STRUCTURE AND EXPRESSION OF MAMMALIAN
ZINC FINGER GENES

by

VINCENT TREVOR CUNLiffe

A thesis submitted for the degree of Doctor of Philosophy in the
University of London

Imperial Cancer Research Fund
44 Lincoln's Inn Fields
LONDON

and

University College
Gower Street
LONDON	March 1990
ACKNOWLEDGEMENTS

I would like to thank all the friends I have made at the Imperial Cancer Research Fund over the last three years for their good will, enthusiasm, and critical support. I am especially grateful for the friendship and careful supervision of John Trowsdale. Adrian Kelly, David Wilkinson, and Sarah Baxendale were indispensable sources of optimism, encouragement, and the occasional good joke. I am indebted to Pat Miller, John Young, Ruth Lovering, Isabel Hanson, Jiannis Ragoussis, Danny Altmann, Ian Mockridge, David Sansom, Hitoshi Ikeda, Hidetoshi Inoko, Ian Campbell, Will Foulkes, Chris Lock, Gillian Hynes, Victoria Lewis, Nathan Ellis, Sarah Williams, and Peter Koopman, whose individual contributions helped to make the last three years an enjoyable and challenging experience. Peter Goodfellow, Keith Willison, Anne McLaren, and Walter Bodmer each provided invaluable advice and support throughout the course of this work, and my University College supervisor David Latchman made the experience of being an external student quite painless. The foresight and patience of my mother and father made all of this possible.
The zinc finger domain is a sequence-specific DNA binding motif that was originally identified in the Xenopus laevis transcription factor IIIA (TFIIIA), and subsequently observed in the Kruppel segmentation gene of Drosophila melanogaster and the yeast transcription factor ADRI. Such extensive phylogenetic conservation of zinc finger structures in these regulatory genes suggested that genes containing them might also serve important regulatory functions during mammalian development. In this study an oligonucleotide encoding a sequence conserved between TFIIIA, Kruppel, and ADRI was used to screen mammalian cDNA libraries with the objective of identifying zinc finger genes selectively expressed during human B-lymphocyte and mouse germ cell development.

A large number of cDNA clones were isolated from a human B-lymphocyte library. The sequence of a 1.8kb cDNA, ZFP36-1.8, revealed that its corresponding gene contained a minimum of 14 contiguous zinc fingers and an N-terminal non-finger region. At high stringency of hybridisation in Southern analysis this cDNA identified a large family of 20-30 highly related genomic DNA EcoRI fragments. Pulsed field gel analysis suggested that these fragments were clustered in the genome. The subfamily of sequences defined by ZFP36-1.8 was differentially transcribed in a wide range of cell types, although the transcription patterns in Northern analyses did partially overlap.

Zinc finger cDNA clones were also isolated from an 11.5 day post coitum (p.c.) mouse urogenital ridge library. One was derived from a 2.4kb mRNA encoding a protein with a block of 18 zinc finger domains and an N-terminal region of 79 amino acids rich in acidic residues. Such acidic domains are a characteristic feature of the activation domains of many transcription factors. The corresponding gene, Zfp-35, was shown to be a single copy gene at high stringency of hybridisation. The 2.4kb mRNA was selectively expressed in adult testis by comparison with embryonic gonads. The expression of this mRNA was analysed in a variety of adult tissues, whole testes of prepuberal adult XY mice, and germ cells isolated at different stages of development from XY testes. Combined with in situ hybridisation to mRNA in sections from adult XY testes, these studies showed that a large increase in Zfp-35 expression was restricted to spermatocytes at the pachytene stage of meiotic prophase.

Cosmid clones covering the Zfp-35 gene were characterised. The first exon of the 2.4kb transcription unit contained 53 nucleotides of the 5' untranslated region, exon 2 contained the remaining 5' untranslated region plus the first 17 codons of the open reading frame, and the third exon corresponded to the remainder of the mRNA including the acidic and 18 zinc finger domains of the open reading frame. The cognate gene could be identified in rat, human, whale, and horse genomic DNA. In situ hybridisation to mouse metaphase chromosomes revealed that Zfp-35 localised to a region of chromosome 18 encompassing bands B3 to C.
## CHAPTER ONE

### INTRODUCTION

<table>
<thead>
<tr>
<th>Section</th>
<th>Page Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 TRANSCRIPTION FACTOR IIIA - AN ARCHETYPAL SEQUENCE-SPECIFIC NUCLEIC ACID-BINDING PROTEIN</td>
<td>16</td>
</tr>
<tr>
<td>1.1A Structural analysis of the TFIIIA polypeptide and its gene</td>
<td>17</td>
</tr>
<tr>
<td>Biochemical analysis of TFIIIA suggests that the protein contains protease-resistant zinc-binding domains</td>
<td>17</td>
</tr>
<tr>
<td>The TFIIIA gene has an open reading frame containing nine tandem repeats which possess zinc co-ordinating ligands</td>
<td>18</td>
</tr>
<tr>
<td>The zinc finger domain has several conserved features</td>
<td>18</td>
</tr>
<tr>
<td>Genomic organisation of the Xenopus laevis TFIIIA gene</td>
<td>20</td>
</tr>
<tr>
<td>reveals that many of the zinc fingers reside on individual exons</td>
<td>20</td>
</tr>
<tr>
<td>1.1B The interaction of TFIIIA with DNA</td>
<td>20</td>
</tr>
<tr>
<td>Nuclease sensitivity and methylation interference studies of the Internal Control Region (ICR) and ICR-TFIIIA complex</td>
<td>20</td>
</tr>
<tr>
<td>Models of the TFIIIA-ICR complex suggest periodic interactions along the DNA in the major groove</td>
<td>21</td>
</tr>
<tr>
<td>High resolution hydroxyl radical footprinting studies of the TFIIIA-ICR complex</td>
<td>23</td>
</tr>
<tr>
<td>Mutational analysis of the ICR defines three discontinuous domains which are required for TFIIIA binding</td>
<td>25</td>
</tr>
<tr>
<td>1.1C The interaction of TFIIIA with RNA</td>
<td>25</td>
</tr>
<tr>
<td>The cytoplasmic ribonucleoprotein particle of Xenopus laevis oocytes is a stable complex of TFIIIA with 5S RNA</td>
<td>25</td>
</tr>
<tr>
<td>The mode of interaction of TFIIIA with 5S RNA may be similar to that of TFIIIA with the ICR DNA</td>
<td>26</td>
</tr>
</tbody>
</table>
CONTENTS (continued)

1.1D Mechanism of transcriptional activation by TFIIIA

1.1E The tertiary structure of the zinc finger domain

Three dimensional molecular models of the zinc finger
NMR studies of the zinc finger tertiary structure

1.2 GENES CONTROLLING PATTERN FORMATION IN DROSOPHILA MELANOGASTER

Discovery of the homeobox - an evolutionarily conserved DNA binding domain

Genes controlling segmentation in the Drosophila embryo fall into distinct phenotypic classes

Segmentation genes encode proteins with phylogenetically conserved DNA-binding motifs

The gap genes control metameric subdivision in discrete regions of the embryo

1.2A The Kruppel gene encodes a zinc finger protein

Gradients of maternal gene products control Kr gene expression along the antero-posterior axis

Kruppel protein is located in the nucleus and binds to regulatory sequences in pair-rule and homeotic genes

A point mutation in the zinc finger region of Kr produces a severe phenotype

1.2B The hunchback gene encodes a protein with two blocks of zinc finger domains

The hunchback protein interacts with multiple copies of a specific target sequence in the promoter of a pair rule gene and in its own promoter

The Kruppel protein negatively regulates hb by binding to specific target sequences in the hb promoter
1.2C The *knirps* gene is a member of the steroid hormone receptor family.

The steroid hormone receptor family of transcription factors contain zinc co-ordinating polypeptide domains.

*knirps* is regulated at the transcriptional level by the Kr protein which binds to target sequences in the *kni* promoter.

1.2D Summary of the structure - function relationships amongst the gap genes.

1.2E Drosophila zinc finger genes are components of many developmental regulatory mechanisms.

The *Serendipity* genes encode zinc finger proteins and are expressed during oogenesis and very early embryogenesis.

A blastoderm-specific zinc finger gene is transiently expressed during development of the posterior alimentary tract.

The specification of dorso-ventral polarity requires the zinc finger gene *snail*.

A zinc finger protein which binds to the *ftz* pair rule gene enhancer negatively regulates *ftz* gene expression.

The *suppressor of hairy wing* gene encodes a protein with 12 zinc finger domains which binds to regulatory sequences in the gypsy transposon.

1.3 YEAST TRANSCRIPTION FACTORS CONTAIN ZINC FINGER DOMAINS.

1.3A The *ADR1* gene

*ADR1* encodes a trans-acting factor with two zinc finger domains.

The zinc finger region of *ADR1* is essential for normal gene function.
1.3B The SWI5 gene

The mother cell-specific transcription factor SWI5 contains
three zinc fingers.

The zinc finger domain in the SWI5 protein binds to
regulatory elements in the HO gene.

1.4 VERTEBRATE ZINC FINGER GENES

Families of nuclear proteins containing the TGEKPY/F
sequence can be detected in both vertebrates and invertebrates.
Large families of mammalian zinc finger genes are
ubiquitously and differentially expressed, and map to many
regions in the genome.

1.4A Xenopus zinc finger genes are differentially expressed
during early embryogenesis.

The Xfin gene is expressed maternally and during early
zygotic stages, and encodes a protein with 37 zinc fingers.
A large family of zinc finger genes is expressed in Xenopus
maternal mRNA.

Some Xenopus zinc finger genes encode proteins with
higher order zinc finger repeats and possess similar
non-finger domains.

1.4B Mouse zinc finger genes are differentially regulated
during embryonic development.

1.4C Two mammalian zinc finger genes which are growth
factor-inducible.

The Krox-20 gene is activated by serum and encodes a
protein with a block of three zinc fingers.
Krox-24 is serum-inducible and encodes a protein distinct
from Krox-20 but with highly related zinc fingers.
Transcription factor Spl contains three tandem zinc fingers similar to those of Krox-20 and Krox-24.

Sp1 and Krox-24 recognise similar DNA sequences.

Krox-20 expression is spatio-temporally restricted in a segmental pattern during development of the nervous system.

Krox-24 is induced in synaptic networks within the adult brain via neuron-specific cell surface receptors.

1.4D Some zinc finger genes are aberrantly expressed in tumours.

The GLI gene is rearranged in several tumour types and encodes a zinc finger protein.

The GLI protein is a sequence-specific DNA-binding protein.

Retrovirus insertion into a mouse zinc finger gene causes myeloid leukaemia.

1.4E A genetically defined sex-determining region of the human Y chromosome contains a gene which encodes a zinc finger protein.

Mouse Zfy genes are expressed in germ cells.

Redefinition of the sex-determining region in humans excludes ZFY as a candidate for TDF.

Mouse Zfy genes may be involved in germ cell development.

1.4F Experimental objectives
CHAPTER TWO
MATERIALS AND METHODS

2.1 Bacterial cell culture 74
2.2 Bacterial transformation 75
2.3 Plasmid and cosmid DNA minipreparations 76
2.4 Plasmid and cosmid DNA maxipreparations 78
2.5 Restriction enzyme cleavage and agarose gel electrophoresis of DNA 80
2.6 Preparation of radioactively labelled DNA probes 81
2.7 Southern transfer of DNA 83
2.8 Hybridisation of DNA fixed on nylon membranes 83
2.9 Plating of bacteriophage cDNA libraries 84
2.10 Screening bacteriophage cDNA libraries 85
2.11 Isolation of bacteriophage recombinants 86
2.12 Characterisation of cDNA clone inserts 86
2.13 Mice, embryos, and urogenital ridge dissections 89
2.14 Isolation of total RNA from whole organs and cultured cells 89
2.15 Spermatogenic cell separation and RNA isolation 91
2.16 Northern transfer and hybridisation of RNA 91
2.17 Primer extension analysis 92
2.18 Anchor Polymerase Chain Reaction cloning of 5'-located cDNA sequences 93
2.19 DNA sequence determination 94
2.20 In situ hybridisation to mRNA in tissues 99
2.21 Isolation of genomic clones from cosmid libraries 100
2.22 Characterisation of cosmid genomic clones 100
2.23 In situ hybridisation to mouse metaphase chromosomes 102
2.24 Pulsed field gel analysis 102
CHAPTER THREE
A FAMILY OF HIGHLY RELATED HUMAN ZINC FINGER GENES

3.1 Introduction 104
3.2 A zinc finger oligonucleotide identifies a large family of sequences in mammalian genomes 106
3.3 The gene encoding ZFP36-1.8 is a member of a large subfamily of closely related human zinc finger genes 109
3.4 ZFP36-1.8 corresponds to an mRNA encoding a protein with at least fourteen zinc fingers 111
3.5 Members of the ZFP-36-related gene family have highly homologous zinc finger domains 114
3.6 The ZFP-36-related gene family has a complex transcription pattern and is differentially expressed 116
3.7 Pulsed field gel analysis suggests that the ZFP-36-related genes are clustered in the human genome 116
3.8 Discussion 118
CHAPTER FOUR

A MOUSE ZINC FINGER GENE WHICH IS TRANSIENTLY
UP-REGULATED DURING SPERMATOGENESIS

4.1 Introduction - mouse male germ cell development 122
4.2 Isolation of a cDNA clone from an 11.5 day p.c.
urogenital ridge library which corresponds to an
abundant mRNA in adult testis 124
4.3 Molecular analysis of the 2.4kb Zfp-35 transcript 126
4.4 The Zfp-35 gene encodes a polypeptide with 18
zinc fingers and an acidic N-terminus 128
4.5 Regionalised expression of Zfp-35 in adult tissues 132
4.6 Expression of Zfp-35 in testes of Sex-reversed
mutant mice 135
4.7 Time of onset of Zfp-35 up-regulation during
testis development 135
4.8 Expression of Zfp-35 in fractionated germ cells 137
4.9 In situ analysis of Zfp-35 expression in the testes
of adult XY mice 139
4.10 Zfp-35 is not up-regulated in pachytene oocytes 148
4.11 Discussion 150

Zfp-35 is a unique member of the TGEKPYE zinc finger
gene family 150
Stage-specific control of Zfp-35 expression during
spermatogenesis 153
The Zfp-35 polypeptide as a regulatory protein 154
CHAPTER FIVE
GENOMIC ANALYSIS OF THE Zfp-35 GENE

5.1 Introduction 156
5.2 Isolation of cosmids containing the Zfp-35 gene 156
5.3 Intron - Exon structure of the Zfp-35 gene 158
5.4 Definition of the site of transcription initiation and identification of 5' and 3' flanking sequences 160
5.5 The Zfp-35 gene is conserved in mammals and is a member of a large mammalian zinc finger gene family 162
5.6 Zfp-35 is located on mouse chromosome 18 164
5.7 Discussion 167
The genomic organisation of Zfp-35 and that of other members of the zinc finger gene superfamily are similar 167
The 5' flanking regions of genes selectively expressed in testis have a limited number of common features 169
Zfp-35 as a common regulatory component of spermatogenesis in placental mammals 169

CONCLUDING REMARKS

BIBLIOGRAPHY
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Amino acid sequence of TFIIIA from <em>Xenopus laevis</em> oocytes</td>
<td>19</td>
</tr>
<tr>
<td>1.2</td>
<td>Schematic folding pattern for the repeated &quot;zinc finger&quot; units</td>
<td>19</td>
</tr>
<tr>
<td>1.3</td>
<td>Overall features of TFIIIA structure</td>
<td>22</td>
</tr>
<tr>
<td>1.4</td>
<td>Models for the interaction of the zinc fingers of TFIIIA with DNA</td>
<td>22</td>
</tr>
<tr>
<td>1.5</td>
<td>Alignment of TFIIIA along the ICR of a 5S RNA gene</td>
<td>24</td>
</tr>
<tr>
<td>1.6</td>
<td>Interaction of TFIIIA with 5S RNA</td>
<td>24</td>
</tr>
<tr>
<td>1.7</td>
<td>Three-dimensional models of the zinc finger domain</td>
<td>29</td>
</tr>
<tr>
<td>1.8</td>
<td>The homeobox motif</td>
<td>32</td>
</tr>
<tr>
<td>1.9</td>
<td>Mode of interaction of Cro protein with operator DNA</td>
<td>32</td>
</tr>
<tr>
<td>1.10</td>
<td><em>Drosophila</em> segmentation genes</td>
<td>34</td>
</tr>
<tr>
<td>1.11</td>
<td>Common features of TFIIIA and Kruppel</td>
<td>37</td>
</tr>
<tr>
<td>1.12</td>
<td>The <em>Kr</em>²⁹ mutant contains a single point mutation in the <em>Kr</em> gene</td>
<td>40</td>
</tr>
<tr>
<td>1.13</td>
<td>Common features of Kruppel and hunchback</td>
<td>40</td>
</tr>
<tr>
<td>1.14</td>
<td>The TGEKPYE sequence is conserved in a broad spectrum of eukaryotes</td>
<td>52</td>
</tr>
<tr>
<td>1.15</td>
<td>The <em>Krox-20</em> and <em>Krox-24</em> genes</td>
<td>62</td>
</tr>
<tr>
<td>1.16</td>
<td><em>ZFY, Zfy</em>, and <em>ZFX</em> genes</td>
<td>69</td>
</tr>
<tr>
<td>3.1</td>
<td>Development of the haemopoietic system</td>
<td>105</td>
</tr>
<tr>
<td>3.2</td>
<td>Southern analysis of genomic DNA with the finger 1 probe</td>
<td>107</td>
</tr>
<tr>
<td>3.3</td>
<td>Isolation of finger 1-positive B-LCL cDNA clones</td>
<td>108</td>
</tr>
<tr>
<td>3.4</td>
<td>Southern analysis of genomic DNA with the ZFP36-1.8 probe</td>
<td>110</td>
</tr>
<tr>
<td>3.5</td>
<td>Nucleotide sequence of cDNA clone ZFP36-1.8</td>
<td>112</td>
</tr>
<tr>
<td>3.6</td>
<td>Alignment of the ZFP36-1.8 zinc finger domains</td>
<td>113</td>
</tr>
<tr>
<td>3.7</td>
<td>Southern analysis of the ZFP36-1.8-related gene family</td>
<td>115</td>
</tr>
</tbody>
</table>

with restriction fragments of the ZFP36-1.8 cDNA
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.8</td>
<td>Northern analysis of mRNA from the ZFP36-1.8-related gene family</td>
</tr>
<tr>
<td>3.9</td>
<td>Pulsed field gel analysis of the ZFP36-1.8-related gene family</td>
</tr>
<tr>
<td>4.1</td>
<td>Cell types in the mouse testis</td>
</tr>
<tr>
<td>4.2</td>
<td>Analysis of RNA from mouse embryonic gonads and adult testis for Zfp-35 expression</td>
</tr>
<tr>
<td>4.3</td>
<td>Structure of Zfp-35 cDNA clones</td>
</tr>
<tr>
<td>4.4</td>
<td>Nucleotide sequence of Zfp-35 cDNA</td>
</tr>
<tr>
<td>4.5</td>
<td>Predictions of secondary structure of the non-finger region of Zfp-35</td>
</tr>
<tr>
<td>4.6</td>
<td>Alignment of Zfp-35 zinc finger domains</td>
</tr>
<tr>
<td>4.7</td>
<td>Northern blot analysis of mouse adult organ RNA for Zfp-35 expression</td>
</tr>
<tr>
<td>4.8</td>
<td>Analysis of mouse adult brain and colon RNA for Zfp-35 expression</td>
</tr>
<tr>
<td>4.9</td>
<td>Analysis of testis RNA from Sex-reversed mice for Zfp-35 expression</td>
</tr>
<tr>
<td>4.10</td>
<td>Analysis of testis RNA from prepuberal mice for Zfp-35 expression</td>
</tr>
<tr>
<td>4.11</td>
<td>Spermatogenic cell separation by velocity sedimentation</td>
</tr>
<tr>
<td>4.12</td>
<td>Analysis of gene expression in fractionated adult mouse testis cells</td>
</tr>
<tr>
<td>4.13</td>
<td>Localisation of Zfp-35 transcripts in adult testis by in situ hybridisation</td>
</tr>
<tr>
<td>4.14</td>
<td>Expression of Zfp-35 mRNA in mouse foetal testes and ovaries</td>
</tr>
<tr>
<td>4.15</td>
<td>Sequence similarity of the non-finger region in Zfp-35 to a region of the glucocorticoid receptor</td>
</tr>
<tr>
<td>Section</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>5.1</td>
<td>Genomic organisation of the mouse Zfp-35 gene</td>
</tr>
<tr>
<td>5.2</td>
<td>Nucleotide sequence of the Zfp-35 gene including 5' and 3' flanking regions</td>
</tr>
<tr>
<td>5.3</td>
<td>Definition of the site of transcription initiation by primer extension</td>
</tr>
<tr>
<td>5.4</td>
<td>Conservation of Zfp-35 and related sequences in eukaryotes</td>
</tr>
<tr>
<td>5.5</td>
<td>Chromosomal mapping of the Zfp-35 gene</td>
</tr>
<tr>
<td>5.6</td>
<td>Histogram of grain distribution along chromosome 18 following in situ hybridisation with cosmid Zfp35-1</td>
</tr>
<tr>
<td>6.1</td>
<td>Sequences of synthetic peptides used to raise antibodies to the Zfp-35 protein</td>
</tr>
</tbody>
</table>
INTRODUCTION

One of the challenges of modern biology is to gain an understanding of the mechanisms of animal development, and in particular to identify regulatory genes controlling cellular differentiation. Such genes should themselves be differentially expressed, and possess structural characteristics which permit the proteins they encode to execute regulatory decisions. A number of evolutionarily conserved structural motifs have been recognised within the DNA binding regions of a variety of eukaryotic and prokaryotic transcription factors. One of these well conserved nucleic acid-binding domains is the "zinc finger" motif, a structure that was first recognised in the *Xenopus laevis* Transcription Factor IIIA, and found more recently in the coding sequences of what is becoming an increasingly large superfamily of yeast, *Drosophila, Xenopus*, and mammalian genes.

1.1 TRANSCRIPTION FACTOR IIIA - AN ARCHETYPAL SEQUENCE-SPECIFIC NUCLEIC ACID-BINDING PROTEIN

The genes encoding 5S ribosomal RNA (rRNA) in *Xenopus laevis* are regulated at the transcriptional level by a complex of proteins which associates with RNA polymerase III (Ng et al., 1979). There are two classes of 5S rRNA genes: the somatic-type which are expressed in most cell types, and the oocyte-type which are expressed only in oocytes. In the oocyte, Transcription Factors (TF) -IIIA, -IIIB, and -IIC, assemble onto the internal control region (ICR) of the 5S rRNA gene which is located between nucleotides 45-96, and positively regulate the rate of transcription initiation. TFIIIB and TFIIIC are also required for RNA polymerase III - dependent transcriptional activation of the tRNA^Met^ and adenovirus VA 1 RNA genes (Lassar et al., 1983). TFIIIA is 5S rRNA gene-specific, however, and this 40 kilodalton (K) polypeptide was purified to homogeneity and shown to make a
sequence-specific interaction with the ICR DNA (Engelke et al., 1980). Once assembled, the interaction of TFIIIA with DNA is stable (Bogenhagen et al., 1982).

1.1A Structural analysis of the TFIIIA polypeptide and its gene

Biochemical analysis of TFIIA suggests that the protein contains small, protease-resistant, zinc-binding domains.

