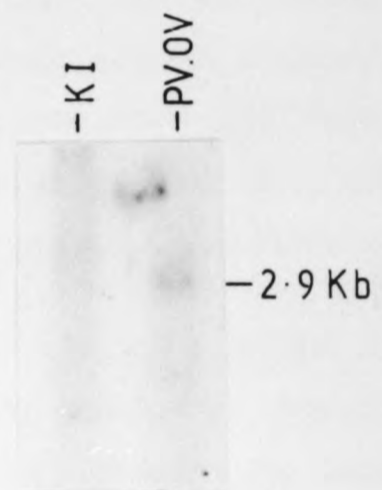
KINASED RNA
a



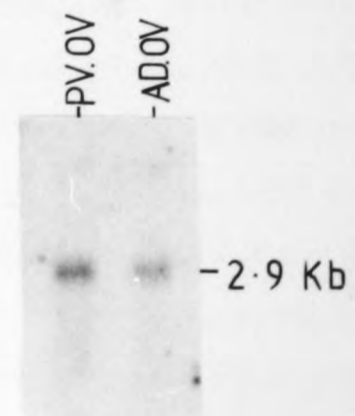
CHICKEN MYC
b



CYTOSKELETAL ACTIN
c



HUMAN MYC
d



CHICKEN MYC
e

FIGURE 3.4

Expression of oncogene related sequences in *X.borealis* RNA. (a) southern blots of digested oncogene plasmids probed with ^{32}P kinased labeled adult ovary poly(A)⁺ RNA (specific activity $\sim 2.6 \times 10^6$ cpm/ μg). Plasmid DNA had $\sim 1.0 - 2.0$ μg insert (which should be in excess for the hybridisation since $< 1\mu\text{g}$ RNA was used). Mito = *X.laevis* mitochondrial genome plasmid *Hind3* digested, other plasmids are as described in chapter 2. Stringency 2x SSPE, 37⁰C, 7 day exposure. (b), (c), (d), and (e) are northern blots with various probes. 4 - 5 μg of poly(A)⁺ RNA was used per track. Bl = blood, Li = liver, Ki = kidney, He = heart, Lu = lung, Mu = muscle, Ad.ov = adult ovary, Pv.ov = previtellogenic ovary, Te = testes and Sp = spleen. All probes had specific activities of $\sim 7 \times 10^8 - 1.5 \times 10^9$ cpm/ μg . Sizes were estimated from *X.borealis* 18S and 28S rRNAs. (b) chicken *c-myc* second exon probe, stringency 0.5x SSPE, 50⁰C, 14 day exposure; (c) cytoskeletal actin probe, stringency 0.2x SSPE, 65⁰C, 1 day exposure; (d) human *c-myc* third exon probe, stringency 0.5x SSPE, 50⁰C, 4 day exposure; (e) chicken *c-myc* second exon probe, stringency 0.5x SSPE, 50⁰C, 6 day exposure.

band in *sis* is due to an unknown trace contaminant in that plasmid preparation. Bearing these reservations in mind the results suggest that *myc*, *abl*, *ha-ras*, *ki-ras*, *rel* and *fos* are quite highly expressed in the ovary whereas *src*, *fms*, *erb A/B* and *sis* are expressed at a low level or not at all. These results provide some guidelines in selecting sequences for further study.

To further characterise these putative oncogene transcripts a series of northern blots were performed using poly(A)⁺ RNA from a variety of different *X.borealis* tissues. It was found when total RNA was used under non-stringent conditions there was a very high level of background bands especially due to the ribosomal RNAs therefore poly(A)⁺ RNA was used to reduce this problem. By using poly(A)⁺ RNA mRNAs also have been effectively concentrated by ~100 fold thereby making the detection of lowly expressed mRNAs easier.

Figure 3.4(b) shows the expression of *c-myc* related sequences as detected by a chicken *c-myc* second exon probe. The same filter was then probed with a cytoskeletal actin probe and is shown in figure 3.4(c) and shows the integrity of the RNA preparations. The 2.3kb band is due to cytoskeletal actin, the 1.6kb band is due to cardiac and skeletal muscle actins detected by cross-reaction of the probe and the intermediate bands are probably other cytoskeletal actins (Mohun *et al* 1984). In this experiment approximately equal amounts of poly(A)⁺ RNA were used (~4-5µg). The differences in actin intensity can be explained partly in terms of expression. Cytoskeletal actin is expressed in all adult tissues but in muscle it is expressed 50 times less and in oocytes it is highly expressed (Mohun *et al* 1984, Gurdon *et al* 1985). Therefore this suggests that roughly similar amounts of RNA have been loaded except maybe for blood and heart RNA.

Using the chicken *c-myc* second exon probe I can detect a ~2.9kb sequence quite strongly in adult ovary RNA. Kidney RNA shows a weaker signal and very weak signals can be detected on the original autoradiograph in blood, liver, lung, testes and spleen RNA. I cannot detect any signal from heart RNA. These differences in intensities may be due to different expression levels and differences in amount of RNA loaded. Figure 3.4(e) shows that the level of expression of this gene appears to be similar in adult and previtellogenic ovary RNA.

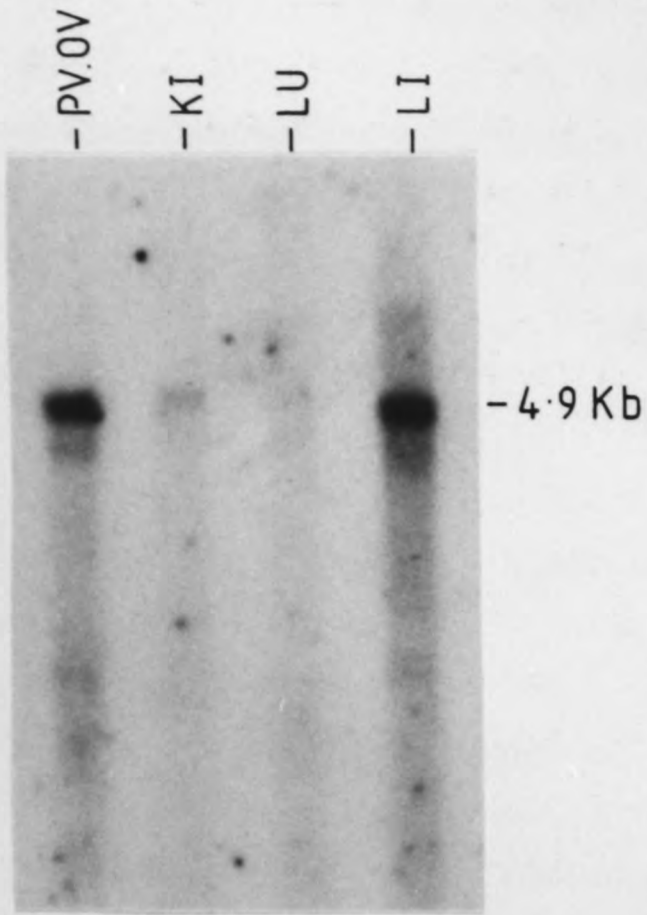
To reinforce the fact that this ~2.9kb band is due to *c-myc* related

sequences a northern blot was performed with a human *c-myc* third exon probe. Figure 3.4(d) shows that this probe detects a similar band in previtellogenic ovary RNA. The ~2.9kb RNA is probably due to the expression of a *c-myc* related RNA since it is detected with two different exon probes. Using these heterologous probes requires exposures of 1-2 weeks compared with the homologous probes which need exposures of 3-5 days (compare northern blots in Chapter 4 with these). Hence the sensitivity is dependent on the level of expression and the probe used (for example the cytoskeletal actin filter was only exposed for 1 day compared with the 2 weeks exposure of the chicken *myc* filter).

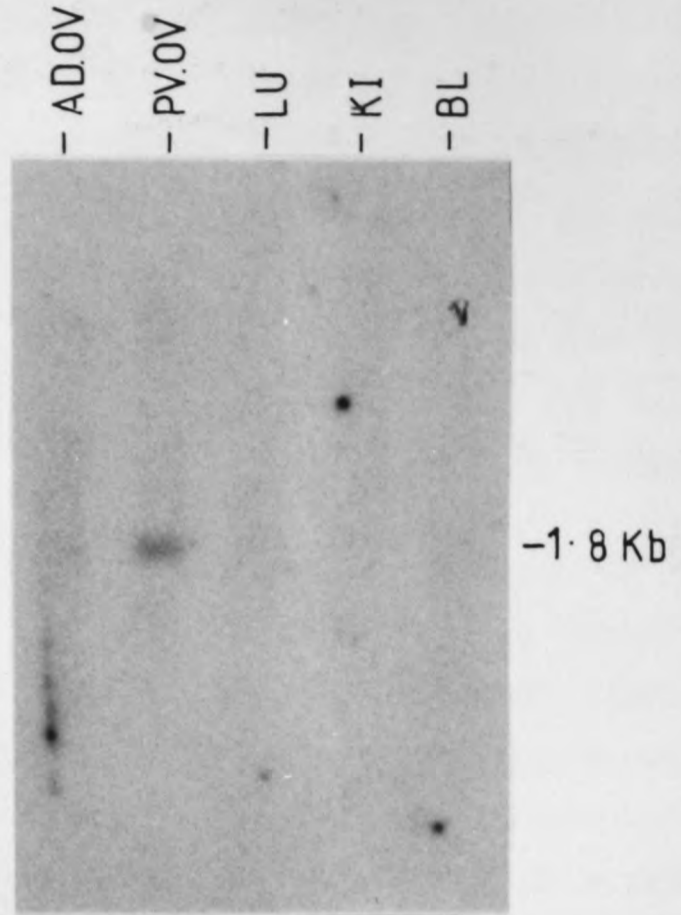
C-myc RNA can be detected in a wide variety of adult tissues and organs (Chapter 1, Gonda *et al* 1982, Stewart, Bellve and Leder 1984, Zimmerman *et al* 1986). Therefore it is not surprising that I can detect weak *c-myc* related transcripts with a heterologous probe in most tissues examined. The use of a homologous probe makes this analysis simpler (see Chapter 4). It is interesting that a high level of *c-myc* transcripts has been found in mouse ovary, uterus and thymus which may be correlated with cellular proliferation (Stewart, Bellve and Leder 1984) and I also detect a higher level of *c-myc* related transcripts in *Xenopus* adult and previtellogenic ovary.

The *Xenopus c-myc* related transcript is ~2.9kb which is larger than the 2.4kb mouse, 2.5kb chicken 2.4-2.5kb human and 2.3kb trout *c-myc* transcripts although there is some disagreement about these sizes in the literature (Zimmerman *et al* 1986, Gonda *et al* 1982, Rabbitts *et al* 1985, Papas *et al* 1985, Van Beneden *et al* 1986). If the *Xenopus* protein product is roughly the same size as in other species then this implies that ~1.5Kb of the mRNA is non-coding (this is confirmed by sequence analysis described in Chapter 5). Other bands detected with the *c-myc* probes may be non-specific as they can vary between individual experiments but it is possible they are due to other *c-myc* related species.

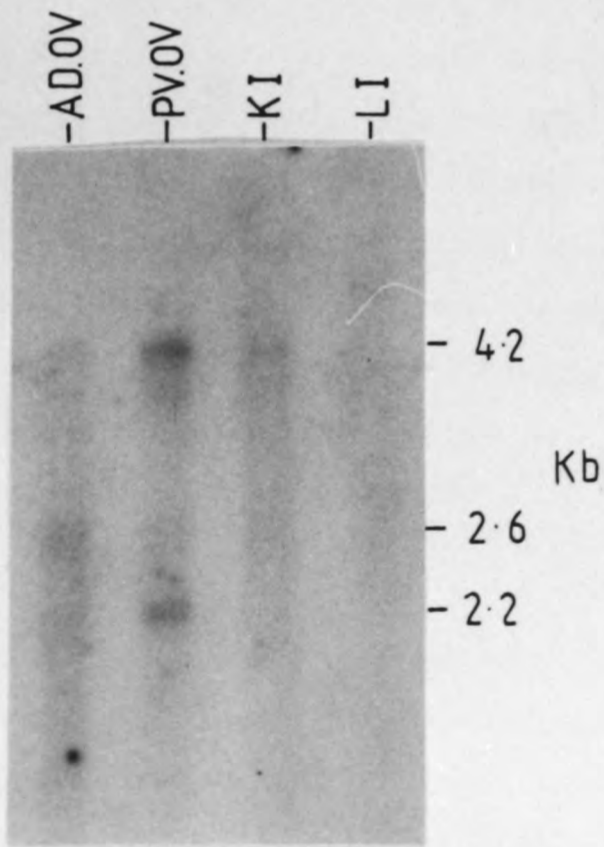
Figure 3.5(a) shows the detection of *N-myc* related transcripts in *X.borealis* RNA. High levels of a ~4.9kb transcript appear to be expressed in previtellogenic ovary and liver whereas lower levels can be detected in kidney and lung. As described in Chapter 1 *N-myc* expression is much more restricted than that of *c-myc* (Zimmerman *et al* 1986, Jakobovits *et al* 1985). In normal mouse tissues *N-myc* mRNA was greatest in newborn forebrain,



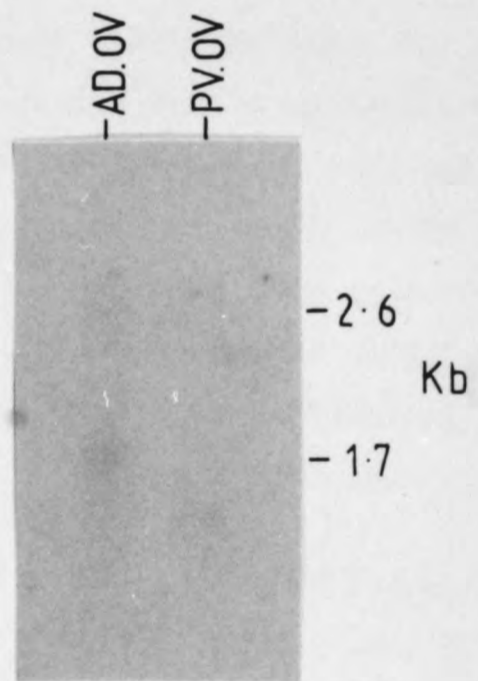
N-MYC
a



FOS
b



KI-RAS
c



HA-RAS
d

FIGURE 3.5

Northern blots using heterologous oncogene probes on *X.borealis* RNA. 4 - 5 μg of poly(A)⁺ RNA were used per track. Ad.ov = adult ovary, Pv.ov = previtellogenic ovary, Ki = kidney, Lu = lung, Li = liver. All probes had specific activities of $\sim 7 \times 10^8 - 1.5 \times 10^9$ cpm/ μg . Sizes were estimated from *X.borealis* 18S and 28S rRNAs. (a) *N-myc* probe, stringency 2x SSPE, 65⁰C, 3 day exposure; (b) *v-fos* probe, stringency 1x SSPE, 50⁰C, 21 day exposure; (c) *v-ki-ras* probe, stringency 1x SSPE, 50⁰C, 21 day exposure; (d) *v-ha-ras* probe, stringency 1x SSPE, 50⁰C, 7 day exposure.

hindbrain, kidney and intestine, but was reduced at least 20 fold in adult tissues. *N-myc* mRNA is expressed in the early stages of the differentiation of brain, kidney and haematopoietic cells in the mouse fetus. It is interesting that *N-myc* related expression in *Xenopus* is quite high in liver since the adult liver in *Xenopus*, unlike mouse, is still an active site of haematopoiesis. Again the *Xenopus N-myc* related transcript appears to be larger than that of mouse or humans which is ~3.0-4.0kb (Jakobovits *et al* 1985, Michitsch and Melera 1985, Zimmerman *et al* 1986).

The expression of a *fos* related sequence is shown in figure 3.5(b). Weak signals can be detected in blood, kidney, lung, previtellogenic ovary and adult ovary. the size of this transcript is ~1.8kb which is slightly smaller than the mouse 2.2kb *c-fos* transcript (Muller and Verma 1984) but is still large enough to code for a protein of similar size to that of mouse (380aa). This 1.8kb band is very similar to the 18S ribosomal gene size but I do not think the sequence detected by the *v-fos* probe is due to non-specific hybridisation since no other tissue shows a strong *fos* transcript and yet 18S rRNA is present in each. The *fos* gene is rapidly induced in response to mitogens and differentiation agents (see Chapter 1, Verma 1986). In fetal mice high expression of the *fos* gene is found in amnion, endoderm, mesoderm and visceral yolk sac (Muller, Verma and Adamson 1983, Muller and Verma 1984). Elevated levels of *c-fos* mRNA are also found in adult bone and skin. Also high mRNA levels are found during the late differentiation of myelomonocytic cells and during macrophage proliferation (Muller, Muller and Guilbert 1984, Gonda and Metcalf 1984, Muller *et al* 1985). It is difficult to assess the expression of the *Xenopus c-fos* gene in terms of these results since the signals, apart from previtellogenic ovary, are weak and since the tissues contain a wide variety of cell types.

The detection of *ki-ras* and *ha-ras* related transcripts in *Xenopus* RNA is shown in figure 3.5(c and d). Multiple transcripts are observed with both probes. Transcripts of 4.2, 2.6 and 2.2kb are found with *v-ki-ras* probe in adult ovary, previtellogenic ovary and kidney although there appear to be differences in the level of expression of each transcript. The *v-ha-ras* detects transcripts of 2.6kb and 1.7kb. The 2.6kb transcript is present in both adult and previtellogenic ovary but the 1.7kb transcript is present only in adult ovary. In mice and humans *c-ha-ras* and *c-ki-ras* seem to be expressed ubiquitously with multiple transcripts (Muller and Verma 1984, Hall 1984). The mouse *c-ha-ras* transcripts are 1.4kb and 4.3kb whereas the human *c-ha-ras*

transcript is 1.2kb. The mouse *c-k1-ras* transcripts are 2.0kb and 4.4kb whereas the human *c-k1-ras* transcripts are 5.5, 3.8 and 1.2kb. The human *N-ras* transcripts are 5.2 and 2.2kb (Hall 1984). The multiple *Xenopus ras* transcripts which appear independent of each probe used (except the 2.6kb band) is entirely consistent with this data as is the expression in many tissues. It is interesting that both probes detect some different bands in northern blots yet the pattern of bands in southern blots is similar.

When probes for *fms*, *rel*, *abl*, *erb B* and *src* are used to probe northern blots of *Xenopus* RNA either very weak signals could be found (that could not be photographed) or no signal was detected. These signals may be non-specific although using *v-src* as a probe I detect a very weak ~3.4kb band in ovary RNA which is in agreement with Steele's results (Steele 1985). The discrepancies between the kinased labeled RNA experiment and the northern blotting experiments are not understood but it is likely the former results are subject to more artifacts.

3.4 Discussion

By using highly labeled probes and non-stringent hybridisation conditions I have extended the initial observations of Shilo that cellular oncogenes can be detected in a wide variety of organisms (Shilo and Weinberg 1981) to include amphibians. Table 3.1 summarises the data I have obtained using different oncogene probes on three species of *Xenopus* DNA. As table 3.1 indicates I can detect all the oncogenes I have looked for, at least to some extent, in the DNA of all three species of *Xenopus*. It is therefore likely that cellular oncogene homologs exist in *Xenopus* for most other oncogenes that I have not tested. Hence this study reaffirms the belief that cellular oncogenes are very highly conserved throughout evolution and provides a basis (albeit teleological) for saying cellular oncogenes play an important role in normal cell growth. This is supported by the fact, described previously, that cellular oncogenes have been isolated from yeast and *Drosophila*.

From this study a general impression is gained that *X.tropicalis* DNA may be simpler than *X.laevis* or *X.borealis* DNA (with possible exceptions such as *ras*) and provides evidence for the genome doubling hypothesis discussed earlier. How this genome doubling event took place is unclear. It may have arisen by a failure of mitosis or meiosis leading to non-reduction generating

Oncogene probe	X.b	X.l	X.t	Ov RNA
<i>N-myc</i> (human)	++	++	++	+++
<i>v-erb B</i>	+++	+++	+++	-
<i>c-myc</i> (human 3rd exon)	+	+++	+++	++
<i>c-myc</i> (chicken 2nd exon)	+++	+++	+++	+++
<i>v-ki-ras</i>	+++	+++	+++	+
<i>v-ha-ras</i>	+++	+++	+++	+
<i>v-src</i>	+++	+++	+++	+/-
<i>v-abl</i>	+++	+++	+++	+/-
<i>v-rel</i>	++	++	++	-
<i>v-fms</i>	+	+	++	-
<i>v-fos</i>	+	+	++	+
<i>v-sis</i>	+	+	++	-

Table 3.1

Summary of heterologous probing southern and northern blotting data from figures 3.1, 3.2, 3.3, 3.4 and 3.5. X.b=*X.borealis* DNA, X.l=*X.laevis* DNA, X.t=*X.tropicalis* DNA, Ov.RNA=ovary RNA, +++=strong signal, ++=moderate signal, +=weak signal, -=no signal.

autopolyploids. Alternatively genome doubling might have occurred by the crossing of two related species to generate allopolyploids. Mating with a haploid followed by further chromosome increases or mating with another tetraploid would give rise to the new tetraploid species. Gene conversion could then take place making two related genes similar and/or one gene could be silenced especially if there was pressure on the cell not to have two non-identical competing functional proteins.

A comparison of vitellogenin genes in *Xenopus* also supports a genome doubling of some kind (Wahli and Ryffel 1985). *X.laevis* has 4 vitellogenin genes. A1 and A2 are related by ~7% divergence as are B1 and B2. The A family are diverged ~20% from the B family suggesting a gene duplication before genome doubling. *X.tropicalis* has similarly two A genes but only one B gene suggesting the A gene duplicated soon after the *laevis-tropicalis* divergence and that they have been separated for the same period of evolutionary time. Therefore the divergence between *X.laevis* and *X.tropicalis* and the genome duplication in *X.laevis* probably occurred during the same period. Maybe the genome doubling led to the separation of the *X.laevis* lineage from *X.tropicalis*.

I can also detect expression of many cellular oncogenes using heterologous probes in *X.borealis* RNA. In general, the kinased RNA experiment in figure 3.4(a) has borne out the results obtained with northern blotting with a few exceptions (e.g. *rel* and *abl* do not appear to be as highly expressed as suggested by the kinased RNA experiment). Therefore I think this kind of experiment can only be used as a guide to test a battery of sequences for expression. Since I am able to detect expression of at least some cellular oncogene sequences this makes me confident that at least some of the homologies I detect are real. It may be some are not real because the stringency is low and one might begin to detect families of functionally related genes such DNA binding proteins or protein kinases but rarely do non-identical probes which may have evolved from the same gene (e.g. *src*, *abl*, *erb B*) detect similar sized bands on southern blots.

Each oncogene sequence, that I find, expressed seems to be in many tissues and at different levels. Such an observation might be expected as it is true of other systems and since these genes probably have very specific functions in certain cell types (and may have more than one function or

different functions in different cells). But it is interesting to note that the best expression I detect, with many oncogene probes, tends to be in ovary RNA, maybe because *Xenopus* oocytes accumulate a large amount of mRNA for use during embryogenesis (see Chapter 1, Rosbash and Ford 1974). If cellular oncogene mRNAs are stored by the oocyte for subsequent development then it implies their protein products are involved in those processes of embryogenesis that do not require zygotic genome transcription (i.e. cell division, protein synthesis and cell lineage determination) although this does not preclude any possible function in later development.

CHAPTER 4

ISOLATION AND EXPRESSION OF C-MYC RELATED SEQUENCES IN *X. BOREALIS*

4.1 Introduction

Using heterologous probes to study cellular oncogenes in *Xenopus* is limited because the range of possible experiments is constricted by the sequence of divergence of the probes which may lead to artifacts. Therefore to pursue any further study of *Xenopus* cellular oncogenes requires the isolation of homologous sequences. Because conditions have been described that detect oncogene related sequences in southern and northern blots, it should be possible to use similar conditions to isolate related sequences from a *Xenopus* gene library. This rationale has already been proved by the isolation of *c-src* from *X.laevis* (Steele 1985) and by the isolation of other oncogene sequences from yeast and *Drosophila*.

A number of reasons influenced the decision to try and isolate *c-myc* related sequences from *X.borealis*. Firstly it seemed sensible to choose the *c-myc* gene because I had good southern blotting data suggesting it was closely related to higher vertebrate *c-myc* genes. Secondly my northern blotting data suggested that it may be moderately expressed in the ovary. This would make the isolation of the gene easier if a cDNA library was employed since less recombinants would have to be screened than a genomic library. Remembering that *Xenopus* has a genome size similar to humans, isolation of a single copy gene from a genomic library would require several million recombinants whereas for a moderately expressed gene (say at least 0.01% of the total message population) only 50,000^{cDNA} recombinants would be required. When using heterologous probes the signals obtained are likely to be weak therefore it is advisable to screen as few recombinants as possible since many false positives may be generated. When required, genomic clones can be obtained using the homologous cDNA clones as probes. Hopefully this strategy also saves on expense and time. Thirdly when I started this work I was unaware of anyone else who was trying to isolate *c-myc* from *Xenopus*.

4.2 Isolation of C-myc Related Sequences from *X.Borealis* cDNA Libraries

4.2.1 Construction of a cDNA Library

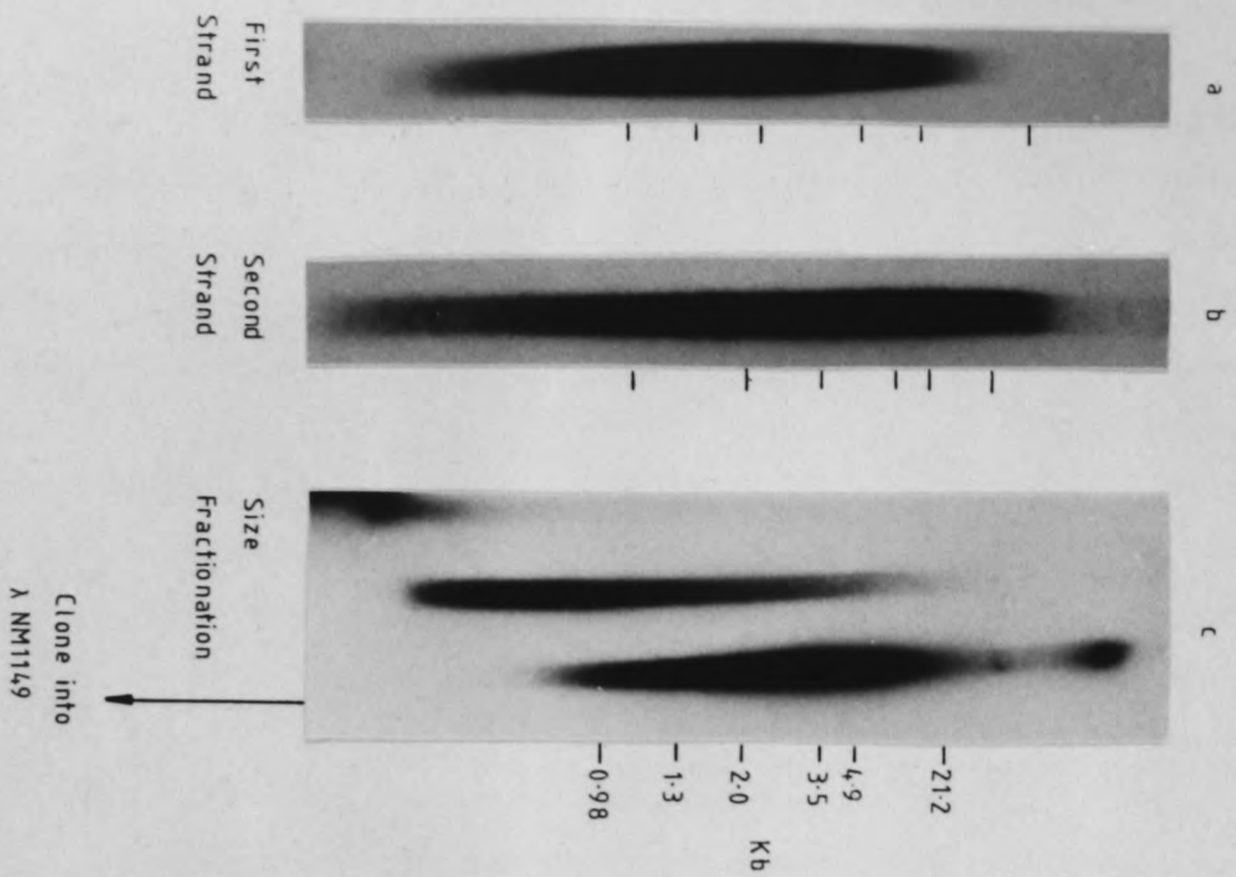
Since it appeared that *X.borealis* ovary had moderate expression of *c-myc* related sequences it was decided to build a previtellogenic ovary cDNA library in the vector λ NM1149 and use heterologous *c-myc* probes from humans and chickens to screen for phages containing a *Xenopus c-myc* related sequence (an adult ovary cDNA library was already under construction in our laboratory). The previtellogenic ovary cDNA library was constructed as described in Chapter 2 and as outlined in figure 4.1. This method has a number of advantages over other approaches (1) nearly full length clones should be generated due to the efficiency of the second strand reaction and the size selection, (2) *in vitro* packaging of λ DNA is generally more efficient than transformation of *E.coli* with plasmids thereby increasing the number of recombinants obtained for equal amounts of cDNA, (3) a positive selection for recombinants exists by using the vector λ NM1149 and plating the phages on *E.coli* NM514 and (4) λ plaques can be screened at a much higher density than *E.coli* colonies.

Figure 4.1 summarises the construction of such a library. (a) shows the first strand products of reverse transcription appear to be quite long assuming the average size of the mRNA is 2.0kb (Rosbash and Ford 1974). Calculations (appendix 2) indicate that 30–50% of the input mass of RNA is converted to first strand products. (b) shows the second strand reactions have apparently gone to completion giving full size products. Calculations (appendix 2) indicate that ~95–100% of the input mass of first strand product is converted to second strand products. (c) shows the size fractionated cDNA that was used for cloning into λ NM1149 was mostly above ~0.7kbp in size and therefore should mostly contain full length cDNA clones. Using these procedures I generated a previtellogenic ovary cDNA library that contained $\sim 4 \times 10^4 - 6 \times 10^4$ recombinants from ~50–100ng of double stranded cDNA.

4.2.2 Isolation of c-myc related cDNA clones

During the construction of the above library an adult ovary cDNA library from *X.borealis* became available (kindly supplied by D.Sleep), therefore it was decided to screen it with a human *c-myc* third exon probe under low stringency conditions similar to that previously described for southern and northern blotting. From $\sim 8 \times 10^4 - 1 \times 10^5$ phages only 1 positive phage was

A



B

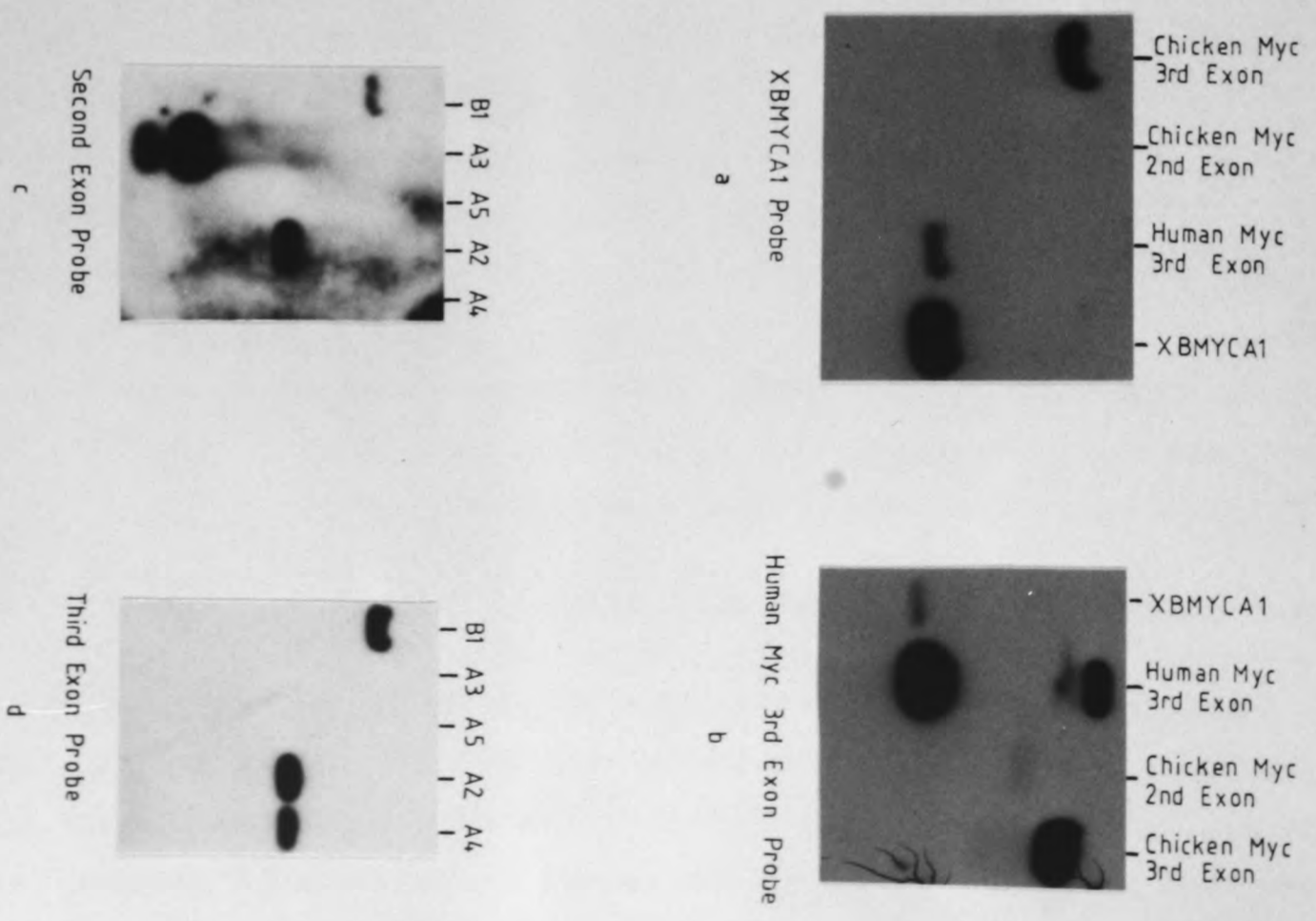


FIGURE 4.1

A: Analysis of the reaction products during the construction of a previtellogenic ovary cDNA library as described in chapter 2. (a) shows the size distribution of the first strand products (labeled with α 32 dCTP in the reaction) run on a 1% alkaline-agarose gel. Similarly (b) shows the size distribution of the second strand products run on a 1% DNA agarose gel. (c) shows the size distribution of the fractionated cDNA from a Biogel A-50 column ran on a 1% DNA agarose gel and indicates the fraction used for cloning. Size markers were λ DNA *Hind3* digested.

FIGURE 4.2

B: Southern blot analysis of potential *X.borealis c-myc* clones. (a) shows the hybridisation of a XbmycA1 probe to various *c-myc* exon clones and (b) similarly shows the hybridisation of a human *c-myc* third exon probe to a duplicate filter. (c) shows the hybridisation of a chicken *c-myc* second exon probe to various potential *X.Borealis c-myc* clones and similarly (d) shows a duplicate filter hybridised with a human *c-myc* third exon probe. All filters were hybridised under low stringency conditions with probes of specific activity of $\sim 5 \times 10^7$ cpm/ μ g. Final washing stringency was 1x SSPE, 65⁰C with an overnight exposure. B1 = XbmycB1, A1 = XbmycA1, A2 = XbmycA2, A3 = XbmycA3, A4 = XbmycA4 and A5 = XbmycA5.

isolated. It was noted that on primary screening of the library the signals obtained were very weak which resulted in picking a high number of false positives (on subsequent rescreening these failed to produce any signal). The positive phage was called XbmycA1 and contains an insert of ~ 1.3 kbp. Initially to show this insert contained *c-myc* related sequences southern blotting analysis was performed. Figure 4.2(a) shows that the human *c-myc* third exon probe hybridises with the XbmycA1 insert and figure 4.2(b) shows that a XbmycA1 insert probe hybridises with the human *c-myc* third exon and chicken *c-myc* third exon sequences but not the chicken *c-myc* second exon sequences. It was concluded that the XbmycA1 sequence contained the 3' coding sequences of a *X.borealis c-myc* gene. This was subsequently confirmed by northern and southern blotting (see later) since the XbmycA1 probe hybridised to similar bands that were picked up by the heterologous probe. DNA sequence analysis (chapter 5) showed that the XbmycA1 insert contained sequences highly related to other *c-myc* genes (in fact it had 300bp of the 3 coding sequence).