The abundance of TFIIIA in *Xenopus* oocytes is extremely high, of the order of $10^{12}$ molecules per cell (1% - 10% total cell protein). Most of this protein exists in a 7S cytoplasmic ribonucleoprotein complex with 5S rRNA (Pelham and Brown, 1980), in a 1:1 molar ratio (Picard and Wegnez, 1979). Proteolytic cleavage of TFIIIA demonstrated that its DNA binding region spanned a 30K domain, and the adjacent 10K C-terminal domain, although not required for DNA binding, interacted with other components of the transcription apparatus and was therefore necessary for gene-specific activation (Smith et al., 1984, Vrana et al., 1987). During the course of proteolysis of the 40K protein a family of digestion intermediates were observed, some of whose molecular weights were multiples of 3K, and prolonged digestion with trypsin produced limit digestion products of 6K, 4K, and 3K (Miller et al., 1985). The finding of such periodic intermediates and a final digestion product equal in molecular weight to the unit of periodicity suggested that the protein might contain a tandemly repeated series of protease-resistant domains of size 3K. Further biochemical studies on the purified TFIIIA polypeptide revealed that each protein contained between 7-11 zinc ions. This value was similar to the number of protease resistant 3K repeats, and suggested that each domain might contain one zinc ion.
The TFIIIA gene has an open reading frame containing nine tandem repeats which possess zinc co-ordinating ligands

The high concentration of TFIIIA in oocytes facilitated its purification in sufficient quantities for the amino acid sequences of cyanogen bromide cleavage products to be determined by gas-phase microsequencing. Using degenerate oligonucleotides encoding these peptide sequences, cDNA clones corresponding to TFIIIA mRNA were isolated from an oocyte cDNA library and sequenced (Ginsberg et al., 1984). Analysis of the longest open reading frame within the TFIIIA cDNA sequence confirmed the hypotheses that had been forwarded on the basis of the proteolytic cleavage analyses, revealing the presence of nine tandemly repeated units of 30 residues in the N-terminal 30K region (Figure 1.1). Each unit contained two characteristically spaced pairs of cysteines and histidines (Brown and Argos, 1985; Miller et al., 1985), which are the most common ligands for tetrahedral co-ordination of zinc in metalloenzymes. From these investigations it was suggested that the 30K DNA binding domain contained nine modules, each of which co-ordinated an ion of zinc with the sulfhydryl and imidazole groups of its conserved cysteines and histidines (Miller et al., 1985).

The zinc finger domain has several conserved features

Extended X-ray absorption fine structure analysis confirmed that the co-ordination sphere of each zinc site in the TFIIIA molecule consisted of two cysteine and two histidine residues (Diakun et al., 1986). Conserved phenylalanine and leucine residues were found in each loop between the cysteine and histidine pairs, and it was suggested that these amino acids could form a hydrophobic core around which the loop might nucleate. Each metal co-ordinating polypeptide module was therefore predicted to fold independently into a particularly compact structure, and was christened the "zinc finger" (Figure 1.2). The zinc fingers in TFIIIA contained an abundance of basic residues, possibly for stabilising the interaction with DNA by contacting the phosphate backbone. The structures of the
Figure 1.1

Amino acid sequence of TFIIIA from *Xenopus laevis* oocytes

The alignment shows the repeated units, numbered in the box on the left. The repeat consensus is displayed at the top, numbered as for a length of 30 residues. Ringed residues indicate the conserved amino acids, asterisks (*) mark positions where an insertion sometimes occurs, dots (.) indicate variable positions, and dashes (-) correspond to gaps in the alignment. (From Klug and Rhodes, 1987)

---

Figure 1.2

Schematic folding pattern for the repeated "zinc finger" units

The co-ordination of a zinc ion by the cysteine and histidine pairs is shown. (From Miller et al., 1985)
(WGEKALPVVVYKR)

1. (WGEKALPVVVYKR)

1.1

1.2
DNA binding domains of several prokaryotic and yeast proteins contained an \( \alpha \)-helix whose dimensions allowed it to make sequence-specific DNA contacts with the bases (Pabo and Sauer, 1984). A preliminary secondary structure analysis of TFIIIA indicated that amino acids 25-32 might also form an \( \alpha \)-helix, thereby implicating them as potential DNA recognition residues (Brown and Argos, 1985).

**Genomic organisation of the *Xenopus laevis* TFIIIA gene reveals that many of the zinc fingers reside on individual exons**

The isolation of genomic clones encompassing the TFIIIA gene from *Xenopus laevis* revealed that the coding sequence comprised 9 exons, the first 6 of which each contained a single isolated zinc finger domains (Tso et al., 1986). The introns separating these exons, however, were not in precisely homologous positions, since they fell in all three phases of the reading frame, and at slightly different positions within the coding repeat. This may indicate that the introns were inserted into the coding sequence of the TFIIIA gene after the tandem finger repeats had been assembled, rather than flanking the zinc finger domains prior to their assembly into the TFIIIA gene (Rogers, 1986).

**1.1B The interaction of TFIIIA with DNA**

**Nuclease sensitivity and methylation interference studies of the ICR and ICR-TFIIIA complex**

X-ray crystallography of a 9 base pair DNA sequence from within the ICR demonstrated that its structure was A-form (McCall et al., 1986), and supported the suggestion that the whole ICR might be A-form (Rhodes and Klug, 1986). However circular dichroism analysis of the complete ICR demonstrated that the structure of the ICR in solution was B-form, when either naked or as a complex with TFIIIA. (Gottesfeld et al., 1987). The sensitivity of ICR DNA, both naked
and in a complex with the TFIIIA polypeptide, to DNase I, DNase II, micrococcal nuclease and dimethylsulphate was measured. When free of TFIIIA, the ICR had an intrinsic periodic sensitivity to DNase I and DNase II of 5.5 base pairs, which was argued to reflect the three dimensional structure of the nucleic acid. This repeat corresponded to periodic clusters of G nucleotides in the DNA, whose centres were about 5 base pairs apart (Rhodes and Klug, 1986). The DNase I and DNase II sensitivity of the ICR when complexed with TFIIIA was also periodic, showing protection maxima every 10.5 base pairs (Rhodes, 1986). Furthermore, the results of micrococcal nuclease and dimethylsulphate sensitivity assays on the naked ICR and the TFIIIA-ICR complex revealed a consistently periodic protection every 5.5 base pairs (Fairall et al., 1986).

Models of the TFIIIA-ICR complex suggest periodic interactions along the DNA in the major groove

The studies described above identified nine patches of protection in the major groove on one face of the DNA helix, and it was suggested that they could represent the position of the zinc fingers of TFIIIA (Figure 1.3; Fairall et al., 1986). Two models were proposed to explain the data. The first suggested that the zinc finger domains made a continuous series of contacts with the base pairs along the major groove of the DNA, tracking along the ICR (Figure 1.4). The second model proposed that TFIIIA approached the double helix from one side and intercalated the nine zinc fingers in an alternating pattern within the major groove of every half turn of the helix. It was argued that the nuclease protection and methylation interference data favoured the second model, as the protected residues appeared to be clustered on one face of the double helix (Fairall et al., 1986; Klug and Rhodes, 1987). Interpretations of these data should however bear in mind that similar results might be expected if the DNA was bent upon binding TFIIIA, irrespective of the orientation of the protein on the double helix, and a recent report has indicated that this could indeed be the case (Schroth et al., 1989). Moreover, the nuclease sensitivity assays have limitations such as nonrandom sequence
Figure 1.3

Overall features of TFIIIA structure

Schematic view of its interaction with DNA and other components of the transcription machinery. (From Miller et al., 1985)

Figure 1.4

Models for the interaction of the zinc fingers of TFIIIA with DNA

In model I consecutive finger domains track along the major groove continuously, whereas in model II the protein approaches the DNA helix from one side and the fingers interdigitate alternately in front and behind, along the major groove. (From Fairall et al., 1986)
specificity and size of the enzymes, and the capacity of some nucleases to cleave only across the minor groove, which collectively impose restrictions on the amount of data that can be gathered and produce sampling bias. Nucleases also have a weak, non-specific affinity for DNA and can displace polypeptide regions with a weak but specific affinity by competition for the binding site, and this can further affect the results.

**High resolution hydroxyl radical footprinting studies of the TFIIA-ICR complex**

Hydroxyl radical footprinting gives high resolution information on the accessibility of each nucleotide pair to solvent because every nucleotide is cleaved to some extent by the hydroxyl radical as a result of its interaction with the deoxyribose-phosphate backbone. Sites of low frequency cleavage correspond to places of local shielding of the DNA from the solvent by the bound protein. Hydroxyl radical footprinting was applied to the DNA-protein complexes formed between the ICR with full length and a series of N-terminal and C-terminal deletion mutant TFIIA proteins (Vrana et al., 1987). The pattern of protection for the full length protein suggested that TFIIA wrapped around each end of the ICR in the major groove at two regions of contact each of which extended over approximately 20 base pairs, and also covered a patch of the minor groove across four nucleotide pairs on one face of the double helix at the centre of the ICR (Figure 1.5). Deletions which removed the C-terminal non-finger region did not affect the hydroxyl radical footprint, but progressive deletion of each of the 5 C-terminal zinc fingers resulted in the sequential removal of a specific part of the footprint, although sometimes the de-protected regions partially overlapped.

These data were consistent with earlier studies on the DNA-binding activity of proteolytically cleaved fragments of TFIIA which suggested that the C-terminal fingers interacted with the 5' end of the ICR, and the N-terminal fingers interacted with the 3' end of the ICR, and that there was an extended, colinear projection of the protein along the ICR (Smith et al., 1984). Deletion of the two N-terminal zinc fingers resulted in a rapid loss of overall binding affinity, which
Figure 1.5
Alignment of TFIIIA along the ICR of a 5S RNA gene
The figure shows the nine zinc finger domains (numbered), and the C-terminal region required for transcriptional activation (black). The three regions of the ICR required for TFIIIA binding (Pieler et al., 1987) A, C, and M are shaded, and their correspondence with the hydroxyl radical footprints are indicated. (From Vrana et al., 1988)

Figure 1.6
Interaction of TFIIIA with RNA
(a) Sequence of the ICR of the Xenopus borealis 5S RNA gene
Regions essential for transcription initiation and the region of TFIIIA binding are shown.
(b) Secondary structure of the Xenopus laevis oocyte 5S RNA
The molecule is arranged with coaxial stacking of helices III, II, V, and IV such that residues 39-88 are aligned with the DNA sequence in (a).
(c) Model of the 7S particle
Views from the side (on the left) and the top (on the right) of the three-dimensional model of the 7S particle showing the proposed locations of the zinc finger domains on the oocyte 5S RNA. (From Christiansen et al., 1987)
Figure A: Somatic 5S DNA

```
5'--TCGATTCGGAGCGCAACGAGGCTCGGCCCTGGTTAGCTGGATGGAGACCCGCTG--
3'--AGACTTAGACCTCTCGTCTCCAGCCCGAGACAACTAGAACCACCCCTCCTGGGAC--
```

Protected by TFIIA
Essential for initiation of transcription

Figure B: Oocyte 5S RNA

```
UCAUCUCAGAAGCGUACGGAGUCGGCCUCGGUUGAUAGGCUCCGAGA
```

Protected by TFIIA
Essential for initiation of transcription
Mutational analysis of the ICR defines three discontinuous domains which are required for TFIIA binding

The hydroxyl radical footprinting studies confirmed and extended data from an earlier mutational analysis of the ICR DNA (Pieler et al., 1987). This study established the functional significance of the three regions in the ICR that were protected from hydroxyl radical cleavage, because point mutations in them drastically weakened the footprint with wild type TFIIA (Figure 1.5). One of these three regions also harboured the positions of sequence differences between the ICRs of the somatic-type and oocyte-type 5S genes, and suggested that these differences might result in TFIIA having a differential affinity for the two ICRs, thereby accounting for the differential rates of transcription from the two types of gene.

1.1C The interaction of TFIIA with RNA

The cytoplasmic 7S ribonucleoprotein particle in *Xenopus laevis* oocytes is a stable complex of TFIIA with 5S RNA

After TFIIA-dependent transcription of the 5S rRNA genes, previtellogenic oocytes contain more than a hundred-fold molar excess of 5S rRNA relative to 18S and 28S rRNA (Ford, 1971). About half of the 5S rRNA in these cells is associated in a 1:1 stoichiometry with TFIIA in the stable cytoplasmic 7S particle (Picard and Wegnez, 1979). Whether the dual specificity of TFIIA for both the ICR DNA and the 5S rRNA involves a common mode of TFIIA binding or two distinct mechanisms is of particular interest because DNA-protein interactions generally involve nucleotide sequence recognition in the major groove of the double helix, whereas previous studies of RNA-binding proteins have suggested that some
recognise their targets by virtue of irregular secondary and tertiary structural features in the RNA, such as bulges and loops (Noller, 1984; Romaniuk et al., 1987, and references therein).

The mode of interaction of TFIIIA with 5S RNA may be similar to that of TFIIIA with the ICR DNA

Ribonuclease (RNase) sensitivity and chemical modification analyses of the 5S rRNA, when naked in solution and when complexed in the 7S particle demonstrated that the structure of the RNA was the same in both situations, and moreover, comprised three imperfect double helical segments containing small bulges (Christiansen et al., 1987). The largest two double helical regions collectively contained the same nucleotide sequences as the ICR of the 5S rRNA gene, and this suggested that the three double helical segments were most likely arranged in a T-like structure in which the longest two double helical structures were stacked co-axially (Figure 1.6b). This nucleotide protection data indicated that TFIIIA interacted in the major grooves of three clustered sets of coaxial nucleotide pairs, which corresponded well with the three regions of protection on the ICR identified by hydroxyl radical footprinting (Vrana et al., 1987), and the binding regions of the ICR defined by point mutational analysis (Pieler et al., 1985). Thus it appeared that the conformation of 5S rRNA had similar features to that of the ICR of the 5S gene, i.e. double helical segments, suggesting that the zinc fingers of TFIIIA made similar sets of sequence-specific interactions with DNA and RNA (Figure 1.6c). Moreover, deletion of the unpaired, bulged, nucleotides from the 5S RNA molecule did not alter its affinity of TFIIIA, indicating that these nucleotides were unlikely to contribute significantly to the recognition process (Baudin and Romaniuk, 1989).

Neutron-scattering and hydrodynamic experiments confirmed that the overall shape of the naked 5S rRNA in solution was cylindrical (diameter 34A, length 122A), and moreover, indicated that the 7S particle itself was a cylinder of diameter 59A and length 140A. Such values suggested that the TFIIIA molecule was
wrapped around the 5S rRNA, accounting for the increased diameter of the cylinder, and protruded over one or both ends of the RNA, giving the 7S particle a greater length.

1.1D The mechanism of transcriptional activation by TFIIIA

Gene-specific transcriptional activation of the 5S rRNA requires transcription factors IIIA, IIIB, and IIIC as well as RNA polymerase III. A prerequisite for transcriptional initiation is the interaction of TFIIIA with the ICR, which involves the domain of 9 zinc fingers. The C-terminal 10K non-finger region is then required for further steps of transcription complex assembly, one of which is the formation of a stable pre-initiation complex between TFIIIA and TFIIIC (Lassar et al., 1983). Whether the 10K domain mediates this particular event directly is currently unknown. The precise sequence within this domain that is required for transcription activation was defined using the C-terminal deletion series described earlier (Vrana et al., 1987). A TFIIIA-depleted oocyte nuclear extract was supplemented with each of the TFIIIA deletion mutant series, and the rate of transcription was measured relative to the rate when supplemented with full length TFIIIA. Deletion of the first 31 C-terminal residues produced a product with full length transcriptional activity, but deletion of the next 19 amino acids reduced the rate of transcription to less than 10% of that with full length, in the absence of any noticeable inhibition of the DNA binding activity. This result implied that residues within the 19 amino acid sequence provided the critical structure that allowed TFIIIA to interact with the other components of the transcription machinery.

1.1E The tertiary structure of the zinc finger domain

Although TFIIIA is the prototypic zinc finger protein, its precise tertiary structure, including that of its zinc finger domains is not yet understood. A large
body of sequence information has accumulated for many other zinc finger genes in *Drosophila*, yeast, *Xenopus*, and mammals, which facilitated the construction of the first three-dimensional models for the zinc finger domain. Recent NMR experiments to determine the solution structure of single zinc finger peptides from the coding sequences of other zinc finger genes have confirmed that the molecular models derived by sequence comparisons were essentially valid, and a three-dimensional image of the canonical zinc finger structure has clearly emerged.

**Three-dimensional molecular models of the zinc finger**

Reasonable models for the likely tertiary structure of the ~30 amino acid zinc finger module were constructed from a large battery of zinc finger sequences, using refined secondary structure prediction methods and comparisons with the sequences of zinc co-ordinating domains in other proteins whose structures had been solved (Berg, 1988; Gibson et al., 1988). Both reported attempts at structure prediction produced remarkably similar models of the zinc finger module. The structure consisted of a two-stranded \(\beta\)-hairpin which stabilised a more C-terminally located \(\alpha\)-helix by a combination of hydrophobic interactions and tetrahedral co-ordination of the zinc ion (Figure 1.7). The two cysteine residues were located either side of the turn separating the two \(\beta\)-strands, and the two histidines lay towards the bottom of the \(\alpha\)-helix. At the top of the second \(\beta\)-strand was the conserved phenylalanine, and at the top of the \(\alpha\)-helix was the conserved leucine residue. These two residues were predicted to make hydrophobic interactions with each other. One face of the \(\alpha\)-helix contained many of the positions of variable amino acid identity, which generally tended to be polar or basic. It was suggested that this helix was at least part of the DNA sequence recognition domain and that the variable residues made the sequence-specific interactions with the bases. Using this information, the interaction of the zinc finger module with DNA was rigorously modelled. The results suggested that each finger might contact approximately three base pairs of DNA (Gibson et al., 1988), in contrast to earlier hypotheses (Fairall et al., 1986).
Figure 1.7 Three dimensional model of the zinc finger domain

(a) Schematic drawing of the zinc finger structure

(b) Interactions of tandemly repeated zinc fingers in the major groove
Model for the interaction between a protein containing tandem zinc finger repeats and the DNA, based on the structure of the individual domains. The zinc fingers track along the major groove, in a similar way to model I of Figure 1.4, although the number of zinc fingers per turn of the double helix is lower than that suggested by Gibson et al., 1988. (From Berg, 1988)
NMR studies of the zinc finger tertiary structure

The structure in solution of a synthetic peptide encoded by the first zinc finger region of a yeast gene *ADR1* was determined in the presence and absence of zinc. Nuclear magnetic resonance and circular dichroism spectroscopy established that the formation of a compact structure containing a 10-residue \(\alpha\)-helix required the presence of zinc (Parraga et al., 1988). The data suggested, moreover, that the \(\alpha\)-helix encompassed the region from the conserved leucine to the second histidine, confirming the original prediction of Brown and Argos (Brown and Argos, 1985).

The three dimensional structure of a synthetic peptide encoding one of the zinc finger domains from the *Xfin* sequence was solved by nuclear magnetic resonance (Lee et al., 1989). These results extended the conclusions drawn from the *ADR1* analysis (Parraga et al., 1988), such that the \(\alpha\)-helix was longer and the antiparallel \(\beta\)-sheets could be resolved, and confirmed that the structure predicted by molecular modelling was essentially correct (Berg, 1988; Gibson et al., 1988).

### 1.2 GENES CONTROLLING EMBRYONIC PATTERN FORMATION

**IN DROSOPHILA MELANOGASTER**

Discovery of the homeobox - an evolutionarily conserved DNA binding domain

Systematic identification of genes which determined the body plan in animals began with the work of Lewis, who analysed mutations in the insect *Drosophila melanogaster* which converted the character of one segment into that of another, resulting in a "homeotic" transformation (Lewis, 1978). Classical genetic analysis defined two large clusters of homeotic genes, the Bithorax (BX-C) and the Antennapedia (ANT-C) complexes. Extensive molecular analysis of homeotic loci demonstrated that they consist of clusters of related genes encoding proteins with a highly related protein sequence motif, christened the "homeobox" (Gehring, 1987). The homeobox is a sequence-specific DNA-binding domain of 60 amino acids that
permits the homeotic gene products to interact with target genes and thereby regulate their transcription. It consists of two polypeptide helices separated by a turn (Qian et al., 1989), and in this respect has a remarkable similarity to the DNA binding domains of the yeast mating type transcription factors MATa and MATα, as well as those of several prokaryotic DNA binding proteins (Shepherd et al., 1984; Laughon and Scott, 1984; Pabo and Sauer, 1984; Figures 8 and 9). The discovery that homeotic genes encoded DNA-binding proteins was a seminal observation which substantiated prior hypotheses that pattern formation involved levels of transcriptional control, and raised the possibility that similar molecules might control many processes of development and differentiation.

**Genes controlling segmentation in the Drosophila embryo fall into distinct phenotypic classes**

One of the first sets of spatial coordinates to become established in the Drosophila embryo is the antero-posterior axis, and mutational analysis revealed a number of maternal effect genes responsible for pre-setting these coordinates in the egg (Nusslein-Volhard et al., 1987). After fertilisation, the developing embryo is subdivided into segments along the antero-posterior axis. A systematic screen for "segmentation genes" responsible for metameric subdivision of the Drosophila embryo identified fifteen loci which when mutated altered the segmental pattern of the larva (Nusslein-Volhard and Wieschaus, 1980). The phenotypes of the mutant embryos fell into three classes which reflect three discrete levels of spatial organisation: pattern duplication in each segment (segment polarity mutants), pattern deletion in alternating segments (pair rule mutants), and deletion of a group of adjacent segments (gap mutants) (Figure 1.10). Once a segment has been specified in a particular region of the embryo, its subsequent developmental fate comes under the control of homeotic genes in the Bithorax and Antennapedia complexes.
Figure 1.8

The homeobox motif

Alignment of the homeobox motifs of the *ftz*, *Antp*, and *Ubx* genes of *Drosophila* with the helix-turn-helix DNA binding domains of yeast and prokaryotic DNA binding proteins. The conserved residues are underlined, and residues important for the close three-dimensional alignment of the two helices are boxed. Numbering down the sides indicates the residue position in the polypeptide, and numbering along the top is arbitrary. (From Laughon and Scott, 1984)

Figure 1.9

Mode of interaction of Cro protein with operator DNA

The bacteriophage Cro protein DNA-binding domain has many similarities to the homeobox motif. α-carbons from helices 2 and 3 are indicated in a complex with the operator DNA. (From Pabo and Sauer, 1984)
DNA-binding domains

1.8

1.9
Segmentation genes encode proteins with phylogenetically conserved DNA binding motifs

The identification of the helix-turn-helix motif in transcription factors from a wide variety of organisms indicated that this structure was likely to have a conserved function. A comprehensive analysis of segmentation in the developing *Drosophila* embryo has continued to expand the hierarchy of interacting transcription factor genes which control three dimensional organisation.

As well as the homeotic genes, some segmentation genes also encoded transcription factors with homeobox-type DNA-binding domains (Gehring, 1987). A coherent picture of how components of the pattern forming mechanisms in *Drosophila* interact hierarchically, by cross-regulation at the transcriptional level, is now beginning to emerge. Furthermore, *Drosophila* homeobox genes have proved useful as starting points from which to identify, by cross-hybridisation, hitherto inaccessible genes involved in the regulation of vertebrate development (Graham et al., 1989, and references therein).

The zinc finger domain of TFIIIA also emerged from the analysis of segmentation as a second evolutionarily conserved DNA binding motif because it was present in the products of the gap genes.

**The gap genes control metameric subdivision in discrete regions of the embryo**

The body plan of *Drosophila melanogaster* is metamERICALLY subdivided into the head, three thoracic, and eight abdominal segments. The genetic screen of Nusslein-Volhard and Wieschaus identified three genes, recessive alleles of which produced deletions of a continuous stretch of segments in homozygous *Drosophila* embryos. Mutations in the *Kruppel (Kr)* gene produced embryos which lacked the thoracic and anterior abdominal segments, and had an inverted duplication of the most anterior remaining abdominal segment. Mutations at the *hunchback (hb)* locus resulted in deletion of the lower two thoracic segments (meso- and meta-thoracic), and mutations in the *knirps gene (kni)* gave rise to embryos deleted for six out of
Figure 1.10

*Drosophila* segmentation genes

Diagram of the genetic components of the regulatory hierarchy controlling segmentation in *Drosophila melanogaster*, indicating the classes of phenotypes associated with mutations in these genes
Maternal Effect

Gap

Pair Rule

Segment Polarity

Kruppel
hb
kni

Paired
ftz
h

Gooseberry
en
wg

1.10
eight abdominal segments. These phenotypes implied that each gene was required at a particular position along the antero-posterior axis for directing the subdivision of the continuous body region in which they were expressed, and allowed two predictions to be made. First, that the domain of expression of each gap gene would be under the control of genes which specified position along the antero-posterior axis, namely maternal effect genes. Second, that the proteins encoded by the gap genes would regulate the pair-rule and segment polarity genes genes responsible for the segmental subdivision of the regions in which the gap genes exerted their effect.

1.2A The *Kruppell* gene encodes a zinc finger protein

A total of 30 alleles of the *Kruppell* gene were identified and classified into an allelic series (Wieschaus et al., 1984; Preiss et al., 1985). The strongest alleles produced embryos with the severe deletion of thoracic and abdominal segments described above, and serially weaker alleles produced embryos with progressively more anterior segments. This indicated that the thoracic region was more sensitive to lack of wild type *Kr* activity than the abdominal region, and that *Kr* was differentially active in the segments it affected. The *Kr* gene was cloned by microdissection of a polytene chromosome band spanning the *Kr* locus, and then mapping chromosomal rearrangements within the *Kr* region with microclones thus derived (Preiss et al., 1985). The biologically active sequences were localised by injecting homozygous *Kr* embryos with microcloned DNA fragments, and assaying for the amelioration of a strong mutant phenotype by its modification to a weak or intermediate phenotype. Antisense RNA was then produced from sequences with such biological activity and shown to induce a *Kr* phenocopy in wild type embryos (Rosenberg et al., 1986). Hybridisation of these sequences to Northern blots of RNA from different stages of *Drosophila* development revealed that the gene was specifically expressed at the blastoderm to gastrula stage of embryogenesis, as predicted from the mutant phenotypes.