Unfortunately when a XbmycA1 probe was used to screen this library (a further $\sim 4 \times 10^4$ phages) no more positive phages were found. The reasons for a failure to isolate any more clones are unclear. Therefore I screened my previtellogenic ovary cDNA library, when it was successfully completed, with a XbmycA1 probe. Using this probe meant I could use higher stringency conditions and that positive plaques were much easier to identify.

On screening $\sim 4 \times 10^4$ phages of this library with a XbmycA1 probe I obtained 3 strong (XbmycA2, 4, 5) and 3 not so strong (XbmycB1, 2, 3) positive plaques. Figures 4(c) and (d) showed that XbmycA2 had strong homology to both exon probes but XbmycB1-3 had only strong homology to the third exon probe (only XbmycB1 is shown). Subsequent southern blotting analysis suggests XbmycA and XbmycB type sequences are actually different *c-myc* related genes. XbmycA4 and 5 appear to be short cDNAs from the 3' end of the mRNA and were not analysed any further.

Sequence analysis of XbmycA2 (chapter 5) confirmed that it contained a *X.borealis c-myc* gene that lacked the 5 end of the coding sequence. To obtain more 5' sequences a single strand M13 probe from the 5 end of XbmycA2 was used to screen this library. From $\sim 2 \times 10^4$ unamplified and $\sim 6 \times 10^4$ amplified phages 1 further strong positive (XbmycA3) and 11 less strong

positives (XbmycB4-14) were found. Figures 4(c) and (d) shows that XbmycA3 has strong homology to the second exon probe and sequencing (chapter 5) confirms it contains 5' coding sequences of a *X.borealis c-myc* gene. All XbmycB1-14 phages showed a similar hybridisation described for XbmycB1. Table 4.1 summarises the cDNA clones isolated above.

Clone	Insert size (<i>EcoRI</i> fragment)	Signal with probe	
		chicken <i>c-myc</i> exon 2	human <i>c-myc</i> exon 3
XbmycA1	1.3 kbp	-	strong
XbmycA2	1.4, 0.6 kbp	strong	strong
XbmycA3	0.55, 0.35 kbp	strong	-
XbmycA4	0.65, 0.4 kbp	-	strong
XbmycA5	1.3 kbp	-	strong
XbmycB1-14	3.0 kbp	weak	strong

Table 4.1 Summary of cDNA clones obtained

From the above data and the southern analysis described later it appears that I have isolated two different families of *c-myc* genes. The XbmycA class are closely related to the human *c-myc* third exon and chicken *c-myc* second exon sequences but the XbmycB class are only closely related to the human *c-myc* third exon sequence.

It may be that the XbmycA mRNA is difficult to reverse transcribe because I have isolated no complete cDNAs. The first strand reaction appears to have gone to completion since all the XbmycB cDNAs are nearly full length. Secondary structure in the XbmycA mRNA may block the passage of enzyme. During the initial screen I obtained roughly equal numbers of phages from each class which roughly agrees with the expectation from northern blotting in 4.3.1 suggesting both types of genes may be expressed about the same level. The second screen with a 5' probe mainly isolated B type genes thereby suggesting that many A type genes lack a 5' end, again pointing to a difficulty in the reverse transcription of the XbmycA mRNA.

XbmycA3 does contain 5' sequences yet it has no 3' sequences. It does not contain a poly(A) tail (chapter 5) hence it is not a cDNA clone for a short

mRNA. Probably this clone was generated by an incomplete second strand synthesis or by an internal priming event from the mRNA during first strand synthesis. XbmycA2 has a 5' end which is a naturally occurring *EcoRI* site (chapter 5). Possibly this site was not methylated properly before *EcoRI* digestion during library construction but it is not impossible that the cDNA actually extends beyond this site by a very short sequence which I am unable to detect on an agarose gel.

Most of the further work I describe concerns the XbmycA type genes because I was confident that they contained *c-myc* related sequences. The XbmycB types gene obviously are *c-myc* related sequences but they seem to be further diverged from the probes I used to isolate them with than the XbmycA type genes.

4.3 Characterisation of C-myc Related cDNA Clones

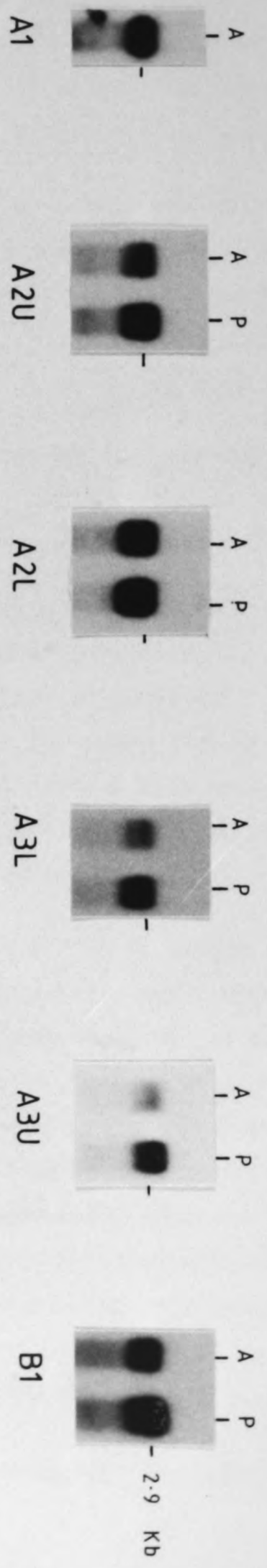
4.3.1 Northern blotting analysis

From heterologous probing I knew that the *c-myc* related transcript was ~2.9kb in ovary RNA therefore if the cDNA clones I had isolated showed ~2.9kb transcripts in northern blotting experiments then it would provide further evidence that the clones were derived from true *c-myc* mRNA. Any cDNA clone that contains more than one *EcoRI* fragment must also be tested to show that it is not derived from two independent cDNAs that have been ligated into the same vector molecule.

The results of such experiments are shown in figure 4.3. It would appear each cDNA clone does hybridise to a similar sized ~2.9kb transcript as was detected by the heterologous probes. Furthermore both *EcoRI* fragments of XbmycA2 and 3 seem to derive from the same transcript. Close inspection suggests that the band is in fact a doublet. This may point to there being two independently transcribed *c-myc* genes or it may be due to transcription from two independent promoters on each gene or to post-transcriptional events such as differential splicing. It is known that the mouse *c-myc* gene is transcribed from two promoters 167bp apart (Stewart, Bellve and Leder 1984) and also that there is some transcription from further upstream of these promoters (Nepveu and Marcu 1986).

The XbmycB type gene I have isolated also hybridises to similar sized

A



B

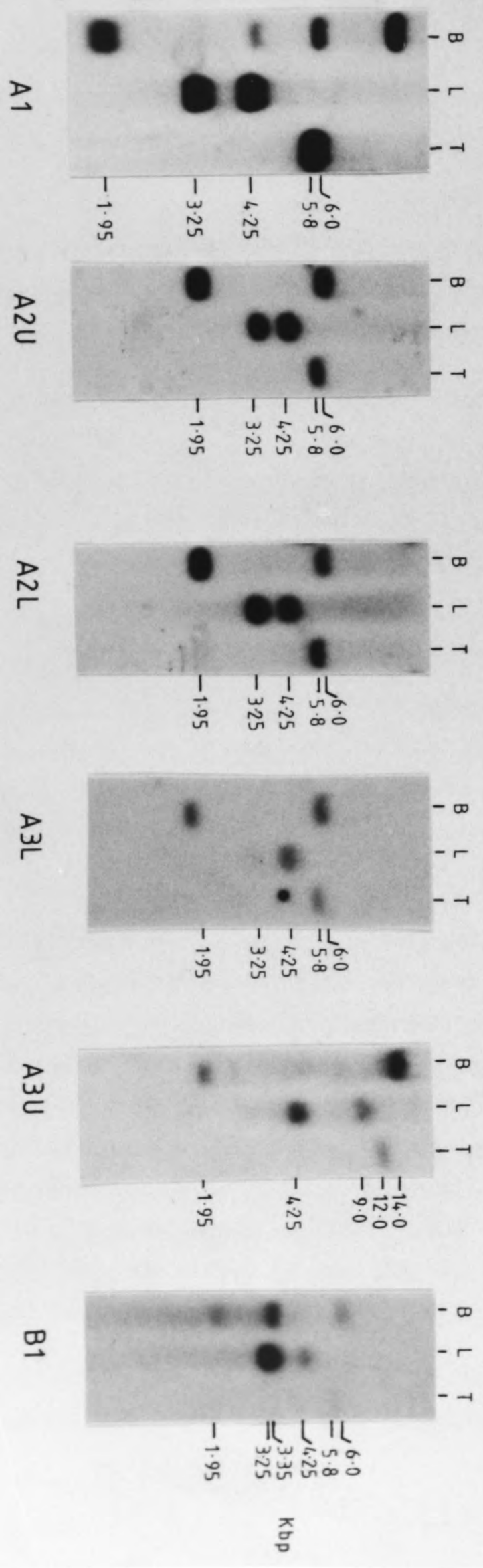


FIGURE 4.3

A: Northern blot analysis of the various *X.borealis* *c-myc* clones under high stringency conditions with probes of specific activity of $\sim 7 \times 10^8 - 1.5 \times 10^9$ cpm/ μ g. About 5 μ g of previtellogenic ovary (P) or adult ovary (O) poly(A)⁺ RNA was used per track. Final stringency of washing was 0.2x SSPE, 50⁰C with a 1 - 2 day exposure though the B1 filter had been exposed for 7 days since it had been reused a number of times. Sizes were estimated from *X.borealis* 18S and 28S rRNA. A1 = XbmycA1, A2U = XbmycA2 (large *EcoRI* fragment), A2L = XbmycA2 (small *EcoRI* fragment), A3U = XbmycA3 (large *EcoRI* fragment), A3L = XbmycA3 (small *EcoRI* fragment) and B1 = XbmycB1.

B: Southern blot analysis of the various *X.borealis* *c-myc* clones under high stringency conditions with probes of specific activity of $\sim 7 \times 10^8 - 1.5 \times 10^9$ cpm/ μ g. 10 μ g of *X.borealis* (B), *X.laevis* (L) or *X.tropicalis* (T) *EcoRI* digested DNA was used per track. Final stringency of washing was 0.2x SSPE, 65⁰C with a 3 - 4 day exposure. Sizes were estimated from λ DNA *Hind3* digested. Probes used are as described in A.

transcripts. I do not know to what extent the XbmycA and B probes cross-hybridise but southern analysis in 4.3.2 suggests this might be ~10% at the most. From figure 4.3 I roughly estimate the XbmycB genes are expressed at least 50% of the level of the XbmycA genes. This assumes that the probes were roughly the same specific activity and allows for the fact the XbmycB filter had been reused and the exposure is ~5x greater. Note that equal numbers of cDNA clones from each class were isolated from the initial library screen using a 3' probe. S1 nuclease protection assays are required to confirm the above estimates.

The differences in signals between previtellogenic and adult ovary are probably due to differences in the amount of RNA loaded. The filter for XbmycA3 had also been reused and probably explains the weaker bands. It must be emphasised that these exposures using homologous probes are 5-15x shorter than those using heterologous probes hence I have now greatly increased the sensitivity of my experiments.

4.3.2 Southern blotting analysis

To determine the relationship of the cDNA clones to each other and to try and work out a genomic structure for them a series of southern blots were performed as shown in figures 4.3. Each probe detects bands that are in common with the heterologous probes (figure 3.2) affirming the belief that the clones are *c-myc* related. Unfortunately it appears that in the filter probed with XbmycA1 the *X.borealis* DNA was not completely digested as determined by band intensity and by comparison with the XbmycA2 (lower) filter since both probes span the same region hence the 4.3 and 9.4kbp bands should be ignored. The *EcoRI* site that splits XbmycA2 into two fragments is either due to an artifact generated during cloning or it is due to a polymorphism since the frog used for southern analysis was different from those used to make the cDNA library. Another cDNA clone, XbmycA4, has an *EcoRI* site roughly in a similar position therefore this may be a real polymorphism although XbmycA1 and 5 do not possess an *EcoRI* site in this position.

My results show that there are at least two classes of *c-myc* genes in *X.borealis* and *X.laevis*. The hybridisation of XbmycA to *X.borealis* and *X.laevis* DNA is clearly different from that of XbmycB. There may be some cross-reaction of the probe since the probe hybridises weakly to the same

sized bands as XbmycA1,2 and 3 (lower). Although it might be possible that there is a second B type gene, split by an *EcoRI* site and having the same fragment sizes as those detected by the A type probe. I think these other bands are probably due to cross-hybridisation for two reasons: 1) the XbmycA type genes pulled out the XbmycB type genes from the library and 2) there are no *EcoRI* sites in all 14 cDNA clones of the B type genes. I do not understand why the A type genes do not weakly hybridise to the B type gene fragments on these southern blots. It is interesting that the XbmycB probe does not hybridise at all well with *X.tropicalis* DNA. This suggests that the XbmycB gene has evolved since the split of *X.borealis* and *X.laevis* from *X.tropicalis* and may have arisen by gene duplication followed by divergence in the *X.borealis* and *X.laevis* lineage. A B type gene might exist in *X.tropicalis* but it might be too diverged from the *X.borealis* probe to detect it under the hybridisation conditions I used. Also because the B type gene hybridises to a ~3.0kb transcript then it cannot contain a very large intron since it appears to be contained within a single ~3.35kbp genomic fragment.

The simplest model for the XbmycA class genes is discussed below and takes into account some of the sequencing data described in chapter 5. Figure 4.4 shows a hypothetical structure of a XbmycA gene and from where each probe is derived. It is assumed there is an intron present in the *Xenopus* gene at the approximate position indicated as one exists in the *c-myc* gene of all other species at approximately this position.

Probes XbmycA2 (upper and lower), XbmycA1, XbmycA3 (lower) hybridise to the same sized fragments therefore this suggests that if an intron is present it does not contain an *EcoRI* site. The XbmycA3 (upper) probe hybridises to one band that is similar in size (except *X.tropicalis*) and one band that is of a different size compared with the other probes. This different sized fragment is due to the presence of a natural *EcoRI* site in the 5' coding sequence which has apparently been conserved in *X.laevis* and *X.tropicalis*. I suggest that the 14.0kbp and 6.0kbp fragments in *X.borealis* and the 9.0kbp and 3.25kbp fragments in *X.laevis* are physically linked to form on XbmycA type gene. I think that 2.0kbp fragment(s) in *X.borealis* and the 4.25kbp fragment in *X.laevis* form a second but different XbmycA type gene since they hybridise to all probes used. Therefore this would mean there are at least two different kinds of highly related XbmycA type genes in *X.borealis* and *X.laevis* but only one kind in *X.tropicalis* and would imply that in addition to genome doubling there

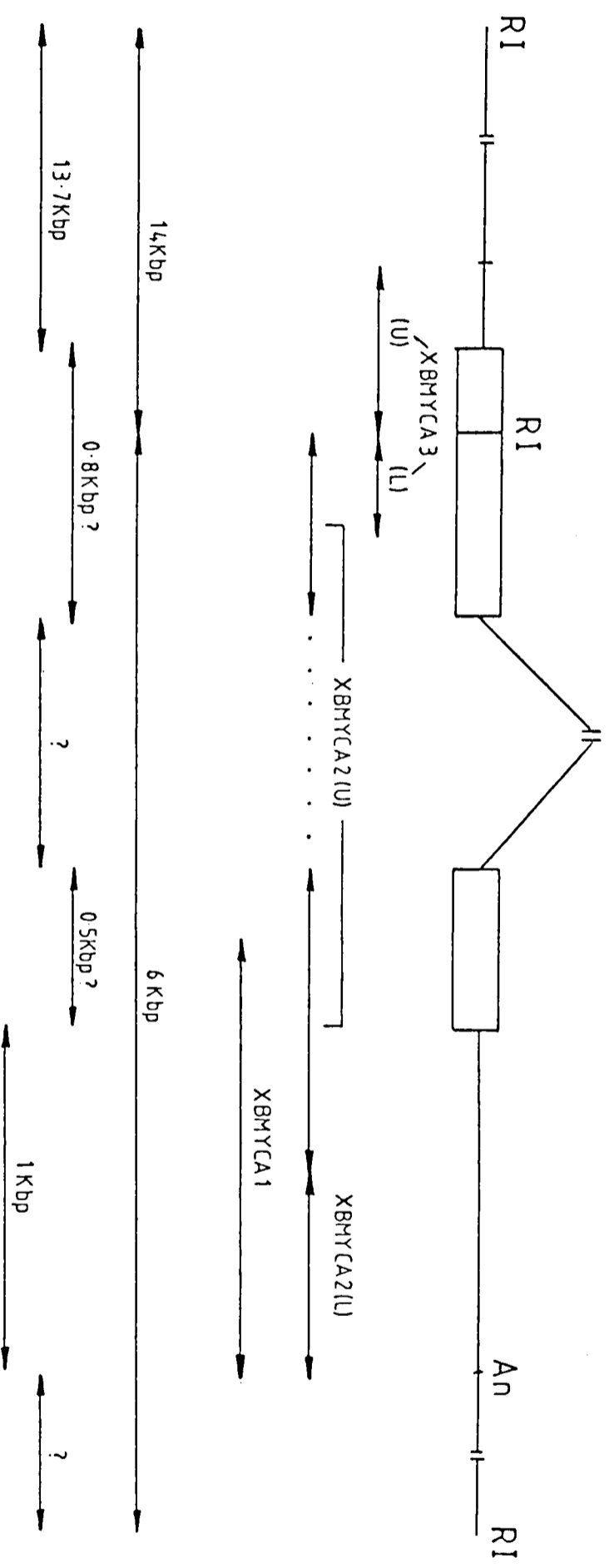


FIGURE 4.4

A hypothetical structure of a XbmycA class gene (see text) from southern blotting data in Figure 4.3 and sequencing data described in chapter 5. The origin of each XbmycA cDNA clone is marked, U and L referring to the large and small fragments generated by *EcoRI* digestion respectively of XbmycA3 and XbmycA2. The dotted line represents sequences not present in the XbmycA2 cDNA clone (i.e. a potential intron sequence represented by a thin diagonal line). Boxes represent coding sequences, An represents the polyadenylation site and RI represents *EcoRI* sites. Sizes are in kilobase pairs and the diagram is not to scale. Question marks refers to sizes that are impossible to estimate from the data.

has also been a gene duplication to generate both A type and B type genes in *X.borealis* and *X.laevis*

If the the above model is correct then it has to explain how the A type gene that is contained in the 2.0kb fragment in *X.borealis* can be shorter than the mRNA size. It could be that it is an unexpressed pseudogene; it could be the 5' end of the mRNA is completely unrelated to the other gene and is split by an *EcoRI* site from this fragment (this means if an intron is present it must be very small) or it is possible that the gene is divided into two *EcoRI* fragments of a similar size. The XbmycA3 (lower) probe overlaps the XbmycA2 (upper) probe and is highly homologous to it but it does not detect both bands in *X.laevis* with equal intensity yet it does in *X.borealis*. If one assumes that the region spanned by the XbmycA3 (lower) probe has diverged in one of the *X.laevis* genes then this would explain the results since this probe is much shorter than the XbmycA2 (upper) probe and might hybridise less well, also the rest of the XbmycA2 (upper) probe might be closely related to both *X.laevis* genes.

If there was one A type gene in *X.tropicalis* and two A type genes in *X.borealis* and *X.laevis* then it would be expected that the intensity of signal from *X.tropicalis* would be double compared with that from each of the *X.borealis* and *X.laevis* genes since twice as many genome equivalents *X.tropicalis* DNA was used. From figure 4.3 it is obvious this is not true and the intensities are roughly equal. Possibly this is a reflection of poorer hybridisation of the *X.borealis* probes to *X.tropicalis* DNA (as is true of XbmycB) or that the genes have become duplicated in *X.borealis* and *X.laevis* such that they give fragments of exactly the same size.

One could imagine other models that would fit the data in figure 4.3 for example the fragments I have suggested that are linked to each other may not be which would lead to other interpretations of the results. More extensive southern blotting analysis is needed to determine the exact genomic structure of these genes. Isolation of genomic clones and heteroduplex analysis would also be required to define the precise gene structure.

4.4 Expression of *X.borealis* C-myc Genes

Most of the following work describes the expression of the XbmycA type genes since these are highly related to the *c-myc* genes of other species.

Sometimes filters were reused for probing with the XbmycB sequences. This was not always totally successful but the results are shown since they can be informative.

4.4.1 Tissue specific expression

XbmycA1 was used to probe a northern blot of different tissue poly(A)⁺ RNAs similar to that already described for heterologous probing. Figure 4.5 shows that *c-myc* RNA can be detected in most tissues. Adult ovary RNA shows the strongest signal; kidney shows a moderate signal; muscle, testes, spleen, blood and lung show weak signals whereas liver and heart do not show a detectable signal. Figure 4.3 shows that both adult and previtellogenic ovary have similar expression levels of *c-myc* RNA. These results confirm those already described for heterologous probes in figure 3.3(c) i.e. *c-myc* RNA is expressed in a wide variety of tissues. Note that by using the homologous probe the exposure required for a good signal has been reduced 5 fold. Bearing in mind that different amounts of poly(A)⁺ RNA might have been used and the fact there is heterogeneity of cell populations in tissues my results suggest that most tissues express *c-myc* RNA at a low level which is agreement with the expression of *c-myc* in other systems (Stewart, Bellve and Leder 1984). My results also suggest that *c-myc* levels are quite high in ovary. This might point to a storage of *c-myc* RNA in the protracted period of oogenesis for subsequent use by the egg and embryo during the rapid cleavage stages of development that are transcriptionally inactive.

4.4.2 Level of expression in stage II oocytes

To quantitate the level of expression of the XbmycA type genes in stage II oocytes a northern blotting experiment was performed that used *in vitro* synthesised antisense XbmycA RNA to compare with *in vivo* levels of XbmycA RNA. A short 205 bp *EcoRI*/*PstI* fragment of the coding sequence of XbmycA1 was subcloned into the transcription vector pGEM1. Truncated ³H antisense message was synthesised with T7 polymerase and known amounts were mixed with the same amount of stage II oocyte total RNA (5 oocytes worth). Mixed RNAs were run on a formaldehyde gel, transferred to hybond-N and probed with the complete XbmycA1 fragment.

The result of such an experiment is shown in figure 4.6. Assuming that the signal is proportional to the length of the probe then densitometry shows that

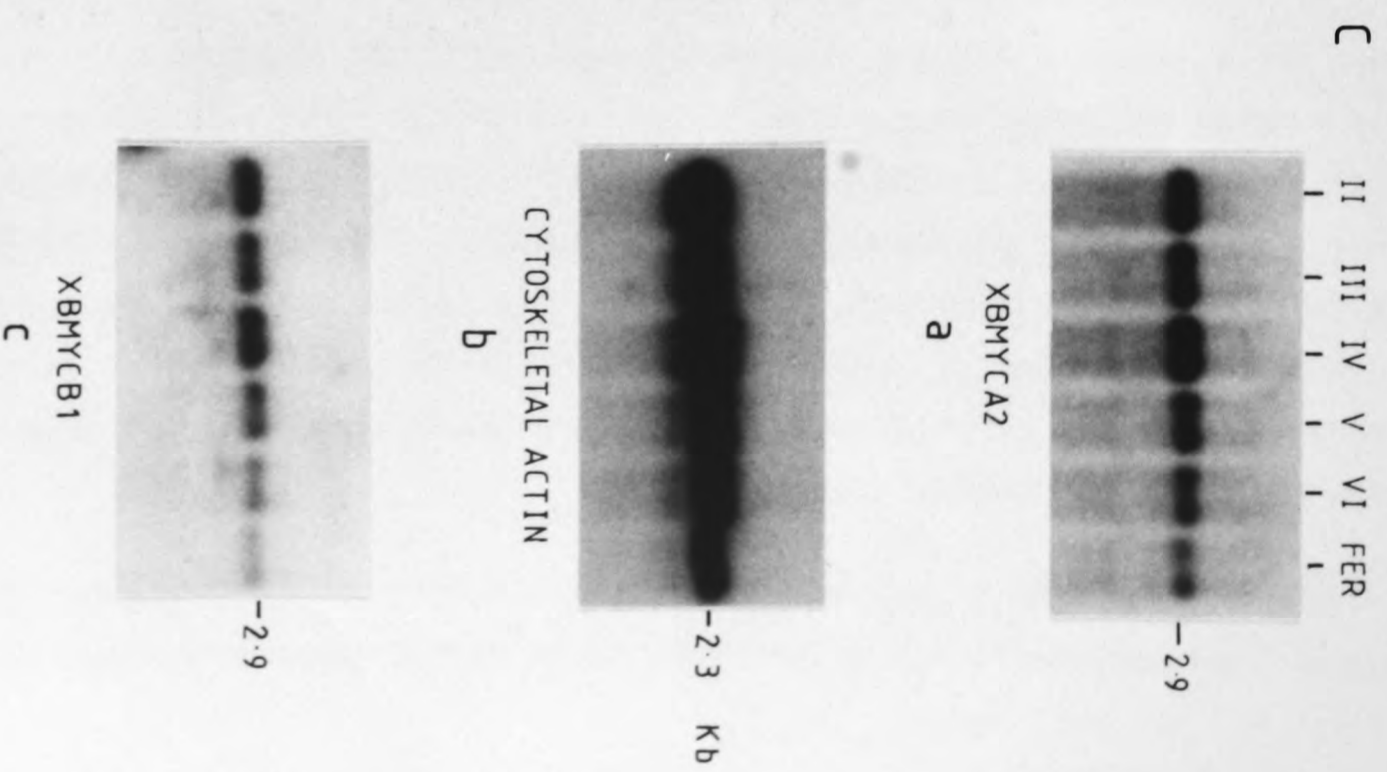
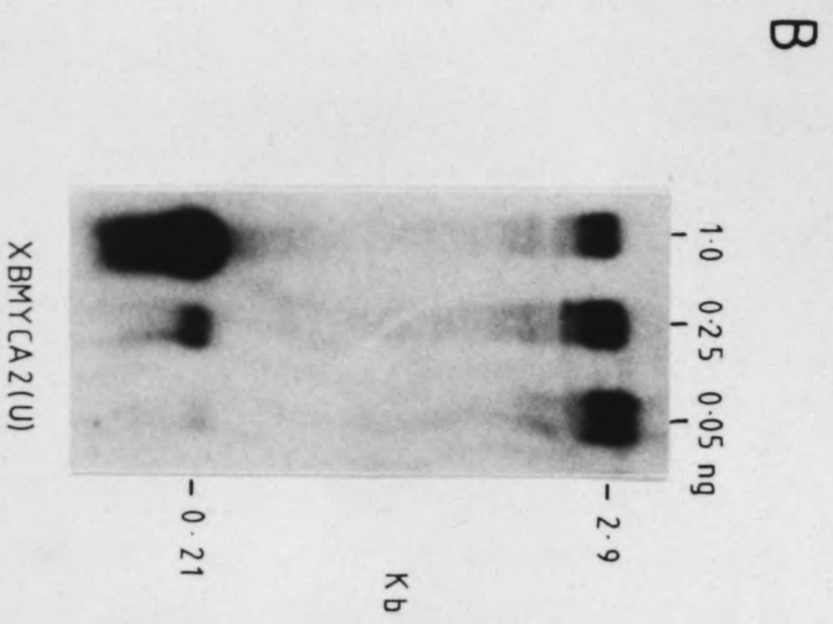
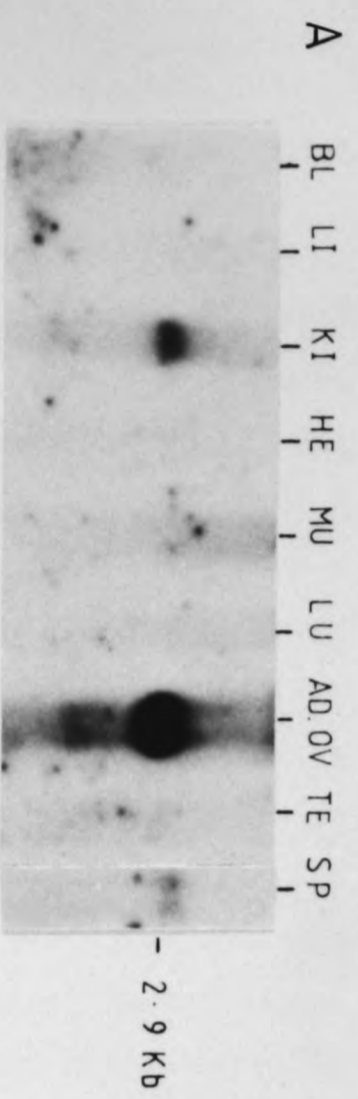


FIGURE 4.5

A: Expression of XbmycA1 in different tissue RNA. 4 - 5 μg of *X.borealis* poly(A)⁺ RNA was used per track. Bl = blood, Li = liver, Ki = kidney, He = heart, Mu = muscle, Lu = lung, Ad.Ov = adult ovary, Te = testes and Sp = spleen. The probe had a specific activity of $\sim 1 \times 10^9$ cpm/ μg . Final stringency of washing was 0.5x SSPE, 50⁰C with a 4 day exposure. Sizes were estimated from *X.borealis* 18S and 28S rRNA.

FIGURE 4.6

B: Level of expression of XbmycA RNA in stage II oocytes. 5 oocytes worth of total RNA were mixed with known amounts (indicated above each lane) of an *in vitro* synthesised 205bp ³H XbmycA1 *EcoRI/PstI* sense RNA. Samples were then ran on an RNA gel transfered to nitrocellulose and probed with the complete XbmycA1 probe (specific activity 1×10^9 cpm/ μg) under high stringency conditions. Final washing conditions were 0.2x SSPE, 60⁰C with a 3 day exposure. Densitometry was then performed and calculations indicate that one stage II oocyte contains $\sim 19\text{pg}$ of XbmycA RNA.

FIGURE 4.7

C: Expression of XbmycA during oogenesis. For each stage (as described by Dumont 1972) the total RNA from 5 oocytes was loaded per track. Fer represents fertilised eggs. All probes had roughly the same specific activity of $\sim 1 \times 10^9$ cpm/ μg and were used under high stringency conditions. Final washing conditions were 0.2x SSPE, 60⁰C and each filter was exposed for 4 days. (a) XbmycA2 probe. (b) the filter from (a) stripped and reprobed with a cytoskeletal actin probe. (c) the same filter stripped and reprobed with XbmycB1.

5 stage II oocytes contain ~95pg of XbmycA RNA. There is 1.6µg of total RNA in 5 stage II oocytes as judged by optical density and 140ng of poly(A) containing RNA as judged by ³H poly(U) hybridisations which agrees well with previous reports (Rosbash and Ford 1974). Therefore XbmycA type RNA is expressed at a moderately high level of ~0.07% of the maternal mRNA population in stage II previtellogenic oocytes.

4.4.3 Expression during oogenesis and development

Gene expression of XbmycA during oogenesis and development was examined by northern blotting using manually staged oocytes and embryos. Filters first were probed with XbmycA2 and then reprobed with cytoskeletal actin (and then XbmycB1). The results of these experiments are shown in figures 4.7 and 4.8. Densities of band intensities from overnight exposures of the autoradiographs were plotted versus stage of development as shown in figure 4.9 calibrating each experiment for the fertilised egg values. The relative expression of XbmycA against cytoskeletal actin is only a guide and assumes the specific activity of the probes were the same and hybridisation rates similar. Since the experimental scatter of the points is quite high the cytoskeletal actin graph was drawn first bearing in mind the previously determined developmental expression of the gene (Mohun *et al* 1984) then the XbmycA graph was drawn as a best fit of the experimental points. Hopefully this represents a true reflection of the expression of these genes.

Both XbmycA and cytoskeletal actin are made early in oogenesis, cytoskeletal actin always being expressed at a higher level (maybe at least by 2 fold). The maximal level of expression is at the beginning of vitellogenesis (stage II oocytes) which has dropped by ~50% in stage VI oocytes. Golden reports similar results that most mRNAs accumulate early in oogenesis but then remain at a constant steady state level for the rest of oogenesis (Golden, Schafer and Rosbash 1980). These results are surprising since lampbrush chromosomes are maximally extended in stage III oocytes (Macgregor 1980) yet the highest levels of expression precede this event. It has been suggested that poly(A)⁺ RNA synthesised early is conserved to the end of oogenesis and eventually inherited by the embryo whereas the poly(A)⁺ RNA synthesised in the latter months of oogenesis turns over (Ford, Mathieson and Rosbash 1977, Dawid, Kay and Sargent 1983). Therefore it may mean that XbmycA and cytoskeletal actin maternal RNA do not turn over during oogenesis though it

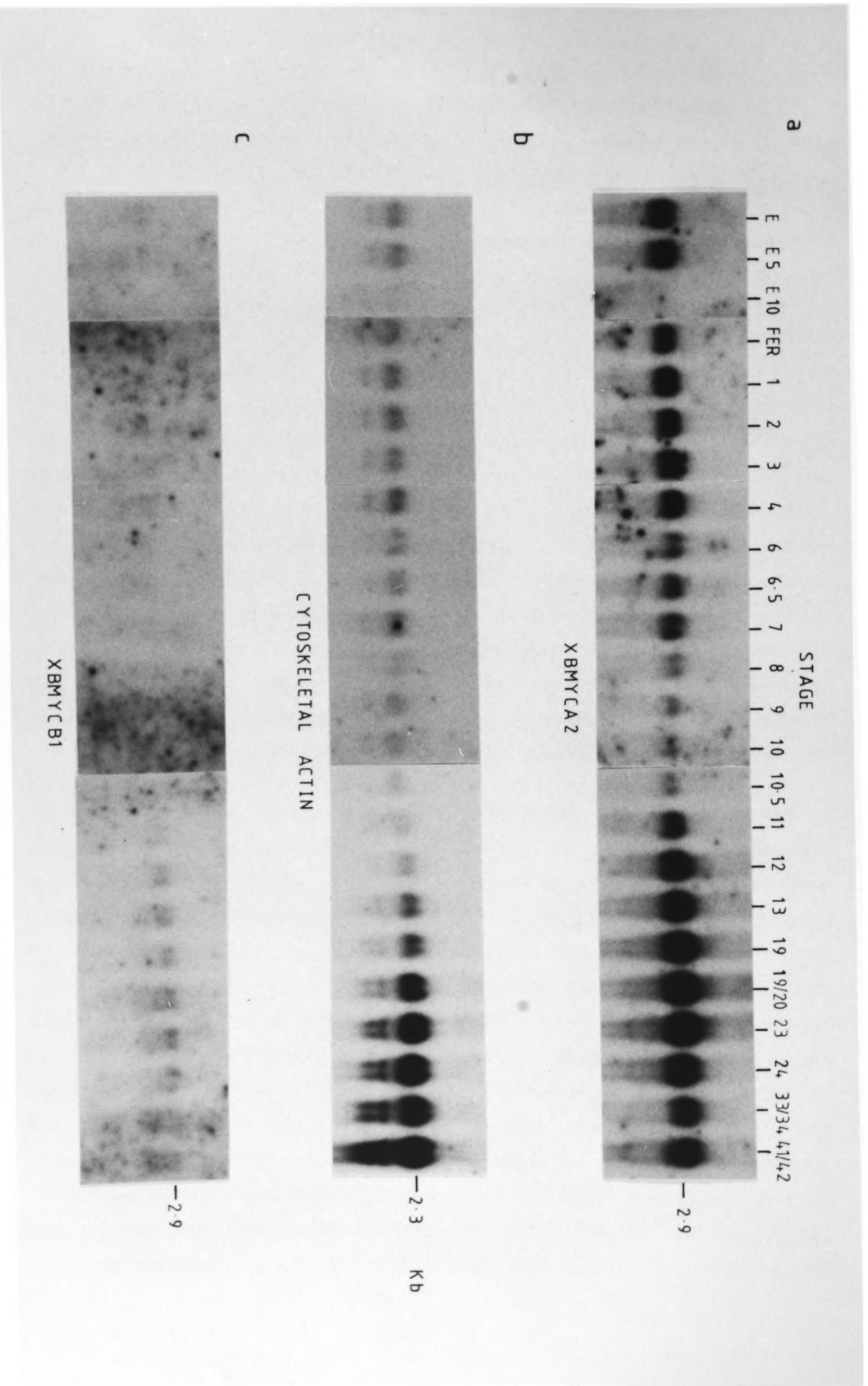


FIGURE 4.8

Expression of XbmycA during development. Embryo stages are as described by Nieukoop and Faber 1956. E represents ovulated eggs; E5 and E10 represent ovulated eggs stored for 5 and 10 hours respectively without being fertilised. Fer represents fertilised eggs. Development proceeded at $\sim 22^{\circ}\text{C}$ and the first cleavage took ~ 1 hour with subsequent cleavages taking ~ 30 minutes each. Probes had roughly the same specific activity of $\sim 1 \times 10^9$ cpm/ μg and were used under high stringency conditions. Final washing conditions were 0.2x SSPE, 60°C . (a) XbmycA2 probe, 7 day exposure. (b) the filter from (a) stripped and reprobbed with a cytoskeletal actin probe, 1 day exposure. (c) the same filter stripped and reprobbed with a XbmycB1 probe, 7 day exposure.

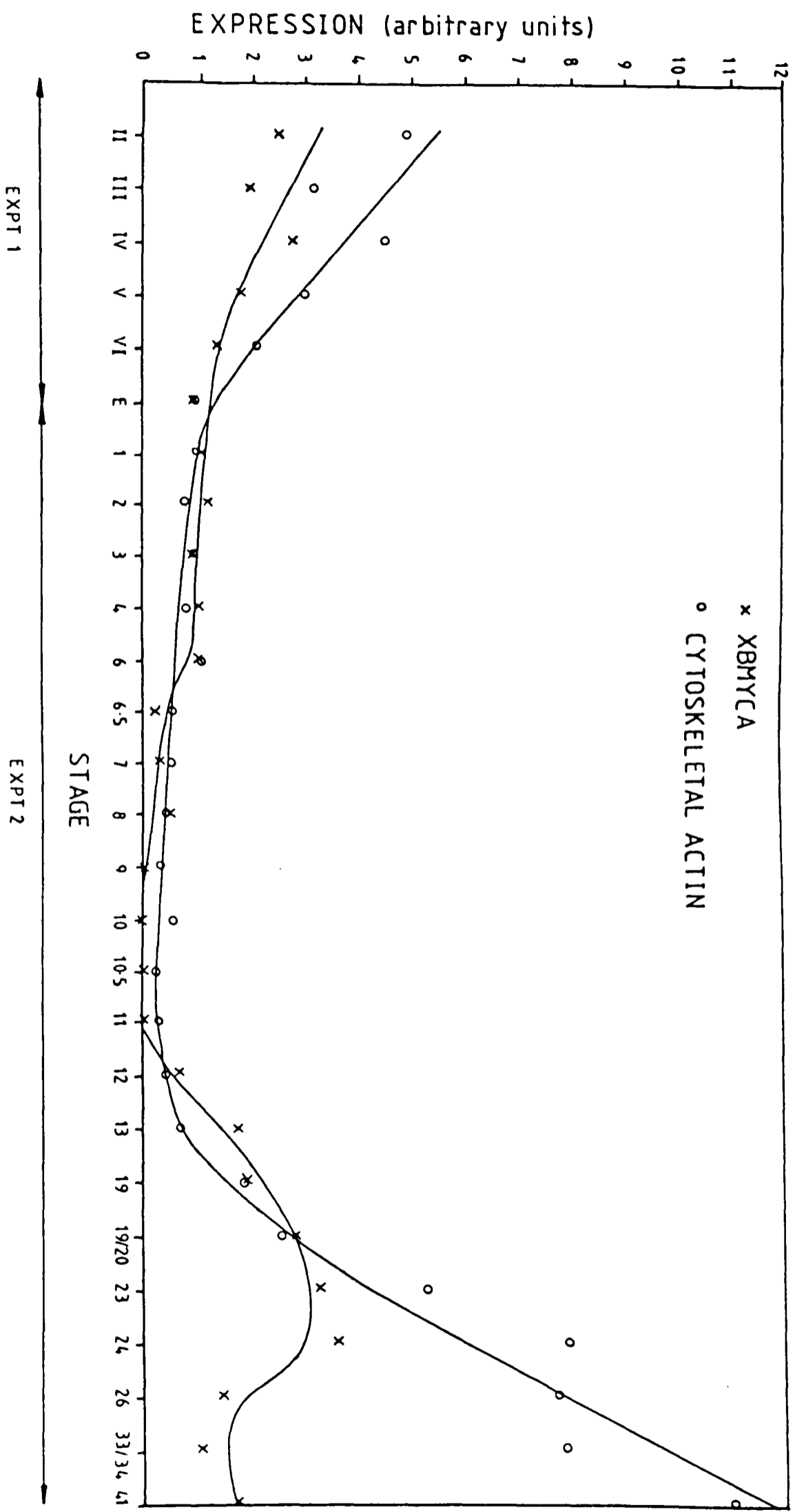


FIGURE 4.9

Expression of XbmycA during oogenesis and development. Densities from 1 day exposures of the filters from Figure 4.7 and 4.8 were plotted (in arbitrary units) against each stage of oogenesis (experiment 1) or development (experiment 2). E refers to newly fertilised eggs and since it was present in both experiments it was used to calibrate experiment 1 with experiment 2 so that one graph could be plotted. Oocyte stages are as described By Dumont 1972: stage II oocytes are late previtellogenic oocytes and stage VI oocytes are fully grown oocytes. Embryo stages are as described by Nieukoop and Faber 1956: stages 1 - 9 are blastula, stages 10 - 13 are gastrula, stages 14 - 24 neurula and stages 25 onwards are tailbud through to swimming tadpole. The midblastula transition occurs about stage 8 of development at $\sim 22^{\circ}\text{C}$ and no new transcription occurs until this stage. From Figure 4.6 I estimate a stage II oocyte contains $\sim 19\text{pg}$ of XbmycA RNA. Similarly the developmental expression of cytoskeletal actin was plotted using the densities from overnight exposures of the filters in Figure 4.7 and 4.8. This profile is almost the same as that described by Mohun *et al* 1984 given the limits of experimental variation. Although both probes had roughly the same specific activity it is difficult to compare the results quantitatively between the two different probes due to many factors such as differences in probe length, hybridisation rates and conditions etc. and since the filters had been reused but a general picture is gained through the comparison.

would be difficult to reconcile this with the fact *c-myc* RNA is highly unstable in other systems.

Cytoskeletal actin mRNA is inherited by the embryo and is steadily degraded until just after gastrulation (stage 13–14) when new transcription of cytoskeletal actin mRNA begins. Transcripts accumulate as cell number increases probably because most cells in the embryo begin to synthesise cytoskeletal actin mRNA. Cell numbers double between stage 14 (gastrula) and stage 26 (tailbud) embryos whereas there is at least an 8 fold increase in cytoskeletal actin levels. This developmental profile is similar to that already described (Mohun *et al* 1984), therefore one can assume that the general activation of RNA synthesis at the midblastula transition (Newport and Kirschner 1982) happens at a similar stage of embryogenesis in these experiments (stage 8, 10,000 cells).

XbmycA mRNA is also inherited by the embryo but seems to be degraded in two waves till at stage 9 over 90% of maternal mRNA has disappeared. This may reflect specific degradation of XbmycA mRNA in different germ layers therefore it is possible that specific factors control the degradation of XbmycA RNA. New XbmycA transcription starts around late gastrula which is slightly earlier than cytoskeletal actin transcription. Since so much of the maternal XbmycA2 mRNA has been degraded it may mean new XbmycA transcription has to start sooner to allow proper cell division.

The accumulation of XbmycA mRNA during later development is quite different from that for cytoskeletal actin mRNA. Maximal levels of XbmycA mRNA are reached by late neurula to tailbud (stage 20–24) which then drop by ~50% and remain roughly constant during later development. The main body organs have been mapped by stage 20 therefore maybe XbmycA is involved in organ differentiation. At stages 20–24 this would correspond to nervous system differentiation. The drop in expression of XbmycA later in development may be due to a smaller set of differentiating cells requiring XbmycA or it may be it is no longer involved in differentiation *per se* but rather is required in low amounts for cell division. Note in adult tissues XbmycA mRNA can be detected in many tissues. It is likely the XbmycA protein has many functions and will be involved in both processes.

The maximal level of XbmycA expression during development is of the

same order as that in stage II oocytes. It seems that the embryo inherits enough XbmycA mRNA to last until it reaches ~50,000 cell stage. Because there is no transcription until stage 8 it should be possible to make an estimate of the half life of XbmycA mRNA but this is confused by the apparently biphasic nature of the XbmycA mRNA degradation. Also figure 4.8 shows that unfertilized eggs that have been incubated for 5 hours have only lost 25% of their XbmycA mRNA but when incubated for a further 5 hrs XbmycA mRNA becomes undetectable. It also seems cytoskeletal actin is much more stable over the first five hour period. There is not enough data to estimate a half life of XbmycA mRNA but it can be said it appears to be a lot longer than other systems (hours rather than minutes) but this may partly reflect differences in temperature as well as cell type.

Another point to note from these experiments and those in 4.4.1 is that the XbmycA transcripts are always the same size i.e. the oocyte XbmycA transcript is the same size as the somatic and embryonic XbmycA transcripts. Anderson has reported that 70% of oocyte poly(A)⁺ RNA carries interspersed repeats (Anderson *et al* 1982). Therefore it might be expected that oocyte XbmycA mRNA would be larger than somatic or embryonic transcripts: this is obviously not true and supports the results of Golden (Golden, Schafer and Rosbash 1980). I am actually unaware of any cloned sequence from *Xenopus* that has a larger transcript in oocytes.

The experiments using XbmycB as a probe are not very good since the filters had been reused several times prior to probing but from figure 4.8 one can see that XbmycB follows a similar developmental profile as XbmycA. Therefore both XbmycA and B might be regulated in a similar fashion.

4.4.4 Expression in dissected embryos

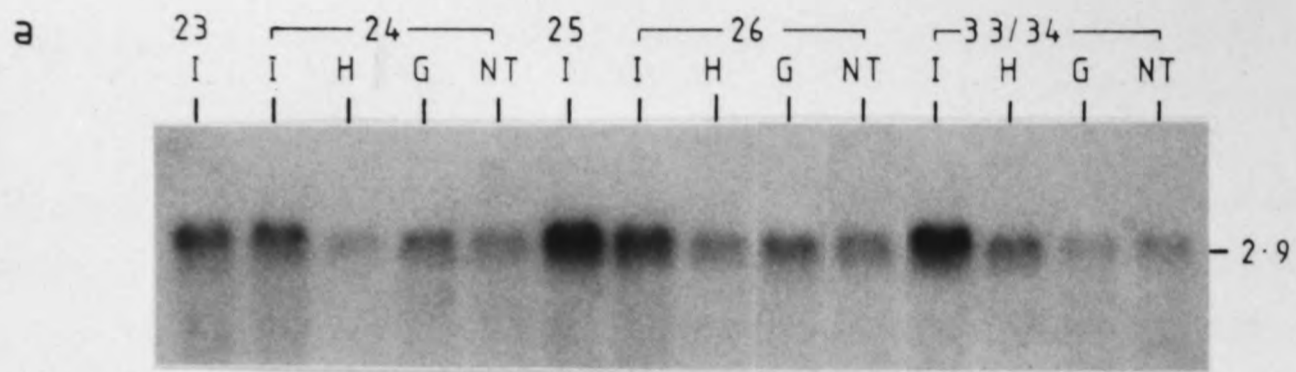
The previous results show XbmycA mRNA peaks during the late neurula to tailbud stages. This might mean XbmycA is involved in the development of the nervous system and the brain. During the post-natal development of the murine cerebellum high levels of *c-myc* mRNA have been found in proliferating granule cells and differentiating Purkinje cells (Ruppert, Goldowitz and Wille 1986). To see if there is any localised expression of XbmycA mRNA in embryos a series of northern blots were done using dissected embryos from stages 24, 26 and 33/34. The embryos were cut into 3 pieces : the head piece containing

the developing brain and eyes ; a gut piece containing primitive body organs such as liver, heart, kidney, gut etc ; and a neural tube piece containing the developing spinal chord and somites. Note that each of these pieces do not derive from a single germ layer (i.e. either, endoderm, ectoderm or mesoderm alone). The results of such experiments are shown in figure 4.10 Densitometry was performed on overnight exposures of the autoradiographs and graphs were plotted of band intensity (in arbitrary units) versus stage for each tissue piece (see figure 4.11).

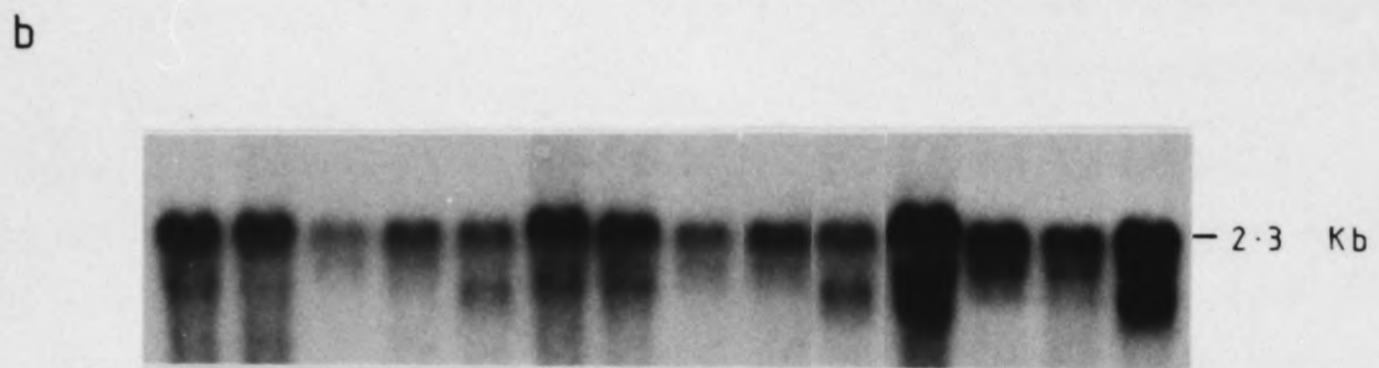
The total expression pattern for complete embryos (when the individual values of head, gut and neural tube are add together for each stage) is similar to previous results except that XbmycA expression turns down slightly later at stage 26 rather than stage 24. This may be due to batch variation or other variables therefore care must be taken when interpreting these results. Also intact undissected embryo results were ambiguous therefore the experiments should be repeated and complementary techniques such as *in situ* hybridisation should be performed to localise the expression of XbmycA precisely in tissues during development.

For each tissue piece cytoskeletal actin expression increases as the embryo stage increases which might be expected since the cell number is increasing. The increase between stages 24 and 26 might not be as dramatic as that between stages 26 and 33/34 since the former development takes ~3 hours whereas the latter takes ~15 hours and therefore has had more time to accumulate cytoskeletal actin mRNA. This increase in expression only approximates to what was previously described.

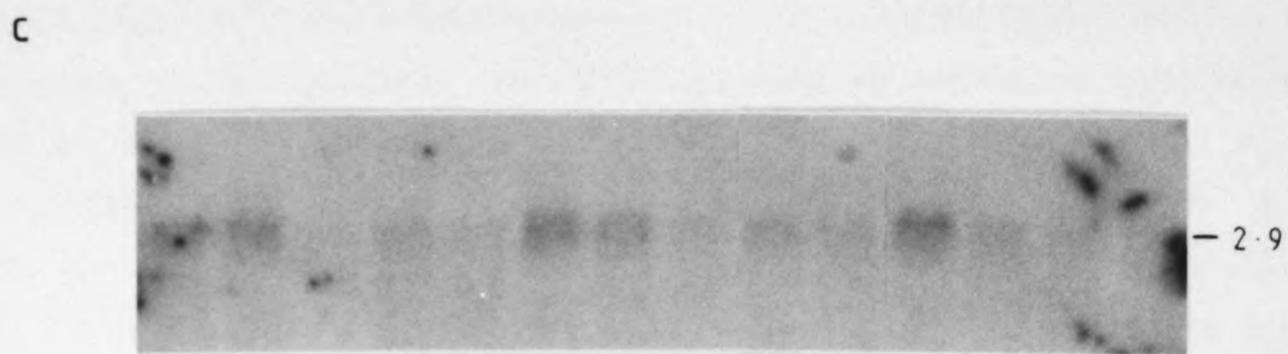
For gut and neural tube tissue XbmycA expression increases between stage 24 and 26 then decreases between stage 26 and 33/34 as well as between stage 24 and 26. Since the total expression of XbmycA is similar to the previous experiment (figure 4.9) and since cytoskeletal actin expression increases at all stages it seems likely this is a true result. Therefore this suggests that many cell types in late neurula/early tailbud require XbmycA for their growth but as the embryo grows older only a limited set of cells in the head require XbmycA. Thus it may be that XbmycA is involved in the early proliferation and differentiation of many cell types in specifically in the early differentiation and/or proliferation in *Xenopus* brain which may be analogous to that described for the mouse cerebellum (Ruppert, Goldowitz and Wille 1986).



XBM YCA 2



CYTOSKELETAL ACTIN



XBM YCB 1

FIGURE 4.10

Expression of XbmycA in dissected embryo pieces. Embryos were grown to the indicated stage of development (Nieukoop and Faber 1956) then dissected into a head (H), gut (G) and neural tube (N) piece or left intact (I). Total RNA from 5 embryo pieces or 5 intact embryos were used for each lane in northern analysis. All probes had roughly the same specific activity of $\sim 1 \times 10^9$ cpm/ μ g and were used under high stringency conditions. Final washing conditions were 0.2x SSPE, 60⁰C. Sizes were estimated from *X.borealis* 18S and 28S rRNA. (a) XbmycA2 probe, 1 day exposure. (b) the filter from (a) stripped and reprobed with cytoskeletal actin, 1 day exposure. (c) the same filter stripped and reprobed with XbmycB1, 3 day exposure.

— CYTOSKELETAL ACTIN
--- XBMYCA2

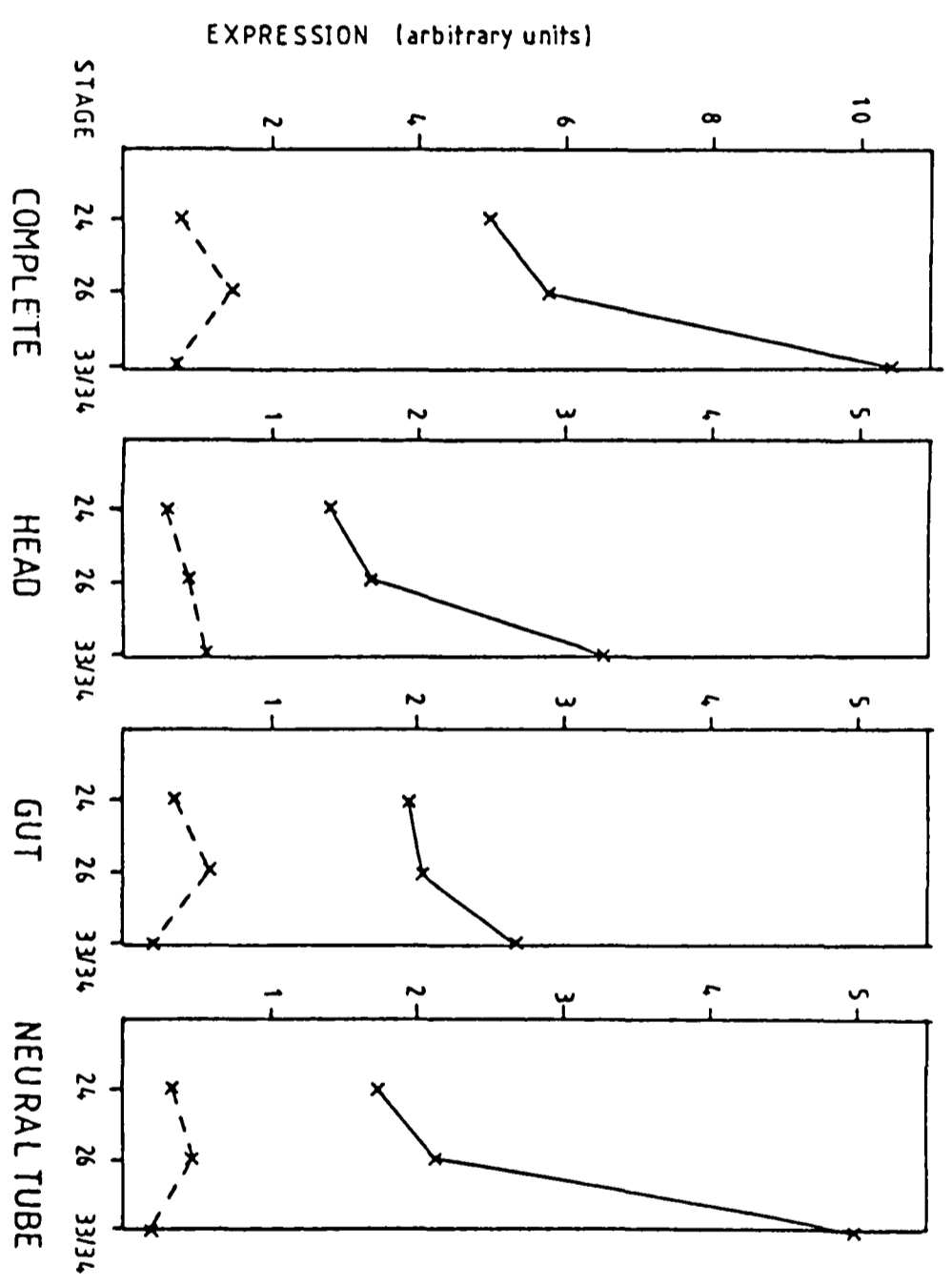


FIGURE 4.11

Expression of XbmycA in dissected embryos. Densities from overnight exposures of the filters from Figure 4.10 were plotted (in arbitrary units) for each tissue piece at each embryonic stage indicated. Stage 24 is late neurula, stage 26 is tailbud and stage 33/34 is a swimming tadpole stage (Nieukoop and Faber 1956). The total densities of all three separate embryo pieces at each stage is also shown. Similarly the expression of cytoskeletal actin is also plotted. For both probes the expression of all three embryo pieces together roughly parallels that shown in Figure 4.9 although it seems to be delayed slightly which is probably due to batch variation. Again for the reasons described previously in Figure 4.9 in and the text it is difficult to quantitate the absolute expression levels between the two probes used.

Once again experiments on the above filters using XbmycB as a probe have not been very successful but examination of figure 4.10 suggest it follows a similar pattern to XbmycA. Again this may point to both genes being regulated in a similar manner.

4.5 Attempts to Analyse XbmycA Protein Expression

In order to study the expression of XbmycA protein I tried to make antibodies to the N-terminal region of the coding sequence for use in western blotting analysis. The 1.3kbp XbmycA1 *EcoRI* fragment was end repaired and then subcloned into the *SmaI* site of pEX2 (Stanley and Luzio 1985) giving a construction which has the 100 amino acids of XbmycA1 fused, in frame, onto the N-terminus of β galactosidase. When ^{E.coli}pop 2136 is transformed it expresses a XbmycA1- β galactosidase fusion protein under the control of a temperature sensitive cro repressor.

Figure 4.12 indicates that the XbmycA1- β galactosidase fusion protein has an increase in molecular weight of roughly 12kD compared with β galactosidase alone which would be expected from sequence of XbmycA1. This figure also shows the products to be inducible and that cells alone do not express these products. Finally this figure shows that when the XbmycA1 insert is cloned into pEX2 in the reverse orientation it does induce a fusion protein which is slightly bigger than β galactosidase but seems to be more unstable probably due the poly(A) tail giving rise to a poly-lysine tract in the protein.

The fusion protein was used to immunise three rabbits as described in Chapter 2 and sera obtained were used in western blotting analysis. All rabbits used reacted in an apparently similar manner hence representative western filters are shown in figure 4.12. Pre-immune sera did not react with any *Xenopus* ovary proteins. When an anti- β galactosidase serum (supplied by A.Khan) was used this reacted with both the vector and fusion *E.coli* proteins as expected but this serum also reacted with two *Xenopus* protein bands weakly (~55, ~90kD) therefore this had to be remembered when judging the results of the anti-XbmycA1- β galactosidase experiments.

All the anti-fusion sera when used in western blot analysis reacted with *E.coli* fusion and vector proteins in a similar manner to the anti- β galactosidase serum. No specific reaction was found in the *Xenopus* ovary

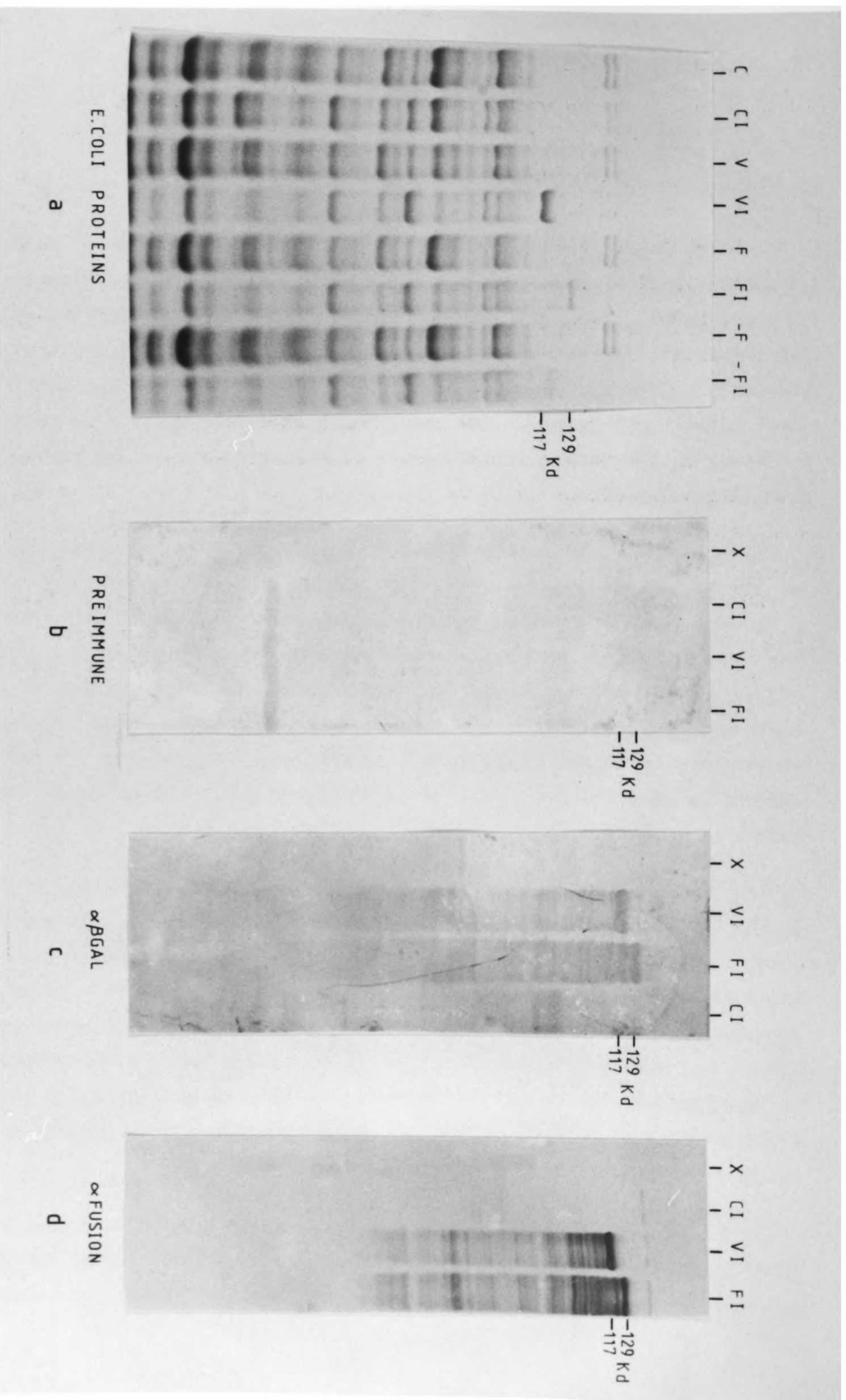


FIGURE 4.12

Generation of a XbmycA - β galactosidase fusion protein and western blotting analysis with antisera generated against it. All proteins and antisera were prepared as described in chapter 2. (a) Coomassie blue stained 7.5% protein gel of pop2136 cell proteins uninduced (C) and induced (CI); pop2136 containing pEX2 (vector) cell proteins uninduced (V) and induced (VI); pop2136 containing pEX2 - XbmycA1 in frame fusion (see text) cell proteins uninduced (F) and induced (FI); pop2136 containing pEX2 - XbmycA in the reverse orientation (see text) cell proteins uninduced (-F) and induced (-FI). The vector pEX2 generates a β galactosidase protein of ~117kD and the pEX2 - XbmycA construct generates a ~129kD β galactosidase - XbmycA fusion protein. Sizes were estimated from BRL protein molecular weight markers. (b) Western blot analysis of the proteins described above plus a *X.borealis* oocyte protein extract (X). 1:200 dilution of rabbit preimmune serum. Antibody - antigen interactions were visualised as described in chapter 2. (c) Similarly western blot analysis using a rabbit anti β galactosidase antiserum (kindly supplied by A. Khan) at a 1:200 dilution. (d) Similarly western blot analysis using a rabbit anti β galactosidase - XbmycA1 antiserum generated as described in chapter 2. This antiserum was affinity purified against the fusion protein before use as described in chapter 2.

protein tracks; in fact many non-specific bands were present in the 60–70kD range. To overcome this problem the anti-fusion sera were affinity purified against the fusion protein as described in chapter 2. When such affinity purified sera were used reaction with the *E.coli* vector and fusion proteins was observed, but not with any *Xenopus* ovary proteins (figure 4.12). The presence of many bands is due to breakdown products of these large fusion proteins.

The reasons why I have not obtained an anti-XbmycA1 antiserum are unclear. It could be that the rabbits were not injected enough times. Since the epitope I am interested in is so much smaller than the β galactosidase moiety (which is an excellent antigen) rabbits may take longer time to produce a response also if the rabbit *c-myc* protein sequence is very similar to the *Xenopus* XbmycA1 protein sequence this too may mean that it is less immunogenic. I would expect that the *Xenopus* sequence to be sufficiently different to be a good immunogen since good antisera to the closely related human *c-myc* protein has been raised in rabbits previously (Persson *et al* 1984, Ramsay, Evan and Bishop 1984). It may be that the XbmycA1 carboxyl terminus is not large enough or not well exposed in the fusion protein to be available for immunisation, though a similar region in the protein was used to generate a human antisera (Persson *et al* 1984). The production of other fusion constructs might solve this problem. Another possibility is that I have generated anti-XbmycA1 antibodies but they are either not high titre or have a low avidity. Therefore if the western technique was used with greater sensitivity (e.g. using alkaline phosphatase or streptavidin amplification systems) it may then detect the XbmycA protein. Possibly immunoprecipitation instead of western blotting would detect the XbmycA protein but since the antisera were raised against denatured proteins I would expect western blotting to have worked well.

4.6 Discussion

In this chapter I described the successful isolation of two different types of *c-myc* genes from *X.borealis*. The XbmycA gene appears to be the equivalent of *c-myc* in other species whereas the relationship of XbmycB to other *myc* genes is unclear although it is *c-myc* related at its 3' end. The *X.borealis* frogs are not genetically inbred therefore it is vital to obtain the complete coding sequence of XbmycA on a single clone in case genetic variation complicates sequence analysis. Indeed chapter 5 indicates the XbmycA clones I have

obtained do differ in sequence which may be due to different alleles or due to different genes and will be discussed in chapter 5.

The southern analysis I describe suggests the genomic structure of XbmycA in *X.borealis* and *X.laevis* may be similar to other *c-myc* genes. This may imply that this kind of genomic structure is important for the regulation of this type of gene. Indeed an analysis of *c-myc* genes that have been altered in genetic structure by chromosomal translocations suggests the post-transcriptional stability of the mRNA is determined by the genomic organisation of the *c-myc* locus (Piechaczyk *et al* 1985, Eick *et al* 1985, Piechaczyk *et al* 1986). The genomic structure of XbmycA would become readily available by the isolation of a clone from a genomic library.

Interestingly it appears that *X.tropicalis* does not have an equivalent to XbmycB. This may point to an event creating a gene duplication in the *X.borealis* and *X.laevis* lineage which may have occurred as described in the previous chapter. Sequence analysis of XbmycB will show how similar the protein products of XbmycA and B are and will show if it is actually a functional gene. Understanding the relationship of these genes is important because if they are very similar then why have two genes evolved or if they are quite distinct then what is the functional basis for these differences. Also, since *X.tropicalis* develops in an apparently similar fashion to *X.borealis* and *X.laevis*, it is important to determine if both XbmycA and B are fully functional to help understand the selection pressure on genes apparently involved in cell division control. Both XbmycA and B give rise to similar sized transcripts which implies both genes are functional in transcription but XbmycB might not be as highly expressed as XbmycA. Sometimes these transcripts appear as doublets therefore it is necessary to see if both size classes come from the same gene with more than one promoter or from different genes and what factors govern their regulation. Assessment of the regulation of these genes will be difficult due to post-transcriptional controls apparently involved (see chapter 1).

XbmycA is expressed in many tissues in agreement with results on *c-myc* expression in other vertebrates (Stewart, Bellve and Leder 1984). The highest level of expression is in the ovary which is probably a reflection on storage of mRNA in the oocyte for embryogenesis or the extremely high levels of gene expression in these cells. Expression levels of *c-myc* are low in most other

tissues which may reflect the fact that only a small number of cells in these tissues are actually dividing and expressing XbmycA in a cell cycle dependent manner. *C-myc* has been regarded as a cell-cycle "enabling" signal whose presence is required either continually or else during a sensitive temporal window to allow a cell to progress through the cell cycle (Evan *et al* 1986) and hence *c-myc* transcripts would only be present in dividing cells. The tissue specific expression of XbmycB remains to be determined.

XbmycA is made early in oogenesis and some of this maternal message is inherited by the dividing embryo. XbmycA is probably also required for the growth of the oocyte since its expression drops during oogenesis. It remains to be determined if XbmycA transcripts turns over or if it is made only once during oogenesis. If it was made only once then it would be of great interest to find what factors stabilize the message such that it can be inherited by the embryo. Many factors govern the maturation of stage VI oocytes into eggs. Progesterone is the *in vivo* signal and exposure to progesterone *in vitro* will also cause oocyte maturation (Wasserman, Houle and Samuel 1984). Interestingly another oncogene protein, the *ha-ras* protein, can induce maturation of *Xenopus* oocytes upon microinjection (Birchmeier, Broek and Wigler 1985). It would also be interesting to determine if high levels of XbmycA protein induce maturation thereby implying a role in signalling division control.

The amount of XbmycA maternal message inherited by the dividing embryo is just enough to last until the gene is transcriptionally reactivated at stage 11-12. This suggests that dividing cells in the embryo only require low amounts of the XbmycA product hinting at a role of XbmycA in the actual cell division process. If one could block the translation of XbmycA by the injection of antisense message then it would be exciting to see if the cells in the embryo are still competent to divide or if there are other determinants governing the decision to divide (antisera to XbmycA may also have to be injected to block the function of any maternal protein that is inherited). Since XbmycA mRNA is not detectably accumulated soon after the midblastula transition when a general activation of RNA synthesis occurs (Newport and Kirschner 1982) then this supports the idea that dividing embryonic cells only require low amounts of *c-myc* protein.