**cDNA clones corresponding to two *Kr* transcripts were cloned and sequenced. The N-terminal region of the open reading frame was 30% rich in**
proline, alanine, and serine, potentially forming a globular domain. However the next 112 amino acids formed four complete copies of an internal repeat which bore remarkable similarity to the zinc finger domains of TFIIIA (Figure 1.11). Each repeat had two conserved pairs of cysteines and histidines, spatially arranged such that they could coordinate a zinc ion, and the loop of polypeptide between them contained a phenylalanine and a leucine residue in the same positions as they were found in the TFIIIA domains. This homology was clearly indicative of a DNA binding function for the Kr gene product, suggesting that it could function as a transcription factor, a prediction entirely consistent with its genetically defined role as a developmental regulatory gene. Between the last histidine of one zinc finger and the first cysteine of the next was a short stretch of conserved amino acid sequence, TGEKPY/F. Sequences very similar to this hexapeptide were also present in analogous positions in the two most N-terminal zinc finger regions of TFIIIA, which made the strongest contributions to the specific interaction with the ICR DNA sequence. The amino acids at the other positions within each repeat unit were quite variable from repeat to repeat, as well as from gene to gene, suggesting a possible role in sequence-specific binding.

Gradients of maternal gene products control Kr gene expression along the antero-posterior axis

In situ hybridisation to mRNA in developing embryos revealed that the Kr gene was first expressed in the syncytial blastoderm in a broad, centrally located band. The spatial restriction in the pattern of Kr expression is known to be imposed primarily by maternal effect genes which act during oogenesis to define the antero-posterior axis of the prospective embryo (Gaul and Jackle, 1987). In embryos with homozygous mutations in the anteriorly expressed maternal effect gene bicoid, the expression domain of Kr spreads more anteriorly. In embryos homozygous for mutations in oskar, a gene which specifies the posterior region of the embryo, the expression domain of Kr spreads more posteriorly. These observations demonstrated that maternal effect genes repressed the Kr gene at each
Figure 1.11

Common features of TFIIIA and Kruppel

Alignment of the zinc finger domains in TFIIIA and the open reading frame of the Kruppel gene to show the conserved residues. Dashes (-) indicate gaps in the alignment.
footprinting assays with recombinant \textit{bicoid} protein revealed that it made sequence-specific interactions with regulatory elements in the \textit{Kr} gene (Driever et al., 1988).

\textbf{Kruppel protein is located in the nucleus and binds to regulatory sequences in pair-rule and homeotic genes}

In situ hybridisation experiments initially revealed that the expression domains of many genes in the pair rule class were altered in homozygous null \textit{Kr} embryos, suggesting that the \textit{Kr} gene product would influence their spatio-temporal pattern of transcription in normal embryos (Ingham et al., 1986, Frasch and Levine, 1987). The expression of \textit{Antp} and \textit{Ubx} homeotic selector genes is also dependent on the normal domain of \textit{Kr} expression (Ingham et al. 1986; Akam, 1989). Transfection of the \textit{Kr} gene into tissue culture cells indicated that it encoded a nuclear protein (Ollo and Maniatis, 1988), and in vitro experiments with \textit{Kr} protein demonstrated that this polypeptide interacted directly with regulatory sequences in the promoters of pair rule genes (Stanojevic et al., 1989). \textit{Even-skipped (eve)} is a pair-rule gene expressed in a characteristic pattern of stripes during segmentation, and encodes a homeodomain-containing protein. Promoter fusion studies showed that the expression of each \textit{eve} stripe was regulated by a discrete \textit{eve}-promoter element, and some of these required \textit{Kr} gene activity for their effect (Harding et al., 1989; Goto et al., 1989). A DNase I footprinting assay was employed to systematically search for sites in \textit{Kr}-dependent elements of the \textit{eve}-promoter that could be bound by recombinant \textit{Kr} protein. The element controlling the stripe number 2 (stripe-2) contained six specific 10 base pair sites of \textit{Kr}-protein interaction, which were closely related to the consensus sequence AACGGGTTAA. The size of this binding site is consistent with each of the four zinc fingers binding to 2-3 basepairs of target DNA as was previously suggested (Gibson et al., 1988).
A point mutation in the zinc finger region of Kr produces a severe phenotype

One of the strongest mutant alleles at the Kr locus, Kr9, completely deleted thoracic and anterior abdominal segments. Analysis of the Kr9 sequence revealed a single nucleotide transversion from T to A which converted the second cysteine codon of the second zinc finger domain to a serine codon (Redemann et al., 1988; Figure 1.12). Although a relatively conservative amino acid substitution, the loss of a sulfhydryl group destroyed the normal biological activity of the protein. A reasonable explanation for the phenotypic effect of this mutation was that the tertiary structure of this finger domain was changed because of ineffective co-ordination of its zinc ion, and this prevented the sequence-specific interaction with cognate targets in the eve promoter and other segmentation genes regulated by Kr.

1.2B The hunchback gene encodes a protein with two blocks of zinc fingers

Drosophila embryos homozygous for strong mutant alleles of hunchback are deleted for thoracic segments, have fused posterior abdominal segments, and an abnormal telson. There is a phenotypic series of weaker alleles which show less severe deletions in the pattern of segmentation, indicating a requirement for a gradient of hb activity in the developing embryo. Some alleles additionally induce homeotic transformations of head into abdominal segments (Lehmann and Nusslein-Volhard, 1987). The hb gene was isolated from the wild type genome by a chromosomal walk into the cytogenetically defined region harbouring the locus (Tautz et al., 1987). DNA fragments thus derived were injected into homozygous hb− embryos to test for phenotypic rescue and used on Northern blots to detect blastoderm stage-specific transcripts. DNA satisfying both criteria was identified, and inferred to contain the hb gene. In fact, two transcripts could be detected with these probes, a maternal mRNA deposited in the egg and a zygotic transcript.
Figure 1.12

The $Kr^9$ mutant contains a single point mutation in the *Kruppel* gene. A nucleotide transversion results in the substitution of a serine codon for a conserved cysteine codon in the second zinc finger domain of the *Kruppel* open reading frame, thereby inactivating the gene. (From Redemann et al., 1988)

Figure 1.13

Common features of Kruppel and hunchback

Alignment of the zinc finger domains in the *Kruppel* and *hunchback* open reading frames, in which the cysteine and histidine pairs are boxed, and other conserved residues are connected by vertical lines. (From Tautz et al., 1987)
was distributed homogeneously throughout the egg and was not required for embryogenesis (Hulskamp et al., 1989; Irish et al., 1989), but the zygotic transcript was localised anteriorly and was under positive control by the morphogen bicoid (Tautz, 1988), in contrast to Kr which was negatively regulated by bicoid in the anterior of the embryo (Gaul and Jackle, 1987).

The DNA sequence of hb revealed that the open reading frame contained two sets of zinc finger domains and each domain was characterised by the presence of appropriately spaced cysteines and histidines (Figure 1.13). The central portion harboured a block of four tandemly repeating zinc finger domains and the C-terminus contained a further pair of zinc fingers, immediately after which came the stop codon. The TGEKPY/F sequence conserved between TFIIIA and Kr was not present, and although the conserved leucines in the loops between the zinc-coordinating residues were replaced conservatively by phenylalanine, leucine, or tyrosine residues, the conserved phenylalanine residue, thought to interact with the conserved leucine in TFIIIA and Kr, was was substituted by a variety of non-conservative amino acids. These features therefore placed hb squarely in the structural class of zinc finger genes, although the lack of particular conserved residues indicated that it belonged to a discrete subclass of zinc finger genes, perhaps with discrete functional properties. Moreover, the possession of two sets of zinc finger domains within the single open reading frame suggested that one molecule of hb protein could interact with two well-separated regulatory sites in either the same or different genes, thus forming chromatin loops and perhaps even co-regulating two genes simultaneously.

The hunchback protein interacts with multiple copies of a specific target DNA sequence in the promoter of a pair rule gene and in its own promoter

Promoter-fusion studies demonstrated that the eve promoter was under the positive control of hb, because stripes 2 and 3 of the eve expression pattern were hb-dependent (Frasch and Levine, 1987; Harding et al., 1989; Goto et al., 1989).
that were activated by hb protein. DNase I footprinting assays with recombinant hb protein and the hb-responsive eve promoter elements demonstrated that the positive regulation of eve by hb was likely to occur by direct interaction of hb protein with promoter sequences. These assays identified three high affinity hb protein target sites in the element controlling stripe 2, the consensus sequence of which was G/C/TCCAAAAAAA (Stanojevic et al., 1989). A total of eighteen hb protein binding sites were similarly identified in the element controlling stripe 3, many of which had a relatively low affinity for the protein. Some, however, consisted of multiple tandem copies of the 10 base pair sequence, and these had a high affinity for the hb protein, suggesting that closely apposed hb polypeptides might bind to their targets co-operatively. The overall consensus of these hb protein target sequences was G/AA/CATAAAAAAA. A separate study of the DNA target sequences in the hb promoter region, defined eight regions of hb protein-mediated protection from DNase I digestion (Treisman and Desplan, 1989). These sites were found in both promoters that had been defined earlier (Tautz et al., 1987), and some overlapped with binding sites that had previously been defined for the Kr and bicoid proteins. Although autoregulation has not yet been demonstrated for any zinc finger gene, it is suggested for hb by this result, and other developmentally regulated transcription factors do exhibit such behaviour. The overlap with bicoid and Kr recognition sites might also indicate that the autoregulation could alter the controls imposed by these proteins.

The Kruppel protein negatively regulates hb by binding to specific target sequences in the hb promoter

The mutant phenotype of hb- homozygous embryos suggested that hb expression was required in the anterior of the embryo, and the bicoid-dependent gradient of expression of hb protein in the anterior confirmed this (Tautz, 1988). There is little detectable posterior hb protein and, moreover, ectopic expression of hb protein in the posterior of the embryo disrupts development. The pattern of
expression of \(hb\) mRNA is positively controlled by \(bicoid\) and negatively by \(Kr\), both of which are known to bind to discrete promoter elements in the \(hb\) promoter (Driever et al., 1989; Treisman and Desplan, 1989). The Kruppel protein interacts with two sequences TGCGGGACTTAA and AAGGGGCATTAA, located at nucleotides -359 and -676 in the \(hb\) gene proximal promoter (Tautz et al., 1987). The target sites to which Kr binds to activate the \(eve\) promoter are very similar to these two sequences (Stanojevic et al., 1989), although their function in the \(hb\) promoter seems to be the Kr-mediated repression of \(hb\) (Jackle et al., 1986), in contrast to their activating roles in \(eve\). Presumably the context of these sites, or their orientation in the promoter has consequences for the specific transcriptional role, either positive or negative, that the Kr zinc finger protein attached to it can perform. This could involve distinct sets of protein-protein interactions. Interestingly, the Kr-binding sites in the \(hb\) promoter contain an extra two nucleotides at their centre in comparison to those of the \(eve\) promoter, and this may change the precise juxtaposition of the Kr protein with either its target DNA or any adjacent proteins.

1.2C The \textit{knirps} gene is a member of the steroid hormone receptor family

The posterior region of the embryo is organised by the third gap gene, \textit{knirps} (\textit{kni}). Mutant alleles of this gene form an allelic series of progressively stronger phenotypes which gradually diminish the number of abdominal segments, the strongest of which deletes 6 of the 8 abdominal segments. The \textit{kni} gene was isolated by microcloning DNA from a region that was deleted in \textit{kni} genotypes, and candidate sequences were identified by looking for transcripts expressed in the blastoderm-gastrula stages of embryogenesis (Nauber et al., 1989). The identity of \textit{knirps} sequences was confirmed by phenotypic rescue of \textit{kni}\textsuperscript{-} alleles whereupon cloned DNA was able to ameliorate the severity of the deletion when injected into \textit{kni} homozygotes. Sequence analysis of the biologically active sequences revealed
an open reading frame containing a series of nine precisely spaced cysteine residues and this region matched with the DNA binding region of a distinct class of vertebrate transcription factors known as the steroid hormone receptor family (Evans, 1989).

**The steroid hormone receptor family of transcription factors contain zinc co-ordinating polypeptide domains**

This group of proteins includes receptors for the oestrogen, glucocorticoid, progesterone, androgen hormones as well as those for vitamin D, thyroid hormone, and the morphogen retinoic acid (Evans, 1988). These proteins contain an N-terminally located DNA binding domain and a C-terminally located ligand binding domain. The DNA binding domains consist of two tandemly repeated units. Each of these units co-ordinates an ion of zinc tetrahedrally with two pairs of invariantly spaced cysteines, although there are no other sequence similarities with the zinc finger domains of the TFIllA/Kruppel type (Freedman et al., 1988). Nevertheless, between the two zinc co-ordinating centres are ten residues with the potential to form an amphipathic α-helix for DNA binding, and mutational analysis has confirmed that this region, together with residues within the most N-terminal of the two zinc finger domains, harbours amino acids which impart specificity to the interaction with DNA (Danielsen et al., 1989; Umesono and Evans, 1989; Berg, 1989). Receptor dimerisation is essential for binding to target DNA, and the apparent role of the second zinc co-ordinating centre is to stabilise the protein-protein interactions which permit this dimerisation. Interestingly, the receptor proteins can only regulate their target genes when their ligand binding domains are occupied, and the transcriptional activity of their target genes is thus directly dependent on the concentration of hormone, vitamin D, or retinoic acid. The knirps protein may therefore represent a ligand-dependent gap gene, which functions only when complexed with an as yet unidentified molecule. These proteins thus constitute a class of transcription factors with structural and functional similarities to the TFIllA/Kruppel-type zinc finger proteins, but which are nevertheless quite distinct.
The importance of zinc finger genes as interacting components of developmental mechanisms found further emphasis in the discovery that the promoter of *kni* contained two Kr protein target sites of sequence AAAAGGGTTAA located 5 nucleotides apart (Pankratz et al., 1989). The functional significance of these sites was confirmed when a promoter fusion construct that deleted these sequences failed to show appropriate expression in germline transformation experiments. The close proximity of the two sites suggested that the Kr polypeptides might bind to them cooperatively. Although the concentration of Kr protein in the posterior region of the embryo was known to be very low, these observations indicated that it was still functionally important for normal pattern formation, which included transcriptionally activating the *kni* gene.

1.2D Summary of the structure - function relationships amongst the gap genes

1. The proteins encoded by every member of the gap class of segmentation genes contain zinc co-ordinating DNA binding domains with similar, yet distinct structural features.

2. The *Kr* gene contains four complete tandem repeats of a TFIIIA-like zinc finger domain each with a pair of cysteine and histidine residues, and conserved phenylalanine and leucine residues. A conserved amino acid sequence TGEKPY/F links one finger to the next, and this is also found twice in the zinc finger region of TFIIIA.

3. The *hb* gene contains two blocks of zinc finger domains and each domain contains pairs of appropriately spaced cysteine and histidine residues. Other sequence features of the TFIIIA prototype are less well conserved.

4. The *kni* gene encodes a protein with a steroid hormone receptor-type zinc finger
5. All gap genes respond differentially to gradients of positional information along the antero-posterior axis, the rudiments of which are laid down by the maternal effect genes.

6. The functions of $Kr$ are manifold. These include positive regulation of $kni$, pair rule, $Ubx$, and $Antp$ genes. In addition, $Kr$ negatively regulates the $hb$ gene. Many of these effects are mediated by sequence-specific binding of the $Kr$ protein with target DNA in the promoters of the regulated genes. Each zinc finger may contact 2-3 base pairs of the $Kr$ target site.

7. The known functions of $hb$ include positive control of the pair rule genes and negative control of $Ubx$. The $hb$ protein is a sequence-specific DNA binding protein, consistent with its function as a master regulator.

8. The $kni$ protein functions to determine the posterior of the embryo and its expression is dependent on $Kr^+$ activity.

1.2E Drosophila zinc finger genes are components of many developmental regulatory mechanisms

The previous sections demonstrated why *Drosophila melanogaster* is an excellent model organism in which to elucidate the molecular basis for pattern formation, and revealed how zinc finger genes involved in this process perform their regulatory roles. Zinc finger genes other than those of the gap class of segmentation genes are now known to contribute to a wide spectrum of developmental processes in *Drosophila*.

The *serendipity* genes encode zinc finger proteins and are expressed during oogenesis and very early embryogenesis

The *serendipity* (sry) $\beta$ and $\delta$ genes are located within a densely transcribed gene cluster which was originally identified by virtue of the blastoderm-specific
expression of some of its components (Vincent et al., 1985). The transcripts of sry β and sry δ, however, are predominantly transcribed during oogenesis, although up-regulation of sry mRNA does occur zygotically prior to cellularisation of the blastoderm (Vincent et al., 1988). The two genes encode proteins with zinc finger domains which are organised into a single block of six for sry β, and there is a similar block of six in sry δ as well as an isolated zinc finger at the C-terminus (Vincent et al., 1985). The sequences of the central four domains in the block of six zinc fingers in both sry β and sry δ open reading frames share 70% sequence similarity, indicating a common evolutionary origin or nucleic acid binding function.

An additional sry-like gene, sry h-1, was isolated by virtue of the conserved region (Vincent et al., 1988). This gene contains a block of eight zinc finger domains, and has a similar pattern of expression to the sry α and sry β genes, being predominantly expressed with a homogeneous distribution in the oocyte and having a second burst of expression prior to cellularisation of the blastoderm.

The sry proteins are distributed uniformly throughout the embryo, and they are associated with blastoderm nuclei after 12-13 cycles of cell division prior to cellularisation of the blastoderm (F. Payre. personal communication). This observation may indicate a regulatory function for the proteins in the control of early zygotic genes required for the cellularisation process.

A blastoderm-specific zinc finger gene is transiently expressed during development of the posterior alimentary tract

The terminus (ter) gene was isolated in a screen for blastoderm-specific genes (Baldarelli et al., 1988). The sequence of its open reading frame revealed a single C-terminally located zinc finger domain. Zygotic transcription commenced during the 10th mitotic cycle, which coincided with the onset of terminus expression throughout the embryo. The pattern became rapidly restricted to the posterior blastoderm, and was subsequently localised around the amnioproctodeal invagination prior to its development into the posterior midgut and proctodeum. By
the end of gastrulation, ter transcripts could no longer be detected. This discrete spatio-temporal pattern of transcript distribution, together with the sequence of its open reading frame, suggested that the encoded protein probably controlled development of the posterior alimentary tract.

The specification of dorso-ventral polarity requires the zinc finger gene snail

Nineteen Drosophila genes have been identified whose activity, either maternal or zygotic, is required for the normal development of structures along the dorso-ventral axis. Twelve of these act maternally, and mutant alleles of most of them convert regions of the embryo in a ventral or lateral position into more dorsal structures, suggesting that normal embryos possess a ventralising system. Snail (sna) is a zygotically acting gene and mutant alleles prevent normal dorso-ventral specification including formation of the mesoderm (Boulay et al., 1987). The gene was cloned by chromosomal walking and biologically active sequences at the locus were identified by phenocopying mutations with antisense RNA. The sequence of this DNA revealed an open reading frame of 43K which contained putative glycosylation and phosphorylation sites towards the N-terminus. The most striking feature was the presence of five tandem repeats of the TFIIIA/Kruppel-type zinc finger domain with many of the canonical features. The third and fourth zinc finger domains were connected by the sequence TGEKPFQ, although the sequence of the connecting peptides between the other zinc fingers was quite different. These data suggested that, by analogy with Kr and hb, sna was likely to be a transcription factor with sequence-specific DNA-binding activity. As a zygotic gene, its most likely function may be to sense the positional information laid down along the dorso-ventral axis by maternal genes, and to positively control the development of ventral tissues including the mesoderm, by either activating other ventralising genes or repressing dorsalising genes.
A zinc finger protein which binds to the \textit{ftz} pair rule gene enhancer negatively regulates \textit{ftz} gene expression.

The characteristic series of stripes which pair rule gene transcripts delineate in the blastoderm are controlled by the gap genes. The expression of pair rule genes such as \textit{ftz} declines during subsequent embryogenesis, and a candidate repressor of \textit{ftz} was described biochemically (Harrison and Travers, 1989). This DNA-binding protein contains a tandem pair of zinc finger domains with pairs of zinc co-ordinating cysteine and histidine amino acids. The protein recognises multiple sequences in \textit{ftz} regulatory regions controlling stripe formation, of consensus ANAGTTATCCG. The transcripts of the gene are deposited in regions of the blastoderm where \textit{ftz} transcripts are absent, and in late stage embryos they form a characteristic series of tram-tracks extending antero-posteriorly in the mesoderm and epidermis between which the \textit{ftz}-expressing nervous system develops (hence the name \textit{tramtrack, ttk}). Perhaps somewhat surprisingly, the map position of \textit{ttk} does not correspond to that of any known segmentation gene whose mutant alleles affect the structure of the larval cuticle (Nusslein-Volhard and Wieschaus, 1980), indicating that its mutant phenotype may be more subtle.

\textbf{The suppressor of hairy wing gene encodes a protein with 12 zinc finger domains which binds to regulatory sequences in the gypsy transposon}

Insertion of the gypsy retroposon into the \textit{achaete-scute} locus produces a gain of function mutation whose phenotype includes hairy wings. The effects of this mutation can be reversed by mutations in a second unlinked gene, \textit{suppressor of hairy wing, su(Hw)}. \textit{su(Hw)} was cloned by chromosomal walking and sequence analysis revealed an open reading frame which encoded a polypeptide of 109K with a block of 12 TFIIIA-type zinc finger domains (Parkhurst et al., 1988). Three of the linking sequences between the second histidine of one centrally located finger and the first cysteine of the next closely matched the \textit{Kruppel} consensus sequence TGEKPY/F. The coding sequences were organised into six exons and there was a
single non-coding 5' untranslated exon. The positions of the introns within the zinc finger region were located in non-homologous positions and occurred in all three phases of the reading frame, much like those in the gene for TFIIIA, and suggested that the 12 zinc finger domains in su(Hw) were assembled prior to the insertion of introns.

When recombinant su(Hw) protein was incubated with gypsy DNA, DNase I footprinting and gel retardation analyses showed that the protein bound to six 55 base pair units within a 367 base pair fragment in the 5' untranslated region of the gypsy transposon (Spana et al., 1988). This fragment contained 12 copies of the sequence PyPuTTGCATACCPy separated by AT-rich stretches, many of which were located within the regions of protection from DNase I. The core element was quite similar to the octamer sequence of mammalian enhancers (ATTTGCAT, Bohmann et al., 1987), and known to be a cis-acting regulatory element for gypsy because revertants of the Hw- phenotype often contained secondary insertions of other transposable elements which destroyed the integrity of this region. This implied that the function of the su(Hw) protein was to bind to the gypsy regulatory element thereby repressing its transcriptional activity. In the absence of such repression at the site of gypsy insertion within the achaete-scute locus, achaete-scute mRNA could be produced because of readthrough from the gypsy element. This in turn could lead to ectopic expression of the achaete-scute protein, and thereby cause the observed developmental abnormalities.

In situ immunolocalisation of su(Hw) protein to Drosophila polytene chromosomes revealed 100-200 specific sites of protein detection, which suggested that su(Hw) regulated a large number of cellular genes because there were only seven euchromatic copies of gypsy in the fly strain analysed.
1.3 YEAST TRANSCRIPTION FACTORS CONTAIN ZINC FINGER DOMAINS

The discovery of the zinc finger motif in proteins of *Xenopus* and *Drosophila* suggested that this polypeptide domain could be of ancient evolutionary origin and conserved in regulatory proteins of much simpler organisms. This was confirmed by sequence analysis of transcription factor genes from the budding yeast *Saccharomyces cerevisiae*.

1.3A The *ADR1* gene

*ADR1* encodes a trans-acting factor with two zinc finger modules.