The highest levels of XbmycA expression are at mid-late neurula stages

20-24 which may imply a role in neural development. Experiments on dissected embryos suggest that XbmycA is expressed at similar levels in each piece analysed (head, gut and neural tube). This may be because there is a rapid burst of transcription to replace the XbmycA that was lost in earlier development and therefore XbmycA might not be involved in the processes of differentiation at these stages but rather in cell division. As the embryo grows older there does appear to be a greater level of XbmycA expression in the head piece hence it may be involved in the later differentiation of the brain.

C-myc RNA levels are increased during the differentiation of mouse Purkinje and granule cells in the brain (Ruppert, Goldowitz and Wille 1986) therefore this correlates with the localisation of XbmycA expression in the *Xenopus* head at stage 33/34. Purkinje cells have undergone their last cell division (Miale and Sidman 1961) hence this implies that *c-myc* may also be involved in differentiation as well as cell division. Indeed in neuronal-pheochromocytoma (PC12) and erythroleukemia tissue culture cells that have been induced to differentiate by NGF and DMSO respectively there is a transient increase in *c-myc* RNA levels before repression of *c-myc* RNA synthesis (Lachman and Skoultchi 1984, Curran and Morgan 1985). Therefore *c-myc* might be involved in signalling pathways that are involved in brain and erythroid cell development. *In situ* hybridisation of XbmycA to early *Xenopus* embryos would provide a clear insight to the functions of *c-myc* in the differentiation of the brain and other tissues.

The isolation of a genomic clone of XbmycA would allow an analysis of the regulatory regions of the gene. Injection of plasmid constructs into *Xenopus* eggs would help to determine which sequences are responsible for the transcriptional activation of these genes at stage 11-12 and may help elucidate what factors are responsible for this. Expression of XbmycA under the control of other promoters that are differentially activated during development may help us to understand the role of XbmycA. For example by putting the XbmycA gene under control of the promoter for GS17 would see what effect the early transcriptional activation of XbmycA at gastrulation (stage 9) would have on the dividing embryo (Krieg and Melton 1985).

I have preliminary evidence that XbmycB is expressed in a similar fashion to XbmycA though this needs further testing. If both genes appear to be regulated in the same manner then the question remains why both genes

appear to be different from each other. It is possible that they both carry out similar functions but they may respond to slightly different stimuli in a large regulatory network. Possibly having two slightly different proteins makes them more sensitive to changes in environment than having only one protein hence growth control decisions can be accurately controlled.

The expression of XbmycA could be further analysed by obtaining antibodies to the protein. This would open up a whole series of experiment which would help us understand XbmycA function. The expression of *myc* protein could be followed during development, the localisation of protein within cells and embryos could be determined, the blocking of XbmycA function could be attempted, the isolation of XbmycA protein from cells would be possible and the isolation XbmycA complexes with other proteins and nucleic acids could be done as well as many other experiments.

I described my attempts to make antibodies to the XbmycA protein which were unsuccessful, but other ways of detecting the *Xenopus c-myc* protein exist. For example it is quite possible that mouse or human anti-*c-myc* antisera will react with the XbmycA protein indeed preliminary evidence suggests this may be the case (Persson *et al* 1984). Alternatively it is conceivable that an anti-peptide antiserum to a region that is conserved in all mammalian *myc* genes (*c*, *N* and *L-myc*) which has been prepared (Evan *et al* 1986) will react with *Xenopus myc* proteins as this region is highly conserved in XbmycA (see chapter 5).

CHAPTER 5

SEQUENCE ANALYSIS OF C-MYC RELATED CLONES FROM X.BOREALIS

5.1 Introduction

Chapters 3 and 4 describe evidence that the cDNA clones from *X.borealis* I have obtained are related to *c-myc* genes of other species. The best proof that these clones are *c-myc* related comes from sequence analysis described in this chapter. Sequencing these genes will provide a basis for determining which parts of the protein coding domain are conserved from species to species thereby implying these regions play an important functional role in the protein. Indeed two such domains have been previously defined in *N-myc*, *L-myc* and *c-myc* in a variety of species (Schwab *et al* 1983, Nau *et al* 1985, Schwab 1985, Van Beneden *et al* 1986). Comparison of the different clones with each other and with a *X.laëvis c-myc* sequence which I recently obtained from M.King, Seattle may provide clues to the evolution of these sequences in *Xenopus*. Also having the sequence of these clones provides information that can be used in the design of future experiments.

5.2 XbmycA1

Initially part of the XbmycA1 clone was sequenced to show it was *c-myc* related before going on to use it to isolate the other Xbmyc clones from my previtellogenic cDNA library. 500 bp of the 5' end of this clone were sequenced using the strategy shown in figure 5.1. Figure 5.2 shows the sequence determined from the XbmycA1 clone and shows that it contains 301 bp of coding sequence and therefore ~1.0kb of 3' non-coding sequence (although this remains to be proved conclusively for this clone).

To determine if the XbmycA1 protein and nucleic acid sequences are related to the *c-myc* genes of other species they were compared to the chicken and human *c-myc* protein and nucleic acid sequences using the program BESTFIT. Figure 5.3 shows the 100 carboxyl terminal amino acids encoded by this clone are highly homologous to the respective region in the chicken and human sequences. The human and XbmycA1 sequences are ~71% homologous and the chicken and XbmycA1 sequences are ~79% homologous at the amino acid level with no insertions or deletions. Of 29 substitutions

between the human and XbmycA1 sequences 11 are conservative changes and of 21 substitutions between the chicken and XbmycA1 sequences 14 are conservative changes (this assumes amino acids are similar as defined by the UWGCG programs where P,A,G,S,T are neutral, weakly hydrophobic; Q,N,E,D,B,Z are hydrophilic, acid amine; H,K,R are hydrophilic, basic; L,I,V,M are hydrophobic; F,Y,W hydrophobic aromatic and C is cross-link forming).

When the nucleic acid sequence of XbmycA1 is compared to the human and chicken sequences as shown in figure 5.4 and 5.5 they are also found to be highly related especially over the coding regions. The human and XbmycA1 sequences are 68% homologous over the first 304 by of coding sequence (i.e. to the stop codon). This homology breaks down after the stop codon although they are still related there are multiple deletions or insertions in the sequences. Similarly the chicken and XbmycA1 sequences are 68% homologous over the coding region and the non-coding region is less homologous due to multiple insertions and deletions. The non-coding region of XbmycA1 is slightly more related to that of chickens than humans since it has 126 matches compared with 104 matches over this region.

From the above data I was confident that I had isolated a portion of a *bona fide c-myc* gene from *X.borealis* which seems more related to chicken *c-myc* than human *c-myc*. As described in chapter 4 this clone was used to isolate all the other Xbmyc clones I obtained in this study hence it was highly likely that they would also be *c-myc* related. The sequence analysis of XbmycA2 and 3 is described in the following sections. No sequence of the XbmycB clones has been determined as of yet.

5.3 XbmycA2 and XbmycA3

Sequencing strategies for XbmycA2 and 3 are shown in figure 5.1. An oligonucleotide was used to help sequence XbmycA2 on both strands. Note that the sequencing does not span the internal *EcoRI* site of XbmycA2 therefore the presence of a small *EcoRI* fragment cannot be ruled out. Sequences of XbmycA2 and 3 and their putative translation products are presented in figures 5.6 and 5.7.

The internal *EcoRI* site of XbmycA2 may be due to a polymorphism at this locus or due to a cloning artifact maybe created during reverse transcription. Since both fragments of XbmycA2 detect the same bands on southern blots

(figure 4.3) and XbmycA1 does not have an internal *EcoRI* site then this suggests that the internal XbmycA2 site is due to a cloning artifact. Evidence against this being a cloning artifact is that XbmycA2 and XbmycA1 were isolated from different libraries and that each library was made from many frogs. Support for the internal *EcoRI* site in XbmycA2 being a polymorphism also comes from the fact XbmycA4 has an internal *EcoRI* site and that only a single frog was used in the southern analysis (remember the frogs are not inbred). Further southern analysis on different frogs from different matings would help resolve this problem.

Both XbmycA1 and 3 DNA sequences overlap the XbmycA2 DNA sequence therefore they were compared to see if they had the same DNA sequences using the BESTFIT program. Figures 5.8 and 5.9 show that the regions of overlap are highly related but they are different. Comparing XbmycA1 with XbmycA2 shows that the coding regions differ by 7.2% but the non-coding regions differ by 13.8% including 4 small insertions. XbmycA3 and 2 only overlap in coding sequences and have a sequence difference of 7.4% including a 6 bp insertion. Figure 5.10 shows that the putative protein coding sequences of XbmycA1 and 2 have 4% difference whereas the XbmycA3 and 2 protein coding sequences have 6.6% difference including an insertion of two amino acids. Some of the nucleotide changes therefore must be silent and also 3 of 4 amino acid changes between XbmycA1 and 2 and 3 of 6 amino acid changes between XbmycA3 and 2 are conservative substitutions.

From the above data it is obvious that these clones do not arise from the same gene. For XbmycA1 and 2 this might be partly explained by the fact that they come from different libraries and that each library was made from frogs from different matings. It is not so easy to explain the differences between XbmycA3 and 2 as they were isolated from the same library. The library that these clones came from was made using the RNA of many frogs which were from the same mating. Considering that *X.borealis* is tetraploid and assuming that XbmycA is a single copy gene then it is possible that there will be a maximum of 8 different kinds of XbmycA gene in my library. I think it is unlikely that XbmycA3 and 2 are different alleles from the same frog since a nucleic acid difference of 6.6% is quite a high value for pairs of alleles.

Another possibility exists that there could be two XbmycA genes. Indeed the southern data presented in the previous chapter suggests there may be

two highly related XbmycA genes although their genomic organisation remains to be resolved. The sequence divergence between XbmycA2 and 3 is 6.6% which is similar to the sequence divergence between two A type vitellogenin (A1,A2) and the two B type vitellogenin genes (B1,B2) in *X.laevis* which is ~7% for both types (Wahli and Ryffel 1985) hence it is possible that there are two highly related XbmycA genes. Isolation of genomic clones and further sequence and southern analysis should resolve how many XbmycA genes exist.

5.4 XbmycA and XlmycA

Recently I was very kindly send the nucleic acid sequence of a *X.laevis* *c-myc* cDNA clone (which I have named XlmycA) from M.King, Seattle. I have used the program BESTFIT to compare this sequence with the XbmycA2 and 3. To use the BESTFIT program I had to split the sequences into small pieces before the comparison since the program can only compare two 1000 bp sequences at the most. The combined data of these comparisons is shown in figure 5.11 and 5.12.

The nucleic acid sequences of XbmycA3 and XlmycA show 6.6% divergence over the region they overlap. Homology between these two sequences breaks down rapidly upstream of the initiating methionine codon. Over the coding region XbmycA2 and XlmycA show 8.2% divergence and over the non-coding region they show 14.1% divergence which has multiple insertions and deletions. Therefore this analysis shows that the *X.borealis* and *X.laevis* *c-myc* genes are highly related especially across the coding regions. The 3' non-coding regions are also highly conserved but not as much as the coding regions. It is interesting that the 3' non-coding regions are so highly conserved since it is generally thought that non-coding sequences evolve rapidly accumulating many changes (this appears to be true of the 5' non-coding sequences) therefore this may mean these sequences are important in the regulation of these genes in this system. It is also interesting that the divergence between the XbmycA type genes and the XlmycA gene is as high as the divergence between the XbmycA type genes. A result which might indicate that a XbmycA gene duplication event took place around the same time as the divergence of the *X.borealis* and *X.laevis* lineages. This also suggests, because the sequence divergence is small, that this event was relatively recent, estimated to by ~30MY by Wahli and Ryffel, 1985.

Figure 5.13 shows that the putative translation products of XbmycA2 and 3 and XlmycA are very highly related. The XbmycA3 and XlmycA peptide sequences differ by 5.2% including a 4 amino acid insertion in XlmycA. The XbmycA2 and XlmycA peptide sequences have 6.2% divergence with a deletion of a single amino acid in XbmycA2. Again many of the changes are conservative substitutions (3 from 11 changes between XbmycA2 and XlmycA and 7 from 19 changes between XbmycA3 and XlmycA). Hence these sequences are very closely related with the divergence between species similar to that within species. Therefore this also supports the *X.borealis* and *X.laevis* lineages arising from the same event and that this event was relatively recent.

The extra 4 amino acids in the XlmycA sequence compared with the XbmycA3 sequence are right at the N-terminus such that the initiating methionine of XbmycA3 matches a methionine residue at amino acid position 5 of XlmycA. It is not known which methionine is used *in vivo* by the XlmycA mRNA. Computer analysis of many eukaryotic mRNAs suggests the 5 proximal AUG triplet generally serves as the initiation codon and a consensus of CC A/G CCAUG has been proposed in which the A 3 bp upstream of the AUG is very highly conserved (Kozak 1984). If true, this might suggest that the second in frame AUG of XlmycA serves as the initiating codon because only the second AUG has an A 3 bp upstream. Note the initiating AUG in XbmycA3 also has an A at a similar position. See figure (5.12) for the positions of the initiation AUG codons.

5.5 Features of XbmycA Nucleic Acid Sequences

5.5.1 Poly(A) addition site

Most eukaryotic mRNAs are post-transcriptionally modified by the addition of a poly(A) tail of 100–250 nucleotides to their 3' ends by the action of a poly(A) polymerase. Part of the recognition sequence for this process has been defined as AAUAAA which lies about 20 nucleotides from the 3' end of the mRNA and has been found in most poly(A) containing mRNAs (Proudfoot and Brownlee 1976). The XbmycA2 sequence contains one copy of this sequence which lies 16 nucleotides from the 3' end of the mRNA (figure 5.6); also XlmycA has one such sequence 19 nucleotides from the 3' end of the mRNA. Therefore it seems that the XbmycA gene(s) have structures that typify them as normal eukaryotic mRNAs.

5.5.2 Rapid degradation signals

Very recently it has been reported that the mRNA of transiently expressed genes contain a conserved AU-rich sequence in their 3' untranslated regions (Shaw and Kamen 1986). The sequence AUUUA surrounded by AU-rich nucleotides has been defined as a possible sequence that mediates selective mRNA degradation and is found in the 3' untranslated regions of several lymphokine and cellular oncogene mRNAs. As described in chapter 1 *c-myc* mRNA in a variety of cell types is highly unstable with a half life of 20-30 minutes (Dani *et al* 1984, Rabbitts *et al* 1985). Shaw and Kamen note that *c-myc* mRNA does contain such an AUUUA consensus in an AU-rich region which may reflect the instability of *c-myc* mRNA.

It would appear that the *X.borealis c-myc* mRNA during embryogenesis has a longer half life than mammalian *c-myc* mRNAs (see chapter 4) but this may reflect the temperature difference between organisms and also the mode of embryogenesis that *Xenopus* exhibits. Also the *X.borealis c-myc* mRNA might be more stable during oogenesis (see chapter 6). Interestingly I have found six copies of the AUUUA consensus in the 3' untranslated region of XbmycA2 (figure 5.6). All of these AUUUA sequences lie in relatively AU-rich regions. Therefore it seems that the *X.borealis c-myc* gene may contain signals that make the mRNA potentially highly unstable. The actual half life of *c-myc* mRNA in *Xenopus* remains to be determined both in embryonic and somatic cells. Note that the XlmycA mRNA also has six AUUUA sequences in its 3' untranslated region.

5.5.3 Comparison with other *c-myc* sequences

To compare the XbmycA sequences with the *c-myc* sequences from other species the programs COMPARE and DOTPLOT were used to provide a graphical representation of the homologies between the sequences. The XbmycA2 and 3 sequences were assembled together to form one XbmycA sequence (bases 1-564 of XbmycA3 and bases 1-2011 of XbmycA2) to help simplify the analysis. This is justified on the basis that because the differences between XbmycA2 and 3 are so small (where they overlap) then they would have little effect when comparing the large differences that occur between species. I have used the COMPARE and DOTPLOT programs rather than the BESTFIT program multiple times because it is much easier to obtain an overall

impression of the divergence between sequences using COMPARE and DOTPLOT especially since the sequences are so large. Also COMAPRE and DOTPLOT may show regions that are repeated within sequences and may show homologies that are present in regions such as introns. For this analysis I have used trout, chicken and human *c-myc* sequences (Van Beneden *et al* 1986, Waston *et al* 1983, Shih *et al* 1984, Colby *et al* 1983, Gazin *et al* 1984) to compare with the XbmycA sequence. No extra information is gained by using the mouse and viral *c-myc* sequences since they are so similar to human and chicken *c-myc* sequences respectively. Each comparison was performed using a window of 21 and a stringency of 14.

Figures 5.14, 5.15 and 5.16 show the output generated by comparing XbmycA with human, chicken and trout *c-myc* sequences respectively. Where known the mRNA start sites, introns, exons and mRNA finish sites are shown on the diagram. From all these diagrams it can be seen that the XbmycA coding sequences are highly related to the *c-myc* coding sequences of other species. To obtain a rough estimate of the extent of homology between each sequence the length of the breaks in the diagonal that is plotted where the homologies occurs was divided by the total length of the diagonal. This only provides a rough guide to the divergence between the two sequences as a window of 21 was used with a stringency of 14. Table 5.1 summarises the results of this comparison and the data of Van Beneden *et al* 1986 for the trout vs human and chicken vs human comparisons and indicates that the way I have measured the divergence between two sequences is, as expected, an overestimate. So I have corrected all the values by multiplying by a factor of 0.82 which accounts for the average discrepancy between the values measured from the DOTPLOT figures and the data of Van Beneden *et al* 1986. Again I must emphasise this is only an estimate of divergence and comparisons between amino acid sequences discussed later provide better estimates of this parameter.

The first point to note from this comparison is that the *c-myc* genes from many species are highly homologous which is extremely strong evidence that these genes play an important function that has been evolutionarily conserved. Secondly the third exon shows the highest conservation in all comparisons and hence may be extremely important for *c-myc* function. Indeed the third exon protein sequences are implicated in DNA binding as it contains far more basic amino acids than the second exon protein sequences (Persson and Leder 1984).

	XbmycA vs human	XbmycA vs chicken	XbmycA vs trout	Trout vs human	Chicken vs human
total	70(57)	72(59)	65(53)	73(60*)	70*
exon2	63(52)	65(53)	63(52)	71(55*)	67*
exon3	80(66)	79(65)	67(55)	76(66*)	74*

Table 5.1

Homologies between *c-myc* nucleic acid sequences using the output from figure 5.14, 5.15 and 5.16. * Data from Van Beneden *et al* 1986. Figures in brackets are values corrected for the discrepancies between the values obtained from dotplots and the values from Van Beneden *et al* 1986. All figures are the percentage homology between the two sequences under comparison.

This analysis also suggests that the *Xenopus c-myc* nucleic acid sequence is more closely related to the chicken and human *c-myc* nucleic acid sequences than to the trout *c-myc* nucleic acid sequence. It is not really possible to derive a phylogeny of the various species using the divergence, as derived above, of the *c-myc* genes but one can say that the *Xenopus* sequence is apparently more related to mammalian and avian *c-myc* sequences than fish *c-myc* sequences.

The intron/exon boundaries of the human, chicken and trout *c-myc* genes are clearly delimited as shown in figures 5.14, 5.15 and 5.16. It is interesting to note close to the end of exon 2 of the human and chicken *c-myc* genes and the presumptive second exon of the trout *c-myc* gene that the sequence of XbmycA diverges significantly from the other sequences. Also it is of interest that at the beginning of the third exon that there appears to be a cluster of short simple repeats which might suggest slippage-like mechanisms are a source of genetic variation at this position (Tautz, Trick and Dover 1986). It may also be that both the above points suggest that the protein sequences encoded by each exon form two different domains of the protein.

It would appear from figures 5.14, 5.15 and 5.16 that the 3' non-coding region of XbmycA is related most closely of the 3' non-coding regions of chickens but it has diverged by the presence of multiple insertions and deletions. The human and trout *c-myc* 3' non-coding sequences are also related to the same region of XbmycA but it seems they have diverged significantly. The XbmycA sequence has a 3' non-coding sequence of ~1.0kb which is significantly longer than the same region in other *c-myc* genes (~200 bp). It is not clear what this increase in size of the 3' non-coding sequence of XbmycA means but it might be involved in the regulation of stability of *c-myc* mRNA in oocytes compared with somatic cells if indeed any difference does exist.

From figure 5.14, 5.15 and 5.16 it can be seen that the 5' non-coding sequences of XbmycA are unrelated to any other *c-myc* 5' non-coding sequences. Note size estimation of XbmycA mRNA by northern blotting in chapter 4 suggested that the XbmycA sequence is lacking ~400bp of 5' sequence information therefore it is possible that within this missing region the XbmycA gene does have 5' sequence information related to that of other *c-myc* genes. The *c-myc* genes of humans, mice and chickens contain an

upstream exon (exon 1) which does not contribute to the coding information of the *c-myc* protein but may be involved in the regulation of the *c-myc* genes (see chapter 1, Bernard *et al* 1983, Gazin *et al* 1984, Shih *et al* 1984). This exon is ~76% homologous between humans and mice but not related at all to any chicken sequences (Bernard *et al* 1983, Shih *et al* 1984) therefore it could be possible that XbmycA has three exons but the first exon is completely unrelated to any other sequences (indeed a WORDSEARCH of the upstream sequence of XbmycA on the GENBANK and EMBL data bases finds no significantly related sequences). It has been reported that the human *c-myc* first exon sequence has a provocative open reading frame which could encode a protein of 30KD (Gazin *et al* 1984). Whether this frame is translated or not *in vivo* remains to be determined but it is possible that translation of this exon modulates the translation of exon 2 and 3. Interestingly the first ~161 bp of the XbmycA sequence is an open reading frame as shown in figure 5.17 which shows all possible open reading frames (note the large open reading frame which encodes the putative *c-myc* protein). It remains to be determined if the 5' open reading frame extends further upstream. Although it is not clear how the *c-myc* gene is regulated these results might suggest that a common mechanism exists for the regulation of the *c-myc* genes in many species.

5.6 Features of XbmycA Protein Sequences

5.6.1 Comparison with other *c-myc* protein sequences

The putative translation products of the XbmycA sequences were compared with the putative translation products of human, mouse, chicken, viral, trout and XlmycA *c-myc* sequences by using the program GAP reiteratively with the command line option out until no further gaps are introduced then using the program PRETTY to align the sequences. The alignment was then manually edited to increase the matches. Results of such a comparison are shown in figure 5.18, note that a consensus has been derived using a plurality of 5, to highlight the conservation between sequences (note this will be slightly biased where the XbmycA sequences overlap).

It can be seen from figure 5.18 that all the *c-myc* protein sequences are highly related to each other and often at positions of difference the changes are conservative. Residues 379-449 on this figure show the highest conservation between all the *c-myc* sequences and contains 26% basic amino

acids which might reflect a DNA binding function. Residues 450-491 are also rich in basic amino acid: but there are many conservative changes between the protein sequences. Other regions are also very highly conserved, for example, the regions which span the proposed *myc* boxes (Schwab *et al* 1983, Nau *et al* 1985, Van Beneden *et al* 1986) are also very homologous (this is discussed further in section 5.7).

It is of interest that each *c-myc* sequence has inserted regions that are unique to themselves in the second exon. For example residues 82-99 and 242-262 are unique to the mammalian *c-myc* sequences and residues 174-193 are specific for the chicken, viral and *Xenopus c-myc* sequences. The trout *c-myc* sequence also has residues that are specific for it alone at positions 210-216 (Van Beneden *et al* 1986) but the alignment in figure 5.18 does not show this well. Indeed this region is actually specific for each type of *c-myc* sequence i.e. either mammalian, avian, amphibian or fish *c-myc* sequences are unique at this position. Other than these insertions and discounting other small insertions and deletions it appears that the *c-myc* coding sequences have gradually evolved by the accumulative acceptance of point mutations.

On comparing each amino acid sequence using the program GAP a value of the percentage similarity is generated which I have used to estimate the sequence divergence between *c-myc* sequences and is shown in table 5.2. These values agree quite closely with the data of Van Beneden *et al* 1986 therefore this analysis may give clues to the evolution of the *c-myc* protein sequences. In this analysis I have used the assembled XbmycA protein sequence for reasons outlined earlier.

As expected (and as described earlier) the *X.borealis c-myc* protein sequence is most closely related to the *X.laevis c-myc* protein sequence again suggesting both lineages arose from the same, relatively recent, event. The XbmycA sequence is next most closely related to the chicken, human, mouse and viral *c-myc* sequences and is least related to the trout *c-myc* sequence. Since the divergence between the XbmycA and human, XbmycA and chicken and the human and chicken comparisons are similar, each protein may have large blocks of homology interspersed with regions that are unique and free to mutate. As discussed earlier this is apparently the case. This might suggest that all the *c-myc* proteins play a common and important function in each species.

XbmycA vs human	66.9	human vs mouse	92.0
XbmycA vs mouse	66.9	human vs chicken	70.9(68)
XbmycA vs chicken	67.0	human vs trout	63.2(62)
XbmycA vs viral	66.2	chicken vs trout	61.0(64)
XbmycA vs XlmycA	94.7		
XbmycA vs trout	64.5		

Table 5.2

Homologies between *c-myc* protein sequences showing the percentage similarity generated from the program gap when two sequences are aligned. Figures in brackets refer to the data of Van Beneden *et al* 1986.

5.6.2 Nuclear location signals

The *c-myc* protein has been shown to localise to the cell nucleus though it is not clear if it is associated with the nuclear matrix or not (Persson and Leder 1984, Ramsay, Evan and Bishop 1984, Eisenman *et al* 1985, Evan *et al* 1986). *C-myc* protein made in *E.coli* and microinjected into fibroblasts localises to the nucleus (Sullivan *et al* 1986). Therefore it is likely that the *c-myc* protein sequence contains motifs that localise it to the nucleus. The nuclear transport signal of SV40 large T antigen comprises of a highly basic stretch of amino acids (Pro-Lys-Lys-Lys-Arg-Lys-Val) with an absolute requirement for a positive charge at the second lysine (Smith *et al* 1985). Indeed a carrier protein cross-linked to a synthetic peptide with the above sequence when microinjected into cells transports to the nucleus (Lanford, Kanda and Kennedy 1986).

A computer search of sequences related to the SV40 T antigen nuclear location signal has revealed this sequence may act as a nuclear location signal in other proteins including the chicken *c-myc* sequence which has the related sequence Lys-Glu-Gln-Leu-Arg-Arg-Arg-Arg-Glu (Smith *et al* 1985). It is questionable whether this is a true nuclear location signal because it does not contain a Pro or Gly preceding the basic amino acids but it is interesting that this sequence is partially conserved in the mouse and human p53 protein sequences which are also known to be located in the nucleus (Crawford 1985). I have analysed *Xenopus c-myc* protein sequence to see if it contains a proposed nuclear location signal. Figure 5.18 (marked NL1) shows the location of such a sequence in the *Xenopus c-myc* protein. Crawford suggests two other areas of the *c-myc* protein might be important in *c-myc* function (nuclear location?) because of the conservation of some amino acids between the human and mouse p53, adenovirus E1a and mouse *c-myc* proteins and since these are all located in the nucleus (Crawford 1985). These are marked NL2 and NL3 on figure 5.18 and lie in areas of good conservation between all the *c-myc* proteins with only a few conservative changes hence these could be important for the nuclear location of *c-myc*.

5.6.3 Rapid degradation signals

The human *c-myc* protein has a short half life of ~30 minutes in cells (Ramsay, Evan and Bishop 1984, Rabbitts *et al* 1985) which may indicate that

the *c-myc* protein level is capable of being rapidly modulated in normal cells. Recently it has been reported that there are amino acid sequences common to rapidly degraded proteins and the human *c-myc* protein has such sequences (Rogers, Wells and Rechsteiner 1986). The amino acid sequences that signify rapid degradation are rich in proline, glutamic acid, serine and threonine (PEST) residues and these regions begin and end with positively charged residues, have no internal basic amino acids and tend to have clustering of negatively charged residues. Also the 'PEST' stretches seems to be extremely hydrophilic when analysed by hydropathy plots. In figure 5.18 I have underlined the 'PEST' regions that have been identified in the human *c-myc* protein (RD 1-7). Many of the regions are quite highly conserved between all the *c-myc* proteins which might suggest the proteins are degraded in a similar manner.

5.6.4 The myc boxes

Sequence analysis of the *c-myc*, *N-myc* and *L-myc* genes has revealed that two short regions are highly conserved between all the *myc* genes (Schwab *et al* 1983, Nau *et al* 1985, Schwab 1985). These so-called *myc* boxes are also conserved in the *c-myc* genes from different species (Van Beneden *et al* 1986 and figure 5.18). This suggests that the protein domains encoded by the *myc* boxes serve an important conserved function between all *myc* genes. It is interesting that the A box overlaps one of the proposed rapid degradation signals described earlier which may hint at the function of these domains since the *N-myc* protein also has a short half life (Evan *et al* 1986).

It would be expected from the above data that the *Xenopus c-myc* proteins would also have highly conserved *myc* boxes. Figure 5.18 shows that for all the *c-myc* proteins that the A box has only 1 deviation in the viral sequence, and that for the B box there is 1 deviation in the mammalian, trout and *Xenopus c-myc* sequences. Figure 5.19 shows that the consensus sequence for the proposed A box *c-myc* sequence is ~95% and ~79% homologous to the proposed *N-myc* and *L-myc* A box sequences respectively and that the consensus sequence for the proposed B box *c-myc* sequence is ~71% and 79% homologous to the proposed *N-myc* and *L-myc* B box sequences. Hence this shows there is strong evolutionary conservation of this protein domain which is likely to have a common function in all the *c-myc* proteins. Note that for the *c-myc* protein sequences alone there are other areas in the protein that have greater conservation than the *myc* box sequences, for example residues

391-449 are very highly conserved (figure 5.19), which probably points to a very important domain in the *c-myc* protein unique for its function.

5.7 Discussion

In this chapter I describe that the DNA and putative protein sequences of the cDNA clones I have obtained from *X.borealis* cDNA libraries are almost definitely portions of *c-myc* genes from *X.borealis*. The initial clone I obtained, XbmycA1, is strongly homologous to the *c-myc* genes from other species and was used to isolate XbmycA2 and 3 which are also very homologous to other *c-myc* genes. Chapter 4 describes reasons why I might not have isolated full length cDNA clones but XbmycA2 and 3 do overlap to give a putative full length protein product although they are encoded by two slightly different genes. Since the XbmycA2 and 3 DNA and putative protein sequences are so similar to the *X.laevis* and other *c-myc* DNA and putative protein sequences it is likely that the full length *X.borealis c-myc* gene will have a coding sequence similar to the assembled XbmycA2 and 3 sequences though it is difficult to say if the non-coding regions will be similar (note the 5' region of XbmycA3 is not so homologous to the corresponding region of XlmycA). If the divergence between XbmycA2 and 3 is as high as ~6.6% at the protein level then this raises the possibility that there are two very similar *c-myc* genes which are not alleles. Such a situation might reflect the tetraploid nature of the *X.borealis* genome or it might reflect an extended function of *c-myc* in *Xenopus* which may have evolved a requirement for two different protein products. It is interesting that in humans there are multiple forms of *c-myc* proteins but these are thought to involve post-translational modifications (Ramsay, Evan and Bishop 1984, Persson *et al* 1984, Persson *et al* 1986), however maybe there are as yet undiscovered *c-myc* genes in this system.

When the *X.borealis c-myc* DNA and putative protein sequences are compared with other available *c-myc* sequences it is found they are highly conserved over the coding sequences. The *X.borealis* and *X.laevis* genes and protein are greater than 90% homologous showing that both species recently evolved from the same common ancestor. Compared with the other *c-myc* genes the *X.borealis c-myc* DNA and protein sequences are about 60 and 67% homologous respectively with the trout *c-myc* sequences being slightly less homologous. This demonstrates the existence of a conserved gene product which spans greater than 400 million years of evolutionary time and allows a

confident extrapolation of the results obtained with *Xenopus* to mammals. Again such conservation implies the *c-myc* gene plays an important role in cell function. Some regions of extreme conservation exist between many kinds of *myc* proteins (for example the *myc* boxes) which is probably due to some common function between all the proteins. It is interesting that *myc* box A overlaps a potential sequence for rapid degradation and both *c-myc* and *N-myc* are rapidly degraded proteins (Evan *et al* 1986). Probably the diversity between the many different kinds of *myc* proteins arose through divergent evolution from a single progenitor gene, to those seen today. Therefore it is not surprising that areas of good conservation can be found.

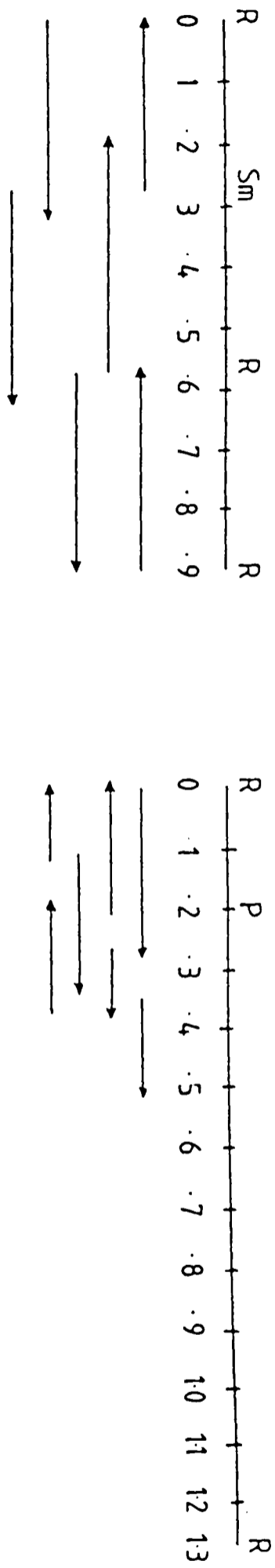
Comparing the *X.borealis c-myc* non-coding DNA sequences with other *c-myc* non-coding sequences shows they are much more diverged than the coding regions with many insertions or deletions and substitutions. The highest conservation is between the *X.borealis* and *X.laevis* sequences but even in these closely related species the non-coding regions have diverged hence there may be less selection pressure to maintain the (probably) untranslated sequences. This might suggest that the non-coding sequences are not important for regulation of the *c-myc* gene. Sequence analysis of genomic *c-myc* clones from *X.borealis* would help in determining how the gene is regulated and whether it has features upstream of transcription initiation in common with other *c-myc* genes. It would be extremely interesting to see if the stability of *c-myc* mRNA in *Xenopus* is similar in oocytes and somatic cells. If the oocyte *c-myc* mRNA is more stable it may mean that specific proteins interact with and control the degradation of the *c-myc* message. A class of poly(A)⁺ mRNA binding proteins has been shown to exist in oocytes (Darnborough and Ford 1981) which may provide stabilising functions for mRNAs. Also it will be interesting to determine whether the putative rapid degradation signals of the *c-myc* mRNA play a role in the stability of the message in both oocytes and somatic cells. Conservation of a polyadenylation signal in the *X.borealis* DNA sequence suggests that 3' processing of the *c-myc* mRNA is similar in many systems. It is not surprising that a polyadenylation signal was found since the clones I isolated were from a cDNA library and that most poly(A) containing mRNAs have such a signal.

As described the *X.borealis* protein sequences are highly conserved. The probable *X.borealis* protein would have roughly 419 amino acids with a molecular weight of ~47.5kD (taken from the assembled XbmycA2 and 3

sequences). The human *c-myc* protein is a phosphoprotein and would have a molecular weight of ~48kD but has at least two forms of ~62 and 66kD when analysed on SDS gels (Ramsay, Evan and Bishop 1984). It would be interesting to see if the *Xenopus* protein is also a phosphoprotein and whether it also runs anomalously on SDS gels.