The glucose-repressible alcohol dehydrogenase gene of *S. cerevisiae* ADH2, is activated by the trans-acting regulatory gene *ADR1*. The *ADR1* gene was cloned by complementation, and sequence analysis of the DNA revealed an open reading frame of 1323 amino acids encoding a protein of 151K. The N-terminal region was highly basic, containing 21% lysine plus arginine between residues 62-256 (Hartshorne et al., 1986). Two tandem zinc finger repeats were located between amino acids 98-155, and exhibited extensive similarity to the zinc finger domains of TFIIIA and the Kruppel protein (Figure 1.14). Each repeat contained appropriately spaced zinc co-ordinating pairs of histidine and cysteine residues. Conserved phenylalanine and leucine residues were located between the second cysteine and first histidine of each domain. In addition, the seven amino acids between the last histidine of the first repeat and the first cysteine of the second repeat (TNEKPYP) had considerable similarity to the corresponding regions between every Kr zinc finger, and the most N-terminal three finger regions of TFIIIA.
Figure 1.14

The TGEKPYE sequence is conserved in a broad spectrum of eukaryotes.

Alignment of the TFIIIA, Kruppel, and ADR1 zinc finger domains to show the conserved residues.
The zinc finger region of \textit{ADR1} is essential for normal gene function

C-terminal deletion mutants of the \textit{ADR1} gene revealed that its essential region was located near the N-terminus because a mutant polypeptide consisting of residues 1-302 was able to trans-activate the \textit{ADH2} gene. In contrast, a deletion derivative containing residues 1-150 which removed amino acids at the C-terminus of the second zinc finger as well as a basic region, failed to stimulate \textit{ADH2} activity (Hartshorne et al., 1986). Of nineteen independently isolated mis-sense point mutations in the \textit{ADR1} gene, seventeen were located in the zinc finger region (Blumberg et al., 1987). None were defective in nuclear targeting. Twelve of the seventeen clustered mutations changed zinc co-ordinating cysteine and histidine codons to tyrosine codons, thus confirming the critical role of these residues in \textit{ADH2} activation.

1.3B The \textit{SWI5} gene

\textit{The mother cell-specific transcription factor SWI5 contains three zinc fingers}

The \textit{HO} endonuclease initiates the mating type switch during the life cycle of \textit{S. cerevisiae} (Nasmyth, 1983). The activity of the \textit{HO} gene is controlled at the transcriptional level by the trans-acting locus \textit{SWI5}. The \textit{SWI5} gene was cloned by complementation, and the sequence of its open reading frame revealed a contiguous block of three zinc finger domains each with a pair of cysteine and a pair of histidine residues, as well as the conserved hydrophobic phenylalanine and leucine amino acids in the region between the zinc co-ordination sites (Stillman et al., 1988).
The zinc finger domain of the SWI5 protein binds to regulatory sequences in the HO gene.

The regulatory region in the HO promoter with which the SWI5 protein interacted was defined in a deletion analysis. This DNA fragment was used in DNase I footprinting assays to show that an 89 residue deletion derivative of the SWI5 protein, containing only the three zinc finger domains, bound approximately eighteen base pairs of specific DNA sequence in the presence of zinc (Nagai et al., 1988). This was the first direct demonstration that isolated zinc finger domains, other than those of TFIIIA, were able to interact sequence-specifically with the target DNA sequence whose activity they had been postulated to control. The sequence of the target site was CCAGCATGCTATAATGCT, and had no precise or general similarities to the binding sites of TFIIIA or the Drosophila gap proteins. Moreover, this site lacked the clustering of G nucleotides that had been noted earlier for TFIIIA, in the same way that the hunchback protein recognition sequences were G-cluster deficient. The size of the protected fragment suggested that if one molecule of protein bound to the protected region, then each zinc finger might interact with approximately six base pairs, consistent with the original models in Figure 1.4. Nevertheless, the data did not exclude the possibility that more than one protein molecule could bind within the protected region, and as the sequence ATGCT occurs twice within the binding site, it is tempting to speculate that two molecules of protein might bind to the DNA sequence. If this were true, then each finger might interact with 2 or 3 base pairs of DNA, in close agreement with predictions described earlier (Gibson et al., 1988).
Families of nuclear proteins containing the TGEKPYE sequence can be detected in many vertebrate and invertebrates.

The discovery of zinc fingers in the DNA binding domains of yeast transcription factors strengthened the arguments that this motif, originally identified in *Xenopus* TFIIIA, was of widespread significance for the regulation of eukaryotic gene expression, and generally functioned as a sequence-specific DNA-binding domain.

An anti-peptide antiserum raised against a synthetic peptide containing the TGEKPYE sequence identified proteins in the nuclei of a variety of eukaryotes (Schuh et al., 1986). Blastoderm and yolk nuclei of the *Drosophila* embryo stained positive with the antiserum, as did the nuclei of chicken, calf, rat, mouse and human tissue sections or cell lines. The staining in all cases was restricted to nuclei, and cytoplasm was characteristically negative. Such strong evolutionary conservation of this sequence was confirmed by subsequent low stringency hybridisation analyses of vertebrate genomes for the presence of zinc finger genes similar to the *Kruppel* gene, and two mouse genomic clones were initially identified this way (Chowdhury et al., 1987). mKr1 (now Zfp-1) was a mouse partial genomic clone encoding eight tandemly repeated zinc finger domains. Each module contained appropriately spaced cysteine, histidine, phenylalanine, and leucine residues in a similar distribution to those in the *Kruppel* gene, and between every module was a seven amino acid sequence similar or identical to TGEKPYE. A second mouse partial genomic clone, mKr2 (now Zfp-2), was similar to Zfp-1 in all these respects, but the other residues were quite distinct from those of Zfp-1, and there were 9 tandem repeats of the zinc finger motif. Probes from both sequences detected a single gene in high stringency hybridisation analyses of mouse, hamster and human genomic DNA, but at low stringency these specific signals became obscured by smears, indicating the presence of many weakly related genomic sequences.
Large families of mammalian zinc finger genes are both ubiquitous and differentially expressed, and map to many regions of the genome.

The procedure of identifying mammalian zinc finger genes by virtue of their sequence similarity to Kruppel and each other has uncovered large families of such genes, some of which are differentially transcribed, and others whose mRNAs appear to be uniformly distributed without prejudice for any cell type (Chavrier et al., 1988a, 1988b; Chowdhury et al., 1988a, 1988b). Some are members of gene families that can be detected by genomic Southern analysis, whereas others appear to represent single copy genes by the same criterion, although their sequences contain the features shared with other zinc finger genes of the same organism. Preliminary studies of the chromosomal locations of several mammalian zinc finger genes reveals that they map to many distinct locations throughout the genome (Ashworth et al., 1989a; Chavrier et al., 1989; Janssen-Timmen et al., 1989; Sukhatme et al., 1988; Joseph et al., 1989; Ruppert et al., 1988).

1.4A *Xenopus* zinc finger genes are differentially expressed during early embryogenesis

The *Xenopus* embryo is a suitable subject for experiments on early vertebrate embryogenesis because of its large size, its ability to withstand microsurgical manipulations, and its complete and relatively rapid development outside the mother. Many maternal products are deposited in discrete regions of the oocyte, and these localised determinants make essential contributions to the development of the embryo (Slack, 1983). The first 12 cell cycles are very rapid, and zygotic gene transcription cannot be detected until a point around the 4000-cell stage, known as the mid-blastula transition (MBT, Newport and Kirschner, 1982). At this time the cell cycle time becomes longer, the ratio of nuclei to amount of egg cytoplasm seems to reach a threshold value, and newly synthesised zygotic RNA appears.
The Xfin gene is expressed maternally and during early zygote stages, and encodes a protein with 37 zinc fingers.

An initial low stringency screen of a *Xenopus* genomic library with the *Drosophila* Kruppel probe identified a single positive clone, *Xfin*, and Northern analysis of polyadenylated mRNA from a series of embryonic stages demonstrated that the gene was most strongly expressed in the ovary (Ruiz i Altaba et al., 1987). Only poly(A) transcripts could be detected in the mature ovum, although poly(A) addition to *Xfin* mRNA recommenced again after the MBT through the blastula, gastrula and neurula stages. Sequence analysis of oligo-dT-primed cDNA clones from an ovary library revealed that the gene encoded a 4.4 kb mRNA with a total of 37 Kruppel/TFIIIA-type zinc finger modules. Each module contained the characteristic pairs of cysteines and histidines, the invariant phenylalanine and leucine residues, and the linking sequences were close if not identical to the TGEKPY/F consensus. The zinc fingers were organised into 6 "hands" each containing between 3 to 8 tandem repeats, which might represent independent, alternative domains for nucleic acid binding. The change in polyadenylation state of *Xfin* maternal mRNA was intriguing and may reflect a post-transcriptional control on gene activity, since polyadenylation of mRNAs in another organism was correlated with their presence in polysomes (Rosenthal et al., 1983).

A large family of zinc finger genes is expressed in *Xenopus* maternal mRNA.

Further screening of ovary and gastrula cDNA libraries at low stringency with an oligonucleotide probe encoding the TGEKPY/F sequence revealed the existence of a further 42 distinct cDNA clones encoding a total of 342 finger repeats, all of which possessed the same Kruppel/TFIIIA/Xfin-type canonical sequence motifs (Koster et al., 1987; Nietfeld et al., 1989). Many of the cDNAs were expressed as maternal mRNA and either switched off or re-expressed after the MBT.
Some *Xenopus* zinc finger genes encode proteins with higher order zinc finger repeats and possess similar non-finger domains.

Analysis of these cDNA sequences revealed that within many cDNA clones, pairs of alternate zinc finger domains were often more closely related at the amino acid sequence level than pairs of adjacent zinc fingers, such that a higher order repeat of structure (ab)$_n$ could be defined (Nietfeld et al., 1989). Another higher order repeat of structure (abcdx)$_n$ could be discerned within the open reading frames of some cDNAs, where a, b, c, and d, were four distinct finger sequences that occurred in a tandemly repeating quartet. They were separated by a single zinc finger domain x whose sequence was neither highly related to those of a, b, c, or d, nor consistently repeated. Such higher order repeats in the zinc finger domains may reflect an alternating repeat in the sequence of potential target sites. Regions of the open reading frames which did not encode zinc fingers contained other modules that were shared between cDNA clones, that were detectable in human genomic DNA, and thus represented evolutionarily conserved "finger-associated-boxes", or "FAX" domains (Knochel et al., 1989). It is conceivable that the FAX domains interact with other proteins, such as components of the transcription machinery or localised determinants in the developing oocyte.

1.4B Mouse zinc finger genes are differentially regulated during embryonic development

Characterisation of cDNA clones from the mouse Zfp-1 gene described earlier revealed that two closely related mRNAs of 2.0kb each encoded proteins with the same block of seven zinc fingers at the C-terminus (Chowdhury et al., 1989). The N-terminal non-finger regions of the two mRNAs differed in that one had an insertion of 14 extra codons, but were otherwise identical. The F9 teratocarcinoma cell line has some of the characteristics of mouse early embryonic stem cells (Hogan et al., 1981) and treatment with the known morphogen retinoic acid converts them
Zfp-1 mRNA was present in F9 cells but its steady state level decreased significantly upon administration of retinoic acid. This in vitro observation was recapitulated with an analysis of Zfp-1 mRNA in embryonic, extra-embryonic, and adult tissues. A very low level of expression was detected in all embryonic and extra-embryonic stages, except for day 12 post coitum (p.c.) embryo, where the steady state level of Zfp-1 mRNA peaked by at least ten-fold with respect to earlier and later stages. At this stage the formation of somites is complete, most organs have developed, and they contain many differentiated cell types. These facts, however, do not suggest an obvious role of Zfp-1 in embryogenesis other than a possible function in the differentiation of some cell type present at this time. A sample of adult tissues expressed approximately uniform quantities of Zfp-1 mRNA, although at much lower steady state levels than the day 12 p.c. embryo.

An initial characterisation of the Zfp-2 gene revealed that a 2.8kb mRNA encoded the C-terminally located block of 9 zinc finger domains and was preceded by a non finger region of 99 amino acids (Chowdhury et al., 1987). This transcript was selectively expressed in F9 cells prior to treatment with retinoic acid, whereafter it became undetectable. In situ hybridisation and northern analysis demonstrated that Zfp-2 mRNA was selectively expressed in neurons throughout the developing nervous system of embryos between days 10.5 and 17.5 p.c., as well as in adult brain and spinal cord (Chowdhury et al., 1988a). Many different types of neurons hybridised to the probe, suggesting that Zfp-2 might be involved in an early developmental decision for neuronal cells. Expression in glial cells, by way of contrast, was not detectable.

1.4C Two mammalian zinc finger genes are growth factor-inducible

Extracellular signals in the form of growth factors typically control the decisions to proliferate or differentiate. Cell surface growth factor receptors interact with their ligands and transmit a signal via intracellular pathways to the nucleus.
Specific genes become activated as part of the immediate early transcriptional response to growth factors and serum, many of which encode transcription factors themselves. The Serum Response Factor (SRF), and the proto-oncogenes c-fos, c-myc, and c-jun are activated during the G0/G1 transition in cultured mammalian cells (Norman et al., 1988; Ryder et al., 1988, and references therein). Zinc finger genes represented a class of regulatory genes whose existence in mammalian genomes had been previously demonstrated (Chowdhury et al., 1987), and it was conceivable that some might be serum-inducible.

The Krox-20 gene is induced in response to serum and encodes a protein with a block of three zinc fingers

A collection of 70 distinct serum-inducible cDNAs was screened at low stringency with the Drosophila Kruppel probe (Chavrier et al., 1988a), and a single positive cDNA clone was identified. The sequence of this clone revealed that it encoded a 48K polypeptide with three tandemly repeated zinc finger modules located near the C-terminus. The overall organisation of each finger motif was similar to those of the Kruppel, Zfp-1, and Zfp-2 genes, and closely related variants of the TGEKPY/F sequence linked the last histidine of one finger to the first histidine of the next. The gene from which this cDNA was derived, Krox-20 (also known as Egr-2, Joseph et al., 1988), was rapidly induced by serum in the absence of de novo protein synthesis, although the induction was transient and the mRNA disappeared within 30 minutes. This suggested that the expression of Krox-20 was contingent upon the presence of growth and/or differentiation factors in serum. Northern analysis of Krox-20 mRNA in a range of adult tissues revealed that the gene was transcribed at detectable levels in thymus, spleen, and testis. These tissues contain a large proportion of cycling stem cells and proliferating descendants, and in view of the fact that Krox-20 was induced in the G0/G1 transition, the apparent adult tissue specificity of Krox-20 may reflect the mitotic index of the constituent cell types. Alternatively, it could indicate Krox-20 is involved in pathways of differentiation that may be controlled by extracellular
Krox-24 is serum-inducible and encodes a protein distinct from Krox-20 but with highly related zinc fingers.

A second serum-inducible zinc finger cDNA, Krox-24, was isolated by low stringency hybridisation to a Krox-20 probe (Lemaire et al., 1988). The same gene has been isolated by other groups and is also known as NGFI-A, Egr-1 or zif268 (Milbrandt, 1987; Sukhatme et al., 1988; Christy et al., 1988). Krox-24 encoded a protein of 54K with three zinc finger modules that were practically identical to those of Krox-20, similarly located towards the C-terminus (Figure 1.15a). Remarkably, the remaining non-finger coding sequences were completely distinct from those of Krox-20. Although the two zinc finger genes were co-regulated in fibroblasts by serum, and most likely interacted with similar target DNA sequences, their divergent polypeptide sequences indicated that they probably made quite different sets of protein-protein interactions with other cellular components. This feature could be envisaged to permit the two proteins to function in distinct, yet partially overlapping networks of gene interactions during the serum response and thus co-ordinately regulate a multiplicity of cellular functions.

Despite their close similarity the two genes were located on distinct autosomes, and mapped to regions of known synteny in the human genome. Krox-20 mapped to both human and mouse chromosomes 10 (Chavrier et al., 1989; Joseph et al., 1988), whereas Krox-24 was located on mouse chromosome 18 and human chromosome 5 (Janssen-Timmen et al., 1989; Sukhatme et al., 1988). The promoters of both genes were serum-inducible, and contained multiple Serum Response Elements (SREs) located in their 5' flanking regions (Treisman, 1986). Antisera against the two polypeptide revealed that both proteins were localised in the nuclei of cultured cells, consistent with their putative functions as transcription factors.
Figure 1.15

The *Krox-20* and *Krox-24* genes

(a) Sequence similarities between the proteins encoded by the *Krox-20* and *Krox-24* genes

Regions of strong sequence similarity are boxed.

(b) Amino acid comparison of Kruppel, Krox-20, Krox-24, and Sp1 zinc finger domains

The canonical features are boxed, dashes (•) represent gaps in the alignment, strictly conserved amino acids are in uppercase letters, and the zinc finger region conserved between Krox-20, Krox-24, and Sp1 are in bold. (From Lemaire et al., 1988)
Transcription factor Sp1 contains three zinc finger domains similar to those of Krox-20 and Krox-24.

The mammalian transcription factor Sp1 recognises a sequence GGGCGG (GC box) that was originally discovered in multiple copies within the 21 base pair repeat elements of the SV40 early promoter, and subsequently identified as a functional cis-acting sequence in a variety of mammalian cellular genes (McKnight and Tjian, 1986). The Sp1 polypeptide was purified to homogeneity in sufficient quantities to be microsequenced, and degenerate oligonucleotides encoding Sp1 were used to clone its corresponding gene (Kadonaga et al., 1987). Surprisingly, Sp1 turned out to be a zinc finger protein with a C-terminally located block of three zinc fingers similar in sequence to those in the Krox-20 and Krox-24 polypeptide (Figure 1.15b).

Sp1 and Krox-24 recognise similar DNA sequences

Since the target site of Sp1 was known and its zinc fingers containing the DNA binding amino acids were also similar to those of Krox-20 and Krox-24, it was possible that the target sites for these two proteins might be related. The Krox-24 protein was subsequently found to interact specifically with two copies of a 9 base pair sequence within its own promoter, of consensus GCGG/TGGGCG (Christy and Nathans, 1989). This sequence is similar to the Sp1 binding site, and is located in the promoters of other serum-inducible transcription factor genes, which indicates that as for the segmentation genes of Drosophila, mammalian serum-inducible genes may cross modulate the transcriptional activity of each other. Given that the Krox-20 zinc finger region is so similar to that of Krox-24, it might be predicted that the target sequence of this protein would also be very closely related if not identical to GCGG/TGGGCG.
Krox-20 expression is spatio-temporally restricted in a segmental pattern during development of the nervous system

The formation of somites by mesoderm demonstrates that vertebrates have a segmented body plan, and this is further supported by the fact that the developing brain is composed of a series of reiterated units of bulged epithelium arranged consecutively along the antero-posterior axis (Lumsden and Keynes, 1988). Each segmental bulge, or neuromere, gives rise to neurons that innervate specific regions of the body, and therefore represent important organisational determinants of the body plan.

In situ hybridisation to developing mouse embryo sections demonstrated that the spatio-temporal pattern of Krox-20 expression in the developing hindbrain was delimited by the anatomically defined segmental boundaries in the neuroepithelium (Wilkinson et al., 1989). Krox-20 mRNA was transiently restricted to two alternate developing neuromeres, between days 7.5 and 10.5 p.c., which indicated that the gene may be part of the segmentation mechanism operating within the hindbrain. The expression did not correlate with cell division in the neuromeres, and it was stronger in the more posterior of the two, possibly reflecting a gradient of antero-posterior stimulation. By 10.5 days p.c. no expression in either neuromere could be detected. At day 14.5 p.c. a second burst of expression arose in dorso-lateral hindbrain ganglia which disappeared by day 16.5 p.c. Krox-20 was also expressed throughout the period between days 8.5 and 14.5 p.c. in small groups of neural crest-derived cells of the developing peripheral nervous system, in a non-segmental pattern. The basis for expression in these derivatives was unclear, although it could reflect their mitotic activity, by analogy with the expression in cultured fibroblasts.

It could be concluded that Krox-20 was transiently expressed in functionally distinct populations of cells within segmental structures of the developing hindbrain and non-segmented structures of its associated peripheral nervous tissues. This suggested that in the mouse Krox-20 was a component of several distinct programs of gene expression controlling both pattern formation and cellular differentiation.
Such a multiplicity of gene interactions involving Krox-20 clearly paralleled the pleiotropic roles of the gap class of Drosophila segmentation genes, although a precisely homologous relationship could not be inferred.

**Krox-24 is induced in synaptic networks within the adult brain via neuron-specific cell surface receptors**

Northern analysis of mRNA from a variety of adult mouse tissues revealed that Krox-24 was expressed in brain, superior cervical ganglia, adrenal gland, lung, heart, and thymus (Milbrandt, 1987; Lemaire et al., 1988; Sukhatme et al., 1988; Christy et al., 1989). When rat pheochromocytoma cells (PC12) were treated with nerve growth factor (NGF), Krox-24 mRNA was induced within 15 min. Expression of Krox-24 mRNA was also induced by treatment with phorbol ester or calcium ionophore, indicating that a system similar to that mediating the serum response of fibroblasts was responsible for signal transduction (Sukhatme et al., 1988). The embryonal carcinoma cell line P19S1801A1 differentiated stably into either beating cardiac muscle when treated with DMSO, or neurones and glia when treated with retinoic acid. A concomitant and stable induction of Krox-24 mRNA was observed upon differentiation to either non-mitotic cell type, which was in contrast to its growth factor-dependent expression in proliferating fibroblasts and PC12 cells, suggesting that like Krox-20, Krox-24 expression was exquisitely dependent on the state of cellular differentiation, and did not correlate solely with a proliferative response.

When rat brains were electrically stimulated with convulsants such as Metrazole or picrotoxin, northern analysis revealed that Krox-24 was induced within 15 min and peaked after 60 min (Sukhatme et al., 1988). To ascertain which cell types were responding to such seizure-inducing depolarisation, in situ hybridisation to rat brain sections was performed (Saffen et al., 1988). This analysis revealed that neuronal populations within the hippocampus, including the dentate gyrus as well as the pyriform, cingulate, and neo-cortex all expressed Krox-24 mRNA within 15 min of administration, although by 120 min the
expression had returned to an undetectable level.

Interestingly the \textit{c-fos}, \textit{c-jun}, and \textit{junB} proto-oncogenes were co-ordinately expressed with \textit{Krox-24}, suggesting that a complex transcriptional response was mounted upon this intense drug-induced electrical stimulation. None of these genes were induced in the cerebellum, brain stem, or midbrain, despite being densely packed with neurons, demonstrating the highly cell-type specific nature of this response. Electrical stimulation of perforant path neurons induced expression of \textit{Krox-24} mRNA in neurons with which they synapsed in the dentate gyrus (Cole et al., 1989). This response was blocked by antagonists of the N-methyl D-aspartate (NMDA) receptor, suggesting that NMDA receptors in the dentate gyrus were transducing the synaptic signal from the perforant path neurons into an intracellular signal which activated \textit{Krox-24} transcription. Transcriptional activation of the \textit{Krox-24} gene in the dentate gyrus was thus synapse-dependent and mediated via a neuron-specific cell surface receptor.

\textbf{1.4D Some zinc finger genes are aberrantly expressed in tumours}

The \textit{GLI} gene is rearranged in several tumour types and encodes a zinc finger protein

The previous discussion focused on the properties of two zinc finger genes involved in transcriptional control during the G0/G1 transition in cultured cells, and their co-regulation with a group of nuclear proto-oncogenes. The involvement of other zinc finger genes in mechanisms of growth control and/or malignant transformation has been suggested because of their abnormal rearrangement in tumour DNA. The \textit{GLI} gene was originally discovered by virtue of its amplification in human glioblastomas, liposarcomas, osteosarcomas, and teratocarcinomas, and was mapped to a region frequently translocated in other tumours of similar developmental origin (Kinzler et al., 1988). Sequence analysis indicated that the gene encoded a protein of 118K with a block of 5 \textit{Kruppel}/TFIII A-type zinc finger domains, each possessing the conserved cysteines, histidines, phenylalanine and
leucine residues. Moreover, the linking sequences at the C-termini of the three centrally located zinc fingers closely matched the TGEKPY/F consensus.

**The GLI protein is a sequence-specific DNA binding protein**

Recombinant GLI protein was used in the first reported experiment to identify previously unknown target sequences for a zinc finger protein by screening genomic DNA fragments with bacterially expressed GLI-lacZ fusion protein (Kinzel et al., 1989). Protein was incubated with sheared genomic DNA, then specific DNA-protein complexes were purified, and the bound DNA fragments were subcloned and sequenced. Three GLI-binding sites were identified in the human genome, two of which were single copy sequences. The third was part of a variable number tandem repeat (VNTR), located at a unique site within the genome, and the GLI binding site was represented once within each of the 63 base pair repeat units. All fragments were protected over a 23-24 base pair region, the core sequence of which was GACCACCCA. The clustering of G nucleotides on the opposite strand of this consensus sequence was once more reminiscent of the binding sites of other zinc finger proteins, such as those for TFIIIA, Sp1, Kruppel, and Krox-24, and may reflect the intrinsic preference of some zinc finger domains for G-rich sequences postulated earlier (Rhodes and Klug, 1986). The identity of the genes to which these target site are linked is currently unknown. Antisera against the recombinant protein confirmed that the GLI polypeptide was a nuclear protein, consistent with its sequence-specific genomic DNA binding activity.