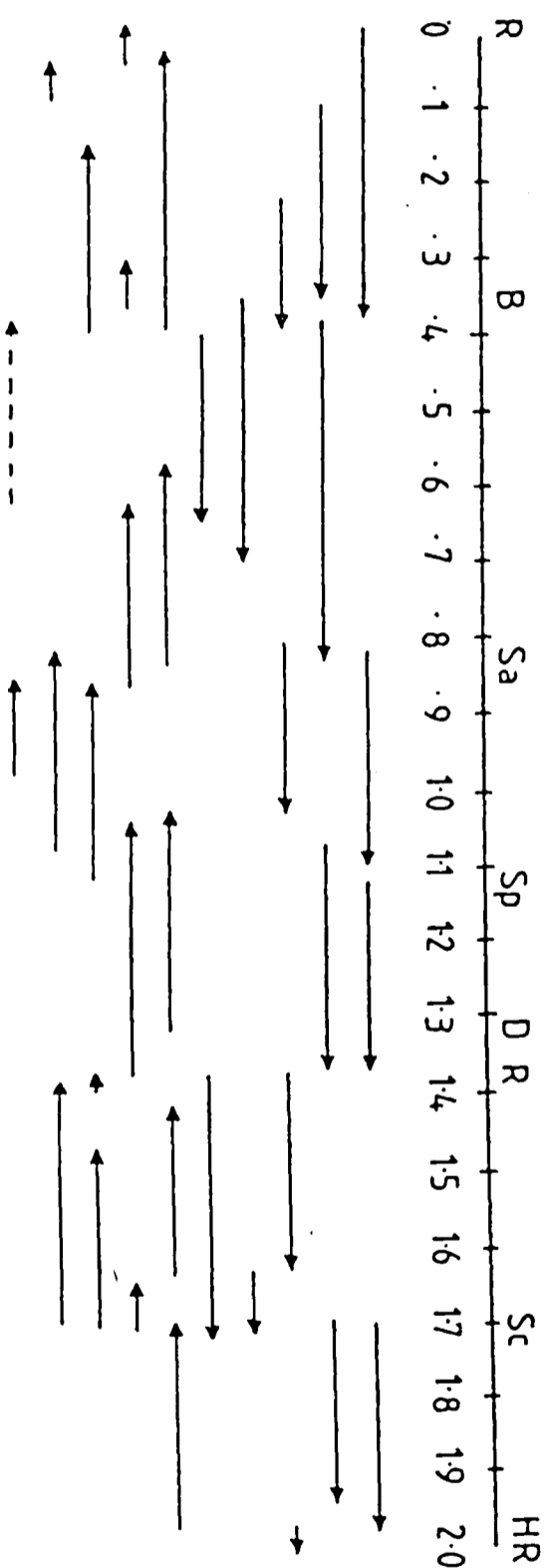
Like other *c-myc* proteins the *Xenopus c-myc* proteins are rich in basic amino acids in the carboxyl half of the protein. This region of the protein spans the supposed DNA binding domain of *myc* proteins (Persson and Leder 1984). Indeed deletions of portions of this domain in *v-myc* fail to bind DNA but still localise to the nucleus and have a reduced capacity to transform macrophages but not fibroblasts when incorporated into MC29 virus (Enritto and Hayman 1985). This may suggest *myc* proteins have more than one function some of which require DNA binding and others do not. Interestingly these deletions do not disrupt the potential nuclear location signals of the proteins.

The identification of potential nuclear location signals and rapid degradation signals in the *X.borealis c-myc* proteins provides a starting point in testing their function directly. Since the *c-myc* protein could be made in *E.coli* with various mutations of these signals, then it is a relatively simple matter to microinject a *Xenopus* oocyte and analyse the effect of these mutations with the availability of a good antiserum. If more sophisticated apparatus was available then injection of oocyte and somatic cells could be compared to see, for example, if there are factors which stabilise *c-myc* protein in the oocyte. Hence powerful analysis of *c-myc* protein function can be carried out quite simply in the *Xenopus* oocyte.



XBMYYCA3

XBMYYCA1



XBMYYCA2

FIGURE 5.1

Sequencing strategy of the XbmycA clones. Sequences were determined by the dideoxynucleotide sequencing method of Sanger as described in chapter 2. Both strands were sequenced as indicated by arrowheads, the dotted arrow refers to the sequence determined using an oligonucleotide primer as described in chapter 2. Six base pair recognition restriction enzymes used for cloning into M13 vectors are shown. R = *EcorI*, Sm = *SmaI*, P = *PstI*, B = *BglII*, Sa = *SacI*, Sp = *SphI*, D = *DraI*, Sc = *Scal* and H = *HindIII*. All sizes are in kilobase pairs.

XBMCA1 DNA SEQUENCE AND PROTEIN TRANSLATION

```

CCGCAAGTGCGCCAGTCCCAGGTCCTCTGATTCCGAGGAGAACGACAAGAGGAAGACGCA
1  -----+-----+-----+-----+-----+-----+-----+ 60
GGCGTTCACGCGGTCAGGGTCCAGGAGACTAAGGCTCCTCTTGCTGTTCTCCTTCTGCGT

  R  K  C  A  S  P  R  S  S  D  S  E  E  N  D  K  R  K  T  H  -

CAACGTTCTGGAGCGTCAGAGGCGGAACGAGCTCAAGTTGAGTTTTTTCGCCTTGCGCGA
61  -----+-----+-----+-----+-----+-----+-----+ 120
GTTGCAAGACCTCGCAGTCTCCGCCTTGCTCGAGTTCAACTCAAAAAGCGGAACGCGCT

  N  V  L  E  R  Q  R  R  N  E  L  K  L  S  F  F  A  L  R  D  -

TCAGGTACCGGAGGTGGCGAACACGAGAAAGCCCCAAAGTCGTCATCCTCAAAAAGGC
121 -----+-----+-----+-----+-----+-----+-----+ 180
AGTCCATGGCCTCCACCGCTTGTTGCTCTTTCGGGGGTTTCAGCAGTAGGAGTTTTTCCG

  Q  V  P  E  V  A  N  N  E  K  A  P  K  V  V  I  L  K  K  A  -

AACGGAATACGCCATTTCTCTGCAGGAGGATGAGCGACGGCTCATACGTGAAACAGAACA
181 -----+-----+-----+-----+-----+-----+-----+ 240
TTGCCTTATGCGGTAAAGAGACGTCCTCCTACTCGCTGCCGAGTATGCACTTTGTCTTGT

  T  E  Y  A  I  S  L  Q  E  D  E  R  R  L  I  R  E  T  E  Q  -

GTTAAAGTTCAGGAAAGAGCAGTTGAAACAGAGACTCCAACAGCTAAGGAACTTGGTTGT
241 -----+-----+-----+-----+-----+-----+-----+ 300
CAATTTCAAGTCCTTTCCTCGTCAACTTTGTCTCTGAGGTTGTCGATTCCTTGAACCAACA

  L  K  F  R  K  E  Q  L  K  Q  R  L  Q  Q  L  R  N  L  V  V  -

CTAAATTCATAAACACTTTATTATAAACTGTTGACAGTCTATATCAAGGTTGTCTCTCCT
301 -----+-----+-----+-----+-----+-----+-----+ 360
GATTAAGTATTTGTGAAATAATATTTGACAACCTGTCAGATATAGTTCCAACAGAGAAGGA

CGAGACTACAAAGCTGAATTTGTAGAACATGGACAACCTGCATGCAGATGATTTCACAACT
361 -----+-----+-----+-----+-----+-----+-----+ 420
GCTCTGATGTTTCGACTTAAACATCTTGTACCTGTTGACGTACGTCTACTAAAGTGTTGA

ACACAACCTTGGCTTGGTCTCAAACAGATTCTGCCAGCACCTTAAACTGCCTCAATACT
421 -----+-----+-----+-----+-----+-----+-----+ 480
TGTGTTGGAACCGAACCAGAGTTTGTCTAAGACGGTCGTGGAATTTTGACGGAGTTATGA

GGGATTTTGGGTATTATGGG
481 -----+-----+ 500
CCCTAAAACCCATAATACCC

```

FIGURE 5.2

XbmycA1 DNA sequence and putative protein translation. The potential stop codon is boxed.

COMPARISON OF XBMYCA1 AND HUMAN C-MYC PROTEIN SEQUENCES

```

      .           .           .           .
1  RKCASPRSSDSEENDKRKTHNVLERQRRNELKLSFFALRDQVPEVANNEK 50
   ||| ||||| ||| || ||||| ||||| ||||| ||||| ||| |||
340 RKCTSPRSSDTEENVKRRTHNVLERQRRNELKRSFFALRDQIPELENNEK 389

      .           .           .           .
51  APKVVILKKATEYAISLQEDERRLIRETEQLKFRKEQLKQRLQQLRN 97
   ||||| ||||| || | | | | | | | | | | | | | | | | | | |
390 APKVVILKKATAYILSVQAEEQKLISEEDLLRKRREQLKHKLEQLRN 436

```

A

COMPARISON OF XBMYCA1 AND CHICKEN C-MYC PROTEIN SEQUENCES

```

      .           .           .           .
1  RKCASPRSSDSEENDKRKTHNVLERQRRNELKLSFFALRDQVPEVANNEK 50
   ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
317 RKCSSPRTSDSEENDKRRTHNVLERQRRNELKLSFFALRDQIPEVANNEK 366

      .           .           .           .
51  APKVVILKKATEYAISLQEDERRLIRETEQLKFRKEQLKQRLQQLRN 97
   ||||| ||||| || | | | | | | | | | | | | | | | | | | |
367 APKVVILKKATEYVLSIQSDEHRLIAEKEQLRRRREQLKHKLEQLRN 413

```

B

FIGURE 5.3

A: Comparison of XbmycA1 and human *c-myc* protein sequences using the program BESTFIT with a gap weight of 5.0 and a length weight of 0.3. The top sequence is XbmycA1 and the lower sequence is human *c-myc*. The human *c-myc* sequence is from a translation of the human *c-myc* DNA sequence given in the GENBANK database release 40.0. % similarity = 73.196, quality = 68.4 and quality ratio = 0.705.

B: Comparison of XbmycA1 and chicken *c-myc* protein sequences using the program BESTFIT with a gap weight of 5.0 and a length weight of 0.3. The top sequence is XbmycA1 and the lower sequence is chicken *c-myc*. The chicken *c-myc* sequence is from a translation of the chicken *c-myc* DNA sequence given in the GENBANK database release 40.0. % similarity = 82.474, quality = 78.3 and quality ratio = 0.807.

COMPARISON OF XBMYCA1 AND HUMAN C-MYC DNA SEQUENCES

```

      .           .           .           .
1  CCGCAAGTGCGCCAGTCCCAGGTCCTCTGATTCCGAGGAGAACGACAAGA 50
   ||| || ||| |||| ||||| ||||| ||| ||||| ||||| | |||||
1587 CCGAAAATGCACCAGCCCCAGGTCCTCGGACACCGAGGAGAATGTCAAGA 1636

      .           .           .           .
51  GGAAGACGCACAACGTTCTGGAGCGTCAGAGGCGGAACGAGCTCAAGTTG 100
   ||  || ||||| ||||| ||||| ||||| ||||| |||  |
1637 GGCGAACACACAACGTCTTGGAGCGCCAGAGGAGGAACGAGCTAAAACGG 1686

      .           .           .           .
101 AGTTTTTTTCGCCTTGCGCGATCAGGTACCGGAGGTGGCGAACACGAGAA 150
   || ||||| ||| |||| || ||| | ||||| ||| ||||| || ||
1687 AGCTTTTTTGGCCTGCGTGACCAGATCCCGGAGTTGGAAAACAATGAAAA 1736

      .           .           .           .
151 AGCCCCCAAAGTCGTCATCCTCAAAAAGGCAACGGAATACGCCATTTCTC 200
   ||||| || || ||||| ||||| || || | |||| | | ||
1737 GGCCCCCAAGGTAGTTATCCTTAAAAAGCCACAGCATACATCCTGTCCG 1786

      .           .           .           .
201 TGCAGGAGGATGAGCGACGGCTCATACGTGAAACAGAACAGTTAAAGTTC 250
   | || | || |||| | ||||| |||| | || ||| |
1787 TCCAAGCAGAGGAGCAAAGCTCATTTCTGAAGAGGACTTGTTGCGGAAA 1836

      .           .           .           .
251 AGGAAAGAGCAGTTGAAACAGAGACTCCAACAGCTAAGGAACTTGGTTGT 300
   | ||| ||||| ||||| | ||| ||||| ||||| ||
1837 CGACGAGAACAGTTGAAACACAACTTGAACAGCTACGGAACCTTGTGTC 1886

      .           .           .           .
301 CTAATTCA.....TAAACACTTTATTATAAACTGTTGACAGTCTATA 342
   ||| | |||| || || |||
1887 GTAAGGAAAAGTAAGGAAAACGATTCCTTCTAAC.....A 1921

      .           .           .           .
343 TCAAGGTTGTCTCTTCCTCGAGACTACAAAGCTGAATTTGTAGAACATGG 392
   || || | ||| || |||| ||||
1922 GAAATGT.....CCTGAGCAATCACCTATGAACTTGT.....T 1954

      .           .           .           .
393 ACAACTGCATG..CAGATGATTCACAACCTACACAACCTTGGCTTGGTCT 440
   ||| ||||| || ||| |||| ||||| ||||| ||||
1955 TCAAATGCATGATCAAATG.....CAACCTCACAACCTTGGCTGAGTCT 1998

      .           .           .           .
441 CAA.....ACAGATTCTGCCAGCACCTTAAACTGCCTCAATACTGGGAT 485
   | ||||| |||| | | ||||| |||| | |||
1999 TGAGACTGAAAGATTTAGCCATAATGT..AAACTGCCTCA..AATTGGAC 2044

      .
486 TTTGGGTATTA 496
   ||||| || |
2045 TTTGGGCATAA 2055

```

FIGURE 5.4

Comparison of XbmycA1 and human *c-myc* DNA sequences using the program BESTFIT with a gap weight of 5.0 and a length weight of 0.3. The top sequence is XbmycA1 and the lower sequence is human *c-myc*. Boxed are the potential stop codons. The human *c-myc* DNA sequence is from the GENBANK database release 40.0. % similarity = 70.705, quality = 139.2 and quality ratio = 0.297.

COMPARISON OF XBMVCA1 AND CHICKEN C-MYC DNA SEQUENCES

```

      .       .       .       .
1  CCGCAAGTGCGCCAGTCCCAGGTCCTCTGATTCCGAGGAGAACGACAAGA 50
   ||| || ||| ||||| | | || || ||||| |||||
3215 CCGAAAATGCTCCAGTCCCCGCACGTCAGACTCAGAGGAGAACGACAAGA 3264

      .       .       .       .
51 GGAAGACGCACAACGTTCTGGAGCGTCAGAGGCGGAACGAGCTCAAGTTG 100
   || ||||| ||||| ||||| ||| | |||| ||||| ||| ||
3265 GGCGAACGCACAACGTCTTGGAGCGCCAGCGAAGGAATGAGCTGAAGCTG 3314

      .       .       .       .
101 AGTTTTTTCGCCTTGCGCGATCAGGTACCGGAGGTGGCGAACAACGAGAA 150
     ||||| || ||| |||| | | ||| ||||| ||||| |||||
3315 AGTTTCTTTGCCCTGCGTGACCAGATACCCGAGGTGGCCAACAACGAGAA 3364

      .       .       .       .
151 AGCCCCAAAGTCGTCATCCTCAAAAAGGCAACGGAATACGCCATTTCTC 200
     || ||||| || ||||| ||||| ||||| ||||| |||||
3365 GGCGCCCAAGGTTGTCATCCTGAAAAAGCCACGGAGTACGTTCTGTCTA 3414

      .       .       .       .
201 TGCAGGAGGATGAGCGACGGCTCATACGTGAAACAGAACAGTTAAAGTTC 250
     | || ||| |||| | ||| || ||| ||||| ||||| | |
3415 TCCAATCGGACGAGCACAGACTAATCGCAGAGAAAGAGCAGTTGAGGCGG 3464

      .       .       .       .
251 AGGAAAGAGCAGTTGAAACAGAGACTCCAACAGCTAAGGAACTTGGTTGT 300
     |||| ||| ||||| ||||| | ||| | ||||| ||||| ||
3465 AGGAGAGAACAGTTGAAACACAACTTGAGCAGCTAAGGAACTCTCGTGC 3514

      .       .       .       .
301 CTA.....ATTCATAAACACTT.....TA 319
     || || | ||||| ||
3515 ATAGGAACTCTTGGACATCACTTAGAATACCCCAAAGTACTGAAACTA 3564

      .       .       .       .
320 TTATAAACTGTTGACAGTCTATATCAAGGTTGTCTCTTCCTCGAG.ACTA 368
     | ||||| | || ||| | | | | | | | ||| | ||||
3565 TGATAAAATATT...AGTGTTTCTAA.....TATCACTCATGAACTA 3603

      .       .       .       .
369 CAAAGCTGAATT...TGTAGAACATGGACAAGTGCATGCAG.ATGATTTTC 414
     || | ||| | | |||| ||||| ||||| | ||| ||
3604 CATCAGTCCATTGAGTATGGAAGTATTGCAACTGCATGCTGTGCGACTTA 3653

      .       .       .       .
415 AC....AACTACACAACCTTGGCTTGGTCTCAAACAGATTCTGCCAGCAC 460
     || ||||| ||||| |||| | | || ||||| ||
3654 ACTTGAGACTACACAACCTTGGCCGAATCTCCGAACGGTTTGGCCAGAAC 3703

      .       .       .       .
461 CTTAAACTGCCTCAATACTGGGATTTTGGGTATTATGG 499
     || ||||| ||||| | || | ||||| || | ||
3704 CTCAAAACTGCCTCATAATTGATACTTTGGGCATAAGGG 3742

```

FIGURE 5.5

Comparison of XbmycA1 and chicken *c-myc* DNA sequences using the program BESTFIT with a gap weight of 5.0 and a length weight of 0.3. The top sequence is XbmycA1 and the lower sequence is chicken *c-myc*. Boxed are the potential stop codons. The chicken *c-myc* DNA sequence is from the GENBANK database release 40.0. % similarity = 71.926, quality = 172.4 and quality ratio = 0.345.

XBMYCA2 DNA SEQUENCE AND PROTEIN TRANSLATION

```

GAATTCCTGGGAGGGGACATGGTCAACCAGAGCTTTATCTGCGAGGCGGATGACGAAGCC
1  -----+-----+-----+-----+-----+-----+ 60
CTTAAGGACCCTCCCCTGTACCAGTTGGTCTCGAAATAGACGCTCCGCCTACTGCTTCGG

E F L G G D M V N Q S F I C E A D D E A -

TTGCTGAAGTCCATCGTCATACAGGACTGTATGTGGAGTGGCTTCTCGGCGGCGCCAAG
61  -----+-----+-----+-----+-----+-----+ 120
AACGACTTCAGGTAGCAGTATGTCCTGACATACACCTCACCGAAGAGCCGCCGCCGGTTC

L L K S I V I Q D C M W S G F S A A A K -

CTGGAGAAAGTGGTGTCTGGAGAAGCTGGCATCGTACCAGGCTTCTAGGAAAGAGAGTGCT
121 -----+-----+-----+-----+-----+-----+ 180
GACCTCTTTCACCACAGCCTCTTCGACCGTAGCATGGTCCGAAGATCCTTTCTCTCACGA

L E K V V S E K L A S Y Q A S R K E S A -

CTGTCTTCTTCTCAGTGCCAGAGTCAACCACCACAGAGCCCGATCAAGTCTCCCTCATGC
181 -----+-----+-----+-----+-----+-----+ 240
GACAGAAGAAGAGTCACGGTCTCAGTTGGTGGTGTCTCGGGCTAGTTCAGAGGGAGTACG

L S S S Q C Q S Q P P Q S P I K S P S C -

GATGGGAGCCTGAGTCTCGGAGGGACCAACAGGAGCAGCCATGGATTCCCTCCAGGACCCC
241 -----+-----+-----+-----+-----+-----+ 300
CTACCCTCGGACTCAGAGCCTCCCTGGTTGTCTCGTCCGTACCTAAGGAGGTCCTGGGG

D G S L S L G G T N R S S H G F L Q D P -

AGCTCGGATTGTGTTGACCCATCAGTGGTCTTCCCATACCCACTGAACGACACCATTTCC
301 -----+-----+-----+-----+-----+-----+ 360
TCGAGCCTAACACAACACTGGGTAGTCACCAGAAGGGTATGGGTGACTTGCTGTGGTAAAGG

S S D C V D P S V V F P Y P L N D T I S -

AACGCCAGCTCACCATGCCAAGATCTCATGTTGGAAACACCACCCATCAGCAGTAACAGC
361 -----+-----+-----+-----+-----+-----+ 420
TTGCGGTTCGAGTGGTACGGTTCTAGAGTACAACCTTTGTGGTGGGTAGTCGTCATTGTCTG

N A S S P C Q D L M L E T P P I S S N S -

AGCAGTAGTGAATCAGAAGAGGAACAGGAAGATGATGATGAAGACTGTGACGAGGAAGAG
421 -----+-----+-----+-----+-----+-----+ 480
TCGTCATCACTTAGTCTTCTCCTTGTCTTCTACTACTACTTCTGACACTGCTCCTTCTC

S S S E S E E E Q E D D D E D C D E E E -

GAGATTGATGTGGTCACGGTAGAAAAGAGACAGACTGCGTCCAGGCGGATGGAATCCAGT
481 -----+-----+-----+-----+-----+-----+ 540
CTCTAACTACACCAGTGCCATCTTTTCTCTGTCTGACGCAGGTCCGCCTACCTTAGGTCA

E I D V V T V E K R Q T A S R R M E S S -

```

FIGURE 5.6

XbmycA2 DNA sequence and putative protein translation. The potential stop codon and the potential polyadenylation signal are boxed. Underlined are the potential rapid degradation signals.

TCTCATTTCGCAGCCTTCCCAGCCTCACCACAGCCCATTGGTTCTGAAGCGGTGTCACGTT
 541 -----+-----+-----+-----+-----+-----+ 600
 AGAGTAAGCGTCGGAAGGGCCGGAGTGGTGTCTGGGTAACCAAGACTTCGCCACAGTGCAA

 S H S Q P S R P H H S P L V L K R C H V -

 CCCATTACCAACACAACACTACGCAGCGTCCCCCTCAACCAAAGTGGACTATGTTTCTTCC
 601 -----+-----+-----+-----+-----+-----+ 660
 GGGTAAGTGGTTGTGTTGATGCGTCGCAGGGGGAGTTGGTTTCACCTGATACAAAGAAGG

 P I H Q H N Y A A S P S T K V D Y V S S -

 AAGAGGGCGAAACTAGAAAGCAACGCCAGGGTCCTCAAACAGATCAGCAATAACCGCAAG
 661 -----+-----+-----+-----+-----+-----+ 720
 TTCTCCCGCTTTGATCTTTCGTTGCGGTCCCAGGAGTTTGTCTAGTCGTTATTGGCGTTC

 K R A K L E S N A R V L K Q I S N N R K -

 TGCGCCAGCCCCAGGTCCTCTGATTCCGAAGAGAACGACAAGAGGCGGACGCACAACGTC
 721 -----+-----+-----+-----+-----+-----+ 780
 ACGCGGTCTGGGTCCAGGAGACTAAGGCTTCTCTTGCTGTTCTCCGCCTGCGTGTTCAG

 C A S P R S S D S E E N D K R R T H N V -

 TTGGAGCGTCAAAGGCGGAACGAGCTCAAGTTGAGTTTCTTCGCCTTGCGCGACCAGGTA
 781 -----+-----+-----+-----+-----+-----+ 840
 AACCTCGCAGTTTCCGCCTTGCTCGAGTTCAACTCAAAGAAGCGGAACGCGCTGGTCCAT

 L E R Q R R N E L K L S F F A L R D Q V -

 CCGGAGGTGGCCAACAACGAGAAGGCCCCAAAGTCGTCATCCTCAAAAAGGCAACGGAA
 841 -----+-----+-----+-----+-----+-----+ 900
 GGCCTCCACCGGTTGTTGCTCTTCCGGGGGTTTCAGCAGTAGGAGTTTTTCCGTTGCCTT

 P E V A N N E K A P K V V I L K K A T E -

 TACGTCATTTCTATGCAGGAGGACGAGCGACGGCTCATAACGGGAAACAGAACAGTTAAAG
 901 -----+-----+-----+-----+-----+-----+ 960
 ATGCAGTAAAGATACGTCCTCCTGCTCGCTGCCGAGTATGCCCTTTGTCTTGTCATTTTC

 Y V I S M Q E D E R R L I R E T E Q L K -

 TACAGGAAAGAGCAGTTAAAACAGAGACTCCAACAGTTACGGAAGTTGGTTGTCATTTTC
 961 -----+-----+-----+-----+-----+-----+ 1020
 ATGTCCTTTCTCGTCAATTTTGTCTCTGAGGTTGTCAATGCCTTGAACCAACAGATTAAG

 Y R K E Q L K Q R L Q Q L R N L V V -

 ACAAACTCTTATTTACACTTTATACAAACTGTGACGGTCTATATCAAGGTTGTCTCTTCC
 1021 -----+-----+-----+-----+-----+-----+ 1080
 TGTTTGAGAATAAATGTGAAATATGTTTGACACTGCCAGATATAGTTCCAACAGAGAAGG

 TCGAGACAAGACTACAAAACGAATTTGTAGAAAATGGACAACATGCATGCATGATTTTAC
 1081 -----+-----+-----+-----+-----+-----+ 1140
 AGCTCTGTTCTGATGTTTGTACTTAAACATCTTTTACCTGTTGACGTACGTACTAAAGTG

AACTACACAACCTTGGCATGGTCTCAACCGGATTCTGCCAGCACCTTAAACTGCCTCAA
 1141 -----+-----+-----+-----+-----+-----+-----+ 1200
 TTGATGTGTTGGAACCGTACCAGAGTTGGCCTAAGACGGTCGTGGAATTTTGACGGAGTT

 TCCTGGGATTTTGGGTATTATGGGACTGTTTGTAATAAATAATTTTTTTTTTCTGCTTGT
 1201 -----+-----+-----+-----+-----+-----+-----+ 1260
 AGGACCCTAAAACCCATAATACCCTGACAAACATTTTTTTTAAAAAAGACGAACA

 GGGGGGTATATTTGCCAACTTTTTCTTTGTTTGTGTTGTTATTGTTTTACAACCTTGT
 1261 -----+-----+-----+-----+-----+-----+-----+ 1320
 CCCCCCATATAAACGGTTGAAAAAGAAACAAACAATAACAAAATGTTGGAAACA

ATTTAAACATTTATTTTTCTTATAAACGAATTCCTGAAGTTTGCCTCTGAATAGCTTAA
 1321 -----+-----+-----+-----+-----+-----+-----+ 1380
 TAAATTTTGTAATAAAAAGAATATTTGCTTAAGGACTTCAAACGGAGACTTATCGAATT

 ATATATACATCATCGCTGACTCGTCTATAATATCATTAAATAGGTCGTTATAGAAATTATT
 1381 -----+-----+-----+-----+-----+-----+-----+ 1440
 TATATATGTAGTAGCGACTGAGCAGATATTATAGTAATTATCCAGCAATATCTTTAATAA

 GTATTCACTTCTAACCGCTATGGCGCAAAAAGAGAGAAGCAAACCTTTTTTTGTTTTTT
 1441 -----+-----+-----+-----+-----+-----+-----+ 1500
 CATAAGTGAAGATTGGCGATACCGCGTTTTTTCTCTCTTCGTTTTGAAAAAACAAAAA

 GCCAATTAATACCATGAATGCATTTAACAACTTGTTTTTTTTTATTTTTTGTTATTTT
 1501 -----+-----+-----+-----+-----+-----+-----+ 1560
 CGGTTAATTATGGTACTTACGTAATTGTTTGAACAAAAAATAAAAAACAAATAAAA

 AAGAAGAGAGAATTTAACATTTATCTATCAAGTTTCATTTGGATAATTCTGTATTGGAGC
 1561 -----+-----+-----+-----+-----+-----+-----+ 1620
 TTCTTCTCTCTTAAATTGTAATAGATAGTTCAAAGTAAACCTATTAAGACATAACCTCG

 TAGTCGTACCAGTTTCTGAAGAAGTATGAAGAAGGCACTGACCAACATGTAACCTCAAAA
 1621 -----+-----+-----+-----+-----+-----+-----+ 1680
 ATCAGCATGGTCAAAGACTTCTTGACTACTTCTTCCGTGACTGGTTGTACATTGAGTTTT

 CAAATTACTAAAAGTACTTAATATTATCCCATACCCCTCTATCCTTGTGCATGTGAAAACC
 1681 -----+-----+-----+-----+-----+-----+-----+ 1740
 GTTTAATGATTTTCATGAATTATAATAGGGTATGGGGAGATAGGAACAGTACACTTTTGG

 TCAAACCTGCTGGCTGCTTATGGCGTCCTATTGTAGTAATTGCAGCAGGATTGTGGTCTT
 1741 -----+-----+-----+-----+-----+-----+-----+ 1800
 AGTTTGGACGACCGACGAATACCGCAGGATAACATCATTAAACGTCGTCCTAACACCAGAA

 TCCTAGTAACACAGAAGCAGTCTGCTGTTTCTATATTGAATCTCTTAAAGATATTATAAC
 1801 -----+-----+-----+-----+-----+-----+-----+ 1860
 AGGATCATTGTGTCTTCGTCAGACGACAAAGATATAACTTAGAGAATTTCTATAATATTG

 ATGCTGGTCTGAAAATTCATCATTGTTTATTGTAAGGCTTGTTTTTGCTTAAGATGAA
 1861 -----+-----+-----+-----+-----+-----+-----+ 1920
 TACGACCAGACTTTTTAAGTAGTAAACAATAACATTCCGAACAAAACGAATTCTACTT

 ACTATTCATGTGGGGTCTTGACATCCCATAATAGTACTTTTAAACATAGCAAGCTGTTTG
 1921 -----+-----+-----+-----+-----+-----+-----+ 1980
 TGATAAGTACACCCAGAACTGTAGGGTATTATCACTGAAAATTGTATCGTTTCGACAAAC

 GAATAAGCTTCAGAAATAAAAATCAAGTTT
 1981 -----+-----+-----+-----+-----+-----+-----+ 2011
 CTTATTGAGTCTTTTTATTTTTAGTTCAA

XBMYCA3 DNA SEQUENCE AND PROTEIN TRANSLATION

```

GAATTCcccccccccttcccagagctgccccaaatcacaggggtctcgctccattccag
1  -----+-----+-----+-----+-----+-----+ 60
CTTAAGGGGGGGGGGAAGGGTCTCGACGGGGTTTAGTGTCCCAGAGCGAGGTAAGGTC

CCCCTTATGTGGGACCCCCCTTCCATAAAGTCAAGTATCTCTGAAAGAAAAGACTTTGCA
61  -----+-----+-----+-----+-----+-----+ 120
GGGGAATACACCCTGGGGGAAGGTATTTTCAGTTCATAGAGACTTTCTTTTCTGAAACGT

GACACAGATGAGGAAAACAAACATTCACTAATTGTTAATTAAGCAGGATAAGAGGGAGAC
121 -----+-----+-----+-----+-----+-----+ 180
CTGTGTCTACTCCTTTTGTTTGTAAGTGATTAACAATTAATTCGTCCTATTCTCCCTCTG

TGGGAAATGCGATTGGGTACGAAGACTCCGAGCGAGACTGTTTCAGCTGCTCCGACGCAT
181 -----+-----+-----+-----+-----+-----+ 240
ACCCTTTACGCTAACCCAATGCTTCTGAGGCTCGCTCTGACAAGTCGACGAGGCTGCGTA

CAAAGTGGATGAGCCCATCCTGAGACCTGCATCAGCCCCGGGGACAGCGATAAAGTGAG
241 -----+-----+-----+-----+-----+-----+ 300
GTTTCACCTACTCGGGTAGGACTCTGGACGTAGTCGGGGGCCCTGTGCGTATTTCACTC

ACCAATCGCAAGGGAGGAAAGATGCCTCTTAATGCCAGTTTCCCAGCAAGAACTACGAC
301 -----+-----+-----+-----+-----+-----+ 360
TGGTTAGCGTTCCTCCTTTCTACGGAGAATTACGGTCAAAGGGGTCGTTCTTGATGCTG

                                M P L N A S F P S K N Y D -

TACGACTATGACTTGCAGCCCTGCTTCTTCTTCCCTGGAGGAGGAGAACTTCTACCACCAG
361 -----+-----+-----+-----+-----+-----+ 420
ATGCTGATACTGAACGTCGGGACGAAGAAGAAGGACCTCCTCCTTGAAGATGGTGGTC

Y D Y D L Q P C F F F L E E E N F Y H Q -

CAAAGCCGACTGCACGCGCCGGCCCCCAGTGAGGACATCTGGAAGAAGTTTGAGCTGCTC
421 -----+-----+-----+-----+-----+-----+ 480
GTTTCGGCTGACGTGCGCGGGGGGTCACTCCTGTAGACCTTCTTCAAACCTCGACGAG

Q S R L H A P A P S E D I W K K F E L L -

CCCACCCCGCCCTTGTCGCCCAGCCGCAGATCCAGCCAGTCCAGCCTTTTCCCCTCTACT
481 -----+-----+-----+-----+-----+-----+ 540
GGGTGGGGCGGGAACAGCGGGTCGGCGTCTAGGTCGGTCAGGTCGGAAAAGGGGAGATGA

P T P P L S P S R R S S Q S S L F P S T -

GCCGATCAGCTGGAGATGGTGACCGAATTCCTGGGAGGGGACATGGTCAACCAGAGCTTT
541 -----+-----+-----+-----+-----+-----+ 600
CGGCTAGTCGACCTCTACCACTGGCTTAAGGACCCTCCCCTGTACCAGTTGGTCTCGAAA

A D Q L E M V T E F L G G D M V N Q S F -

ATCTGCGAGGCGGATGACGAAGCCTTGCTGAAGTCCATCGTCATACAGGACTGTATGTGG
601 -----+-----+-----+-----+-----+-----+ 660
TAGACGCTCCGCTACTGCTTCGGAACGACTTCAGGTAGCAGTATGTCCTGACATACACC

I C E A D D E A L L K S I V I Q D C M W -

```

FIGURE 5.7

XbmycA3 DNA sequence and putative protein translation. The potential start codon is boxed.

AGTGGTTTTTCAGCGGCGGCCAAGCTGGAGAAAGTGGTGTCTGAGAAGCTGGCATCATAC
 661 -----+-----+-----+-----+-----+-----+-----+ 720
 TCACCAAAAAGTCGCCGCCGGTTCGACCTCTTTCACCACAGACTCTTCGACCGTAGTATG

 S G F S A A A K L E K V V S E K L A S Y -

 CAGGCTTCTAGGAAAGAGAGTGCTCTGTCTTCTTCTCAGTGTCCAAGTCAACCGCCTCAG
 721 -----+-----+-----+-----+-----+-----+-----+ 780
 GTCCGAAGATCCTTCTCTCACGAGACAGAAGAAGAGTCACAGGTTTCAGTTGGCGGAGTC

 Q A S R K E S A L S S S Q C P S Q P P Q -

 AGCCCACTCAAGTCTCCCTCATGCCACGGGAGTCTCGAAGGGACCCACAGGAGCAGCCAT
 781 -----+-----+-----+-----+-----+-----+-----+ 840
 TCGGGTGAGTTCAGAGGGAGTACGGTGCCCTCAGAGCTTCCCTGGGTGTCCTCGTCGGTA

 S P L K S P S C H G S L E G T H R S S H -

 GCTTTTCTCCAGGACCCAGCTCGGATTGTGTTGACCCTTCCGTGGTCTTCCCATAACCA
 841 -----+-----+-----+-----+-----+-----+-----+ 900
 CGAAAAGAGGTCCTGGGGTTCGAGCCTAACACAACCTGGGAAGGCACCAGAAGGGTATGGGT

 A F L Q D P S S D C V D P S V V F P Y P -

 CTGAACGACAGCATTTCCAACCTCCA
 901 -----+-----+----- 925
 GACTTGCTGTCGTAAAGGTTGAGGT

 L N D S I S N S -

FIGURE 5.8

Comparison of XbmycA1 and XbmycA2 DNA sequences using the program BESTFIT with a gap weight of 5.0 and a length weight of 0.3. The top sequence is XbmycA1 and the lower sequence is XbmycA2. Boxed are the potential stop codons. % similarity = 93.763, quality = 413.0 and quality ratio = 0.826.

FIGURE 5.9

Comparison of XbmycA3 and XbmycA2 DNA sequences using the program BESTFIT with a gap weight of 5.0 and a length weight of 0.3. The top sequence is XbmycA3 and the lower sequence is XbmycA2. % similarity = 93.629, quality = 310.5 and quality ratio = 0.860.

FIGURE 5.10

A: Comparison of XbmycA1 and XbmycA2 protein sequences using the program BESTFIT with a gap weight of 5.0 and a length weight of 0.3. The top sequence is XbmycA1 and the lower sequence is XbmycA2. % similarity = 96.000, quality = 95.6 and quality ratio = 0.956.

B: Comparison of XbmycA3 and XbmycA2 protein sequences using the program BESTFIT with a gap weight of 5.0 and a length weight of 0.3. The top sequence is XbmycA3 and the lower sequence is XbmycA2. % similarity = 94.118, quality = 105.7 and quality ratio = 0.888.

COMPARISON OF XBMCA2 AND XLMCA DNA SEQUENCES

```

      .           .           .           .           .
1 GAATTCCTGGGAGGGGACATGGTCAACCAGAGCTTTATCTGCGAGGCGGA 50
  |||
385 GAATTCCTGGGAGGGGACATGGTCAACCAGAGCTTTATCTGCGAGGCGGA 434

      .           .           .           .           .
51 TGACGAAGCCTTGCTGAAGTCCATCGTCATACAGGACTGTATGTGGAGTG 100
  |||
435 TGACGAAGCCTTGCTGAAGTCCATCGTCATACAGGACTGTATGTGGAGTG 484

      .           .           .           .           .
101 GCTTCTCGGCGGCGGCAAGCTGGAGAAAGTGGTGTCTGGAGAAGCTGGCA 150
  |||
485 GATTTTCGGCTGCGGCAAGCTGGAGAAAGTGGTGTCTGAGAAGCTGGCG 534

      .           .           .           .           .
151 TCGTACCAGGCTTCTAGGAAAGAGAGTGCTCTG...TCTTCTTCTCAGTG 197
  |||
535 TCCTACCAGGCTTCTAGGAAAGAGAGTGCTCTGTCTTCTTCTTCTCCGTG 584

      .           .           .           .           .
198 CCAGAGTCAACCACCACAGAGCCCGATCAAGTCTCCCTCATGCGATGGGA 247
  |||
585 TCAGAGTCAACCACCACCGAGCCCGCTTAAGTCTCCCTCGTGTTCATGGGA 634

      .           .           .           .           .
248 GCCTGAGTCTCGGAGGGACCAACAGGAGCAGCCATGGATTCCTCCAGGAC 297
  |||
635 GCCTGAGTCTGGGAGGGACCCACAGGAGCAGCCATGGTTTTCTCCAGGAC 684

      .           .           .           .           .
298 CCCAGCTCGGATTGTGTTGACCCATCAGTGGTCTTCCCATAACCCACTGAA 347
  |||
685 CCCAGCTCGGATTGTGTTGACCCTTCAGTGGTCTTCCCATAACCCACTGAA 734

      .           .           .           .           .
348 CGACACCATTCCAACGCCAGCTCACCATGCCAAGATCTCATGTTGGAAA 397
  |||
735 CGACAGCATTCCAACGCCAGCTCTCCTTGCCAAGATCTCATTTTGGAAA 784

      .           .           .           .           .
398 CACCACCCATCAGCAGTAACAGCAGCAGTAGTGAATCAGAAGAGGAACAG 447
  |||
785 CTCCGCCATCAGCAGTAACAGCAGCAGCAGTGAATCTGAAGAGGAACCA 834

      .           .           .           .           .
448 GAAGATGATGATGAAGACTGTGACGAGGAAGAGGAGATTGATGTGGTCAC 497
  |||
835 GAAGATGAGGATGAAGACTGCGACGAGGAAGAGGAGATTGACGTTGTCAC 884

      .           .           .           .           .
498 GGTAGAAAAGAGACAGACTGCGTCCAGGCGGATGGAATCCAGTTCTCATT 547
  |||
885 AGTAGAAAAAAGGCAGTCGGCATCCAAGCGGGTGAATCCAGTTCTCATT 934

      .           .           .           .           .
548 CGCAGCCTTCCCGGCCTCACCACAGCCCATTTGGTTCTGAAGCGGTGTCAC 597
  |||
935 CGCAGCCCTCCCGACCCCACTACAGCCCTTTAGTTCTGAAGCGGTGTCAC 984

      .           .           .           .           .
598 GTTCCCATTCACCAACACAACACTACGCAGCGTCCCCCTCAACCAAGTGGA 647
  |||
985 GTTCCCATTCACCAACACAACACTACGCAGCGTCTCCCTCGACCAAGTGGA 1034
  
```

FIGURE 5.11

Comparison of XbmycA2 and XlmycA DNA sequences using the program GAP with a gap weight of 5.0 and a length weight of 3.0. To compare the whole sequences they first were divided into three parts and then each comparison was run (since at a maximum only two 1000 bp sequences can be compared). The combined output of these comparisons is shown in this figure. Boxed are the potential stop codons. The XlmycA DNA sequence was kindly provided by M. King. The average % similarity = 93.763, quality = 413.0 and quality ratio = 0.826 over the whole sequences.

648 CTATGTTTCTTCCAAGAGGGCGAAACTAGAAAGCAACGCCAGGGTCCTCA 697
 |||
 1035 CTATGTTTCTTCCAAAAGGGCGAAACTAGAAAGCAACATCCGGGTCCTCA 1084
 .
 698 AACAGATCAGCAATAACCGCAAGTGCGCCAGCCCCAGGTCCTCTGATTCC 747
 |||
 1085 AACAGATCAGCAACAACCGCAAGTGCGCCAGTCCCAGGTCCTCGGATTCC 1134
 .
 748 GAAGAGAACGACAAGAGGGCGGACGCACAACGTCTTGGAGCGTCAAAGGCG 797
 |||
 1135 GAAGAGAACGACAAGAGGAAGACGCACAACGTCTTGGAGCGCCAGAGGCG 1184
 .
 798 GAACGAGCTCAAGTTGAGTTTCTTCGCCTTGCGCGACCAGGTACCGGAGG 847
 |||
 1185 GAACGAGCTCAAGTTGAGTTTTTTTGCCTTGCGCGATCAGGTACCGGAGG 1234
 .
 848 TGGCCAACAACGAGAAGGCCCCCAAAGTCGTCATCCTCAAAAAGGCAACG 897
 |||
 1235 TGGCGAGCAACGAGAAGGCCCCCAAAGTAGTCATCCTCAAAAAGGCAACG 1284
 .
 898 GAATACGTCATTTCTATGCAGGAGGACGAGCGACGGCTCATAACGGGAAAC 947
 |||
 1285 GAATACGCCATTTCTCTGCAGGAGGACGAACGGCGGCTCATAACGGGAAAC 1334
 .
 948 AGAACAGTTAAAGTACAGGAAAGAGCAGTTAAAACAGAGACTCCAACAGT 997
 |||
 1335 AGAACAGTTAAAGTACAGGAAAGAGCAGTTAAAACAGAGACTCCAACAGC 1384
 .
 998 TACGGAACCTGGTTGTCTAAATTCACAACTCTTATTT.ACACTTTATACA 1046
 |
 1385 TGAGGAACT...TTGTCTAAATTCACAACTCTTATTTAACACTTTATATA 1431
 .
 1047 AACTGTGACGGTCTATATCAAGGTTGTCTCTTCCTCGAGACAAGACTACA 1096
 |||
 1432 AACTGTGACCGTCTATATCAAGGTTGTCTCTTCCTCGAGAC.CGACTACA 1480
 .
 1097 AACTGAATTTGTAGAAAATGGACAACCTGCATGC..ATGATTTACAACCT 1144
 |||
 1481 AACTGAATTTGTAGAACATGGACAACCTGCATGCAGATGATTTACAACCT 1530
 .
 1145 ACACAACCTTGGCATGGTCTCAACCGGATTCTGCCAGCACCTTAAAACCTG 1194
 |||
 1531 ACACAACCTTGGCATGGTCTCAAACAGATTCTGCCAGCACCTTAAAACCTG 1580
 .
 1195 CCTCAATCCTGGGATTTTGGGTATTATGGGACTGTTTGTAAAAAATTTT 1244
 |||
 1581 CCTCAATACTGGGA.TTTGGGTATTATGGGACTTGT.....TTTT 1620
 .
 1245 TTTTTTTTCTGCTTGTGGGGGGGTATATTTGCCAACTTTTTTCTTTGTTT 1294
 |||
 1621 TTTTTTTTCTTGCT.TGAGGGGGTATATTTGCCAACT.....TTTGTTT 1664
 .
 1295 GTTTGTATTG.TTTTACAACC.TTTGTATTTAAAACATTTATTTTCTT 1342
 |||
 1665 GTTTGTATTGTTTTTACAACCTTTTGTATTTAAAACATTTATTTTCTT 1714

```

1343 ATAAACGAATTCCTGAAGTTTGCCTCTGAATAGCTTAAATATATACATCA 1392
||||| ||||||||||||||| ||||||||||||||| |||||
1715 ATAAACAAATTCCTGAAGTTTGGCTCTGAATAGCTTAAATATATATATCA 1764

1393 TCGCTGACTCGTCTATAATATCATTAAATAGGTCGTTATAGAAATTATTGT 1442
||||||||||||||||| |||| | |||||||||||||||
1765 TCGCTGACTCGTCTATAATATCATTGATAGAACATTATAGAAATTATTGT 1814

1443 ATTCACTTCTAACCGCTATGGCGCAAAAAGAGAGAAGCAAACTTTTTT 1492
||||||||||||||||| | | |||||||
1815 ATTCACTTCTAACCGCTATGGCGCAAAA..AAAAGCAAAA..... 1855

1493 TGTTTTTTGCCAATTAATACCATGAATGCATTTAACAACTTGTTTTTT 1542
||||||| |||| | ||||||| ||||||| ||| |
1856 ACGTTTTTGCTAATTCGTACCATGAACGCATTTAACGAACCT..... 1897

1543 TTATTTTTTGTATTATTTAAGAAGAGAGAATTTAACATTT.ATCTATCAA 1591
||||||||||||||| | || | ||||||| ||||| |||||
1898 .....TTTATTTAAGAAGAAAAA...AACATTTAATCTTTCAA 1934

1592 GTTTCATTTGGATAATTCTGTATTGGAGCTAGTCGTACCAGTTTCTGAAG 1641
||||||||||| ||||||||||||||||||| |||||||||||||||
1935 GTTTCATTTGAATAATTCTGTATTGGAGCTAGTTGTACCAGTTTCTGAAG 1984

1642 AACTGATGAAGAAGGCACTGACCAACATGTAACT..CAAACAATTACT 1689
||||||||||||||||| ||||||||||||||| | ||||| ||| |||||
1985 AACTGATGAAGAAGGCACTGACCAACATGTAATTTAAAAAAGAAAATACT 2034

1690 AAAAGTACTTAATATTATCCCATACCCCTCTATCCTTGTCATGTGAAAAC 1739
||||||||||||||| ||||||| |||| | ||||||||||||||| |||||
2035 AAAAGTACTTAATAT..CCCCATA.CCCTGTATCCTTGTCATGTGGAAAC 2081

1740 CTCAAACCTGCTGGCTGCTTATGGCGTCCTATTGTAGTAATTGCAGCAGG 1789
||| ||||||||||||||| ||| ||||| ||||| || | |||||||
2082 CTCTAACCTGCTGGCTGCTCCTGGTGTCTGTTGTAATAGCAGCAGCAGG 2131

1790 ATTGTGGTCTTTCCTAGTAACACAGAAGCAGTCTGCTGTTTCTATATTGA 1839
|| | ||||||||||||||| ||||||||||||||| |||||||||||
2132 ATCTTGGTCTTTCCTAGTAACACAGAAGCAGTCTGCTGTTTCTATATTGA 2181

1840 ATCTCTTAAAGATATTATAACATGCTGGTCTG.....AAAATTCATC 1882
||||||||||||||||| ||||||| ||||| |||||
2182 TTCTCTTAAAGATATTATAACATGCTGGTCTGAAAAAAAAAAAACTCATC 2231

1883 ATTTGTTTATTGTAAGGCTTGTTTTTGCTTAAGATGAACTATTCATGTG 1932
||| ||||||||||||| |||| | ||||||||||||||| ||| |
2232 ATTCGTTTATTGTAAGCTTG.CTTTGCTTAAGATGAACTATTCACATT 2280

1933 GGGTCTTGACATCCATA.....ATAGTGACTTTTAACATAGCAAGCTGT 1977
||| | || ||||| | ||||||| |||||||||||||||
2281 GGGGC.CAACCCCCATAATGATATAGTGAC..TTAACATAGCAAGCTGT 2327

1978 TTGGAATAAGCTTCAGAAAATAAAAAT 2004
||||||||||||||||| |||
2328 TTGGAATAAGCTTCAGAAAATCGGAAT 2354

```


FIGURE 5.12

Comparison of XbmycA3 and XlmycA DNA sequences using the program BESTFIT with a gap weight of 0.5 and a length weight of 0.3. The top sequence is XbmycA3 and the lower sequence is XlmycA. Boxed are the potential start codons. The XlmycA DNA sequence was kindly provided by M. King. % similarity = 94.462, quality = 552.8 and quality ratio = 0.875.

COMPARISON OF XBMYCA2 AND XLMYCA PROTEIN SEQUENCES

```

      .           .           .           .
1  EFLGGDMVNQSFICEADDEALLKSIVIQDCMWSGFSAAAKLEKVVSEKLA 50
  |||||||||||||||||||||||||||||||||||||||||||||||||||
86 EFLGGDMVNQSFICEADDEALLKSIVIQDCMWSGFSAAAKLEKVVSEKLA 135

      .           .           .           .
51 SYQASRKESAL.SSSQCQSQPPQSPIKSPSCDGSLSLGGTNRSSHGFLQD 99
  ||||||||| ||| ||||||| || ||||| ||||||||| |||||||||
136 SYQASRKESALSSSSPCQSQPPPSPLKSPSCHGSLSLGGTHRSSHGFLQD 185

      .           .           .           .
100 PSSDCVDPSVVPYPLNDTISNASSPCQDLMLETPPISSNSSSSESEEEEQ 149
  ||||||||||||||||||||| ||||||||||||| ||||||||||||| |||
186 PSSDCVDPSVVPYPLNDSISNASSPCQDLILETPPISSNSSSSESEEEEP 235

      .           .           .           .
150 EDDDEDCEEEEEIDVVTVEKRQTASRRMESSSHSQPSRPHHSPLVLKRCH 199
  || ||||||||||||||||||||| || | ||||||||||||| |||||||||
236 EDEDEDCEEEEEIDVVTVEKRQASAKRVESSSHSQPSRPHYSPLVLKRCH 285

      .           .           .           .
200 VPIHQHNYAASPSTKVDYVSSKRAKLESNARVLKQISNNRKCASPRSSDS 249
  ||||||||||||||||||||||||||||||||||||| ||||||||||||| |||
286 VPIHQHNYAASPSTKVDYVSSKRAKLESNIRVLKQISNNRKCASPRSSDS 335

      .           .           .           .
250 EENDKRRTHNVLERQRRNELKLSFFALRDQVPEVANNEKAPKVVILKKAT 299
  ||||| ||||||||||||||||||||||||||||||||| ||||||||| |||
336 EENDKRKTHNVLERQRRNELKLSFFALRDQVPEVASNEKAPKVVILKKAT 385

      .           .           .           .
300 EYVISMQEDERRLIRETEQLKYRKEQLKQRLQQLRNLV 337
  || || ||||||||||||||||||||||||||||| || |
386 EYAISLQEDERRLIRETEQLKYRKEQLKQRLQQLRNFV 423

```

A

COMPARISON OF XBMYCA3 AND XLMYCA PROTEIN SEQUENCES

```

      .           .           .           .
1  MPLNASFPSKNYDYDYDLQPCFFLEEENFYHQQSRLHAPAPSEDIWKKF 50
  ||||| ||||||||||||||||||||||||||||||||| |||||||||
5  MPLNANFPSKNYDYDYDLQPCFFLEEENFYHQQSRLQPPAPSEDIWKKF 54

      .           .           .           .
51 ELLPTPPLSPSRRSSQSSLFPSTADQLEMVTEFLGGDMVNQSFICEADDE 100
  ||||||||||||||||||||||||||||||||||||| ||||||||| |||
55 ELLPTPPLSPSRRSSQSSLFPSTADQLEMVTEFLGGDMVNQSFICEADDE 104

      .           .           .           .
101 ALLKSIVIQDCMWSGFSAAAKLEKVVSEKLASIQASRKESAL.SSSQCPS 149
  ||||||||||||||||||||||||||||||||||||| ||||| |||
105 ALLKSIVIQDCMWSGFSAAAKLEKVVSEKLASIQASRKESALSSSSPCQS 154

      .           .           .           .
150 QPPQSPLKSPSCHG..SLEGTHRSSHAFQLQDPSSDCVDPSVVPYPLNDS 197
  ||| ||||||||| || ||||||| ||||||||||||| |||||||||
155 QPPPSPLKSPSCHGSLSLGGTHRSSHGFLQDPSSDCVDPSVVPYPLNDS 204

198 ISN 200
  |||
205 ISN 207

```

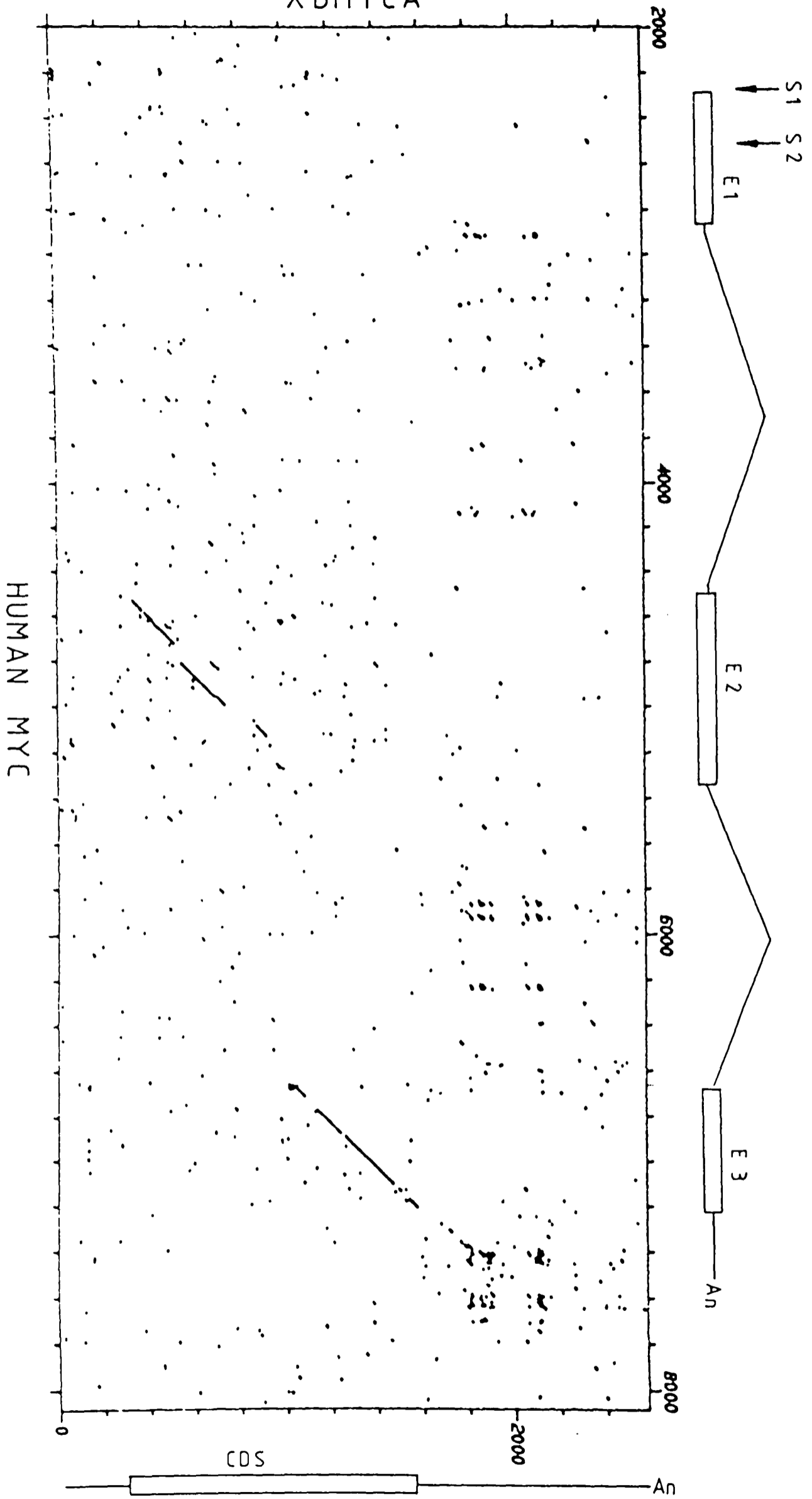
B

FIGURE 5.13

A: Comparison of XbmycA2 and XlmycA protein sequences using the program BESTFIT with a gap weight of 5.0 and a length weight of 0.3. The top sequence is XbmycA2 and the lower sequence is XlmycA. The XlmycA sequence is a translation from the XlmycA DNA sequence kindly supplied by M. King. % similarity = 96.000, quality = 310.8 and quality ratio = 0.922.

B: Comparison of XbmycA3 and XlmycA protein sequences using the program BESTFIT with a gap weight of 5.0 and a length weight of 0.3. The top sequence is XbmycA3 and the lower sequence is XlmycA. The XlmycA sequence is a translation from the XlmycA DNA sequence kindly supplied by M. King. % similarity = 94.362, quality = 180.3 and quality ratio = 0.907.

X BMYCA



S1 S2

E1

E2

E3

An

CDS

An

FIGURE 5.14

Comparison of human *c-myc* and XbmycA DNA sequences. Bases 1 - 564 of XbmycA3 were added to the 5' end of XbmycA2 to form the assembled XbmycA sequence as described in the text. The COMPARE program was used with a window size of 21 and a stringency of 14 to generate the points which were then plotted using the DOTPLOT program. E1, E2 and E3 represent the three exons of the human *c-myc* gene. CDS represents the putative coding sequence of XbmycA. An indicates the polyadenylation site in both genes. S1 and S2 indicate the mRNA start sites of the human *c-myc* gene. Intervening sequences in the human *c-myc* gene are represented by the thin diagonal lines between the exons. The human *c-myc* DNA sequence is from the GENBANK database release 40.0.

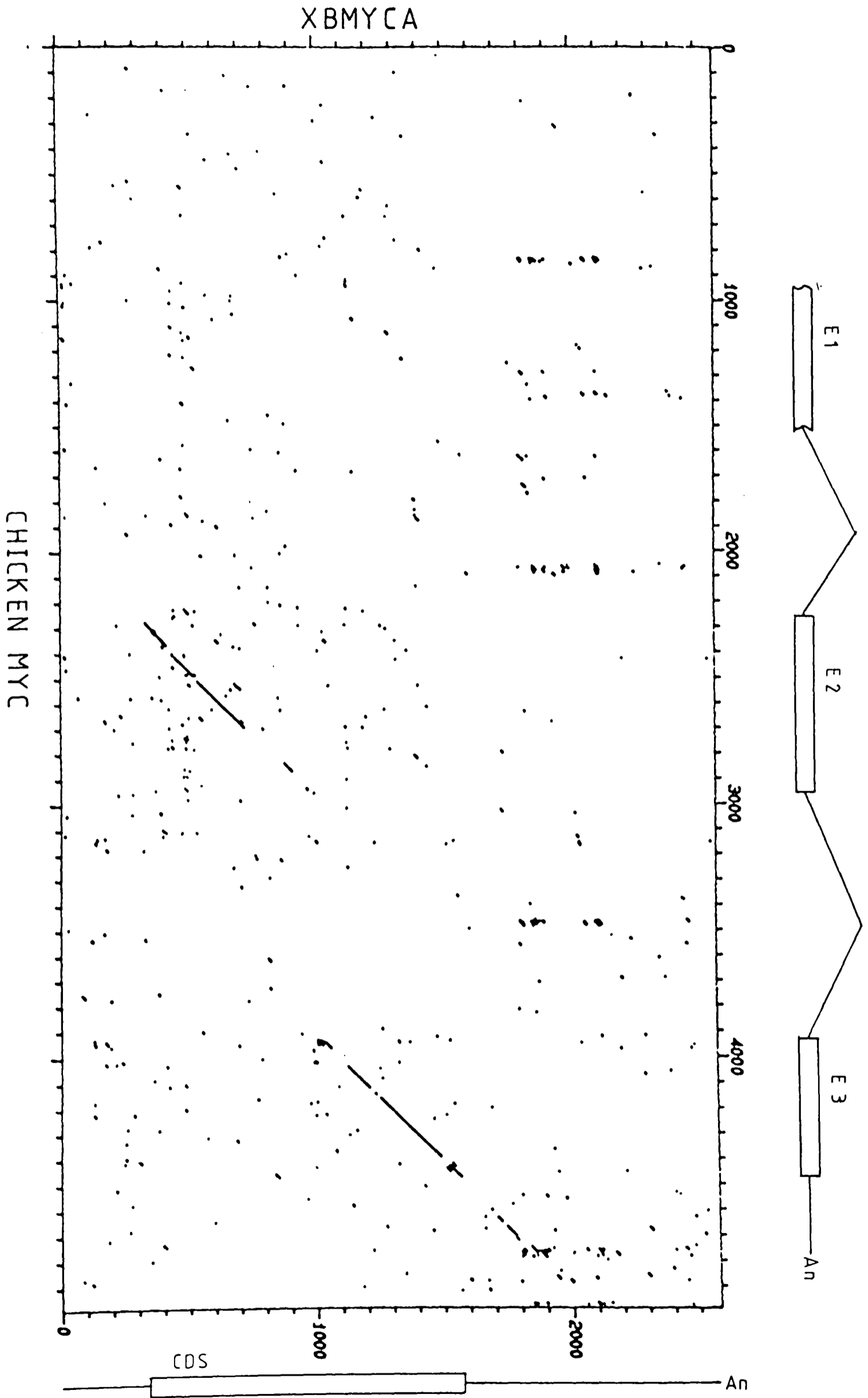


FIGURE 5.15

Comparison of chicken *c-myc* and XbmycA DNA sequences. Bases 1 - 564 of XbmycA3 were added to the 5' end of XbmycA2 to form the assembled XbmycA sequence as described in the text. The COMPARE program was used with a window size of 21 and a stringency of 14 to generate the points which were then plotted using the DOTPLOT program. E1, E2 and E3 represent the three exons of the chicken *c-myc* gene. CDS represents the putative coding sequence of XbmycA. An indicates the polyadenylation site in both genes. Intervening sequences in the chicken *c-myc* gene are represented by the thin diagonal lines between the exons. The limits of the first exon and the mRNA start sites of the chicken *c-myc* gene have not yet been precisely defined. Sizes are marked in base pairs. The chicken *c-myc* DNA sequence is from the GENBANK database release 40.0.

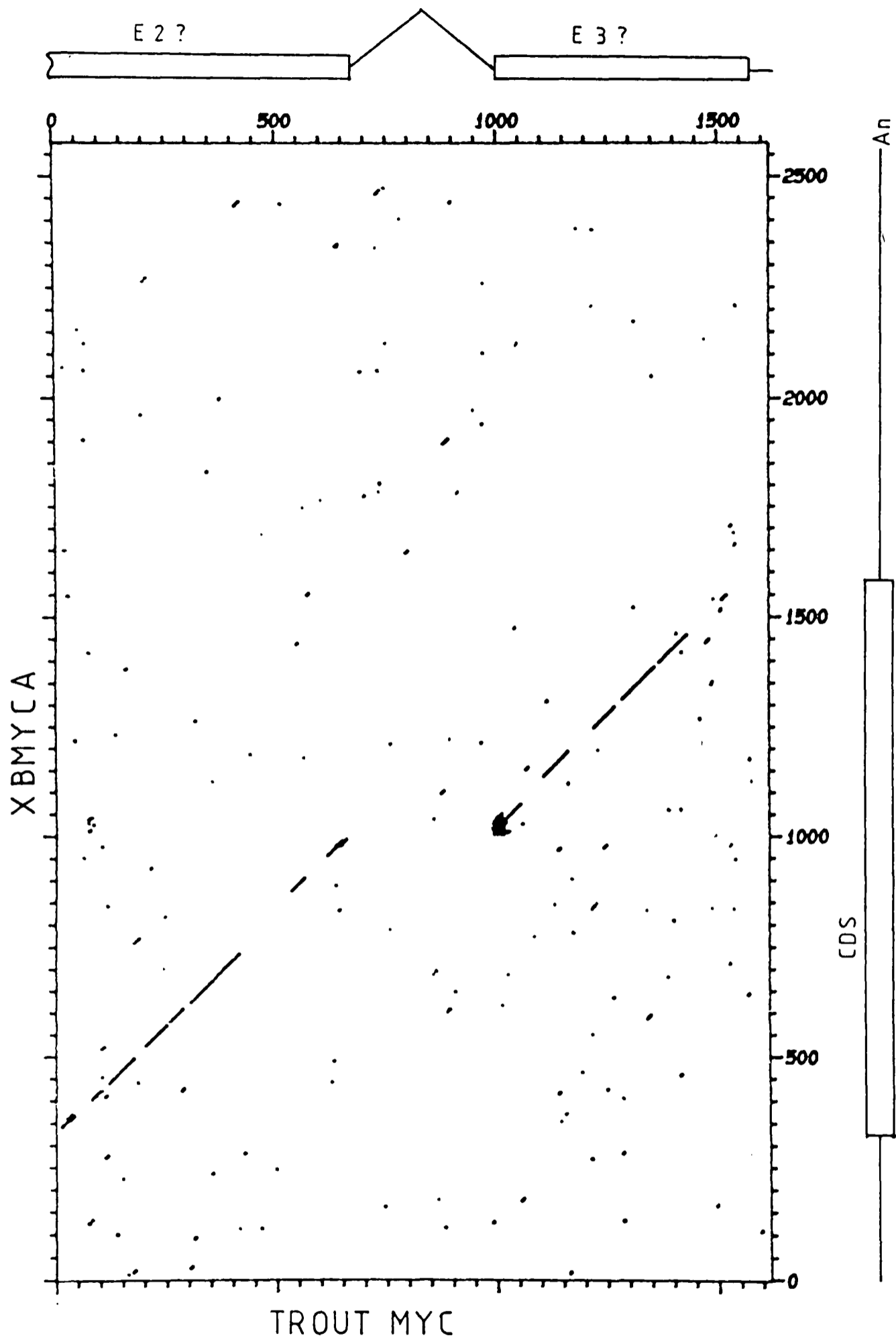
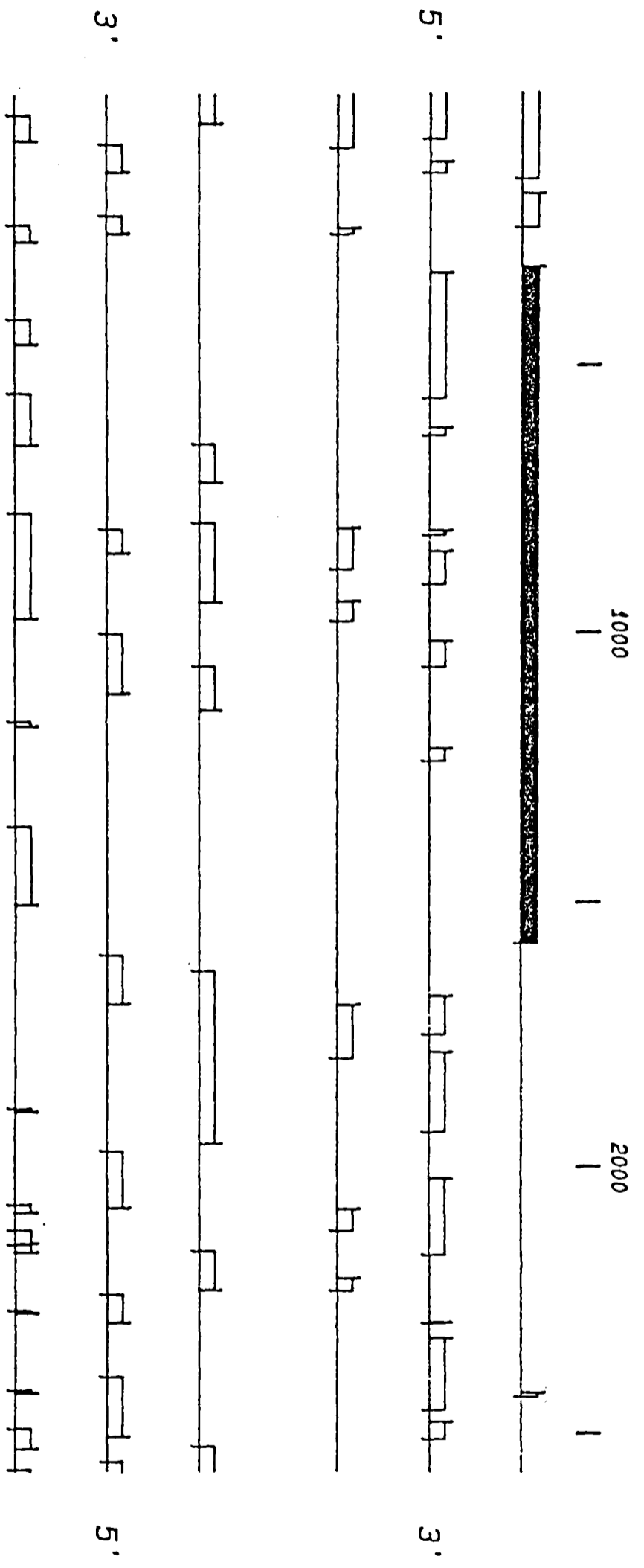


FIGURE 5.16

Comparison of trout *c-myc* and XbmycA DNA sequences. Bases 1 - 564 of XbmycA3 were added to the 5' end of XbmycA2 to form the assembled XbmycA sequence as described in the text. The COMPARE program was used with a window size of 21 and a stringency of 14 to generate the points which were then plotted using the DOTPLOT program. E2 and E3 represent at least two of the exons of the trout *c-myc* gene. CDS represents the putative coding sequence of XbmycA. An indicates the polyadenylation site of XbmycA. Intervening sequences in the trout *c-myc* gene are represented by the thin diagonal line between exons. The number of exons, the mRNA start sites and the polyadenylation site of the trout *c-myc* gene have not yet been defined. Sizes are marked in base pairs. The trout *c-myc* DNA sequence is from Van Beneden *et al* 1986.



OPEN READING FRAMES OF XBM YCA

FIGURE 5.17

Possible open reading frames of the XbmycA DNA sequence. Bases 1 – 564 of XbmycA3 were added to the 5' end of XbmycA2 to form the assembled XbmycA sequence as described in the text. Open reading frames were found using the program FRAMES. Boxes represent open reading frames in each of the six possible frames with start codons represented with ticks above the boxes and stop codons represented by ticks below the boxes. The filled box represents the putative XbmycA *c-myc* coding sequence.