**Retrovirus insertion into a mouse zinc finger gene causes myeloid leukaemia**

Insertional mutagenesis of cellular genes can be mediated by retroviral integration into the DNA. A subset of retroviral integration sites in a group of myeloid leukaemias that were induced by a murine leukaemia virus were cloned from genomic DNA, and this revealed that the integrations occurred at the same locus, termed *Evi-1* (Morishita et al., 1988). The sequence of cDNA clones derived
from the cellular gene at the *Evi-1* locus encoded a protein of 120K with two
separated blocks of *Kruppel*/TFIIIA-type zinc finger domains, and a highly
negatively charged domain rich in acidic residues similar to those in known
transcription factors (Sigler, 1988). The N-terminal zinc finger region contained 7
tandemly arranged modules, and the C-terminal block contained 3. The frequent
activation of the gene at the *Evi-1* locus by retrovirus insertion suggested that its
inappropriate expression contributed to the aetiology of myeloid leukaemias, and
the sequence of its encoded protein indicated that this might occur by altering
programs of gene activity.

### 1.4E A genetically defined sex-determining region of the human Y
chromosome contains a gene which encodes a zinc finger protein

In humans and mice the sex of an individual is under the direct control of the
sex chromosomes, such that normal females are XX and normal males are XY.
Studies with individuals of abnormal sex chromosome constitutions have revealed
that the Y chromosome contains a gene which induces the male phenotype.
Molecular genetic analysis of sex-reversed human individuals with mutant sex
chromosomes, including XX males and XY females indicated that an essential
portion of the sex-determining gene was likely to be located within a 140kb interval
on the short arm of the human Y chromosome (Page et al., 1987). Chromosome
walking was employed to isolate a gene within this region that was conserved
within a wide range of mammals and located in the genetically defined
sex-determining region of the mouse.

The sequence of this gene revealed that it encoded a protein with a block of 13
zinc finger domains with pairs of conserved cysteines and histidines in each finger
module, and one of the linking sequences closely matched the TGEKPY/F motif.
The human gene was named *ZFY* (*Zinc Finger Y*), and the mouse gene was
similarly termed *Zfy*. Analysis of these genes demonstrated that the zinc fingers
formed a two domain repeating unit in that alternate fingers were more similar to
Figure 1.16

The ZFY, Zfy, and ZFX genes

(a) The zinc finger domains of the mouse Zfy-2 and human ZFY proteins

Sequences are aligned to show the two finger higher order repeat, and the conserved cysteine and histidine pairs are boxed. Dots (.) indicate amino acid identity. (From Mardon and Page, 1989)

(b) Alignment of the amino acid sequences of the human ZFY and ZFX gene products

Zinc fingers are boxed, putative nuclear localisation signal is underlined, and positions of sequence differences are indicated by asterisks (*). (From Goodfellow, 1989)
These observations suggested that the genes could have evolved by progressive multiplication and divergence of a primordial two finger unit. Surprisingly, a homologous gene was detected on the X chromosome by cross-hybridisation to human female DNA. The sequence of this gene indicated that it had a similar zinc finger region, and for this reason it was called ZFX (Schneider-Gadicke et al., 1989).

Both ZFX and ZFY were expressed as multiple transcripts in many adult and fetal tissues, the zinc finger domains of both genes were located at the C-terminus of the open reading frame, and they were 96% identical at the amino acid level, implying that their cognate target sequences were very similar (Figure 16b). The N-terminal non-finger regions were 89% identical, and very rich in acidic residues, a characteristic of the transcription activating regions of some transcription factors (Sigler, 1988). A short, highly basic stretch of amino acids KPKKK/RRRP, just prior to the beginning of the zinc finger region, was reminiscent of a nuclear localisation sequence. Although sex in marsupials is determined by the presence or absence of a Y chromosome, the homolog of ZFY/ZFX was located on the autosomes (Sinclair et al., 1988). This was the first suggestion that ZFY might not be the primary sex determinant.

Mouse Zfy genes are expressed in germ cells

The sex-determining region of the mouse was initially defined as a portion of the Y chromosome that was translocated to the tip of the X chromosome in the mutant strain of sex-reversed (Sxr) mice, and thus contained a gene which induced the formation of testes, Tdy (Testis determining Y, McLaren, 1988). Sex determination can therefore be equated with testis determination. This chromosomal region harboured a pair of near identical genes homologous to human ZFY, designated Zfy-1 and Zfy-2 (Mardon et al., 1989). A variant of Sxr, Sxr', was deleted for Zfy-2, yet XX-Sxr' mice, like XX-Sxr mice, possessed testes. This indicated that Zfy-2 was dispensable for testis determination, and was not,
therefore, Tdy. mRNA from both Zfy-1 and Zfy-2 could be detected in adult mouse testis, raising the question of which cell lineage expressed these genes. We/We mice possess testes which lack germ cells due to a defect in their progenitors which prevents their migration into the gonadal primordium, the genital ridge. The formation of testes in these mice, and hence sex determination, was therefore independent of the presence of germ cells. PCR analysis of the expression of both Zfy-1 and Zfy-2 revealed that although Zfy-1 (but not Zfy-2) mRNA was present in normal, differentiating embryonic mouse testes, neither gene was expressed in the embryonic testes of We/We mutant mice (Koopman et al., 1989b). This indicated that detectable expression of Zfy-1 was dependent on the presence of germ cells, and since germ cells were not required for testis determination, Zfy-1 was unlikely to be the sex-determining gene Tdy.

Redefinition of the sex determining region excludes ZFY as a candidate for TDF

Analysis of 14 human XX males who did not possess ZFY sequences revealed that 3 of them, along with an intersex individual, retained Y-specific sequences close to the pseudo-autosomal boundary (Palmer et al., 1989). This redefined the sex-determining region to within 60kb of the pseudo-autosomal boundary on the short arm of the Y-chromosome, and excluded the possibility that ZFY could be the sex determining gene. The XY female which had originally been used to define the region containing the testis determining gene herself contained sequences close to the boundary, yet lacked ZFY, and this was originally offered as evidence that ZFY was testis-determining (Page et al., 1987). However, since the Y-specific sequences in this female were translocated to chromosome 22, a position effect could have inactivated an alternative testis-determining gene located close to the boundary, resulting in a female phenotype independent of the presence or absence of ZFY.
The dependence of testis-specific Zfy-1 expression on the presence of germ cells in the embryonic gonad suggested that Zfy-1 could be involved in spermatogenesis. Moreover, the deletion of Zfy-2 in XX-Sxr' mice coincided with loss of regulation of the H-Y antigen (Burgoyne et al., 1986), suggesting that Zfy-2 could also be involved in controlling an aspect of male sexual differentiation. A recent analysis of the expression of Zfy-1 and Zfy-2 mRNA in the fetal and adult testis supports the view that both genes participate in male germ cell development, and indicates that they may also function in certain fetal somatic tissues (Nagamine et al., 1990). PCR analysis of mRNA from adult brain, heart, kidney, liver, lung, and testis demonstrated that expression of Zfy-1 and Zfy-2 was restricted to testis. Prepuberal mice in which spermatogenesis was just beginning were analysed for the time of onset of Zfy expression. Although there was a low but detectable level of expression in newborn mice, both genes were co-ordinately up-regulated at the initiation of meiotic prophase in the germ line. The bulk of Zfy mRNA in the adult testis, however, was found in the post-meiotic, haploid, round spermatids, and relatively little before this stage. After birth, the level of expression of Zfy-2 mRNA in the testis was greater than that of Zfy-1 mRNA. In contrast, embryonic kidney, liver, heart, lung, and testis mRNA expressed more Zfy-1 mRNA than Zfy-2. Thus both Zfy genes were expressed in several fetal tissues but Zfy-1 mRNA was more abundant, and in the adult their expression was restricted to germ cells, in which Zfy-2 mRNA was more abundant. These results suggested that both Zfy-1 and Zfy-2 were required for the development of a variety of cell types in the 12 day p.c. embryo, but in the adult mouse their functions were specific for germ cells engaged in spermatogenesis.

Consistent with the expression of Zfy-2 mRNA in germ cells throughout many stages of spermatogenesis was the finding that spermatogenesis in XOSxr', Zfy-1-positive, Zfy-2-negative mice was severely defective from an early stage onwards, and thus lacked a primary spermatogenesis gene Spy (Sutcliffe and Burgoyne, 1989). In contrast, XOSxr, Zfy-1 positive, Zfy-2 positive mice were
only affected in late spermatid stages. These observations, together with the mRNA expression analysis, implied that Zfy-2 could be the spermatogenesis gene Spy. A specific genetic counterpart for Zfy-I has yet to be defined.

1.4F Experimental objectives

The goal of the research to be described in the forthcoming chapters was to identify candidate mammalian regulatory genes encoding zinc finger proteins whose expression was cell type-specific and whose functions in the control of cellular differentiation could be investigated. As is clear from this introductory chapter, a number of vertebrate zinc finger genes have been uncovered which are implicated in the control of discrete developmental transitions. The work presented here describes a novel family of highly related human zinc finger genes, and also focuses on a previously undescribed mouse zinc finger gene whose expression is up-regulated at a precise stage of spermatogenesis.
CHAPTER TWO

MATERIALS AND METHODS

2.1 Bacterial cell culture

<table>
<thead>
<tr>
<th>L-broth</th>
<th>Bacto-tryptone</th>
<th>10g/litre</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bacto-yeast extract</td>
<td>5g/litre</td>
</tr>
<tr>
<td></td>
<td>NaCl</td>
<td>10g/litre</td>
</tr>
</tbody>
</table>

L-agar plates

1.5% (w/v) Bacto-agar in L-broth was autoclaved and cooled to 50°C prior to addition of the appropriate antibiotics, then poured into plastic petri dishes and allowed to set.

Antibiotics

Anhydrous ampicillin was redissolved in glass distilled water (GDW) to 50mg/ml, and added to L-broth or L-agar to a final concentration of 50ug/ml.

Solid tetracycline was dissolved in ethanol to a concentration of 12.5mg/ml, and used at a final concentration of 12.5ug/ml.

Solid kanamycin sulphate was dissolved in GDW to 10mg/ml and used at a final concentration of 10ug/ul.

Liquid cultures were propagated in L-broth at 37°C with vigorous shaking; bacterial colonies were grown on inverted L-agar plates at 37°C. Bacterial strains
were preserved by adding glycerol (to 15%, v/v) to aliquots of liquid cultures, then snap frozen and stored in liquid nitrogen.

### 2.2 Bacterial transformation

#### Preparation of competent *E. coli* cells.

<table>
<thead>
<tr>
<th>SOB medium</th>
<th>Bacto-tryptone</th>
<th>20g/litre</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bacto-yeast extract</td>
<td>5g/litre</td>
</tr>
<tr>
<td></td>
<td>NaCl</td>
<td>10mM</td>
</tr>
<tr>
<td></td>
<td>KCl</td>
<td>2.5mM</td>
</tr>
<tr>
<td></td>
<td>MgCl₂</td>
<td>10mM</td>
</tr>
<tr>
<td></td>
<td>MgSO₄</td>
<td>10mM</td>
</tr>
</tbody>
</table>

**TFB I**

- 30mM KAc
- 100mM RbCl₂
- 10mM CaCl₂
- 50mM MnCl₂

Add glycerol to 15% (v/v), adjust pH to 5.8 with 0.2M acetic acid, sterilise by filtration and store at 4°C.

**TFB II**

- 10mM MOPS
- 75mM CaCl₂
- 10mM RbCl₂

Adjust pH to 6.5 with KOH, sterilise by filtration and store at 4°C.

*E. coli* strains DH1 and XL-1 Blue (Stratagene) were routinely made competent for DNA uptake by the following protocol. A single bacterial colony was picked into 5ml SOB medium and grown for 2h at 37°C until the OD₅₅₀ was 0.3. This culture was seeded into 100ml SOB and grown for a further 2h or so until the OD₅₅₀ was
0.48. The culture was chilled on ice for 5 min, then the cells were collected by centrifugation at 6K for 10 min in the Beckman J6B centrifuge. The supernatant was discarded and the pellet was resuspended in 0.4 volumes of TFB I. Cells were incubated on ice for 5 min, then recentrifuged at 1.5K for 10 min at 4°C. The supernatant was discarded, the cell pellet was resuspended in 0.04 volumes of TFB II, and the suspension was incubated on ice for 15 min. 200ul aliquots were dispensed into Nunc vials and snap frozen on dry ice. Competent cells were stored in liquid nitrogen or at -70°C until required.

**Transformation of Competent E. coli cells.**

A frozen aliquot of cells was thawed quickly at room temperature and then placed on ice for 10 min. DNA was added (up to 100 ng in a volume no more than 2/5 volume of cells) and the mixture was incubated on ice for a further 15-45 min. Cells were then heat shocked at 42°C for 90 sec and returned to ice for 2 min. Addition of 2-3 volumes of SOB was subsequently followed by a 50-60 min incubation at 37°C with gentle shaking, and the cells were then plated. Transformation frequencies were usually in the range of 10^6 to 10^8 per ug supercoiled pBR322.

### 2.3 Plasmid and cosmid DNA minipreparations

**Boiling method**

<table>
<thead>
<tr>
<th>STET-lysozyme solution</th>
<th>Na2EDTA</th>
<th>50mM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Triton X-100</td>
<td>5% (v/v)</td>
<td></td>
</tr>
<tr>
<td>Tris.Cl pH8.0</td>
<td>50mM</td>
<td></td>
</tr>
<tr>
<td>sucrose</td>
<td>8% (w/v)</td>
<td></td>
</tr>
<tr>
<td>lysozyme</td>
<td>2g/litre</td>
<td></td>
</tr>
</tbody>
</table>

3ml overnight bacterial cultures were grown overnight in the presence of the appropriate antibiotic. 1.5ml were transferred to an eppendorf, and microfuged for 3 min. The supernatant was aspirated with a drawn out Pasteur pipette, the tube
respun for 5 sec and the remaining supernatant carefully removed from above the pellet. The pellet was thoroughly resuspended in 200ul STET-lysozyme, and placed on ice until all pellets had been similarly treated. The suspensions were boiled for 45 sec, then microfuged at room temperature for 15 min. 150ul of the supernatant was transferred to a clean tube and 150ul propan-2-ol was mixed with it. The DNA was precipitated at -70°C for 20 min. DNA was pelleted by centrifuging at 4°C in a microfuge, then washed with 70% ethanol, dessicated under vacuum, and resuspended in 50ul Tris.Cl pH 8.0, 1mM Na₂EDTA, 0.1mg/ml RNase A by heating at 65°C for 5 min.

This DNA was suitable for most manipulations, and if phenol-extracted 2 to 3 three times could be used as a template for DNA sequence determination.

**Alkaline lysis method**

<table>
<thead>
<tr>
<th>GTE-lysozyme solution</th>
<th>glucose</th>
<th>50mM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Na₂EDTA</td>
<td>10mM</td>
<td></td>
</tr>
<tr>
<td>Tris.Cl pH8.0</td>
<td>25mM</td>
<td></td>
</tr>
<tr>
<td>lysozyme</td>
<td>4g/litre</td>
<td></td>
</tr>
</tbody>
</table>

5M KAc pH 4.8 per 100ml (make fresh):

- 60ml 5M KAc
- 11.5ml glacial acetic acid
- 28.5ml GDW

**PCIA**

- Phenol / chloroform / iso-amyl alcohol (25:24:1), equilibrated with 10mM Tris.Cl, 1mM Na₂EDTA.

**DNase-free RNase**

- Pancreatic RNase A (Sigma) was dissolved in 10mM Tris.Cl pH 7.5, 15mM NaCl, and heated to 100°C for 15min,
5ml overnight cultures of *E. coli* were grown in L-broth with the appropriate antibiotic, and 1.5ml of each suspension was transferred to an Eppendorf tube. Cells were pelleted by centrifugation at 12000g in a microfuge (MSE) for 2min, whilst the remainder of each suspension was stored at 4°C. Supernatant was aspirated with a Pasteur pipette and the bacterial pellet was then resuspended in 100ul GTE-lysozyme and incubated at room temperature for 5min. 200ul of 0.2M NaOH/1% SDS was then added to lyse the bacteria during a further 5min incubation at 4°C. Next, 150ul KAc pH4.8 was added to the lysate. This mixture was incubated for a further 5min at 4°C and subsequently microfuged for 5min at 4°C to pellet the debris. The supernatant containing DNA was removed into a clean tube and extracted once with an equal volume of PCIA. The aqueous phase was transferred to a clean tube and DNA was precipitated from this by adding two volumes of ethanol and incubating the suspension at room temperature for 5min. DNA was pelleted in a microfuge by centrifugation at 4°C for 5min and washed in 500ul 70% ethanol, followed by a final 5min centrifugation at room temperature. The ethanol was removed by aspiration and the pellet was dried briefly under vacuum before resuspending in 50ul 10mM Tris pH8.0/1mM EDTA containing 0.1mg/ml DNase-free RNase. 10ul samples were routinely digested with approximately 10units of restriction enzyme and analysed by agarose gel electrophoresis.

### 2.4 Plasmid and cosmid DNA maxipreparations

<table>
<thead>
<tr>
<th>RNasing Buffer</th>
<th>NaCl</th>
<th>0.1M</th>
</tr>
</thead>
<tbody>
<tr>
<td>Na2EDTA</td>
<td></td>
<td>5mM</td>
</tr>
<tr>
<td>Tris.Cl pH7.9</td>
<td></td>
<td>0.1M</td>
</tr>
</tbody>
</table>

Dehydrated proteinase K was redissolved in 50mM Tris.Cl pH8.0 to 20mg/ml.
incubated at 37°C for 30 min and 1ml aliquots were stored at -20°C.

400ml of L-broth containing the appropriate antibiotic to which the bacterial strain to be propagated was resistant were inoculated with 5ml of overnight culture in a 1l conical flask and then incubated overnight with vigorous shaking at 37°C. Bacteria were pelleted with a Beckman J2-21 centrifuge by centrifugation in the JA-10 rotor at 6K for 10min, then resuspended in 10ml GTE-lysozyme, and the slurry was incubated at room temperature for 5min. Cells were lysed by adding 20ml 0.2M NaOH/ 0.1%SDS and incubating the mixture on ice for 5min. 10ml KAc pH4.8 were subsequently added and the mixture was then rocked on ice for 15min. The precipitated material was pelleted by centrifugation at 8K for 15min with the Beckman J2-21 centrifuge in the JA-10 rotor, and the clarified supernatant was strained through cotton gauze into a clean 200ml glass bottle. 0.6 volumes of propan-2-ol were added, and DNA was precipitated at -20°C for 2h. DNA was recovered by centrifugation at 2K in the J6-B centrifuge at 4°C for 30min, and the pellets were dried then resuspended in 5ml RNasing buffer. 10ul DNase-free RNase was added and the solution was incubated at room temperature for 15min. 125ul 20%SDS and 200ul Proteinase K (20mg/ml) were added and the incubation shifted to 37°C for a further 30min. 5ml PCIA were added to the sample and the mixture was then shaken vigorously. The phases were separated by centrifugation and the aqueous phase was removed into a clean 30ml Corex (Corning) tube. 0.1 volumes 3.5M Na Ac pH4.5 were added, and the DNA was precipitated by adding 2 volumes ethanol and storing at -20°C for at least 2h. The precipitate was collected by centrifugation at 10K and 4°C, with the J2-21 centrifuge in a JS-13 rotor. The solid was then redissolved in enough of the following solution to fill a VTi 65.2 quickseal tube: 8 ml of 10mM Tris.Cl pH8.0, 1mM Na2EDTA to which 8.8g CsCl and 0.8ml 10 mg/ml ethidium bromide solution were added. The samples were centrifuged in the VTi65.2 vertical rotor at 65K for 4 hours (50K overnight), and the lower band of supercoiled plasmid was removed under low frequency UV light
with a syringe and needle. The sample was diluted with 2-3 volumes 10mM Tris.Cl, pH8.0, 1mM Na2EDTA, butanol extracted until all ethidium bromide had been removed, and precipitated with ethanol. The washed, dried pellet was resuspended in 10mM Tris.Cl, pH8.0, 1mM Na2EDTA, and its DNA concentration was estimated by spectrophotometry at OD260.

2.5 Restriction enzyme cleavage and agarose gel electrophoresis of DNA

<table>
<thead>
<tr>
<th>10X New Mix cleavage buffer</th>
<th>Tris.Cl pH7.5</th>
<th>100mM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MgCl2</td>
<td>80mM</td>
</tr>
<tr>
<td></td>
<td>Dithiothreitol (DTT)</td>
<td>10mM</td>
</tr>
<tr>
<td></td>
<td>Gelatin</td>
<td>0.1%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>10X TBE buffer</th>
<th>Tris base</th>
<th>0.89M</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Boric acid</td>
<td>0.89M</td>
</tr>
<tr>
<td></td>
<td>Na2EDTA</td>
<td>20mM</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>10X TEA buffer</th>
<th>Tris.Cl</th>
<th>0.4M</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NaAc</td>
<td>50mM</td>
</tr>
<tr>
<td></td>
<td>Na2EDTA</td>
<td>10mM</td>
</tr>
</tbody>
</table>

DNA was cleaved with restriction enzymes in either an enzyme-specific buffer whose constitution was indicated by the enzyme supplier, or in 1X "New Mix" buffer supplemented with an appropriate concentration of NaCl or KCl (See Sambrook et al., 1989). In all reactions the volume of enzyme stock added was kept below 10% total volume, to prevent the glycerol from affecting the enzyme activity. A fivefold excess of enzyme was generally adequate for cleaving a DNA sample to completion within 1 hour, when the reaction was incubated at the appropriate temperature. When partial cleavage of DNA samples was required, the optimal quantity of enzyme and time course of incubation were determined.
empirically before a precious sample was used. DNA molecules were resolved by agarose gel electrophoresis, using agarose concentrations of between 0.5% - 1.8% (containing 0.5μg/ml ethidium bromide), depending on the size of the DNA fragments to be resolved. For genomic DNA Southern gels, 1X TEA was a suitable running buffer, and for most other preparative and analytical purposes, 0.5X TBE was effective and provided rapid resolution of DNA fragments. Fragment sizes were usually estimated with φX174/HaeIII, λ/HindIII, or λ/BstEII markers.

2.6 Preparation of radioactively labelled DNA probes

**Random hexamer priming** (after Feinberg and Vogelstein, 1983, 1984)

<table>
<thead>
<tr>
<th>Solution</th>
<th>Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1ml (1.25M Tris.Cl, 0.125M MgCl₂ pH7.5)</td>
</tr>
<tr>
<td></td>
<td>18ul β-mercaptoethanol</td>
</tr>
<tr>
<td></td>
<td>5ul 0.1M dATP</td>
</tr>
<tr>
<td></td>
<td>5ul 0.1M dTTP</td>
</tr>
<tr>
<td></td>
<td>5ul 0.1M dGTP</td>
</tr>
<tr>
<td>B</td>
<td>2M Hepes adjusted to pH7.4 with NaOH</td>
</tr>
<tr>
<td>C</td>
<td>Calf thymus hexadeoxyribonucleotides (Pharmacia) suspended in 10mM Tris.Cl,</td>
</tr>
<tr>
<td></td>
<td>1mM Na₂EDTA pH 7.5, at 90 OD₂60 units/ml.</td>
</tr>
</tbody>
</table>

**OLB solution**

100μl solution A
250μl solution B
150μl solution C
BSA  Bovine serum albumin dissolved in 10mM Tris.Cl, pH7.5, 1mM Na$_2$EDTA to 10mg/ml.

TES  Tris.Cl, pH7.9  10mM
Na$_2$EDTA  10mM
SDS  0.5% (w/v)

25-100ng purified double-stranded DNA was suspended in 32ul GDW and boiled for 5 min then placed on ice. 10ul OLB, 2ul BSA, 5ul α-$^{32}$P dCTP (3000Ci/mmol), and 1ul Klenow fragment (8U/ul, BRL) were added and the mixture incubated at room temperature for 2 hours to overnight. For probes isolated on low melting point agarose gels, a gel slice containing the DNA was cut out, 3 volumes GDW were added and the mixture boiled for 7 min. After cooling 10 min at 37°C, an aliquot containing 50ng DNA was transferred to a fresh tube and the volume adjusted to 32ul. The other reagents were added to a final volume of 50ul, and the reaction incubated as above. Labelled probes were purified from unincorporated nucleotides on a Sephadex G-50 column equilibrated with TES. Incorporation of radioactivity was normally >10$^8$ cpm/ug input DNA. The probe was boiled for 3 min before adding to the hybridisation. DNA in low melting point agarose was re-boiled for 2 min before it was re-used in labelling reactions.