COMPARISON OF C-MYC PROTEIN SEQUENCES

```

1                                     50
HUM          v ftnr l v ycd          q qq e
MOU          vn ftnr l v icd          h qq e
CHK          sa lp v yf. ee la a rg e
VIR aaaa sa lp v yf. ee la a rg e
TRO          ... s la i yvd n ded. h pgq
XLA magk an fp .l c ffl          h . r
XB2          .....
XB3          a fp .l c ffl          h . r ha
XB1          .....
CON ....MPLN-S --SKNYDYDY DS-QPYF--- EEE..NFY.. -QQ.-S-LQP

```

```

51          A BOX          RD1          100
HUM          g c p syvav tpfs lrgdndgggg
MOU          g c p syvavatsfs predddgggg
CHK          s aaa s c
VIR          m s aaa s c
TRO          p . s i
XLA          sqs l
XB2          .....
XB3          sqs l
XB1          .....
CON PAPPEDIWKK FELLPTPPLS PSRRS-L-S. ....-

```

```

101          RD 2          B BOX          150
HUM sf l p t i n
MOU nf m l p t i n
CHK l p s v
VIR l p s v
TRO f d v a y sqt l
XLA f ea all v
XB2 ..... .. f ea all v
XB3 f ea all v
XB1 .....
CON FPSTADQLEM VTE-LGGDMV NQSFICD-D. DE-F-KSIII QDCMWSGFS

```

```

151          RD 3          200
HUM          ... a d g..... . . . pnp rg
MOU          ... a d t..... . . . lsp rg
CHK          t r ggpaaasrp g p g p ppp.agp as
VIR          t r ggpaaasrp g p g p ppp.agp as
TRO t r l ta d avgdnaec p ..... . trln ny
XLA          alssspc qsq p l k pschgs
XB2          alsssqc qsq q i k pscdgs
XB3          alsssqc psq q i k pschg.
XB1          .....
CON AAKLEKVVSE KLASIQASRK ES----- . .-PP-S.P- -..S---A--

```

```

201          250
HUM h vcs ssly lqd saa e ss ks casqdssafs
MOU h vcs ssly lqd taa e ss ks ctssdstafs
CHK ag .....y lhd gaa a .sera ra a
VIR ag .....y lhd gaa a .sera ra a
TRO lqdpn sase cigpnts e ..itet k.
XLA l gg hrss hgf qdps v si nas
XB2 l gg nrss hgf qdps v ti nas
XB3 . eg hrss haf qdps v si ns.
XB1 .....
CON -SL--T---- ---L---ASD CIDPSVVFPY PLND--SP-- .....

```

RD 4

RD 5

FIGURE 5.18

Comparison of *c-myc* protein coding sequences using the program GAP (with command line option OUT) reiteratively, with a gap weight of 5.0 and a length weight of 3.0, until no more gaps are introduced between comparisons. Initial alignments were performed using the program PRETTY then the alignments were manually edited to increase the maximum number of matches. Hum = human, mou = mouse, chk = chicken, vir = viral, tro = trout, XIA = XlmycA, Xb2 = XbmycA2, Xb3 = XbmycA3 and Xb1= XbmycA1 *c-myc* protein coding sequences. Con = a consensus sequence generated using a plurality of 5. The putative *myc* boxes A and B are boxed (see 5.6.4). Underlined are the potential nuclear location signals, NL1 - 3 (see 5.6.2) and the potential rapid degradation signals, RD1 - 7 (see 5.6.3). The human, mouse, chicken and viral sequences are translations from the respective DNA sequences given in the GENBANK database release 40.0. The trout sequence is a translation of the trout *c-myc* DNA sequence given in Van Beneden *et al* 1986 and the XlmycA sequence is a translation of the XlmycA DNA sequence kindly supplied by M. King.

251 300

HUM pssdslsst ess .qgs e p v hee tt s e

MOU pssdsls ss ess .ras e p v hee tt s e

CHK pgan a a . gvd tt s e

VIR pgan a a . gvd tt s e

TRO skap t d a d nsgs ssgs dsedd e d

XLA s cq.... d i e issn sses eepe ed d

XB2 s cq.... d m e issn sses eepe dd d

XB3

XB1

CON P----P- -L-L..-TPP ..--SS....D-EE- nl3

RD 6

301 350

HUM q d s apgk s gsp sag ghsk p st

MOU q d s tpak s gsspsr ghsk p st

CHK q ed laeaneses st tease ehck ni

VIR q ed laeaneses st tease ehck ni

TRO d d a.vk cdp t .se ...tr st

XLA cde sask v . shs qpsr y pi

XB2 cde tasr m . shs qpsr pi

XB3

XB1

CON -E-EEEIDVV TVEKRQ---- R-ESS-.---- ----PHHSPL VLKRCHV--H

RD 7

351 400

HUM p rkd a v d v r t t

MOU p rkd a a d g s t

CHK p kve a l d g s t

VIR p kve a l d g s tl

TRO h rheq v l r ens s s s t t dy

XLA s kvd vss a esni a

XB2 s kvd vss a esna a

XB3

XB1 a

CON QHNYAA-PST ---YPA-KR- KL..-S-RVL KQISNNRKC- SPRSSDSEEN

401 450

HUM v r le a i

MOU r le a i

CHK v

VIR r v

TRO e d a ci

XLA k v s a

XB2 v v

XB3

XB1 k v a

CON DKRRTHNVLE RQRRNELKLS FFALRDOIPE VANNEKAPKV VILKKATEY- nl 2

451 491

HUM l v ae qk s edl k r hk e ca

MOU l i a hk ts kdl k r hk e ga

CHK l i s h a k r r hk e ra

VIR l l s h a k r r hn e ra

TRO y m t q vnk rk s h qk a q clsskr h

XLA i l e r r t ky k qr q fv.

XB2 i m e r r t ky k qr q lvv

XB3

XB1 i l e r r t kf k qr q lvv

CON -S-Q-DE-RL I-E-EQLR-R -EQLK--L-Q LRNS--..... nl 1

COMPARISON OF MYC BOX AMINO ACID SEQUENCES

MYC BOX A

c-myc PSEDIWKKFELLTPPLSP
N-myc .G.....
L-myc .S.....V.S..T..
(M)

MYC BOX B

c-myc IIIQDCMWSGFSAA
N-myc V.L.....RG
L-myc .RR.....R
(V T)

FIGURE 5.19

Alignment of *myc* box A and B amino acid sequences (see 5.6.4). The *c-myc* sequences are from a consensus generated from Figure 5.18 and in brackets is shown the deviation in the *c-myc* consensus: in *myc* box A a methionine is present as shown in the viral sequence and in *myc* box B a valine is present as shown in the *Xenopus* sequences and a threonine is present as shown in the trout sequence. The *myc* boxes are as described by Schwab *et al* 1983, Nau *et al* 1985 and Schwab 1985.

CHAPTER 6

DISCUSSION

6.1 Oncogene Homologies in *Xenopus*

Low stringency hybridisation conditions and highly labeled probes were employed to show the presence of many oncogene related sequences in the genome of three species of *Xenopus* and in some cases in the RNA of *X.borealis*. This result, in itself, is not too surprising since it is expected that genes important for development or normal cell growth might be conserved between a variety of organisms. Indeed the isolation of many oncogene related sequences from *Drosophila* (see chapter 1) supports such a view. Conversely the isolation, from many species, genes related to those in *Drosophila* development (i.e. the homeo box containing genes) also supports this view. Only organisms, such as yeast, which exhibit a radically different life style from insects and vertebrates appear not to have the full complement of oncogene related genes though this might reflect a high degree of divergence between the probe that is used and the yeast gene rather than differences in the mode of cell growth. Obviously some oncogene related sequences that are involved in cell type specific processes or in development might not be required in lower organisms hence they will not be present in their genomes (in a functional form?). Since cellular oncogenes are so highly conserved throughout evolution it seems to me that an analysis of cellular oncogene function is best carried out in the organism that can be best exploited for the particular question one is trying to answer. *Xenopus* is one organism that is ideal (in many respects) for studying cellular oncogene function during development and I have demonstrated the potential for isolating many such sequences from *Xenopus*.

6.2 The *X.borealis* C-myc Genes

I have described the isolation of *X.borealis* cDNA clones that represent genes homologous to the mammalian and avian *c-myc* genes. Nucleotide sequence analysis and preliminary southern blot analysis suggests there are at least two distinct types of *c-myc* genes in *X.borealis* and *X.laevis* (represented by XbmycA and B in *X.borealis*) and only one type in *X.tropicalis*. Also this

analysis showed that the XbmycA genes may have been duplicated to give rise to two related genes. As discussed previously this is consistent with a total genome duplication within the genus *Xenopus*. These studies also showed that the nucleotide and predicted amino acid sequences of the XbmycA type genes have a great degree of homology with other *c-myc* sequences which suggests the gene products have an important common function. Preliminary evidence suggests that at least one of the XbmycA genes may have a similar structure to avian and mammalian *c-myc* genes which may point to the structure of the *c-myc* gene being important in its regulation. Further sequence analysis of both genomic and cDNA clones along with southern blotting and expression data would help determine the precise relationship of the *c-myc* gene family members in *X.borealis* and would help us to understand whether each gene is in fact functional. If more than one *c-myc* gene is functional it would be extremely interesting to see if they play different roles. For example there may be *c-myc* equivalents of somatic and oocyte type genes like in the 5S gene system, but this is unlikely for the *c-myc* genes I have isolated (unless translational regulation or other controls are evoked) since they were all isolated from ovary cDNA libraries. An understanding of the *c-myc* gene family in *Xenopus* might actually be better done using the diploid species *X.tropicalis* since both *X.borealis* and *X.laevis* are tetraploid they should have an increased gene number which might complicate matters.

6.3 Expression of *X.borealis* C-myc Genes

The *c-myc* gene XbmycA of *X.borealis* is an example of a maternal mRNA that is made early in oogenesis and then subsequently inherited by the fertilised egg. The stage II previtellogenic oocyte appears to have maximal levels of synthesis with around 19 pg of XbmycA mRNA which subsequently declines to around 8 pg in the stage VI oocyte. I calculate that the non-dividing stage II oocyte has $\sim 1.2 \times 10^7$ molecules of XbmycA mRNA per cell therefore the stage VI oocyte has $\sim 5 \times 10^6$ molecules of XbmycA mRNA per cell which only decreases slightly on fertilisation. After fertilisation there is a period of extremely active cell division without detectable RNA synthesis (Newport and Kirschner 1982 a) during which the XbmycA mRNA is degraded such that by the time RNA synthesis begins (stage 8) there are only < 40 XbmycA mRNA molecules per cell. New transcription of the XbmycA gene is not detected till late gastrula (stage 11) by which time the inherited XbmycA mRNA has been degraded even further. There is then an increase in XbmycA

mRNA in the embryo during neurulation to about 130 copies of XbmycA mRNA per cell, followed by a decrease to ~30 copies per cell in the stage 33/34 tadpole. These calculations use cell numbers estimated by Gurdon, 1974 and averaged values over the embryo as a whole, which do not allow for the possibility that certain cell types may have localised increased expression of the XbmycA gene(s).

The total amount of poly(A)⁺ RNA has reached its final abundance in previtellogenic oocytes (Rosbash and Ford 1974, Cabada *et al* 1977, Dolecki and Smith 1979), does not turnover significantly within a 18 month period (Ford, Mathieson and Rosbash 1977) and many individual poly(A)⁺ RNAs cease to accumulate after stage II of oogenesis (Golden, Schafer and Rosbash 1980). Such observations suggest that *c-myc* RNA itself in a *Xenopus* oocyte is extremely stable with a half life in the order of several weeks, contrasting starkly with *c-myc* RNA half life of only a few tens of minutes in other systems (Dani *et al* 1984). This is supported by calculations of the type described by Perlman and Rosbash, 1978 which show (assuming that (1) the newly made RNA is stable, (2) there are 8 genes that are fully active, (3) the RNA polymerase transit time is 15 bases/second and (4) there are 10 RNA polymerases loaded/kilobase) that it would take nearly 100 days to synthesise the 19 pg of XbmycA mRNA in a stage II oocyte which is about half the time required to reach sexual maturity. Therefore without direct evidence of the XbmycA mRNA half life in oocytes it seems probable that it is in fact essentially stable. Masking proteins of the type described in chapter 1 may stabilise the RNA. Using *in vivo* labeled oocytes (Ford, Mathieson and Rosbash 1977) in conjunction with a sensitive RNase protection assay (Zinn, Dimaio and Manaitis 1983) might help to demonstrate the half life of XbmycA mRNA in oocytes. This might be not possible since XbmycA mRNA seems to represent one of the low abundance classes of poly(A)⁺ RNAs (Perlman and Rosbash, 1978). Injection of *in vitro* synthesised XbmycA RNA directly into oocytes may also answer the same question and might provide a basis for determining what sequences (or other factors) are important in the stabilisation of XbmycA RNA in oocytes.

Upon fertilisation degradation of XbmycA mRNA is induced. It is difficult to estimate the half life of the XbmycA mRNA during early cleavage due to the kinetics of degradation but it is of the order of 5 hours. These experiments should be repeated to determine this precisely. Even this half life is several

times greater than that described for mammalian systems (Dani *et al* 1984). If in *Xenopus* somatic cells *c-myc* mRNA has a half life (which remains to be determined) similar to mammalian systems and if during embryogenesis mechanisms that degrade *c-myc* mRNA are similar to somatic cells then factors that influence *c-myc* degradation may be limiting early in development. Such factors may be stored in the oocyte then activated on fertilisation or they may become newly made during development. Also it may be that *c-myc* mRNA is only released for translation slowly in embryos and after doing so becomes degraded. After stage 11 XbmycA mRNA assumes a more typical somatic level with slightly greater expression during neurulation. I have preliminary evidence that XbmycA mRNA is expressed to a higher level in the developing brain and nervous system, which might parallel observations in mouse development (Rupport, Goldowitz and Wille 1986). Further northern analysis with more finely dissected embryos in conjunction with *in situ* hybridisation experiments may define the localisation of expression of the XbmycA gene during early development. Also initial experiments suggest that the XbmycA genes are expressed in a variety of tissues at a low level which is similar to results for mammalian *c-myc* (Stewart, Bellve and Leder 1984). This analysis could be extended to other *Xenopus* tissues and known numbers of cells could be used to determine if any adult tissue has an unusually high level of *c-myc* expression; for example it would be of interest to determine *c-myc* expression in the developing brain compared with the adult brain.

As described in chapter 1 there is a strong correlation of *c-myc* mRNA expression with cell proliferation and a downregulation of *c-myc* mRNA expression upon differentiation. Obviously from the experiments I described *c-myc* mRNA expression is uncoupled from cellular proliferation during *Xenopus* oogenesis. Undoubtably this is a reflection on the mode of embryogenesis that *Xenopus* exhibits. There are two possible interpretations of the high levels of *c-myc* mRNA in the early oocyte. Firstly it may be required to sustain oocyte metabolism and therefore may be involved in the regulation of ribosomal RNA synthesis which is a major metabolic event occurring maximally after stage II of oogenesis (Rosbash and Ford 1974). Secondly it may be required to provide protein products to sustain embryo division through its early phase of transcriptional quiescence. This may mean the *c-myc* protein is involved in DNA synthesis, RNA processing or ribosome assembly. Possibly the *c-myc* protein has a variety of functions with different

roles in the developing oocyte or neuron to the proliferating embryonic or somatic cell.

6.4 The *X.borealis* C-myc Protein

In chapter 4 I described my attempts to make an antiserum against the XbmycA protein which were unsuccessful. To further study the expression of the protein in *Xenopus* and to attempt biochemical experiments desperately requires the availability of a homologous antiserum to the protein (I have previously described ways how this might be achieved). An immediate question to be asked with the availability of such an antiserum would be a study of XbmycA protein synthesis during oogenesis. It would be interesting to see if protein synthesis is coupled to mRNA expression or if there is a delay in synthesis due to translational control. Also it could be determined whether the XbmycA protein is synthesised in a stable fashion during oogenesis or does it turnover; this could also be compared with the protein half life in embryonic and somatic cells. Expression of the XbmycA protein during development could also be studied to see if it parallels the expression of XbmycA mRNA.

Many studies with such an antiserum using immunofluorescence *in situ* could be done to determine both the localisation of the XbmycA protein within a cell and its localisation between a variety of cell types. It is expected, providing the XbmycA protein has a function similar to other *c-myc* proteins, that it would be localised in the cell nucleus though possibly some becomes sequestered into the oocyte cytoplasm in a storage form for future use. Previously it has been shown that injection of DNA into the oocyte cytoplasm leads to the production of nucleus like structures suggesting the oocyte is rich in nuclear structural components (Forbes, Kirschner and Newport 1983). *In situ* hybridisation to different embryo stages using an anti-XbmycA antiserum would also help confirm whether the protein product indeed has a role in neural development.

An anti-XbmycA antiserum could also be used to study the characteristics of specifically modified XbmycA proteins. The antiserum could be used to analyse the fate of bacterially synthesised XbmycA protein containing specific deletions injected into oocytes or to analyse the fate of specifically deleted XbmycA proteins synthesised from mRNA made *in vitro* that has been injected

into oocytes. These experiments may address the question of which sequence(s) are responsible for nuclear location of the XbmycA protein and might help understanding of post-translational modifications on the protein structure and function. Similar experiments could be done to see if injection of XbmycA protein has any effect on oocyte maturation for instance by influencing cAMP levels. An early response of *Xenopus* oocytes to progesterone is a drop in cAMP levels which is the possible signal for induction of meiosis (Maller 1983) therefore such experiments may show a link between *c-myc* and meiotic cell divisions. If the oocyte does contain a large store of *c-myc* protein it is difficult to see a role for it in meiosis unless it is stored in an inactive form hence the previous experiment might not be particularly informative.

6.5 Some Other Future Experiments

Apart from the experiments eluded to above and in other chapters an analysis of *c-myc* function could be attempted by trying to block its function in early embryos. There are three possible ways that this might be carried out. Firstly if homologous antiserum was available this could be injected into oocytes and embryos to try and block *c-myc* function. It is known that injected antibodies are relatively stable and retain their antigenic specificity and they can inactivate certain nuclear enzymes such as RNA polymerase I (see Scheer 1986 for review). Hence a careful analysis of the effect of injected anti-*c-myc* antibodies might help elucidate the *c-myc* protein function in early development. Also injection of such antibodies into specific blastomeres might show whether the *c-myc* protein is required for the differentiation of certain cell types (e.g. neural development).

A second approach would attempt to achieve the same outcome as the first but involves injection of antisense message into oocytes and embryos to hopefully block translation of *c-myc* mRNA. This technology appears to work in oocytes as the translation of exogenously introduced globin mRNA can be blocked by the injection of antisense globin RNA (Melton 1985) and the translation of an endogenous ribosomal protein L1 gene can be blocked by the injection of antisense L1 RNA (Wormington 1986). As yet the technology cannot be used to block translation of specific mRNAs during development as many factors appear to influence the stabilisation/degradation and localisation of the injected antisense RNA in cleavage cells (Colman and Drummond 1986,

Melton and Rebagliati 1986), though these problems may be overcome soon. There are also two other points to note: 1) it is not clear how 'leaky' these treatments will be and 2) any *c-myc* protein inherited from the egg may complicate the analysis hence antibodies may also have to be employed.

The third approach is actually a modification of the second type of experiment and is based on the injection of cloned DNA into oocytes and embryos to direct the endogenous synthesis of antisense *c-myc* RNA to try and block the translation of *c-myc* mRNA. Previously it has been demonstrated that the expression of a thymidine kinase gene injected into mouse L cells is significantly reduced by co-injection of a plasmid that directs the synthesis of antisense thymidine kinase RNA (Izant and Weintraub 1984). Plasmid DNAs injected into fertilised *Xenopus* eggs can persist, in an episomal form, beyond the neurula stage of development (e.g. Krieg and Melton 1985, Mohun, Garrett and Gurdon 1986), providing the prospect for injection of plasmids into *Xenopus* fertilised eggs which direct the synthesis of antisense *c-myc* RNA. Such plasmid constructs could be designed to express antisense RNA at particular stages of development or in particular cell types. For example plasmids which direct antisense *c-myc* RNA under the control of the gastrula specific promoter from the GS17 gene (Krieg and Melton 1985) may block the translation of *c-myc* mRNA during early neurulation or under the control of the cardiac actin gene promoter (Mohun, Garrett and Gurdon 1986) may block the translation of *c-myc* mRNA in embryonic muscle cells. Experiments along these lines may be very revealing about the function of *c-myc*. Parallel experiments to try and over-express *c-myc* mRNA could also be done to try and see what effect high levels of *c-myc* protein have at a particular developmental stage or cell type. It should be borne in mind that transgenic mice containing *c-myc* gene constructs showed no developmental defects (Stewart *et al* 1985, Adams *et al* 1985, Leder *et al* 1986) hence experiments using plasmids directing synthesis of sense *c-myc* RNA in *Xenopus* embryos may not show any gross effect on development.

6.6 Recent Publications on the *X.laevis* C-myc Gene

Within the last two weeks of writing the bulk of this thesis three independent papers describing the *c-myc* gene in *X.laevis* have appeared (King, Roberts and Eisenman 1986, Godeau *et al* 1986, Taylor *et al* 1986). This obviously reflects the potential of *Xenopus* for studying cellular oncogene

function and overall the results described therein are in agreement with mine. Two studies have isolated homologous *c-myc* cDNA probes from *X.laevis* cDNA libraries (King, Roberts and Eisenman 1986, Taylor *et al* 1986) and one used heterologous probes (with inherent difficulties) to look at *c-myc* mRNA expression (Godeau *et al* 1986). As described in chapter 5 I have shown that the *X.laevis* and *X.borealis* *c-myc* cDNA sequences are very closely related.

Each of the above studies show a temporal pattern of *c-myc* mRNA expression during oogenesis and development which approximately agrees with my data. King and co-workers do not give any densitometric estimates of expression levels but from the original paper it looks very similar to my data. There is some disagreement on the expression of *c-myc* mRNA during oogenesis. I find a gradual decrease during oogenesis, Taylor *et al* find a gradual increase till stage III of oogenesis which subsequently declines and Godeau *et al* find an almost constant level of expression through all stages. This discrepancy could be due to individual batch variation and would be resolved by repeating the experiments. I find it difficult to believe that *c-myc* mRNA accumulates until stage III of oogenesis since as discussed previously no poly(A)⁺ RNAs have ever been shown to accumulate after stage II of oogenesis. The estimated levels of expression of the *c-myc* gene during oogenesis and development by Taylor *et al* are almost exactly the same as my estimates, this study also showed a small peak of expression during neurulation. King and co-workers report a higher level of *c-myc* mRNA in the ventral mesoderm of stage 13 neurula embryos but they did not analyse any dorsal tissue so this result might not disagree with the observed peak during neurulation described by myself and Taylor *et al*.

All three studies use anti-human *c-myc* antisera to study the expression of *X.laevis* *c-myc* proteins. Taylor and co-workers show the presence, in a *Xenopus* kidney cell line, of two nuclear phosphoprotein species with molecular weights of 61 and 64 kD which have a short half of about 30 minutes. The oocyte appears only to have the 64 kD species which is located both in the nucleus and cytoplasm; also since it was impossible to detect this protein by immunoprecipitation in the oocyte they suggest it has a very long half life. Godeau *et al* show the presence of a single 65 kD species and using immunofluorescence they show in follicle cells it is only located in the nucleus but in oocytes it is located mainly in the nucleus with some cytoplasmic staining which may be due to sequestering of nuclear product in the

cytoplasm. Also they demonstrate that the maximal levels of the putative *c-myc* protein synthesis occur in stage V oocytes hence invoking some kind of translational regulation of the *c-myc* mRNA. Taylor *et al* detect the presence of putative *c-myc* protein species of 62 and 77 kD and show again that maximal synthesis occurs in stage V oocytes. It seems the 77 kD species may be an artifact since it is not detected from proteins synthesised from hybrid selected *c-myc* mRNA and only two out of five antisera used detect the larger product. Hence although there are differences between each protein study, probably due to the different heterologous antiserum used, the overall conclusions appear to be similar: the *c-myc* protein is synthesised from mRNA that is translationally regulated in the oocyte and is mainly localised in the nucleus with some cytoplasmic storage and is apparently stable. In contrast it is only nuclear and has a short half life in *Xenopus* somatic cells.

6.7 Conclusions

My own work and the recent work described in 6.6 has shown that it is possible to undertake an analysis of *c-myc* function during early development in *Xenopus*. This work is basically at the preliminary stages with only a descriptive analysis of the *Xenopus c-myc* gene and product being described so far, hopefully the future will bring some real insight to how this mysterious oncogene functions and possibly *Xenopus* will once again be instrumental in this!

References

1. Abrams H.D., L.R. Rohrschneider, R.N. Eisenman: *Cell* 29 427-439, 1982.
2. Adams J.M., A.W. Harris, C.A. Pinkert, L.M. Corcoran, W.S. Alexander, S. Cory, R.D. Palmiter, R.L. Brinster: *Nature* 318 533-538, 1985.
3. Alitalo K., G. Ramsay, J.M. Bishop, S. Ohlsson-Pfeifer, W.W. Colby, A.D. Levinson: *Nature* 306 274-277, 1983.
4. Alitalo K., M. Schwab, C.C. Lin, H.E. Varmus, J.M. Bishop: *PNAS* 80 1707-1711, 1983.
5. Anderson D.M., M. Chamberlin, R.J. Britten, E.H. Davidson: *J.Mol.Biol* 155 281-309, 1982.
6. Anderson K.V., C. Nusslein-Volhard: *Nature* 311 223-227, 1984.
7. Anderson M.L.M., B.D. Young: In *Nucleic Acid Hybridisation-A Practical Approach*. Eds. B.D. Hames, S.J. Higgins, IRL Press, 1985.
8. Appleyard R.K.: *Genetics* 39 440-452, 1954.
9. Armelin H.A., M.C.S. Armelin, K. Kelly, T. Stewart, P. Leder, B.H. Corcoran, C.D. Stiles: *Nature* 310 655-700, 1984.
10. Aviv H., P. Leder: *PNAS* 69 1408-1412, 1972.
11. Balinsky B.I.: *An Introduction to Embryology*, 3rd edition, 1970. W.B. Saunders and Company, Philadelphia, London, Toronto.
12. Barbacid M.: *TIG* 2 188-192, 1986.
13. Barth L.E., J.L. Barth: *J.Embryol.Exp.Morphol* 7 210-222, 1954.
14. Battey J., C. Moulding, R. Taub, W. Murphy, T. Stewart, H. Potter, G. Lenoir, P. Leder: *Cell* 34 779-789, 1983.
15. Bauer S.R., M. Piechaczyk, K.B. Marcu, R.P. Norden, M. Potter, J.F. Mushinski: *Curr.Top.Microb.Immunol.* 132 339-344, 1986.
16. Beachy P.A., S.L. Helfland, D.S. Hogness: *Nature* 313 545-551, 1985.
17. Beckner S.K., S. Hattori, T.Y. Shih: *Nature* 317 71-72, 1985.
18. Bentley D.L., M. Groudine: *Nature* 321 702-706, 1986.

19. Benton W.D., R.W. Davis: *Science* 196 180-182, 1977.
20. Bernard O., S. Cory, S. Gerondakis, E. Webb, J.M. Adams: *EMBO J* 2 2375-2383, 1983.
21. Biggen M.D., T.J. Gibson, G.F. Hong: *PNAS* 80 3963-3965, 1983.
22. Birchmeier C., D. Broek, M. Wigler: *Cell* 43 651-621, 1985.
23. Birnstiel M.L., J. Speirs, I. Purdom, K. Jones, U.E. Loening: *Nature* 219 454-463, 1968.
24. Bisbee C.A., M.A. Baker, A.C. Wilson, J. Hadzi-Azimi, M. Fischberg: *Science* 195 785-787, 1977.
25. Bishop J.M.: *Ann.Rev.Biochem.* 52 301-354, 1983.
26. Bishop J.M.: *TIG* 7 245-249, 1985.
27. Bishop J.M., B. Drees, A.L. Katzen, T.B. Kornberg, M.A. Simon: *Cold.Spr.Har.Sym.Quan.Biol.* 50 727-732, 1985.
28. Bishop J.M., H.E. Varmus: In *Molecular Biology of Tumour Viruses, RNA Tumour Viruses, 2/Supplements & Appendices*, 2nd Edition, 1985. CSH Press, New York.
29. Blanchard J.M., M. Piechaczyk, C. Dani, J.C. Chambard, A. Franchi, J. Pouysegur, P. Jeanteur: *Nature* 317 443-445, 1985.
30. Blin N., Stafford D.W.: *NAR* 3 2303-2308, 1976.
31. Bogenhagen D.F., S. Sakonju, D.D. Brown: *Cell* 19 27-35, 1980.
32. Bolivar T., R.L. Rodriguez, P.J. Greene, M.C. Betlach, H.L. Heyrecker, H.W. Boyer, J.H. Crosa, S. Falkow: *Gene* 2 95-113, 1977.
33. Boyer H.W., D. Rouland-Dussoix: *J.Mol.Biol* 41 459-472, 1969.
34. Brodeur G.M., R.C. Seeger, M. Schwab, H.E. Varmus, J.M. Bishop: *Science* 224 1121-1124, 1984.
35. Brown D.D.: *Cell* 37 359-365, 1984.
36. Brown D.D., E. Littna: *J.Mol.Biol* 20 81-94, 1966.
37. Brown D.D., E. Littna: *J.Mol.Biol.* 20 95-112, 1966.
38. Brown D.D., M.S. Schlissel: *Cell* 42 759-767, 1985.
39. Bunte T., I. Greiser-Wilke, P. Donner, K. Moelling: *EMBO J* 1 919-927, 1982.

40. Cabada M.O., C. Darnbrough, P.J. Ford, P.C. Turner: *Dev.Biol* 57 427-439, 1977.
41. Call J., M. Righi, C. deTaisne, C. Dissous, A. Gigonne, D. Stehlin: *EMBO J* 2 2189-2194, 1983.
42. Campisi J., H.E. Gray, A.B. Pardee, M. Dean, G.E. Sonenshein: *Cell* 36 241-247, 1984.
43. Capco D., W.R. Jeffrey: *Dev.Biol.* 89 1-12, 1982.
44. Capco D., H. Jackle: *Dev.Biol.* 94 41-50, 1982.
45. Carpenter C.D., W.H.Klein: *Dev.Biol.* 91 43-49, 1982.
46. Carpenter G.: *Cell* 37 357-358, 1984.
47. Carrasco A.E., W. McGinnis, W.J. Gehring, E.M. DeRobertis: *Cell* 37 409-414, 1984.
48. Chen I., T. Mak, J. O'Rear, H. Temin: *J.Viro* 40 800-811, 1981.
49. Colby W., E. Chen, D. Smith, A. Levinson: *Nature* 301 722-725, 1983.
50. Collins S., M. Groudine: *Nature* 298 679-681, 1982.
51. Colman A., D. Drummond: *J.Embryol.Exp.Morphol.* 97 Supplement 197-209, 1986.
52. Cooper J.A., B.M. Sefton, T.Hunter: *Mol.Cell.Biol* 4 30-37, 1984.
53. Cooper G.M.: *Science* 217 801-806, 1982.
54. Coppola J.A., M.D. Cole: *Nature* 320 760-763, 1986.
55. Corcoran L.M., J.M. Adams, A.R. Dunn, S. Cory: *Cell* 37 113-122, 1984.
56. Courtneidge S.A., A.D. Levinson, J.M. Bishop: *PNAS* 77 3783-3787, 1980.
57. Curran T., G. Peters, C. VanBeveren, N.M. Teich, I.M. Verma: *J.Viro* 44 674-682, 1982.
58. Curran T., J. Morgan: *Science* 229 1265-1268, 1985.
59. Crawford L.V.: *Proc.Roy.Soc.Lond B* 226 93-97, 1983.
60. Czernilofsky A.P., A.D. Levinson, H.E. Varmus, J.M. Bishop, E. Tischer, H.M. Goodman: *Nature* 287 198-203, 1980.
61. Dalla-Favera R., E.P. Gelmann, S. Martinotte, G. Franchini, T.S. Papas, R.C. Gallo, F. Wong-Staal: *PNAS* 79 6497-6501,

- 1982.
62. Dani C., M. Blanchard, M. Piechaczyk, S.El-Sabouty, L. Marty, P.H. Jeanteur: PNAS *81* 7046-7050, 1984.
 63. Darnbrough C., P.J. Ford: Dev.Biol *50* 285-301, 1976.
 64. Darnbrough C., P.J. Ford: Eur.J.Biochem *113* 415-424, 1981.
 65. Davidson E.H.: Gene Activity In Early Development, 2nd Edition, Academic Press, Inc, 1976.
 66. Dawid I.B., B.K. Kay, T.D. Sargent: In Gene Structure and Regulation in Development, 171-182. A.R. Liss Inc, New York, 1983.
 67. Dawid I.B., S.R. Hayne, M. Jamrich, E. Jonas, S. Miyatani, T.D. Sargent, J.A. Winkles: J.Embryol.Exp.Morphol. *89* Supplement 113-124, 1985.
 68. Dawid I.B., T.D. Sargent: TIG *2* 47-50, 1986.
 69. Defeo-Jones D., E. Scolnick, R. Koller, R. Dhar: Nature *306* 707-709, 1983.
 70. Denis H., M. Wegnez: Biochimie *55* 1137-1151, 1973.
 71. DeRobertis E.M.: Cell *32* 1021-1025, 1983.
 72. Desplan C., J. Theis, P.H. O'Farrell: Nature *318* 630-635, 1985.
 73. Devereux J., P. Haerberli, O. Smithies: NAR *12* 387-395, 1984.
 74. Dinardo , J.M. Kuner, J. Theis, P.H. O'Farrell: Cell *43* 59-69, 1985.
 75. Dixon L.K., P.J. Ford: Dev.Biol *93* 478-497, 1982.
 76. Dmitrovsky E., W.M. Kuehl, G.F. Hollis, I.R. Kirsch. T.P. Bender, S. Segel: Nature *322* 748-750, 1986.
 77. Dolecki G.J., L.D. Smith: Dev.Biol *69* 217-236, 1979.
 78. Dolecki G.J., S. Wannakrairoji, R. Lum, G. Wang, H.D. Riley, R. Carlas, A. Wang, T. Humpherys: EMBO J *5* 925-930, 1986.
 79. Donis-Keller H., A.M. Maxam, W. Gilbert: NAR *4* 2527-2538, 1977.
 80. Donner L., A. Fedele, C.F. Garon, S.J. Anderson, C.J. Sherr: J.Viro *41* 489-500, 1982.
 81. Donner P., I. Greiser-Wilke, K. Moelling: Nature *296* 262-266, 1982.

82. Doolittle R.F., M.W. Hunkapiller, L.E. Hood, S.G. Devare, K.C. Robbins, S.A. Aaronson, H.N. Antoniades: *Science* 221 275-277, 1983.
83. Dotto G.P. M.Z. Gilman, M. Maruyama, R.A. Weinberg: *EMBO J* 5 2853-2857, 1986.
84. Downward J., Y. Yarden, E. Mayes, G. Scrace, N. Totty, P. Stockwell, A. Ullrich, J. Schlessinger, M.D. Waterfield: *Nature* 307 521-527, 1983.
85. Duesberg P.H.: *Nature* 304 219-226, 1983.
86. Dumont J.N.: *J.Morphol.* 136 152-180, 1972.
87. Dworkin M.B., I.B. Dawid: *Dev.Biol.* 76 435-448, 1980.
88. Dworkin M.B., I.B. Dawid: *Dev.Biol.* 76 449-464, 1980.
89. Dworkin M.B., J.W.B. Hershey: *Mol.Cell.Biol.* 1 983-993, 1981.
90. Dworkin M.B., B.K. Kay, J.W.B. Hershey, I.B. Dawid: *Dev.Biol* 86 502-504, 1981.
91. Dworkin M.B., A. Shrutkowski, M. Baugarten, E. Dworkin-Rastl: *Dev.Biol* 106 289-295, 1984.
92. Dworkin M.B., A. Shrutkowski, E. Dworkin-Rastl: *PNAS* 82 7636-7640, 1985.
93. Dynan W.S.: *TIG* 1 269-270, 1985.
94. Efstratiadis A. J.W. Posakony, T. Maniatis, R.M. Lawn, C. O'Connel, R.A. Spritz, J.K. DeRiel, B.G. Forget, S.M. Weissman, J.L. Slightom, A.L. Blechl, O. Smithies, F.E. Baralle, C.C. Shoulders, N.J. Proudfoot: *Cell* 21 653-668, 1980.
95. Eick D., M. Piechaczyk, B. Henglein, J.M. Blanchard, B. Traub, E. Koffler, S. Wiest, G. Lenoir, G. Barnkam: *EMBO J* 4 3717-3725, 1985.
96. Einat M., D. Resnitzky, A. Kimchi: *Nature* 313 597-600, 1985.
97. Eisenman R.N., C.Y. Tachibana, H.D. Abrams, S.R. Hann: *Mol.Cell.Biol* 5 114-126, 1985.
98. Ellis R.W., D. Defeo, T.Y. Shih, M.A. Gonda, H.A. Young, N. Tsuchida, D.R. Lowy, E.M. Scolnick: *Nature* 292 506-511, 1981.
99. Engelke D.R., S.Y. Ng, B.S. Shastry, R.G. Roeder: *Cell* 19 717-728, 1980.
100. Enrietto P.J., M.J. Hayman: *Proc.Roy.Soc.Lond B* 226 83-92,

1985.

101. Evan G.I., D.C. Hancock, T.D. Littlewood, N.S. Gee: *Curr.Top.Microb.Immunol.* 132 362-374, 1986.
102. Fahrlander P.D., M. Piechaczyk, K.B. Marcu: *EMBO J* 4 3195-3202, 1985.
103. Feinberg A.P., B. Vogelstein: *Anal.Biochem.* 132 6-13, 1983.
104. Feinberg A.P., B. Vogelstein: *Anal.Biochem.* 137 266-267, 1984.
105. Feramisco J.R., D.M. Helfman, J.E. Smart, K. Burridge, G.P. Thomas: *J.Biol.Chem* 257 11024-11031, 1982.
106. Forbes D.J., M.W. Kirschner, J.W. Newport: *Cell* 34 13-23, 1983.
107. Ford P.J.: *Nature* 233 561-564, 1971.
108. Ford P.J., T. Mathieson, M. Rosbash: *Dev.Biol.* 57 417-426, 1977.
109. Ford P.J., E.M. Southern: *Nature* 241 7-12, 1973.
110. Gazin C., S. Dupont de Dinechin, A. Hampe, J.M. Masson, P. Martin, D. Stehelin, F. Galibert: *EMBO J* 3 383-387, 1984.
111. Gazin C., M. Rigolet, J.P. Briand, M.H.V. Van Regenmortel, F. Galibert: *EMBO J* 5 2241-2250, 1986.
112. Gehring W.J.: *Cell* 40 3-5, 1985.
113. Gergen J.P., R.H. Stern, P.C. Wensink: *NAR* 7 2115-2136, 1979.
114. Ginsberg A.M., B.O. King, R.G. Roeder: *Cell* 39 479-489, 1984.
115. Ghrist B.F.D., R.P. Ricciardi: *J.Mol.Evol.* 2 177-181, 1986.
116. Godeau F., H. Persson, H.E. Gray, A.B. Pardee: *EMBO J* 5 3571-3577, 1986.
117. Golden L., U. Schafer, M. Rosbash: *Cell* 22 835-844, 1980.
118. Gonda T.J., D.K. Sheiness, J.M. Bishop: *Mol.Cell.Biol.* 2 617-624, 1982.
119. Gonda T.J., D. Metcalf: *Nature* 310 249-251, 1984.
120. Gough J.A., N.E. Murray: *J.Mol.Biol.* 166 1-19, 1983.
121. Goustin A.S., C. Betsholtz, S. Pfeifer-Ohlsson, H. Persson, J. Rydnert, M. Bywater, G. Holmjren, C-H. Heldin,

- B. Westermark, R. Ohlsson: *Cell* 47 301-312, 1985.
122. Graf T., H. Beug: *Biochim.Biophys.Acta.* 516 269-299, 1978.
123. Graham M., J.M. Adams: *EMBO J* 5 2845-2851, 1986.
124. Graziani Y., J.L. Maller, Y. Sugimoto, R.L. Erikson: In *Cancer Cells* 2 27-35, 1984. Cold Spring Harbour Press, New York.
125. Green M.R., R. Treisman, T. Maniatis: *Cell* 35 137-148, 1983.
126. Green S., P. Chambon: *Nature* 324 615-617, 1986.
127. Greenberg M.E., E.B. Ziff: *Nature* 311 433-438, 1983.
128. Groffen J., J.R. Stephenson, N. Heisterkamp, A. deKlein, C.R. Bartram, G. Grosveld: *Cell* 36 93-99, 1984.
129. Gubler U., B.J. Hoffman: *Gene* 25 263-269, 1983.
130. Gurdon J.B.: *The Control of Gene Expression in Animal Development.* Harvard University Press, 1974.
131. Gurdon J.B., T.J. Mohun, S. Brennan, S. Cascio: *J.Embryol.Exp.Morphol.* 89 Supplement 125-136, 1985.
132. Gurdon J.B., T.J. Mohun, S. Fairman, S. Brennan: *PNAS* 82 139-143, 1985.
133. Gurdon J.B., S. Fairman, T.J. Mohun, S. Brennan: *Cell* 41 913-922, 1985.
134. Gurdon J.B., S. Fairman: *J.Embryol.Exp.Morphol.* 97 Supplement 75-84, 1986.
135. Hall A.: In *Oxford Surveys of Eukaryotic Genes* 1 111-145, 1984. Oxford University Press.
136. Hall A.: In *Genetic Engineering* 5 61-117, 1986. Academic Press Inc.
137. Hann S.R., H.D. Abrams, L.R. Rohrschneider, R.N. Eisenman: *Cell* 34 789-798, 1983.
138. Hann S.R., C.B. Thompson, R.N. Eisenman: *Nature* 314 366-369, 1985.
139. Hart C.P., A. Awgulewitsch, A. Fainsod, W. McGinnis, F.H. Ruddle: *Cell* 43 9-18, 1985.
140. Harvey R.P., C.J. Tabin, D.A. Melton: *EMBO J* 5 1237-1244, 1986.
141. Hauser C.A., A.L. Joyner, R.D. Klein, T.K. Learned, G.R. Martin, R. Tijan: *Cell* 43 19-28, 1985.

142. Hayman J.M.: *Curr.Top.Microb.Immunol.* 98 109-126, 1983.
143. Hayman J.M.: *TIG* 2 260-262, 1986.
144. Hayward W.S. B.G. Neel, S. Astrin: *Nature* 296 475-480, 1981.
145. Heisterkamp N., J.R. Stephenson, J. Groffen, P.H. Hansen, A. deKlein, C.R. Bartram, G. Grosveld: *Nature* 306 239-242, 1983.
146. Heldin C-H., B. Westermark: *Cell* 37 9-20, 1984.
147. Hoffman F.M., L.D. Fresco, H. Hoffman-Falk, B. Shilo: *Cell* 35 393-401, 1983.
148. Hoffman-Falk H., P. Einat, B. Shilo, F.M. Hoffman: *Cell* 32 589-598, 1983.
149. Hohn B., K. Murray: *PNAS* 74 3259-3263, 1977.
150. Holmes D.S., M. Quigley: *Anal.Biochem.* 114 193-197, 1981.
151. Hu N., J. Messing: *Gene* 17 271-277, 1982.
152. Huebner R.J., G.J. Todaro: *PNAS* 64 1087-1094, 1969.
153. Hunter T.: *TIB* 10 275-280, 1985.
154. Huynh T.V., R.A. Young, R.W. Davis: In *DNA Cloning* 1, 49-78, 1985. Editor D. Glover, IRL Press.
155. Izant J., H. Weintraub: *Cell* 36 1007-1015, 1984.
156. Jakobovits A., M. Schwab, J.M. Bishop, G.R. Martin: *Nature* 318 188-191, 1985.
157. Jansen H.W., B. Ruckert, R. Lurz, K. Bister: *EMBO J* 2 1969-1976, 1983.
158. Jamrich M., T.D. Sargent, I.B. Dawid: *Cold.Spr.Har.Sym.Quant.Biol.* 50 31-31, 1985.
159. Jonas E., T.D. Sargent, I.B. Dawid: *PNAS* 82 5413-5417, 1985.
160. Kamata T., J.R. Feramisco: *Nature* 310 147-150, 1984.
161. Kan N.C., C.S. Flordellis, C.F. Gason, P.H. Deusberg, T.S. Papas: *PNAS* 80 6566-6570, 1983.
162. Kataoka T., S. Powers, C. McGill, O.Fasono, J. Strathern, J. Broach, M. Wigler: *Cell* 37 437-445, 1984.
163. Keath E.J., P.G. Caimi, M.D. Cole: *Cell* 39 339 , 1984.
164. Kelly K., B.H. Corcoran, C.D. Stiles, P. Leder: *Cell* 35 603-610,

- 1983.
165. Kelly K., U. Siebenlist: *Ann.Rev.Immunol.* 4 317-338, 1986.
 166. King M.L., E. Barklis: *Dev.Biol.* 112 203-212, 1985.
 167. King M.W., J.M. Roberts, R.N. Eisenman: *Mol.Cell.Biol.* 6 4499-4508, 1986.
 168. Kingston R.E., A.S. Baldwin, P.A. Sharp: *Nature* 312 280-282, 1984.
 169. Kirby K.S.: *Biochem.J.* 96 266-269, 1965.
 170. Kirschner M., J. Newport, J. Gerhart: *TIG* 1 41-47, 1985.
 171. Knight E., E.D. Anton, D. Fahey, B.K. Friedland, G.J. Jonak: *PNAS* 82 1151-1154, 1985.
 172. Kohl N.E., N. Kanada, R.R. Schreck, G. Bruns, S.A. Latt, F. Galibert, F.W. Alt: *Cell* 35 359-367, 1983.
 173. Kohl N.E., E. Legouy, R.A. DePinho, P.D. Nisen, R.K. Smith, C.E. Gee, F.W. Alt: *Nature* 319 73-77, 1986.
 174. Kozak M.: *NAR* 12 857-872, 1984.
 175. Krieg P.A., D.A. Melton: *EMBO J* 4 3463-3471, 1985.
 176. Krieg P.A., D.A. Melton: *Meth.Enz.* In Press. Academic Press Inc.
 177. Krontiris T.G., G.M. Cooper: *PNAS* 78 1181-1184, 1981.
 178. Lachman H., A. Skoultchi: *Nature* 310 592-594, 1984.
 179. Lachman H., G. Cheng, A. Skoultchi: *PNAS* 83 6480-6484, 1986.
 180. Laemmli U.K.: *Nature* 227 680-685, 1970.
 181. Land H., L.F. Parada, R.A. Weinberg: *Nature* 304 596-602, 1983.
 182. Lanford R.E., P. Knanda, R.C. Kennedy: *Cell* 46 575-582, 1986.
 183. Laskey R.A., A.D. Mills: *FEBS Lett.* 82 314-316, 1977.
 184. Laskey R.A., A.D. Mills, J.B. Gurdon, G.A. Partington: *Cell* 11 345-351, 1977.
 185. Laughon A., M.P. Matthew: *Nature* 310 25-31, 1984.
 186. Lautenberger J.A., R.A. Schultz, C.F. Garan, P.N. Tsiichlis, T.S. Papas: *PNAS* 78 1518-1522, 1981.

187. Leder A., P. Pattengale, A. Kuo, T.A. Stewart, P. Leder: *Cell* 45 485-495, 1986.
188. Leder P., J. Battey, G. Lenoir, C. Moulding, W. Murphy, H. Potter, T. Stewart, R. Taub: *Science* 222 766-770, 1983.
189. Lee W.H., A.C. Murphee, W.F. Benedict: *Nature* 309 458-460, 1984.
190. Lehrach H.D., D. Diamond, J.M. Wozney, H. Boedtker: *Biochemistry* 16 4743-4751, 1977.
191. Leonard D.A., J.M. LaMarca: *Dev.Biol.* 45 199-202, 1975.
192. Levine M., G.M. Rubin, R.Tijan: *Cell* 38 667-673, 1984.
193. Levinson A.D.: *TIG* 2 81-85, 1986.
194. Lewin B.: *Gene Expression 2 Eukaryotic Chromosomes*, Wiley & Sons Inc, 1980.
195. Linal M., M. Groudine: *PNAS* 82 53-57 1985.
196. Little C.D., M.M. Nau, D.N. Carney, A.F. Gazdar, J.D. Minna: *Nature* 306 194-196, 1983.
197. Liveh E., L. Glazer, D. Segal, J. Schlessinger, B.Shilo: *Cell* 40 599-607, 1985.
198. MacDonald P.M., P. Ingham, G. Struhl: *Cell* 47 721-734, 1986.
199. MacGregor H.C.: *Heredity* 44 3-35, 1980.
200. Madhavan K., D. Bilodeau-Wentworth, S.C. Wadsworth: *Mol.Cell.Biol.* 5 7-16, 1985.
201. Mairy M., H. Denis: *Dev.Biol.* 24 143-165, 1971.
202. Maller J.: *Adv.Cyclic.Nucleotide.Res.* 13 295, 1983.
203. Mandel M., A. Higa: *J.Mol.Biol.* 53 154-162, 1970.
204. Maniatis T., E.F. Fritsch, J. Sambrook: *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbour Press, New York, 1984.
205. Mattaj I., R. Zeller, A. Carrasco, M. Jamrich, S. Lienhard, E.M. DeRobertis: *In Oxford Surveys Of Eukaryotic Genes* 2 120-140, 1985. Oxford University Press.
206. McDonnell M.W., M.W. Simon, F.W. Studier: *J.Mol.Biol.* 110 119-146, 1977.
207. Mellon P.A., K. Pawson, K. Bister, G.S. Martin, P.H. Deusberg: *PNAS* 75 5874-5878, 1978.

208. Melton D.A.: PNAS 82 144-148, 1985.
209. Melton D.A., P. Krieg, M.R. Rebagliati, T. Maniatis, K. Zinn, M.R. Green: NAR 12 7035-7056, 1984.
210. Melton D.A., M.R. Rebagliati: J.Embryol.Exp.Morphol. 97 Supplement 211-221, 1986.
211. Messing J.: Meth.Enz. 101 20-79, 1983. Academic Press Inc.
212. Miake-Lye R., M.W. Kirschner: Cell 41 165-175, 1985.
213. Miale I.L., S.L. Sidman: Exp.Neurol. 4 277-296, 1961.
214. Michitsch R.W., P. Melera: NAR 13 2545-2558, 1985.
215. Miller A., T. Curran, I.M. Verma: Cell 36 51-60, 1984.
216. Moen T.L., M. Namenwirth: Dev.Biol. 58 1-10, 1977.
217. Mohun T.J., S. Brennan, N. Dathan, S. Fairman, J.B. Gurdon: Nature 311 716-720, 1984.
218. Mohun T.J., N. Garret, J.B. Gurdon: EMBO J 5 3185-3193, 1986.
219. Mrockowski B., G. Mosig, S. Cohen: Nature 309 270-273, 1984.
220. Muller M.M., A.E. Carrasco, E.M. DeRobertis: Cell 39 157-162, 1984.
221. Muller R., I.M. Verma, E.D. Adamson: EMBO J 2 679-684, 1983.
222. Muller R., D. Muller, L. Guilbert: EMBO J 3 1887-1890, 1984.
223. Muller R., R. Bravo, J. Burckhardt, T. Curran: Nature 312 716-720, 1984.
224. Muller R., I.M. Verma: Curr.Top.Microb.Immuno. 112 75-115, 1985.
225. Muller R., E.F. Wagner: Nature 311 438-442, 1984.
226. Muller R., T. Curran, D. Muller, L. Guilbert: Nature 314 548-549, 1985.
227. Muller R., T. Curran, J. Burckhardt, U. Ruther, E.F. Wagner, R. Bravo: In Cancer Cells 3 289-299, 1985. Cold Spring Harbour Press, New York.
228. Murray N.E.: In Lambda II, Editors R.W Hendrix, J.W. Roberts, F.W. Stahl, R.S. Weisberg, 1983. Cold Spring Harbour Press, New York.

229. Murray N.E., W.J. Brammer, K. Murray: *Mol.Gen.Genet.* 150 53-61, 1977.
230. Nau M.M., D.N. Carney, J. Battey, B. Johnson, C. Little, A. Gazdar, J.D. Minna: *Curr.Top.Microb.Immunol.* 113 172-177, 1984.
231. Nau M.M., B.J. Brooks, J. Battey, E. Sausuille, A.F. Gazdar, I.R. Kirsch, O.W. McBride, V. Bertness, G.F. Hollis, J.D. Minna: *Nature* 318 69-73, 1985.
232. Neel B.G., W.S. Hayward, H.L. Robinson, J. Faj, S. Astrin: *Cell* 23 323-334, 1981.
233. Nepveu A., K.B. Marcu: *EMBO J* 5 2859-2856, 1986.
234. Neumann-Silberberg F.S., E. Schejtr, F.M.Hoffman, B.Shilo: *Cell* 37 1027-1033, 1984.
235. Newport J., M. Kirschner: *Cell* 30 675-686, 1982.
236. Newport J., M. Kirschner: *Cell* 30 687-696, 1982.
237. Nieuwkoop P.D., J. Faber: *Normal Table of Xenopus laevis (Daudin)*, Amsterdam: North-Holland, 1956.
238. Nusse R.: *TIG* 2 244-247, 1986.
239. O'Farrell P.H., E. Kutter, M. Nalcanishi: *Mol.Gen.Genet.* 179 421-435, 1980.
240. Okayama H., P. Berg: *Mol.Cell.Biol* 2 161-170, 1982.
241. Papas T.S., N.K. Kan, D.K. Watson, C.S. Flordellis, M.C. Psallidopoulos, J. Lautenberger, K.P. Samual, P.H. Deusberg: *In Cancer Cells* 2 153-163, 1984. Cold Spring Harbour Press, New York.
242. Payne G.S., J.M. Bishop, H.E. Varmus: *Nature* 295 209-214, 1982.
243. Pelham H.R.B., D.D. Brrown: *PNAS* 77 4170-4174, 1980.
244. Pelham H.R.B., W.M. Wormington, D.D. Brown: *PNAS* 78 1760-1764, 1981.
245. Perlman S., M. Rosbash: *Dev.Biol.* 63 197-212, 1978.
246. Persson H., L. Hennighausen, R. Taub, W. DeGrado, P. Leder: *Science* 225 687-693, 1984.
247. Persson H., P. Leder: *Science* 225 718-721, 1984.
248. Persson H., H.E. Gray, F. Godeau, S. Braunhut, A.R. Bellve: *Mol.Cell.Biol.* 6 942-949, 1986.

249. Perry R.P.: *Cell* 33 647-649, 1983.
250. Pfeifer-Ohlsson S., A.S. Goustin, J. Rydnert, T. Wahlstrom, L. Bjersing, D. Stehelin, R. Ohlsson: *Cell* 38 585-596, 1986.
251. Pfeifer-Ohlsson S., J. Rydnert, A.S. Goustin, E. Larsson, C. Betsholtz, R. Ohlsson: *PNAS* 82 5050-5055, 1984.
252. Picard B., M. Wegnez: *PNAS* 76 241-245, 1978.
253. Piechaczyk M., J-Q. Yang, J-M. Blanchard, P. Jeanteur, K.B. Marcu: *Cell* 42 589-597, 1985.
254. Pierandrei-Amaldi P., N. Campioni, E. Beccari, I. Bozzoni, F. Amaldi: *Cell* 30 163-171, 1982.
255. Powers S., T. Kataoka, O. Fasano, M. Goldfarb, J. Strathern, J. Broach, M. Wigler: *Cell* 36 607-612, 1984.
256. Proudfoot N.J., G.G. Brownlee: *Nature* 263 211-214, 1976.
257. Rabbitts P.H., J.V. Watson, A. Lamond, A. Forster, M.A. Stinson, G. Evan, W. Fischer, E. Atherton, R. Sheppard, T.H. Rabbitts: *EMBO J* 4 2009-2915, 1985.
258. Rabbitts P.H., A. Forster, M.A. Stinson, T.H. Rabbitts: *EMBO J* 4 3727-3733, 1985.
259. Rabbitts T.H., P.H. Hamlyn, R. Baer: *Nature* 306 760-765, 1983.
260. Rabbitts T.H., P. Van Straaten, P.H. Rabbitts, J. Watson: *Proc.Roy.Soc.Lond. B* 226 79-82, 1985.
261. Rabin M., C.P. Hart, A. Ferguson-Smith, W. McGinnis, M. Levine, F.H. Ruddle: *Nature* 314 175-178, 1985.
262. Raibaud O., M. Mock, M. Schwartz: *Gene* 29 231-241, 1984.
263. Raines M.A., W.G. Lewis, L.B. Crittenden, H-J Kung: *PNAS* 82 2287-2291, 1985.
264. Ralston R., J.M. Bishop: *Nature* 306 803-806, 1983.
265. Ramsay G., G.I. Evan, J.M. Bishop: *PNAS* 81 7742-7746, 1984.
266. Ran W., M. Dean, R.A. Levine, J. Campisi: *Curr.Top.Microb.Immunol.* 132 313-319, 1986.
267. Rastl E., I.B. Dawid: *Cell* 18 501-510, 1979.
268. Rebagliati M.R., D.L. Weeks, R.P. Harvey, D.A. Melton: *Cell* 42 769-777, 1985.
269. Reddy E.P., R.K. Reynolds, E. Santos, M. Barbacid: *Nature* 300

149-152, 1982.

270. Remmers E.F., J-Q. Yang, K.B. Marcu: *5* 899-904, 1986.
271. Richter J.D., L.D. Smith: *Cell* *27* 183-191, 1981.
272. Richter J.D., L.D. Smith, D.M. Anderson, E.H. Davidson: *J.Mol.Biol.* *173* 227-241, 1984.
273. Rigby P.W., M. Dieckmann, C. Rhodes, P. Berg: *J.Mol.Biol.* *113* 237-251, 1977.
274. Robbins K.C., S.G. Devare, S.A. Aaronson: *PNAS* *78* 2918-2922, 1981.
275. Robins T., K. Bister, C. Garon, T.S. Papas, P.H. Deusberg: *J. Viro* *41* 635-642, 1982.
276. Rogers S., R. Wells, M. Rechsteiner: *Science* *234* 364-368, 1986.
277. Rosbash M.: *Dev.Biol.* *87* 319-329, 1981.
278. Rosbash M., P.J. Ford: *J.Mol.Biol.* *85* 87-101, 1974.
279. Ross R., E.W. Raines, D.F. Bowen-Pope: *Cell* *46* 155-169, 1986.
280. Rous P.: *J.Am.Med.Assoc.* *56* 198, 1911.
281. Roussel, S. Saule, C. Lagrou, C. Rommers, H. Beug, T. Graf, D. Stehelin: *Nature* *281* 452-455, 1979.
282. Rowely J.D.: *Ann.Rev.Genet.* *14* 17-39, 1980.
283. Rowely J.D.: *Nature* *301* 290-291, 1983.
284. Ruley E.H.: *Nature* *304* 602-606, 1983.
285. Ruppert C., D. Goldowitz, W. Wille: *EMBO J* *5* 1897-1901, 1986.
286. Sagata N., K. Shiokawa, K. Yamana: *Dev.Biol.* *77* 431-448, 1980.
287. Sanger F., S. Nicklen, A.R. Coulson: *PNAS* *74* 5463-5467, 1977.
288. Sanger F., A.R. Coulson; *FEBS Lett.* *87* 107-110, 1978.
289. Sap J., A. Munoz, K. Damn, Y. Goldberg, J. Ghysdael, A. Leutz, H. Beug, B. Vennstrom: *Nature* *324* 635-640, 1986.
290. Sargent T.D., I.B. Dawid: *Science* *222* 131-139, 1983.

291. Schlissel M.S., D.D. Brown: *Cell* 37 903-913, 1984.
292. Schwab M., K. Alitalo, K. Klempnauer, H.E. Varmus, J.M. Bishop, F. Gilbert, G. Brodeur, M. Goldstein, J. Trent: *Nature* 305 245-248, 1983.
293. Schwab M., H.E. Varmus, J.M. Bishop, K-H. Greschik, S.L. Naylor, A.Y. Sakajuchi, G. Brodeur, J. Trent: *Nature* 308 288-291, 1984.
294. Schwab M., G. Ramsay, K. Alitalo, H.E. Varmus, J.M. Bishop: *Nature* 315 345-347, 1985.
295. Schwab M.: *TIG* 1 271-275, 1985
296. Scheer U.: *J.Embryol.Exp.Morphol.* 97 Supplement 223-242, 1986.
297. Sefton B.M.: *TIG* 1 306-308, 1985.
298. Segall J., T. Matsu, R.G. Roeder: *J.Biol.Chem.* 255 11986-11991, 1980.
299. Selten G., H. Cuypers, M. Zijlstra, M. Meliof, A. Berns: *EMBO J* 3 3215-3222, 1984.
300. Shaw G., R. Kamen: *Cell* 46 659-667, 1986.
301. Sheiness D.L., L. Fanshier, J.M. Bishop: *J.Viro.* 28 600-610, 1978.
302. Sheiness D.L., J.M. Bishop: *J.Viro* 31 514-521, 1979.
303. Sheiness D.L., S.H. Hughes, H.E. Varmus, E. Stubblefield, J.M. Bishop: *Virology* 105 415-424, 1980.
304. Shepherd J.C.W., W. McGinnis, A.E. Carrasco, E.M. DeRobertis, W.J. Gehring: *Nature* 310 70-71, 1984.
305. Sherr C.J., C.W. Rettenmier, R. Sacca, M.F. Roussel, A.T. Look, E.R. Stanley: *Cell* 41 665-676, 1985.
306. Shih C.K., M. Linial, M. Goodenow, M. Haywood: *PNAS* 81 4697-4701, 1984.
307. Shilo B.Z., R.A. Weinberg; *PNAS* 78 6789-6792, 1981.
308. Shiokawa K., Y. Misumi, K. Yamana: *Dev.Growth.Differ.* 23 579-587, 1981.
309. Shiokawa K., K. Tashiro, Y. Misumi, K. Yamana: *Dev.Growth.Differ.* 23 589-597, 1981.
310. Siebenlist U., L. Hennighausen, J. Battey, P. Leder: *Cell* 37 381-391, 1984.

311. Simon M.A., B. Drees, T. Kornberg, J.M. Bishop: *Cell* 42 831-840, 1985.
312. Smeland E., T. Godal, E. Ruud, K. Beiske, S. Funderud, E.A. Clark, S. Pfeifer-Ohlsson, R. Ohlsson: *PNAS* 82 6255-6259, 1985.
313. Smeland E.B., T. Godal, K. Beiske, R. Watt, S. Pfeifer-Ohlsson, R. Ohlsson: *Curr.Top.Microb.Immunol.* 132 290-296, 1986.
314. Smith A.E., D. Kalderon, B.L. Roberts, W.H. Colledge, M. Edge, P. Gillet, A. Markham, E. Paucha, W.P. Richardson: *Proc.Roy.Soc.Lond B* 226 43-58, 1985.
315. L.D. Smith, J.D. Richter, M.A. Taylor: In *Molecular Biology of Development*, 129-143, 1984. Alan R. Liss Inc.
316. Sorge L.K., B.T. Ley, T.M. Gilmer, P.F. Maness: In *Cancer Cells* 1 117-122, 1984. Cold Spring Harbour Press, New York.
317. Southern E.M.: *J.Mol.Biol.* 98 503-517, 1975.
318. Spandidos D., N.M. Wilkie: *Nature* 310 469-475, 1984.
319. Spector D.H., H.E. Varmus, J.M. Bishop: *PNAS* 75 4102-4106, 1978.
320. Srinivasan A., E.H. Reddy, S.A. Aaronson: *PNAS* 78 2077-2081, 1981.
321. Staden R.: *NAR* 8 3673-3694, 1980.
322. Stanely K.K., J.P. Luzzio: *EMBO J* 3 1429-1436, 1984.
323. Stanton L.W., P.D. Fahrlander, P.Tesser, K.B. Marcu: *Nature* 310 423-425, 1984.
324. Steele R.E.: *NAR* 13 1747-1761, 1985.
325. Stehelin D., H.E. Varmus, J.M. Bishop, P.K. Vogt: *Nature* 260 170-173 1976.
326. Stern D.F., A.B. Roberts, N.S. Roche, M.B. Sporn, R.A. Weinberg: *Mol.Cell.Biol.* 6 870-877, 1986.
327. Stewart T.A., A.R. Bellve, P. Leder: *Science* 226 707-710, 1984
328. Stewart T.A., P.K. Pattengale, P. Leder: *Cell* 38 627-637, 1984.
329. Struhl G., R.A.H. White: *Cell* 43 507-519, 1985.
330. Sugimoto Y., M. Whitman, L.C. Cantley, R.L. Erikson: *PNAS* 81 2117-2121, 1984.

331. Sullivan N., C. Green, M. Pasdar, R. Watt: *Curr.Top.Microb.Immunol.* 132 355-361, 1986.
332. Suttcliffe J.G.: *Cold.Spr.Har.Symp.Quant.Biol.* 43 77-90, 1978.
333. Tabin C.J., S.M. Bradley, C.I. Bargmann, R.A. Weinberg, A.G. Papageorge, E.M. Scolnick, R. Dhar, D.R. Lowy, E.H. Chang: *Nature* 300 143-149, 1982.
334. Taparowsky E., Y. Suard, O. Fasano, K. Shimizu, M. Wigler: *Nature* 300 762-765, 1982.
335. Tatchell K., D.T. Chaleff, D. DeFeo-Jones, E.M. Scolnick: *Nature* 309 523-527, 1984.
336. Taub R., C. Molding, J. Battey, W. Murphy, T. Vasicek, G. Lenoir, P. Leder: *Cell* 36 339-348, 1984.
337. Tautz D.A., M. Trick, G.A. Dover: *Nature* 322 652-656, 1986.
338. Taylor M.V., M. Gusse, G.I. Evan, N. Dathan, M. Mechali: *EMBO J* 5 3563-3570, 1986.
339. Temin H.M.: *J.Nat.Canc.Inst.* 46 III-VI, 1971
340. Thiebaud C., M. Fischberg: *Chromosoma* 59 253-257, 1977.
341. Thiele C.J., C.P. Reynolds, M.A. Israel: *Nature* 313 404-406, 1985.
342. Thomas P.S.: *Meth.Enz.* 100 255-267, 1983. Academic Press Inc.
343. Thompson C.B., P.B. Challoner, P.E. Neiman, M Groudine: *Nature* 314 363-366, 1985.
344. Toda T., T. Ishikawa, S. Powers, T. Katoka, D. Broek, S. Cameron, J. Broach, K. Matsumoto, M. Wigler: *Cell* 40 27-36, 1985.
345. Towbin H., T. Staehelin, J. Gordon: *PNAS* 76 4350-4354, 1979.
346. Treisman R.: *Cell* 46 567-574, 1986.
347. Turner P.C., H.R. Woodland: *NAR* 10 3769-3780, 1982.
348. Twigg A.J., D.J. Sherratt: *Nature* 283 216-218, 1980.
349. Ullrich A., L. Coussens, J.S. Hayflick, T.J. Doll, A. Gray, A.W. Tam, J. Lee, Y. Yarden, T.A. Libermann, J. Schlessinger, J. Downward, E.L.V. Mayes, N. Whittle, M.D. Waterfield, P.H. Seeburg: *Nature* 309 418-425, 1984.
350. Van Beneden R.J., D.K. Watson, T.T. Chen, J.A. Lautenberger,

- T.S. Papas: PNAS *83* 3698-3702, 1986.
351. Varmus H.E.: Ann.Rev.Genet. *18* 553-612, 1984.
352. Vennstrom B., J.M. Bishop: Cell *28* 135-143, 1982.
353. Verma I.M.: TIG *2* 245-249, 1986.
354. Wahl G.M., M. Stern, G.R. Stark: PNAS *76* 3683-3687, 1979.
355. Wahli W., I.B. Dawid, G.U. Ryffel, R. Weber: Science *212* 298-304, 1981.
356. Wahli W., G.U. Ryffel: In Oxford Surveys of Eukaryotic Genes, 2 96-120, 1985. Oxford University Press.
357. Wakelm M.J.O., S.A. Davies, M.D. Houslay, I. McKay, C.J. Marshall, A. Hall: Nature *323* 173-176, 1986. Wasserman W.J., J.D. Richter, L.D. Smith: Dev.Biol. *89* 152-158, 1982.
358. Wasserman W.J., J.G. Houle, D. Samuel: Dev.Biol. *105* 315-324, 1984.
359. Waterfield M.D., G.T. Scrace, N. Whittle, P. Stroobant, A. Johnsson, A. Wasteson, B. Westermark, C-H. Heldin, J.S. Huan, T.F. Deuel: Nature *304* 35-39, 1983.
360. Waston C.J., J.F. Jackson: In DNA Cloning *1* 79-87, 1985. Editor D. Glover, IRL Press.
361. Waston D.K., M. Psallidopoulos, K. Samuel, R. Dalla-Favera, T.S. Papas: PNAS *80* 3642-3645, 1983.
362. Watson D.K., E.P. Reddy, P.H. Deusberg, T.S. Papas: PNAS *80* 2146-2150, 1983.
363. Watt R., K. Nishikura, J. Sorrentino, A. ArRushdi, C.M. Croce, G. Rovera: PNAS *80* 6307-6311, 1983.
364. Watt R.A., A.R. Shatzman, M. Rosenberg: Mol.Cell.Biol. *5* 448-456, 1985.
365. Weeks D.L., M.R. Rebagliati, R.P. Harvey, D.A. Melton: Cold.Spr.Harb.Sym.Quant.Biol. *50* 21-29, 1985.
366. Weinberger C., C.C. Thompson, E.S. Ong, R. Lebo, D.J. Gruol, R.M. Evans: Nature *324* 641-646, 1986.
367. Weislander L.: Anal.Biochem. *98* 305-309, 1979.
368. Westin E.H., F. Wong-Staal, E.P. Gelmann, R. Dalla-Favera, T.S. Papas, J.A. Lautenberger, A. Eva, E.P. Reddy, S.R. Tronick, S.A. Aaronson, R.C. Gallo: PNAS *79* 2490-2494, 1982.
369. White R.A.H., M. Wilcon: Cell *39* 163-171, 1984.

370. Wolf D.P., J.L. Hedrick: *Dev.Biol.* 25 348-359, 1971.
371. Woodland H.R.: *Dev.Biol.* 40 90-101, 1974.
372. Woodland H.R.: *FEBS Lett.* 121 1-7, 1980.
373. Wormington W.M., D.D. Brown: *Dev.Biol.* 99 248-257, 1983.
374. Wormington W.M.: *PNAS* 83 8639-8643, 1986.
375. Yanisch-Perron C., J. Viera, J. Messing: *Gene* 33 103-119, 1985.
376. Yang J-Q., S.R. Bauer, J.F. Mushinski, K.B. Marcu: *EMBO J* 4 1441-1447, 1985.
377. Zimmerman K.A., G.D. Yancopoulos, R.G. Collum, R.K. Smith, N.E. Kohl, K.A. Denis, M.M. Nau, O.N. Witte, D. Toran-Allerand, C.E. Gee, J.D. Minna, F.W. Alt: *Nature* 319 780-783, 1986.
378. Zinn K., D. Dimaiio, T. Maniatis: *Cell* 34 865-879, 1983.

APPENDIX 1

CALCULATION OF AMOUNT OF POLY(A)⁺ RNA AS ESTIMATED BY ³H POLY(U) HYBRIDISATION

If ³H poly(U) has specific activity of S cpm/ng and

n = cpm poly(U) bound in the reaction - background

then total poly(A) per reaction = n/S ng.

Assuming the average length of a poly(A)⁺ containing RNA is 2.0kb with a poly(A) tail length of 50 nucleotides then the poly(A) content is equivalent to 1/40 of the weight of poly(A)⁺ containing RNA,

Hence ng poly(A)⁺ per reaction containing RNA = 40 x ng poly(A).

APPENDIX 2

CALCULATION OF THE AMOUNT OF cDNA PRODUCTS SYNTHESISED

For both first and second strand reactions the weight of cDNA synthesised in ng is given by the general formula:

ng of product made = (total incorporated - background cpm in the reaction) x total moles NTP (both labeled and unlabeled) x 10^9 / total cpm in the reaction x molecular weight NTP x fractional base composition of the nucleic acid. Where NTP is used as the labeled nucleotide(s) in the reaction.

Example

For first strand reaction described in chapter 2 containing 120 nmoles dCTP and 20 μCi α ^{32}P dCTP (from which the number of moles is insignificant) and the fraction of CTP in the RNA is 0.22 the amount of first strand cDNA made is

$$(5132 - 174) \times (120 \times 10^{-9}) \times (340) \times 10^9 / (194222) \times (0.22) = 4734\text{ng.}$$

Since 10 μg of poly(A)⁺ RNA was used this represents 47% first strand synthesis.

Similarly for second strand reaction described in chapter 2 containing 5 nmoles dCTP, 20 μCi α ^{32}P dCTP and 100ng of first strand cDNA the amount of second strand cDNA made is

$$(2065 - 515) \times (5 \times 10^{-9}) \times (340) \times 10^9 / (129150) \times (0.22) = 92.7\text{ng.}$$

Since 100ng of first strand cDNA was used this represents 93% second strand synthesis.