**End-labelling oligonucleotides**

10X kinase buffer  Tris.Cl pH7.6  0.5M
MgCl$_2$  0.1M
DTT  50mM
Spermidine  10mM
Na$_2$EDTA  10mM

20-200ng oligonucleotide in GDW was mixed with 15ul γ-$^{32}$P rATP
(>5000Ci/mm), 2ul 10X kinase buffer, 1ul T4 polynucleotide kinase (Biolabs, 10U/ul) in a final volume of 20ul, and incubated at 37°C for 30-60 min. The labelled product was purified by native polyacrylamide gel electrophoresis or ethanol precipitation.

2.7 Southern transfer of DNA

<table>
<thead>
<tr>
<th>Denaturing solution</th>
<th>NaOH</th>
<th>0.5M</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NaCl</td>
<td>1.5M</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Neutralising solution</th>
<th>Tris.Cl, pH7.6</th>
<th>0.5M</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NaCl</td>
<td>1.5M</td>
</tr>
<tr>
<td></td>
<td>Na₂EDTA</td>
<td>1mM</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>20X SSC</th>
<th>NaCl</th>
<th>3M</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Na₃Citrate</td>
<td>0.3M</td>
</tr>
</tbody>
</table>

DNA samples were cleaved with appropriate restriction enzymes and electrophoresed in 1X TEA 0.8% agarose gels. After electrophoresis the gel was photographed, then treated for 15 min with 0.15M HCl, and washed in GDW before immersing in Denaturing solution for 30 min. After rinsing once more with GDW the gel was submerged in Neutralising solution, and then capillary blotted onto a nylon membrane (Hybond-N, Amersham) in 20X SSC overnight. The membrane was removed the next day and the DNA fixed to it by baking 2 hours at 80°C, or ultraviolet light cross-linking for 2-5 min.

2.8 Hybridisation to DNA fixed on Nylon membranes

<table>
<thead>
<tr>
<th>100X Denhardt's solution</th>
<th>BSA</th>
<th>2% (w/v)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ficoll</td>
<td>2% (w/v)</td>
</tr>
<tr>
<td></td>
<td>PVP</td>
<td>2% (w/v)</td>
</tr>
</tbody>
</table>

| Hybridisation solution    | SSC     | 6X       |
Membranes were pre-hybridised for 1-2 hours at 65°C in Hybridisation solution, to which denatured sonicated salmon sperm DNA had been added to a final concentration of 100μg/ml. 5ml solution per 100cm² membrane were used. Boiled randomly primed probes or end-labelled oligonucleotides were then added to a final concentration of 5 x 10⁵ - 1 x 10⁶ cpm/ml, and the hybridisation allowed to proceed overnight. Randomly primed probes were hybridised at 65°C, whereas oligonucleotides were generally hybridised at 2-5°C below the calculated Tm.

Membranes were washed non-stringently at between 6X SSC/0.1% SDS and 2X SSC/0.1% SDS, at an empirically determined temperature, whereas stringent washes were done in 0.1X SSC/0.1% SDS for 30-60 min at 65°C, prior to autoradiography at -80°C with intensifying screens.

2.9 Plating of bacteriophage cDNA libraries

<table>
<thead>
<tr>
<th>CY medium</th>
<th>Bacto-casamino acids</th>
<th>10g/litre</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bacto-yeast extract</td>
<td>5g/litre</td>
</tr>
<tr>
<td></td>
<td>NaCl</td>
<td>3g/litre</td>
</tr>
<tr>
<td></td>
<td>KCl</td>
<td>2g/litre</td>
</tr>
<tr>
<td></td>
<td>Tris.Cl pH 7.4</td>
<td>25mM</td>
</tr>
<tr>
<td></td>
<td>MgCl₂</td>
<td>10mM</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>λdil</th>
<th>NaCl</th>
<th>5.8g/litre</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MgSO₄.7H₂O</td>
<td>2g/litre</td>
</tr>
<tr>
<td></td>
<td>Tris.Cl pH 7.5</td>
<td>50mM</td>
</tr>
<tr>
<td></td>
<td>Gelatin</td>
<td>2% (w/v)</td>
</tr>
</tbody>
</table>

The hfl⁺ E.coli strain NM514 was used as host for plating cDNA libraries constructed in the bacteriophage vector λgt10. E. coli strain XL1-Blue was the host used for plating cDNA libraries constructed in the bacteriophage vector λZAP.
20ml overnight cultures of the host were grown in CY medium supplemented with 0.2% maltose. Cells were centrifuged at 2K for 10 min at 4°C in the J6-B centrifuge, and the pellet was resuspended in 8ml 10mM MgCl₂. These infection-competent cells were then stored at 4°C and used within 3 weeks. The titre of a cDNA library was determined by mixing serial dilutions (10⁰, 10⁻², 10⁻⁴) of 1ul of the library stock in λdil with 200ul competent NM514. Each mixture was incubated at 37°C for 15 min then 4ml CY top agarose (CY medium + 0.7% agarose w/v) was added and poured immediately onto a 80mm diameter L-agar plate. Plates were then incubated overnight at 37°C. Plaques were then scored and the titre estimated. To plate the titred library for screening with DNA probes, large 245mm x 245mm "megaplates" (Nunc) containing L-agar were used. In this case approximately 100,000 phages were mixed with 2.4ml plating cells and subsequently poured onto the plate in 35ml CY top agarose. Plates were incubated at 37°C for 5-8 hours, until the plaques were visible and the edges barely touched one another. Plates were placed at 4°C until required for screening (at least overnight).

2.10 Screening bacteriophage cDNA libraries
Megaplates containing a bacteriophage cDNA library were transferred to 220 x 220 Hybond-N Nylon membranes (Amersham) as follows. A dry membrane was placed on the surface of each plate for 1 min then carefully removed and placed on top of a pad of Whatman 3MM paper soaked in Denaturing solution for 7 min, with the phage-containing surface upwards. The membrane was subsequently transferred to a pad of 3MM soaked in Neutralising solution for 3 min, then a second pad soaked again in Neutralising solution for 3 min, and finally immersed briefly in 2X SSC. After air-drying, the DNA was fixed to the membrane by baking at 80°C for 2 hours. A second, duplicate lift was taken and processed in parallel with the first. Hybridisations with α- and γ-³²P-labelled probes were according to the methods described earlier.
2.11 Isolation of bacteriophage recombinants

Only phages which hybridised in duplicate to the probe were isolated. Bacteriophage clones were purified by successive rounds of screening. Either *E.coli* strains NM514 or C600 could be used for secondary screens and isolated recombinant propagation. The region of the primary plate corresponding to the positive hybridisation signal was picked into 1ml λdil with the wide end of a Pasteur pipette. The phages in this primary plug were serially diluted and subjected to the second round of plating and screening. At least 200 bacteriophages from each primary plug were screened in this way, and single, isolated clones which hybridised to the probe were picked.

2.12 Characterisation of cDNA clone inserts

<table>
<thead>
<tr>
<th>Extraction Mix</th>
<th>Tris.Cl pH 8.5</th>
<th>2M</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Na₂EDTA</td>
<td>0.2M</td>
</tr>
</tbody>
</table>

2X YT-broth

<table>
<thead>
<tr>
<th></th>
<th>Bacto-yeast extract</th>
<th>10g/litre</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bacto-tryptone</td>
<td>16g/litre</td>
</tr>
<tr>
<td></td>
<td>NaCl</td>
<td>10g/litre</td>
</tr>
</tbody>
</table>

10X Ligation buffer

<table>
<thead>
<tr>
<th>Tris.Cl pH7.5</th>
<th>0.5M</th>
</tr>
</thead>
<tbody>
<tr>
<td>MgCl₂</td>
<td>0.1M</td>
</tr>
<tr>
<td>DTT</td>
<td>0.1M</td>
</tr>
</tbody>
</table>

Preparation of dialysis tubing

Lengths of tubing were boiled 10 min in a large volume of 2% (w/v) sodium bicarbonate, rinsed in GDW, boiled 10 min in 1mM Na₂EDTA, and stored submerged at 4°C. Rinse tubing with GDW before use.
100X IPTG 7.68g isopropyl thiogalactoside in 16ml GDW. Add to L-agar plates to a final concentration of 1X (4.8mg/ml).

100X Xgal 12mg bromo-chloro-indolyl-β-D-galactoside in 16ml dimethylformamide. Add to L-agar plates to a final concentration of 1X (80ug/ml).

For cDNA libraries constructed in λgt10, purified bacteriophage were incubated in 1ml λdil and 150ul aliquots were mixed with 200ul infection-competent *E.coli* NM514 or C600 cells. 10ml CY Top agarose (at 50°C) were mixed with each sample and poured onto freshly poured 140mm L-agar plates. After incubation overnight at 37°C without inverting the plates, 15ml λdil was added to each plate which was then incubated overnight at 4°C. 10.6ml of phage supernatant was harvested from each plate and 8.4g CsCl₂ was added. Samples were centrifuged in a Beckman Ti70.1 rotor at 55K for 16hours at 25°C, and the bacteriophage band was then removed in a maximum volume of 1ml with a 19-gauge needle and 1ml syringe. An equal volume of λdil was added to each pulled band and the bacteriophage were re-banded on CsCl₂ step gradients. Step gradients were prepared by adding sequentially, 20 drops of paraffin oil, 1ml (0.842g CsCl₂/1ml λdil), and 2ml (0.51g CsCl₂/1ml λdil). The sample was layered on top of this gradient in a centrifuge tube and centrifuged in the SW50.1 rotor at 35K for 1 hour at 25°C. The band of bacteriophages was removed with a needle and syringe, and to this suspension 0.1 volumes Extraction mix and an equal volume of formamide were added. The extraction was incubated at room temperature overnight and then for 1 hour the next day at 65°C. 6 volumes of ethanol were added and the DNA was pelleted by spining at 10K for 20 min. Pellets were washed in 70% ethanol, the DNA was resuspended in 500ul GDW, and stored at 4°C.

cDNA inserts were cleaved from vector arms with EcoRI and electrophoresed in
0.8% agarose 0.5X TBE gels. A gel slice containing the fragment of interest was cut from the gel and placed in dialysis tubing along with 0.5-1.0ml 0.5X TBE, and each end of the tubing was sealed with a clip. The DNA was electroeluted from the slice, and extracted with PCIA 2-3 times, then ethanol precipitated, washed with 70% ethanol, and resuspended in 10mM Tris.Cl, 1mM Na2EDTA. The cDNA was then ligated to EcoRI-cleaved Bluescript plasmid as follows. 50-500ng insert was mixed with 10-50ng vector in 10ul GDW. 2ul 10X Ligation buffer, 2ul 10mM rATP, 1ul T4 DNA ligase (400U/ul), and 5ul GDW were added and the reaction incubated overnight at 4°C. The next day aliquots of the ligation were transformed into XL1-Blue and plasmid DNA was prepared from white, ampicillin-resistant colonies selected on L-agar / ampicillin / Xgal / IPTG plates.

For cDNA libraries constructed in λZAP, each plaque of interest was picked into 1ml λdil and 200ul of this was mixed with 200ul infection-competent XL1-Blue cells and 10ul of R408 helper phage (titre of ~10^6-10^11 per ml, Stratagene, La Jolla, CA). The mixture was incubated at 37°C for 15 min then 5ml 2X YT-Broth was added and the culture then incubated at 37°C with shaking for 4-6 hours. Each culture was then incubated at 70°C for 20 min, then debris was pelleted by centrifugation in the J6-B centrifuge at 2.2K for 5 min. The clear supernatant was carefully decanted into a sterile tube and stored at 4°C. This stock contained single-stranded bacteriophage particles containing the Bluescript phagemid with cDNA clone insert. To rescue the cDNA clone as a covalently closed circular plasmid, 200ul of the aforementioned stock was mixed with 200ul infection-competent XL1-Blue and incubated at 37°C for 15 min. 1ul, 10ul, and 100ul of this mixture were then plated on L-agar / ampicillin plates and incubated overnight at 37°C.

The next day single colonies were picked into 3ml cultures of L-broth / ampicillin and plasmid DNA was prepared according to the boiling protocol for plasmid DNA minipreparations.
2.13 Mice, embryos, and urogenital ridge dissections

MF1 XY adult mice were used for germ cell fractionation experiments and for producing the adult testis cDNA library (kindly provided by Dr Keith Willison). Other XY and XX mice were normal C57Bl/6J individuals, and the XXSxr and XXSxr' individuals had a C57Bl/6J background (Epplen et al., 1988). Urogenital ridge RNA samples from Q and MF1 random-bred embryos were used for northern analysis and construction of the 11.5 day p.c. urogenital ridge library (kindly constructed by Dr C. Watson). Urogenital ridges consisting of the genital ridge and the attached mesonephric region were dissected (collaboration with Dr Anne McLaren, MRC Mammalian Development Unit, London) under sterile conditions from embryos 11.5, 12.5, and 16.5 days post coitum. The ridges were immediately placed in ice cold 6M urea, 3M LiCl, and stored at -70°C prior to RNA isolation.

2.14 Isolation of total RNA from whole organs and cultured cells

All glassware was soaked in chromic acid overnight to remove all traces of ribonucleases, then rinsed with copious quantities of GDW then baked. All solutions were prepared with GDW that had first been autoclaved in bottles thus treated. Wherever possible sterile disposable plasticware was used.

<table>
<thead>
<tr>
<th>GUTC solution</th>
<th>Guanidinium thiocyanate 5M</th>
</tr>
</thead>
<tbody>
<tr>
<td>tri-sodium citrate</td>
<td>5mM</td>
</tr>
<tr>
<td>Sarkosyl</td>
<td>0.5%</td>
</tr>
<tr>
<td>β-mercaptoethanol</td>
<td>0.7%</td>
</tr>
</tbody>
</table>

Whole organs were dissected from the animal and homogenised in 5-10ml GUTC solution with a mechanical blender (Chirgwin et al., 1979). Cultured mammalian cells were pelleted in a bench top centrifuge, washed in PBSA, then homogenised in 5 cell pellet volumes of GUTC solution with a whirlimixer. The cellular DNA was sheared by passing the homogenate 15-20 times through a 19 gauge needle, then 1g of CsCl was added per 2.5ml homogenate and dissolved by whirlimixing. This mixture was then layered onto a cushion of 5.7M CsCl/0.1M Na₂EDTA in a
Beckman ultracentrifuge tube (polyallomer), and the tube filled to the top with GUTC solution. The samples were centrifuged for 16 hours at 18°C in the Beckman L5-65B ultracentrifuge. The volume of the sample governed the size of the centrifuge tube and cushion volume, and thereby determined which rotor was employed and the speed of centrifugation:

<table>
<thead>
<tr>
<th>Cushion Volume</th>
<th>CentrifugeRotor</th>
<th>Spin Speed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.2ml cushion</td>
<td>Beckman SW50.1 rotor</td>
<td>35K spin</td>
</tr>
<tr>
<td>4.0ml cushion</td>
<td>Beckman SW41 rotor</td>
<td>32K spin</td>
</tr>
<tr>
<td>5.0ml cushion</td>
<td>Beckman SW28 rotor</td>
<td>24K spin</td>
</tr>
</tbody>
</table>

After centrifugation, the supernatant was carefully discarded with a pasteur pipette, an the bottom of the tube containing the clear pellet of RNA was cut off with a scalpel. The remaining drops of GUTC solution were drained from the pellet, and the RNA was resuspended in 0.5ml 10mM Tris.Cl pH7.5, 5mM Na₂EDTA, 1% SDS. The solution was extracted twice with PCIA, and RNA was precipitated at -20°C overnight following the addition of 0.1 volumes NaAc pH5.2 and 2.5 volumes ethanol. The precipitate was pelleted in a microfuge by centrifugation at 4°C for 15min, washed with 75% ethanol for 5min at room temperature, then dried under vacuum and resuspended in 200ul GDW. The optical density of the RNA solution was then measured with a spectrophotometer at a wavelength of 260nm, and using the conversion factor of 1.0 OD\textsubscript{260} units = 40μg/ml, the concentration of the RNA solution was calculated.

Polyadenylated mRNA was purified from total cellular RNA using oligo-dT cellulose chromatography (Aviv and Leder, 1972).
2.15 Spermatogenic cell separation and RNA isolation

<table>
<thead>
<tr>
<th>Solution D</th>
<th>guanidinium thiocyanate 4M</th>
</tr>
</thead>
<tbody>
<tr>
<td>Na₃citrate pH7.0</td>
<td>25mM</td>
</tr>
<tr>
<td>sarcosyl</td>
<td>0.5%</td>
</tr>
<tr>
<td>β-mercaptoethanol</td>
<td>100mM</td>
</tr>
</tbody>
</table>

Testis cells from three adult mice were disaggregated and separated into sixteen fractions according to the method of Wolgemuth et al., 1985. The composition of each fraction was evaluated by microscopic examination with a Laser-sharp confocal scanning microscope, and RNA was isolated from each fraction by the method of Chomczynski and Scacchi, 1987. Cells in each 5ml fraction were pelleted and lysed in 400ul solution D. 40ul 2M NaAc pH4.0, 400ul water-saturated phenol, and 80ul chloroform/iso-amyl alcohol (49:1) were added sequentially, mixed by inversion for 10 sec, cooled on ice 15 min, and the samples microfuged 20 min at 4°C. The RNA-containing aqueous phase was removed into 400ul propan-2-ol, and placed at -20°C for 1 hour. RNA was pelleted by centrifugation, redissolved in 100ul solution D, and reprecipitated with an equal volume of propan-2-ol, at -20°C for up to 1 hour. RNA was pelleted by centrifugation, washed in 75% ethanol, dessicated, and resuspended in 200ul 0.5% SDS. RNA yields were quantitated by optical density measurement, and ethidium bromide staining of electrophoresed samples.

2.16 Northern transfer and hybridisation of RNA

<table>
<thead>
<tr>
<th>10X MOPS buffer</th>
<th>MOPS</th>
<th>0.2M</th>
</tr>
</thead>
<tbody>
<tr>
<td>NaAc pH7.0</td>
<td>5mM</td>
<td></td>
</tr>
<tr>
<td>Na₂EDTA</td>
<td>1mM</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>20X SSPE</th>
<th>NaCl</th>
<th>3.6M</th>
</tr>
</thead>
<tbody>
<tr>
<td>NaPO₄ pH7.7</td>
<td>0.2M</td>
<td></td>
</tr>
<tr>
<td>Na₂EDTA</td>
<td>2mM</td>
<td></td>
</tr>
</tbody>
</table>
Eukaryotic cellular RNA was electrophoresed in formaldehyde - agarose gels which were prepared as follows. For a 200ml gel, 3g agarose (BRL, ultrapure), 20ml 10X MOPS buffer, and 150ml GDW were mixed and heated in a microwave oven to melt the agarose. The gel was cooled to 50°C and 34ml 38% v/v formaldehyde was mixed in and the gel poured immediately in a fume hood. RNA samples were dessicated and resuspended 5ul deionised formamide, 1ul 10X MOPS buffer, 1.6ul 38% formaldehyde, and 1.4ul GDW. Each sample was heated to 65°C for 5 min, cooled on ice, and 1ul loading dye (50% glycerol, 0.1mg/ml bromophenol blue) was added prior to loading. The gel was electrophoresed overnight at 30-40mA. The next day the gel was stained in 1mg/ml ethidium bromide, photographed, and RNA was transferred overnight to a Hybond N membrane in 20X SSC. The next day the blot was disassembled, RNA was fixed to the membrane by baking 2 hours at 80°C, and prehybridised in Northern hybridisation solution 3-4 hours at 42°C. Denatured, random hexamer-primed DNA probes were then added to the prehybridisation mix and allowed to hybridise to the RNA overnight at 42°C. The next day the membranes were washed at high stringency in 0.1X SSC/0.1% SDS at 65°C for 20-30 min prior to autoradiography.

2.17 Primer extension analysis

30ug RNA was mixed with 2ng γ-32P(ADP) end labelled mRNA-specific oligonucleotide in 20mM Tris-HCl (pH8.5), 100mM NaAc, 16mM MgCl₂, in a final volume of 10ul. After heating at 90°C for 2 min, the mixture was cooled at 37°C for 15 min, then 10 ul 20mM DTT, 1ul 10mM dNTPs, 20 U RNasin, and 10U AMV Reverse Transcriptase were added and the reaction was incubated for 45 min at 43°C. The reaction was terminated by adding EDTA to 20mM, and after phenol extraction nucleic acids were precipitated with NH₄Ac. cDNA products
2.18 Anchor Polymerase Chain Reaction cloning of 5' located cDNA sequences

Total cellular RNA containing the mRNA species to be analysed was heated to 90°C for 2 min in 25ul of 5mM Tris (pH7.5) and then mixed with 1ug of the mRNA-specific oligonucleotide in 10mM Tris-HCl (pH8.5), 50mM NaAc, 8mM MgCl$_2$, 10mM DTT, 2mM dNTPs, 20U RNasin, in a final volume of 50ul at 43°C. 10U AMV Reverse Transcriptase were then added and the primer was extended at 43°C for 45 min. The reaction was then terminated by adding Na$_2$EDTA to 20mM and cDNA was precipitated with spermine. A poly (dG) tail sequence was introduced at the 3' end of the first strand cDNA with TdT (BRL) in 2mM CoCl$_2$, 1mM dGTP, for 60 min at 37°C. This reaction was terminated by heating to 70°C and the DNA was precipitated in ethanol. Enzymatic DNA amplification with Taq Polymerase (Perkin Elmer Cetus) was performed in a final volume of 100ul using standard reaction conditions of the manufacturer. The primers included an mRNA specific primer, and for the poly-G end a mixture of oligonucleotide AN-polyC 5'-GCATGCGCGCGGAGC-3' and the AN primer 5'-GCATGCGCGCGGAGG-3' were used at a ratio of 1:9 (Loh et al., 1989). Thirty cycles of amplification were performed, with the annealing step in the first five cycles being performed at 45°C, and in the subsequent 25 cycles at 55°C. The denaturation step was at 94°C for 1 min, the annealing step was for 1.5 min, and the extension was at 72°C for 2.5 min. Products were subjected to a second round of amplification with a second mRNA-specific primer located 5' to the first mRNA-specific primer, along with the AN primer, and the discrete fragment of mRNA-specific cDNA was purified from unincorporated primers by electrophoresis in a 1% agarose gel. This DNA was blunt-end ligated, using standard techniques, into the Bluescript plasmid (Stratagene) and transformed into E.coli (XL-1 Blue, Stratagene). White, ampicillin resistant colonies containing the recombinant plasmid were identified on L-agar / ampicillin / IPTG / Xgal plates, and plasmid DNA was isolated according to
2.19 DNA sequence determination and analysis

DNA sequences were determined using the "shotgun" sequencing system (Bankier et al., 1987) after the preparation of libraries of randomly sonicated fragments of DNA inserted into the Sma I site of the Bluescript plasmid. Each nucleotide was sequenced several times in both directions employing the primed synthesis chain termination method. Sequences were compiled using the Intelligenetics GEL Program, and compared to EMBL, GenBank and PIR databases with the Intelligenetics IFIND software, or the PROSRCH program (Drs J Collins and A Coulson, Department of Molecular Biology, University of Edinburgh). Secondary structure prediction was performed with a program by E Eliopoulos, Department of Biophysics, Leeds University. This program combined the methods of Burgess & Scheraga, Dufton & Hider, Chou & Fasman, Garnier & Robson, Kabat & Wu, Lim, McLachlan, and Nagano, to generate a joint prediction.

Preparation of a "shotgun" library in the Bluescript plasmid

<table>
<thead>
<tr>
<th>Buffer Type</th>
<th>Component</th>
<th>Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>10X ligation buffer</td>
<td>Tris.Cl pH 8.0</td>
<td>0.5M</td>
</tr>
<tr>
<td></td>
<td>MgCl2</td>
<td>0.1M</td>
</tr>
<tr>
<td>10X Polishing Buffer</td>
<td>NaCl</td>
<td>0.6M</td>
</tr>
<tr>
<td></td>
<td>MgSO4</td>
<td>67mM</td>
</tr>
<tr>
<td></td>
<td>Tris.Cl pH 7.9</td>
<td>0.1M</td>
</tr>
<tr>
<td></td>
<td>β-mercaptoethanol</td>
<td>5mM</td>
</tr>
</tbody>
</table>

Supercoiled Bluescript plasmid was cleaved with SmaI (1U per ug DNA) for 1 hour, and agarose gel electrophoresis was used to check for complete linearisation of the DNA sample. 20-30 ug of the DNA fragment to be sonicated was isolated from the vector to which it had been previously ligated by restriction enzyme cleavage, preparative gel electrophoresis, and electroelution from the gel slice into standard large and small scale procedures.
dialysis bags. 10ug batches of purified insert were self-ligated under the following conditions:

<table>
<thead>
<tr>
<th>Component</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>10X ligation buffer</td>
<td>2 ul</td>
</tr>
<tr>
<td>DNA</td>
<td>10ug</td>
</tr>
<tr>
<td>0.1M rATP</td>
<td>2 ul</td>
</tr>
<tr>
<td>0.1M dithiothreitol (DTT)</td>
<td>2 ul</td>
</tr>
<tr>
<td>T4 DNA ligase (Biolabs)</td>
<td>20U</td>
</tr>
<tr>
<td>GDW</td>
<td>to 20ul</td>
</tr>
</tbody>
</table>

The reaction was incubated at 4°C overnight, and checked the next day for effective ligation. 180ul GDW was then added and the high molecular weight DNA was sheared randomly with a probe sonicator. Small aliquots were analysed during the course of sonication by gel electrophoresis, and when the mean fragment size was approximately 500bp, the DNA was ethanol precipitated, washed with 70% ethanol and resuspended in 28ul 10mM Tris.Cl pH8.0, 0.1mM Na2EDTA. Fragments were rendered blunt-ended as follows.

<table>
<thead>
<tr>
<th>Component</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA</td>
<td>28ul</td>
</tr>
<tr>
<td>10X Polishing buffer</td>
<td>3 ul</td>
</tr>
<tr>
<td>0.5mM dNTPs</td>
<td>3 ul</td>
</tr>
<tr>
<td>T4 DNA polymerase (10U/ul)</td>
<td>2 ul</td>
</tr>
<tr>
<td>Klenow fragment (5U/ul)</td>
<td>2 ul</td>
</tr>
<tr>
<td>GDW</td>
<td>to 20ul</td>
</tr>
</tbody>
</table>

The reaction was incubated at 15°C for 3-4 hours, then loaded onto a 1.0% agarose minigel and electrophoresed alongside φX174 / HaeIII markers. A trough was cut in the gel and size fractions within the 500bp and 1000bp size range were collected, phenol extracted, precipitated with ethanol, washed with 70% ethanol, and resuspended in 20ul GDW.

These end-repaired products of random sonication were ligated to Smal-cut Bluescript, overnight at 4°C. A range of vector to insert ratios were used (keeping the quantity of vector constant) to achieve optimal efficiencies of intermolecular
Aliquots of the ligation were transformed into XL1-Blue according to the Standard transformation protocol. After overnight incubation at 37°C on L-agar / ampicillin plates white colonies were scored, picked into 3ml liquid cultures, and a minipreparation of plasmid DNA was prepared according to the Boiling method.

**Direct sequencing of double-stranded plasmid DNA**

<table>
<thead>
<tr>
<th>Buffer Type</th>
<th>Composition</th>
<th>Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>5X Sequenase buffer</td>
<td>Tris.Cl pH 7.5</td>
<td>0.2M</td>
</tr>
<tr>
<td></td>
<td>MgCl₂</td>
<td>0.1M</td>
</tr>
<tr>
<td></td>
<td>NaCl</td>
<td>0.25M</td>
</tr>
<tr>
<td>5X Labelling mix</td>
<td>dGTP</td>
<td>7.5μM</td>
</tr>
<tr>
<td></td>
<td>dCTP</td>
<td>7.5μM</td>
</tr>
<tr>
<td></td>
<td>dTTP</td>
<td>7.5μM</td>
</tr>
<tr>
<td>ddG Termination mix</td>
<td>dGTP</td>
<td>80μM</td>
</tr>
<tr>
<td></td>
<td>dATP</td>
<td>80μM</td>
</tr>
<tr>
<td></td>
<td>dCTP</td>
<td>80μM</td>
</tr>
<tr>
<td></td>
<td>dTTP</td>
<td>80μM</td>
</tr>
<tr>
<td></td>
<td>ddGTP</td>
<td>8μM</td>
</tr>
<tr>
<td></td>
<td>NaCl</td>
<td>50mM</td>
</tr>
<tr>
<td>ddA Termination mix</td>
<td>dGTP</td>
<td>80μM</td>
</tr>
<tr>
<td>ddCTP Termination mix</td>
<td>dGTP</td>
<td>80uM</td>
</tr>
<tr>
<td>-----------------------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td></td>
<td>dATP</td>
<td>80uM</td>
</tr>
<tr>
<td></td>
<td>dCTP</td>
<td>80uM</td>
</tr>
<tr>
<td></td>
<td>dTTP</td>
<td>80uM</td>
</tr>
<tr>
<td></td>
<td>ddCTP</td>
<td>8uM</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ddTTP Termination mix</th>
<th>dGTP</th>
<th>80uM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>dATP</td>
<td>80uM</td>
</tr>
<tr>
<td></td>
<td>dCTP</td>
<td>80uM</td>
</tr>
<tr>
<td></td>
<td>dTTP</td>
<td>80uM</td>
</tr>
<tr>
<td></td>
<td>ddTTP</td>
<td>8uM</td>
</tr>
</tbody>
</table>

| Stop solution         | formamide | 95% (v/v) |
|                       | Na2EDTA   | 20mM      |
|                       | bromophenol blue | 0.05% (w/v) |
|                       | xylene cyanol | 0.05% (w/v) |

The DNA prepared as above was extracted twice with an equal volume of redistilled phenol / chloroform / iso-amyl alcohol (PCIA; ratio: 25-24-1, respectively), ethanol precipitated, washed in 70% ethanol, and resuspended in 50ul GDW after dessication under vacuum.

5ul DNA was diluted with 15ul GDW, and 2ul 2M NaOH/2mM was mixed at room temperature and the reaction incubated at room temperature for 5 min. 3ul 3M NaAc was then added, followed by 7ul GDW, and the DNA was precipitated with 75ul ethanol at -70°C for 5-20 min. DNA was collected by centrifugation, the pellet was washed with 70% ethanol, and dessicated under vacuum.
Sequencing reactions were performed with the modified bacteriophage T7 polymerase, "Sequenase" (USB), according to the manufacturers recommendations.

The DNA pellet was resuspended with 7ul GDW, 2ul 5X Sequenase buffer, and 1ul oligonucleotide primer (0.5 pmol/ul ~ 2.5-5ng/ul for 17-mer). Annealing of primer to template was effected by heating the mixture to 65°C for 2 min in a water bath, then allowing to cool slowly to <35°C.

Primers were extended with Sequenase (1ul stock diluted with 7ul ice cold 10mM Tris.Cl pH8.0, 1mM Na2EDTA) as follows:

<table>
<thead>
<tr>
<th>Ingredient</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>Template / primer mix</td>
<td>10ul</td>
</tr>
<tr>
<td>0.1M DTT</td>
<td>1ul</td>
</tr>
<tr>
<td>1X Labelling mix</td>
<td>2ul</td>
</tr>
<tr>
<td>35S-dATP (&gt;600Ci/mmol)</td>
<td>0.5ul</td>
</tr>
<tr>
<td>diluted Sequenase</td>
<td>2ul</td>
</tr>
</tbody>
</table>

The reaction was mixed carefully and incubated at room temperature for 5-10 min. 3.5ul was then added to each of four prewarmed tubes containing 2.5ul of one of the four Termination mixes, and the reactions terminated at 37°C for 5 min. 4ul Stop solution was then added and the reactions stored at -20°C for up to one week. Prior to loading onto an acrylamide gel, the samples were deanatured at 75-80°C for 2 min. 3ul aliquots were loaded per track.

**Polyacrylamide gel electrophoresis**

<table>
<thead>
<tr>
<th>Ingredient</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>40% acrylamide stock</td>
<td></td>
</tr>
<tr>
<td>acrylamide</td>
<td>100g</td>
</tr>
<tr>
<td>N, N'-bismethylene-acrylamide</td>
<td>5g</td>
</tr>
</tbody>
</table>
Dissolve in 250 ml final volume with GDW. Add 5g amberlite resin MB-1, stir 20 min, filter through Whatman paper and store at 4°C.

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>10X sequencing TBE</td>
<td></td>
<td>108g/litre</td>
</tr>
<tr>
<td>Tris base</td>
<td></td>
<td>108g/litre</td>
</tr>
<tr>
<td>boric acid</td>
<td></td>
<td>55g/litre</td>
</tr>
<tr>
<td>Na₂EDTA</td>
<td></td>
<td>9.3g/litre</td>
</tr>
</tbody>
</table>

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>6% acrylamide gel solution</td>
<td></td>
<td>460g/litre</td>
</tr>
<tr>
<td>urea</td>
<td></td>
<td>460g/litre</td>
</tr>
<tr>
<td>10Xsequencing TBE</td>
<td></td>
<td>100ml/litre</td>
</tr>
<tr>
<td>40% acrylamide stock</td>
<td></td>
<td>9ml/litre</td>
</tr>
</tbody>
</table>

Glass gel casting plates were cleaned with ethanol, then acetone, and finally GDW. The front (toothed) plate was coated on the inside face with repelcote. The gel was prepared by adding 70ul 25% ammonium persulfate (APS, Bio-Rad) and 55ul TEMED (Bio-Rad) to 35 ml acrylamide gel mix, swirling, and then pouring between the two plates separated by plastic spacers at the edges and sealed with PVC tape. The gel was run in 1X TBE at 40mA constant current until the desired separation was achieved. Gels were fixed in 10% glacial acetic acid / 10% methanol for 15 min in a fume hood, then dried under vacuum, and autoradiographed. Sequences were compiled using Intelligenetics GEL software, and compared to the EMBL, GenBank, and PIR databases using the Intelligenetics IFIND program.

2.20 In situ hybridisation to mRNA in tissues
(Collaboration with Dr Peter Koopman, National Institute for Medical Research, London)

Tissue samples were fixed in 4% paraformaldehyde in PBS at room temperature for 20 min. They were rinsed thoroughly in PBS, dehydrated through graded alcohol, and embedded in paraffin. 7μm sections were then cut, then prehybridised and hybridised as described in Koopman et al., 1989a.
2.21 Isolation of genomic clones from cosmid libraries
Duplicate Nylon filter lifts from cosmid libraries constructed in pcos2EMBL (Erich et al., 1987; kindly provided by Dr Lisa Stubbs, ICRF), prepared by standard techniques, (Sambrook et al., 1989) were screened with cDNA probes labelled by random hexamer priming. Only clones which were positive in duplicate were isolated. A 5mm x 5mm area containing each clone was cut from the colony master plate and dispersed in 1ml L-broth and incubated 1 hour at 37°C. 10ul of this was diluted in a further 1ml L-broth, then 2ul, 10ul, and 20ul were spread onto L-agar/kanamycin plates, which were then incubated at 37°C overnight. Glycerol was added to the remaining 990ul to 15% (v/v). This was then snap frozen and stored at -80°C for future reference. Kanamycin-resistant colonies were subjected to a second round of screening and isolated positive colonies were identified.

2.22 Characterisation of cosmid genomic clones

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>DPA</td>
<td>Tris.Cl pH8.0</td>
<td>20mM</td>
</tr>
<tr>
<td></td>
<td>MgCl₂</td>
<td>3mM</td>
</tr>
<tr>
<td></td>
<td>Na₂EDTA</td>
<td>1mM</td>
</tr>
<tr>
<td>DPA-DTT</td>
<td>DPA + 5mM DTT</td>
<td></td>
</tr>
<tr>
<td>DPB</td>
<td>Tris.Cl pH7.5</td>
<td>6mM</td>
</tr>
<tr>
<td></td>
<td>MgCl₂</td>
<td>18mM</td>
</tr>
<tr>
<td></td>
<td>spermidine</td>
<td>30mM</td>
</tr>
<tr>
<td>DPB-ATP</td>
<td>DPB + 10mM rATP</td>
<td></td>
</tr>
<tr>
<td>Diluent</td>
<td>Tris.Cl pH 7.5</td>
<td>40mM</td>
</tr>
<tr>
<td></td>
<td>KCl</td>
<td>200mM</td>
</tr>
<tr>
<td></td>
<td>β-mercaptoethanol</td>
<td>0.15% (v/v)</td>
</tr>
<tr>
<td></td>
<td>gelatin</td>
<td>0.1mg/ml</td>
</tr>
<tr>
<td>20X Loening running buffer</td>
<td>Tris base</td>
<td>87g/litre</td>
</tr>
<tr>
<td>NaH₂PO₄·H₂O</td>
<td>106g/litre</td>
<td></td>
</tr>
<tr>
<td>Na₂EDTA</td>
<td>7.1g/litre</td>
<td></td>
</tr>
</tbody>
</table>

| Loening sample buffer | Loening running buffer | 1X |
| Glycerol | 50% |
| Na₂EDTA | 60mM |
| bromophenol blue | 0.1% |

Cosmid DNA was isolated according to the alkaline lysis mini- and maxi-preparation methods. Restriction enzyme cleavage maps were compiled using a partial mapping protocol (Rackwitz et al., 1985). 10μg cosmid DNA was dialysed in a petri dish overnight at 4°C, against DPA buffer on a VSWP 01300 membrane (Millipore). DPA-DTT was added to the DNA to a final volume of 30μl, then mixed with 10μl DPB-ATP and 10μl crude bacteriophage λ terminase extract (a gift from Heinz Himmelbauer, ICRF, London). Following a 30 min incubation at room temperature, 50μl DPC was added, the mixture was extracted with PCIA (equilibrated in 1M NaCl), and precipitated with ethanol. The precipitate was collected in a microfuge, washed with 70% ethanol, and resuspended in 100μl 10mM Tris.Cl pH 8.0, Na₂EDTA. Aliquots of 5μl were partially digested with restriction enzymes (diluted in Diluent) as follows. 1 unit of enzyme was incubated at 37°C in the appropriate buffer with 5μl terminase-treated DNA, in a final volume of 20μl. At 15, 30, and 45 min 6μl aliquots were added to 1μl 0.5M Na₂EDTA in fresh tubes on ice. 2μl each time point was checked for the appropriate partial activity on a minigel, and if all three showed partial activity, they were pooled.

Two oligonucleotides specific for the two complementary strands of the cosmid COS site, (ON-L, and ON-R; Rackwitz et al., 1985) were end-labelled with γ⁻³²P-ATP in a 10μl volume. 1μl each oligonucleotide was diluted in 50μl 1M NaCl, to which 200μl Loening sample buffer was then added. 5μl of the ON-L and ON-R mixture was added to 6μl and 9μl (respectively) of the partially digested DNA, then heated to 68°C for 5 min, and hybridised at 42°C for 1 hour. The
fragments were then analysed on a 0.5% agarose gel in 1X Loening running buffer, run at 30V for at least 24 hours. The gel was then dried under vacuum and autoradiographed. Fragment sizes were measured and the distances of each restriction site from the cos ends were determined. A map was then constructed and fragments of interest were subcloned into the Bluescript plasmid.

2.23 In situ hybridisation to mouse metaphase chromosomes
(Collaboration with Dr Sarah Williams, ICRF)
Mouse bone marrow cultures were incubated for 60 min at 37°C in RPMI 1640 medium containing 10% fetal calf serum and 0.05ug/ml colcemid prior to the preparation of metaphase chromosome spreads. Following G-banding and photography, in situ hybridisation of the biotinylated cosmid to selected metaphases was performed essentially as described by Edwards et al., 1990. To prevent repeated sequences in the cosmid from hybridising to the chromosomes, 80ng of biotinylated probe was competed with 2.5ug sonicated genomic mouse DNA at 37°C for 3 hours before hybridisation to the metaphase spreads. Gene-specific signals were then detected with FITC-conjugated avidin and amplified by successive incubations with biotinylated anti-avidin and FITC-conjugated avidin. After staining with propidium iodide metaphases were re-photographed using Fujicolor HG400 film with a Zeiss filter set 9. Only paired signals located on both chromatids of the same chromosome were scored from these photographs and the distance from centromere to signal was expressed as a proportion of the distance from centromere to terminus. The location of the gene was inferred by comparison of the signal position with the prebanded photographs and the standard karyotype (Evans, 1989).

2.24 Pulsed Field Gel Analysis
(Performed with the assistance of Dr Jiannis Raggousis and Isabel Hanson, ICRF)
Very high molecular weight DNA, from human cells encased in low melting point agarose blocks, was analysed. Each 100ul block contained 3 x 10^6 cells. Each restriction enzyme digest contained one third of a block and 20U enzyme, and was
incubated in 100ul of the appropriate buffer for 6-16 hours. 5mM spermidine was included in each digest. The fragments were resolved in a 1% agarose gel using an LKB Pulsaphor apparatus. Run times were generally 24 hours, and pulse times were 50 or 60 sec. DNA transfer and hybridisation were performed according to the standard procedures described earlier.
3.1 Introduction

The initial aim of the work presented in this thesis was to identify differentially expressed mammalian zinc finger genes whose involvement in the control of discrete developmental pathways could be tested. Systems of mammalian differentiation were selected which possessed characteristics that rendered such an approach feasible.

**Human B lymphocyte development**

Mature human blood cells are a set of highly specialised cells with discrete functions, yet are all derived from a common pluripotent haemopoietic stem cell population which differentiates along alternative developmental pathways (Figure 3.1; Hood et al., 1984). Self-renewing stem cells are found first in the yolk sac, from where they migrate to the foetal liver, and then to the bone marrow. The stem cells produce a variety of progenitors which are responsive only within a particular tissue microenvironment, and after such interactions they become precursors of a differentiating haemopoietic lineage (Dexter and Spooncer, 1987). The committed precursors of each lineage express a subset of cell surface receptors that transduce proliferation and differentiation signals from stromal cells or soluble growth factors. The foetal liver, spleen, and the adult bone marrow contain B lymphoid progenitors. When activated, virgin B cells migrate to specific locations within the peripheral lymphoid organs, where, in response to antigenic stimulation, they proliferate and then differentiate into both memory B cells and the terminally differentiated, circulating plasma cells. At all stages of this pathway of development the cells express a changing pattern of differentiation antigens on their surfaces which permits patterns of gene activity to be related, in many instances, to specific
Figure 3.1

Development of the haemopoietic system

A common stem cell may give rise to all elements of the blood and the lymphoid system. (From Hood et al., 1984)
functional states (McMichael, 1987).

A considerable variety of transformed cell lines are available which represent many of the differentiated cell types in the haemopoietic system (McMichael, 1987). Human B-lymphoblastoid cell lines (B-LCLs) are large, activated, immunoglobulin-producing cells which express a specific set of cell surface antigens. They can be propagated indefinitely in culture whilst retaining their differentiated phenotype, and are therefore accessible sources of potentially lineage-specific zinc finger cDNA clones whose functional properties can be readily assessed.

3.2 A zinc finger oligonucleotide identifies a large family of sequences in mammalian genomes

The experiments described in this chapter aimed to identify human zinc finger genes that were specifically expressed in B-lymphoblastoid cells, as a first step towards their functional analysis. An end-labelled oligonucleotide probe, finger 1, encoding the peptide sequence THTGEKPYECTECG, from the open reading frame of the mouse mKr1 gene (5'-'ACTCACACTGGGGAGAAGCCCTACGAGTGCACGAGTGTGGG-3'), in Chowdhury et al., 1987), was hybridised at low stringency to a Southern blot of human and hamster DNA that had been cleaved with EcoRI. Figure 3.2 demonstrates the existence of a very large family of sequences, in the genomes of both species, that hybridise to the oligonucleotide. A cDNA library constructed in lambda gt10 from the ROF-NL B-lymphoblastoid cell line (Lock et al., 1988), was screened with finger 1. 200 000 clones were screened at low stringency and 110 positive clones were identified. 100 of these clones were isolated, DNA was prepared from them and was spotted in a 10 x 10 matrix on a nylon filter. The matrix was hybridised with the finger 1 oligonucleotide to determine and compare the strengths of the hybridisation signals of the cDNA clones (Figure 3.3a). Over 75% hybridised to the finger 1 oligonucleotide strongly, in a manner that was likely to be dependent on both the number of TGEKPYE
Figure 3.2

Southern analysis of genomic DNA with the finger 1 probe

Low stringency Southern hybridisation of an oligonucleotide encoding the TGEKPYE repeat (see Materials and Methods for nucleotide sequence) to human and Syrian hamster genomic DNA. Samples were cleaved with EcoRI. Tracks 1 and 2, Syrian hamster, 10ug; tracks 3 to 7, human, 10ug. The size (kb) of the phage lambda/HindIII marker fragments are indicated vertically. The membrane was washed for fifteen minutes twice in 6X SSC, 0.1% SDS at room temperature, then incubated in the same solution at 42°C for two minutes. After a final room temperature wash in 6X SSC, 0.1% SDS, the membrane was air dried and autoradiographed.
Figure 3.3

Isolation of finger 1-positive B-LCL cDNA clones

Hybridisation analysis of finger 1-positive cDNA clones from B-LCL ROF-NL library arranged in a 10x10 matrix. (a) Low stringency hybridisation to finger 1 oligonucleotide. (b) High stringency hybridisation to ZFP36-1.8 cDNA clone. The membrane in (a) was washed for fifteen minutes twice in 6X SSC, 0.1% SDS at room temperature, then incubated in the same solution at 42°C for two minutes. After a final room temperature wash in 6X SSC, 0.1% SDS, the membrane was air dried and autoradiographed. The membrane in (b) was washed for 1 hour in 0.1X SSC, 0.1% SDS, at 65°C, prior to autoradiography.
repeats in the clone and their percent similarity to the oligonucleotide. The sequence relationships between several of the cDNA inserts were established by high stringency cross-hybridisation analysis, which permitted their classification into a minimum of 14 subgroups (V. Cunliffe, R. Lovering, and J. Trowsdale, unpublished results). The cDNA clone located at position 59 in the matrix gave the strongest positive signal when hybridised to the finger 1 probe (Figure 3.3a), and was selected for characterisation as a test for the efficacy of the screening protocol. This clone was 1.8kb in length, and hybridised strongly to 15 other clones under stringent conditions (Figure 3.3b), but gave a negligible hybridisation signal with a further 58 which were nevertheless positive with the finger 1 probe (Figure 3.3a). The cDNA clone was named ZFP36-1.8.

3.3 The gene encoding ZFP36-1.8 is a member of a large subfamily of closely related human zinc finger genes

The 1.8kb cDNA insert of the clone at position 59 was subcloned into the Bluescript plasmid and preliminary sequencing of its 5' and 3' ends revealed that the DNA fragment encoded a zinc finger protein. To characterise the genomic sequences from which this cDNA was derived, Southern blots of several human genomic DNA samples digested with EcoRI and HindIII were hybridised under stringent conditions with a $^{32}$P-labelled probe synthesised from the complete 1.8kb insert. The probe identified a ladder of between 20 to 30 genomic DNA fragments under these conditions (Figure 3.4). A major 7kb EcoRI fragment and two HindIII fragments of 4.5kb and 5.0kb, presumably containing the cognate gene, named ZFP-36, were the strongest hybridising fragments in the ladder. From these data it was concluded that the ZFP36-1.8-related family of sequences contained 20-30 members and that this was a subset of the much larger family of sequences identified with the finger 1 oligonucleotide.
Figure 3.4

Southern analysis of genomic DNA with the ZFP36-1.8 probe

Southern blot analysis of 6 human genomic DNA samples cleaved with EcoRI and HindIII, then hybridised at high stringency (0.1X SSC, 0.1% SDS, 65°C, 1 hour) with the cDNA probe ZFP36-1.8. 5ug- 25ug of each sample was loaded as follows: 1, WT (B-lymphoblastoid cell line); 2, LS174T (colon carcinoma); 3, T47D (primary breast carcinoma); 4, ZR (primary breast carcinoma); 5, HFF (human foreskin fibroblasts); 6, PRIESS (B-lymphoblastoid cell line). The markers are as in Figure 3.2. All cell lines are available in the ICRF Cell Production Unit.
3.4 ZFP36-1.8 is derived from an mRNA encoding a protein with at least fourteen zinc fingers

To characterise further the ZFP36-1.8 cDNA, its complete nucleotide sequence was determined by the primed synthesis chain termination method using a shotgun strategy (Figure 3.5). The longest open reading frame extended the complete length of the cDNA and the region between nucleotides 597 and 1748 encoded an uninterrupted block of 14 zinc finger domains. This open reading frame had no in-frame termination codons. It was concluded that ZFP36-1.8 corresponded to only a part of the mRNA from which it was originally derived. The mRNA contained a block of at least 14 zinc finger domains, which was preceded in the open reading frame by a non finger region of at least 199 amino acids. A putative site for N-linked glycosylation (N X S/T) was located at positions 46-48. The N-terminal region was not particularly rich in acidic residues, as has been documented for other mammalian zinc finger genes (Mardon and Page, 1989; Schneider-Gadicke et al., 1989; Cunliffe et al., 1990), and failed to match any known amino acid sequences in the databases. Two in-frame methionine residues were located at positions 35 and 79, and they could be used to initiate protein synthesis, although neither were in an optimal context for translation initiation (Kozak, 1986). A comparison of their amino acid sequences (Figure 3.6) demonstrated that the finger domains in ZFP36-1.8 contained considerable variation in the centrally located portion between the second cysteine and first histidine of each module, whereas much of the remainder was essentially invariant. Interestingly, domain 12 only possessed a single residue between its two cysteines, in contrast to the other thirteen domains which contained two.
Figure 3.5

Nucleotide sequence of cDNA clone ZFP36-1.8

The complete sequence of this cDNA was compiled from a library of random shotgun subclones. An uninterrupted open reading frame containing the zinc finger region spans the 1.8kb fragment. Numbering of DNA sequence is on the left, and the numbering of the amino acids in the open reading frame is on the right. A potential glycosylation site is underlined.
Figure 3.6

Alignment of the ZFP36-1.8 zinc finger domains

The fourteen zinc finger domains are arranged consecutively with the most N-terminal finger at the top of the figure and the most C-terminal at the bottom. An asterisk indicates the absence of a residue in the sequence, a dash indicates a residue identical to that in the ZFP36 consensus. An x indicates positions for which a consensus residue cannot be assigned.
3.5 Members of the ZFP-36-related gene family have highly homologous zinc finger domains

The complicated hybridisation pattern of ZFP36-1.8 to human genomic DNA indicated the existence of a large family of related genes, and so it became important to ascertain whether the similarities were in the zinc finger region or in the non-finger region of the cDNA. Two Rsal sites located at nucleotides 884-887 and 935-938 cleaved the cDNA clone into two large and one very small fragments. The 5' located fragment contained the N-terminal non-finger region plus the first three zinc fingers, whilst the 3' located fragment contained only zinc finger domains. Hybridisation of the 3' located fragment to an EcoRI digest of human genomic DNA identified most of the fragments visualised with the complete cDNA (Figure 3.7). Hybridisation of the 5' located fragment to the same digest, however, only visualised a subset of these fragments. It was therefore concluded that many of the genomic sequences identified by ZFP36-1.8 contained closely related zinc finger modules. This experiment also indicated that many of the genomic sequences were unlikely to be processed pseudogenes produced by reverse transcription of the mRNA from which ZFP36-1.8 was derived, otherwise they would have been expected to hybridise to both probes since the cDNA sequence did not contain any internal EcoRI sites.
Figure 3.7

Southern analysis of the ZFP36-1.8-related gene family with restriction fragments of the ZFP36-1.8 cDNA

Southern analysis of 10μg human genomic DNA cleaved with EcoRI and hybridised with: a 3' located RsaI-EcoRI fragment of ZFP36-1.8 containing the eleven most C-terminal zinc finger domains ("3'"'); and a 5' located EcoRI-RsaI fragment of ZFP36-1.8 containing the non-finger region and the three most N-terminal zinc finger domains ("5'"'). Markers are as in Figure 3.2. Membranes were washed at 65°C for 1 hour in 0.1XSSC, 0.1%SDS, prior to autoradiography.
3.6 The ZFP-36-related gene family has a complex transcription pattern and is differentially expressed

In order to ascertain whether the ZFP-36 and related genes were transcribed, Northern blots of oligo-dT-selected polyadenylated RNA from a variety of human cell lines were hybridised with the ZFP36-1.8 cDNA (Figure 3.8a). The resultant hybridisation patterns were remarkably complicated, and although each cell type analysed had a distinct transcription profile, they partially overlapped with each other. The transcript sizes ranged from 1.4kb to 8kb. In order to demonstrate that approximately equivalent amounts of poly(A)+ RNA were loaded onto the original gel, the blot was re-hybridised with the 7B6 cDNA probe (Figure 3.8b; Kaczmarek et al., 1985). To show that the high molecular weight RNA was not degraded, the blot was re-hybridised with an HLA class I probe (Figure 3.8c; Trowsdale et al., 1989). These data can be interpreted in two ways. Either multiple ZFP36-1.8-related genomic sequences are transcribed, or the profusion of polyadenylated transcripts originates from a single gene or a small number of genes by alternative transcription initiation and/or processing events.

3.7 Pulsed field gel analysis suggests that the ZFP-36-related genes are clustered in the human genome

Many higher eukaryotic gene families lie in clusters (See Bodmer et al., 1986 for a review). Examples include the globin, HLA, T cell receptor, and immunoglobulin genes. In addition, several clusters of mammalian homeobox genes have been recently described (Graham et al., 1989; Acampora et al., 1989). The collagen and actin gene families, by way of contrast, are dispersed throughout the genome.

To investigate whether the family of zinc finger sequences defined by ZFP36-1.8 were linked in the genome, pulsed field gel blots of human genomic DNA that had been cleaved with rare cutter restriction enzymes were hybridised
Figure 3.8

Northern analysis of mRNA from the ZFP36-1.8-related gene family

Northern analysis of human cell lines for expression of the ZFP36-1.8-related gene family. 2μg of poly(A)-selected RNA from each cell line was electrophoresed, transferred, and hybridised sequentially with probes for (a) ZFP36-1.8, (b) 7B6, and (c) HLA class I mRNAs. In each panel the following abbreviations are used: M, monocytic cell line U937; Hp, hepatoma HepG2; H1, cervical carcinoma HeLa; T, T-lymphoma Molt 4; B, B-lymphoblastoid cell line MANN. Numbers to the left of each panel indicate the position and approximate sizes (kb) of 18S and 28S rRNA molecules.
with the ZFP36-1.8 insert. In Figure 3.9a a series of three NotI fragments were detected in the genome, and Figure 3.9b shows that the same probe detects two MluI, two Sall, and three BssHII fragments in the same genomic DNA. Since cleavage with EcoRI and HindIII generates 20-30 fragments (Figure 3.4), and rare cutter enzymes only generate 2-3 fragments (Figure 3.9), the simplest interpretation of these data is that the sequences identified by ZFP36-1.8 comprise from 1 to 3 clusters in the genome.

3.8 Discussion

The involvement of some zinc finger proteins in the regulation of eukaryotic gene expression has been well documented. Although many of the zinc finger sequences that have been structurally defined in mammalian genomes so far have unknown functions, some are known to encode proteins which interact directly with regulatory sequences in the 5' flanking regions of genes (Kadonaga et al., 1987; Christy et al., 1989). The finger 1 oligonucleotide probe identified a very large family of sequences in the human genome (Figure 3.2), and hybridised to 0.05% of cDNA clones in a B-lymphocyte library. At high stringency, ZFP36-1.8 cross-hybridised with approximately 10-20% of this set of cDNA clones, and identified a ladder of fragments in restriction digests of genomic DNA, many of which had closely related zinc finger domains (Figures 3.4 and 3.7). Hybridisation of the same probe to restriction digests of mouse and hamster genomic DNA revealed ladders of homologous genes, which thwarted initial attempts to map the human genes using somatic cell hybrids (data not shown).

The sequence of the open reading frame in ZFP36-1.8 contained a block of at least 14 zinc fingers. The region between the conserved phenylalanine and the first conserved histidine residues of each zinc finger was highly variable from domain to domain. At least part of this region encompasses the putative α-helical region of each finger domain which is believed to bind in the major groove of its DNA or RNA target, and may reflect a corresponding variation along the length of the
Figure 3.9

Pulsed field gel analysis of the ZFP36-1.8-related gene family
(Performed with the assistance of Dr J. Ragoussis and I. Hanson, ICRF)

Pulsed field gel analysis of DNA from the cell line PGF digested in (a) N, with NotI; and in (b) B, BssHII; S, SalI; M, MluI. y indicates the yeast marker lanes in (a). Hybridisation was with the complete 1.8kb ZFP36-1.8 cDNA probe. The sizes of the marker chromosomes are indicated, vertically, in kilobases. Membranes were washed at 65°C for 1 hour in 0.1X SSC, 0.1% SDS, prior to autoradiography.
Most DNA sequence evolution appears to proceed by the divergence of the products of gene duplication events (Ohno, 1970) and therefore the extensive cross-hybridisation of ZFP36-1.8 with between 20 and 30 highly related genomic sequences suggests that several rounds of duplication followed by limited divergence of ancestral zinc finger domains may have occurred successively. To date these sequences constitute the largest, closely related subfamily of mammalian zinc finger genes that has been identified. The products of gene duplication events are sometimes closely linked in the genome, and the simple pattern of genomic restriction fragments identified with the ZFP36-1.8 probe by pulsed field gel analysis suggests that this family of sequences most likely arose through gene duplication.

Other subfamilies of zinc finger genes are beginning to emerge whose finger domains are very similar. The mouse genes *Krox-20*, *Krox-24*, and human *Spl* all contain a DNA-binding region with three zinc fingers, the amino acid sequences of which are closely related. Interestingly, the recognition sites with which *Sp1* and *Krox-24* interact are also similar, and therefore, by inference, it might be expected that the nucleic acid targets to which the ZFP36-1.8-related family of gene products bind would also share common features. The mouse genes *Zfy-1, Zfy-2* (both located on the Y chromosome), *Zfx* (located on the X chromosome), and *Zfa* (autosomal), (Mardon et al., 1989) each possess an almost identical block of nine zinc fingers. The human subfamily of *GLI*-related genes contain tandemly repeating blocks of closely related zinc fingers (Ruppert et al., 1988). In addition, several zinc finger clones have been isolated from a *Xenopus* oocyte cDNA library that each identify, under high stringency hybridisation conditions, many fragments in restriction digests of genomic DNA which may represent highly homologous genes (Koster et al, 1988 Knochel et al., 1989). The number of zinc finger genes in vertebrate genomes appears to reach into the hundreds, and in high stringency hybridisations probes from some of them give single copy signals in the genome (Chowdhury et al., 1987; Ruiz i Altaba et al., 1987; Ruppert et al., 1988; Ashworth et al; 1989a). As the data presented in this chapter demonstrate, however, many
others belong to subfamilies of genes with very closely related zinc finger domains.

These studies revealed that members of the ZFP36-1.8-related subfamily of genes produced so many transcripts that northern analyses revealed smears of mRNA molecules. Such an intricate pattern most likely results from the expression of a small number of transcripts from each of a large number of closely related genes, although it is possible that it is produced by the alternative use of an unprecedentedly extensive series of exons within a one or a few gene(s).

Taken together, the results of our experiments suggest that the family of sequences identified by ZFP36-1.8 are functional genes with extensive similarities in their zinc finger regions, whose expression is, in part, under lineage-specific control, and which appear to be closely linked in the genome. Three distinct genomic cosmid clones have been isolated which hybridise strongly to the ZFP36-1.8 probe, and their zinc finger regions are being sequenced to determine precisely how similar they are to that of the ZFP36 gene. Moreover, these clones have yielded gene-specific probes whose location in the genome is being mapped by in situ hybridisation and somatic cell genetics. In pursuit of the biochemical function(s) of these genes, the ZFP36-1.8 cDNA clone is also being expressed in E. coli. The recombinant protein thus produced will be used for in vitro binding studies to identify the sequence of its putative nucleic acid target(s) (Kinzler and Vogelstein, 1989).

The initial screen with finger 1 identified a large number of cDNA clones in a B-LCL library. Although no evidence can be offered as yet for cell type-specific transcripts from the gene family identified by ZFP36-1.8, the characterisation of several other finger 1-positive cDNA clones which do not cross-hybridise to this cDNA (Figure 3.3) and are lymphoid-specific, is currently in progress (R. Lovering, personal communication).
A MOUSE ZINC FINGER GENE WHICH IS TRANSIENTLY UP-REGULATED DURING SPERMATOGENESIS

4.1 Introduction - mouse male germ cell development

The analysis of the ZFP36-1.8-related human gene family in Chapter Three revealed a genomic and transcriptional complexity that was completely unexpected, and it was possible that this reflected complexity in the number of developmental alternatives available to the haemopoietic stem cell population (Figure 3.1). The differentiation of mouse male germ cells represented a potentially more tractable system in which self-renewing stem cells give rise to differentiating progeny that can follow only one pathway of development, ending with the production of haploid spermatozoa.

The mouse germ line is allocated early during development (Snow, 1981), is experimentally distinguishable from other embryonic and adult somatic lineages, and in XY animals there is a precise spatio-temporal arrangement of its morphologically distinct elements during spermatogenesis. Furthermore, a number of genes which specifically affect germ cell development have been identified. Primordial germ cells arise from primitive ectoderm and by 8 days post coitum (p.c.) they are visible as an alkaline phosphatase positive group of cells at the base of the allantois (Snow, 1981). Over the next three days the germ cells proliferate extensively and locomote from the hindgut region to the mesodermal thickenings of the peritoneum known as genital ridges. The genital ridges develop into the somatic elements of the gonad which include, in XY embryos, the Sertoli cells of the seminiferous epithelium, the interstitial Leydig cells which produce androgens, the peritubular myoid cells, as well as blood vessels and connective tissue.

Once within the embryonic gonad male germ cells cease mitotic proliferation, which is only resumed after birth, when they develop into spermatogonia which are located at the periphery of the testis cords. Spermatogonia embark upon
Figure 4.1

Cell types in the mouse testis

The pathway of male germ cell development is indicated, with $m$ and $M$ representing mitotic and meiotic cell divisions, respectively. Other abbreviations: $In$, intermediate spermatogonium; $Pl$, proleptotene spermatogonium. (From Hecht, 1986)
spermatogenesis (Figure 4.1), which includes both meiosis and spermiogenesis, and during maturation they translocate towards the centre of the testis cords, enlarge to a diameter of 30μm, and differentiate into primary spermatocytes, whilst maintaining cell-cell interactions with the supporting Sertoli cell lineage (Russell, 1980). Thereafter meiosis ensues, and after the second reduction division the haploid round spermatids are transformed into spermatozoa by extensive restructuring of the nucleus, construction of a flagellum, and the assembly of the acrosome (Bellve, 1979). In the course of this radical program of differentiation gene expression can be monitored by several methods that rely on distinguishing cellular properties such as cell size, time of appearance during testis development, and precise location within the testis cords. The aim of the experiments described below was to identify differentially regulated mammalian zinc finger proteins controlling germ cell fate, by isolating zinc finger cDNA clones expressed in the developing germ line, and evaluating their role in germ cell development thereafter.

4.2 Isolation of a cDNA clone from an 11.5 day p.c. urogenital ridge library which corresponds to an abundant mRNA in adult testis

(Collaboration with Dr Anne McLaren, MRC Mammalian Development Unit, London)

Sexual differentiation is first discernable within the embryonic gonad at around 12 days p.c. (McLaren, 1984), and we presumed that the molecular events responsible for triggering germ cell differentiation might occur just before this time. A cDNA library from 11.5 day p.c. urogenital ridges was therefore screened with the finger 1 oligonucleotide. From several clones that were isolated, one of them, designated Zfp35-0.6, was subjected to further analysis.

By sequencing of the 0.6kb insert in Zfp35-0.6 it was established that its longest open reading frame encoded six contiguous zinc fingers completely spanning the 600bp. This fragment was used as a probe for hybridisation to RNA, from 11.5 and 12.5 day p.c. male and female urogenital ridges, and adult testis
Figure 4.2

Analysis of RNA from mouse embryonic gonads and adult testis for Zfp-35 expression

(Collaboration with Dr Anne McLaren, MRC Mammalian Development Unit, London)

Northern blot with 8ug total RNA from (1) 11.5 day p.c. male genital ridges, (2) 11.5 day p.c. female genital ridges, (3) 12.5 day p.c. male genital ridges, (4) 12.5 day p.c. female genital ridges, (5) adult testis. Hybridisation was in (a) to a Zfp-35 probe Zfp35-0.6, and in (b) to an actin probe. Positions of the 18S and 28S rRNA transcripts are marked. The actin probe was a 2.0kb Pst I fragment from a mouse beta actin cDNA clone originally isolated by K. Willison, and provided for these experiments by Dr John Mills, ICRF Molecular Endocrinology laboratory.
A 2.4 kb RNA species was detected, which displayed a striking increase in steady state level within the adult testis as compared to the embryonic gonad samples. With much longer exposure (1 week) a low level of the 2.4kb transcript was visible in the embryonic samples.

To isolate longer cDNA clones, an adult testis cDNA library was selected for screening because of the abundance of Zfp-35 RNA in this organ. The library was screened with the Zfp35-0.6 insert, and of many cDNA clones the largest proved to be 2.0kb. This clone, Zfp35-2.0, contained a similar 0.6kb EcoRI fragment to Zfp35-0.6 (Figure 4.3) and had similar hybridisation properties as Zfp35-0.6 by northern analysis. From these preliminary findings it was concluded that the Zfp-35 gene encoded a potential nucleic acid binding polypeptide whose expression was regulated during gonadal development, and a more thorough analysis of its structure and expression was therefore undertaken.

4.3 Molecular analysis of the Zfp-35 transcript

From the size of the RNA detected by northern analysis (approximately 2.4kb), the Zfp35-2.0 cDNA was shorter than full length. The complete sequence of Zfp35-2.0 was determined and to define the length and sequence of the transcribed region adjacent to the 5' end of the 2.0kb cDNA, an antisense oligonucleotide encoding sequence at the 5' end of the Zfp35-2.0 cDNA was used to prime Zfp35-specific first strand cDNA synthesis 5' to Zfp35-2.0. This cDNA was tailed with deoxyguanidine, then rendered double stranded and amplified using a polymerase chain reaction (PCR) adapted for cloning cDNAs whose 5' sequences are unknown (Loh et al., 1989). The details of the oligonucleotides used are included in the legend to Figure 4.3. This Zfp35-PCR cDNA (Figure 4.3) was subcloned into a plasmid vector and the sequences of twelve independent clones were determined. In addition, genomic sequence covering the 5' end of Zfp35-2.0 and the 3' end of the PCR amplified cDNA was obtained (data not shown), which confirmed the colinearity of the Zfp35-PCR 3' region with the 5' end of Zfp35-2.0.
Figure 4.3

Structure of Zfp-35 cDNA clones

Restriction map of Zfp35-2.0, Zfp35-0.6, and Zfp35-PCR cDNA clones. The thick line in all three maps indicates the position of the longest open reading frame. AAAAA represents the poly(A) tail, GGGGG represents the experimentally incorporated poly (dG) tail. The positions of primer-1 and primer-2 used in Zfp35-specific cDNA synthesis and polymerase chain reaction are indicated.

Sequence of primer-1: 5'-GTGCCTAATGAGACCTGT-3'.
Sequence of primer-2: 5'-CTCTTCAGTCACTGTTCCAATCTTTTCCCTGTCAT-3'.

B, Bam HI; E, Eco RI; P, PstI.
Zfp35-0.6 cDNA

Zfp35-2.0 cDNA

Zfp35-PCR cDNA

4.3
These observations permitted assembly of a composite sequence of the 2.4kb transcription unit identified by northern analysis, which contained sequences from the Zfp35-2.0 cDNA clone and the Zfp35-PCR cDNA clones (Figure 4.4).

4.4 The Zfp-35 gene encodes a polypeptide with eighteen zinc fingers and an acidic N-terminus

The longest open reading frame within the transcription unit was deduced by translation of the cDNA sequences (Figure 4.4). The amino acid sequence of the N-terminal 79 residues did not resemble zinc finger sequence and 20 of these were either glutamic or aspartic acid. This frequency was reminiscent of the highly negatively charged transcription activating regions found in a number of eukaryotic transcription factors, although strict primary sequence similarity was not apparent (Sigler, 1988). Secondary structure prediction analysis of the N-terminal 79 amino acid stretch using a combination of eight independent algorithms predicted the existence of four helices within this region (Figure 4.5). Helix 4 had both negatively charged and apolar faces characteristic of amphipathic helices, and such structures have been shown to confer transcription activating properties when linked to a DNA binding domain (Giniger and Ptashne, 1988). The remainder of the coding sequence was taken up by the uninterrupted stretch of 18 zinc fingers. Figure 4.6 is an alignment of the finger domains in order of their appearance in the gene with the most N-terminal finger at the top and the most C-terminal at the bottom. A consensus based on the most frequent amino acid at each position is shown beneath, and the general finger consensus (Gibson et al., 1988) is shown above. The finger domains show a degree of similarity to each other that is more extensive than their similarity with the general finger consensus. The open reading frame had nine in-frame methionines, only the first of which was in a good sequence context for exclusive translational initiation (Kozak, 1986). The first finger in the sequence and the second to last both had an imperfect overall architecture: there was a tryptophan in place of the threonine-glycine-glutamic acid
Figure 4.4

Nucleotide sequence of Zfp-35 cDNA

The sequence was compiled from a library of random shotgun clones of the Zfp35-2.0 cDNA and 12 PCR-amplified cDNA clones containing the remainder of the transcribed sequences 5' to Zfp35-2.0. The deduced amino acid sequence of the longest open reading frame is shown beneath the DNA sequence. Numbering of amino acid residues starts at the position of the presumed initiator methionine at the start of the longest open reading frame.
Figure 4.5

Predictions of the secondary structure of the non-finger region of Zfp-35

(a) Secondary structure analysis of the N-terminal, 79 residue, non-finger region of the Zfp-35 open reading frame. This is based on a joint prediction from the results obtained with eight standard methods (From a program by E. Eliopoulos, see section 2.19). Upper case letters indicate a prediction made by 5 or more methods, lower case indicate a prediction made by less than 5 methods. H/h: α-helix; B/b: β-sheet; T/t: turn. α-helices are numbered.

(b) Helical wheel representation of predicted helix 4, acidic residues are in bold. Many of the residues in the non-finger region are predicted to reside within one of the three types of secondary structures: α-helix, β-sheet, and turn. This is indicated in (a).
a

MEIQFSYESQDHFLSDGETKIKIGEPATEEMTGKIGTVTEESGSLEEDVPHDSRGKEFREFGEELNDQMLFRRR

1

hhhhhh h

2

hHHHHHHHHh

3

hhhh

4

HHHHHHHHHHHHHHHHh

bBBBB

tt tt t t t t tttt ttTTTTtt

b

R

L

F

G

N

1

4

5

R

E

5

4

2

K

2

K

E

1

Q

E

D

EL

helix 4

4.5
Figure 4.6
Alignment of Zfp-35 zinc finger domains

The eighteen zinc finger domains are arranged consecutively with the most N-terminal finger at the top of the figure and the most C-terminal at the bottom. An asterisk indicates the absence of a residue in the sequence, a dash indicates a residue identical to that in the Zfp-35 consensus. The top line is the general finger consensus (Gibson et al., 1988)
CCKFSLHRHTEKPY

GENERAL
CONSENSUS

-DE-DQ-AW-TG-R-T-*W-E
-EE-A-RM-A-VL-
-WI-R-V-K
-DE-A-MI-Q
-ES-H-GMV-L-R-M
-NH-Y-A-K
-DV-A-R-L-
-A-N-N-R-
-E-A-N-V-L-
-N-T-RL-N-
-S-M-RR-V-Y-E
-DK-T-N-L-
-NS-S-RG-V-T
-NL-S-T-V-S-H
-S-N-A-R-L-E-V-R-A
-T-PR-K-
*M-A-C-AFL
-A-R-VN-V-ADQ-LQM

CSQCGKSFQSSDLIKQRTGKYP

ZfP-35
CONSENSUS

4.6
4.5 Regionalised expression of Zfp-35 in adult tissues

The dramatic increase in expression of Zfp-35 in adult testis compared to that in the embryonic gonad samples indicated a temporal control over gene expression during the development of this organ. It was therefore important to determine the level of transcription in other organs to see whether Zfp-35 was spatially regulated in the adult mouse. A northern blot of RNA from a range of organs was therefore hybridised to probes for Zfp-35, actin, and Tcp-1, whose abundant expression is restricted to spermatogenic cells (Willison et al., 1986).

Figure 4.7 shows that there was a clearly enhanced steady state level of Zfp-35 RNA in the adult testis by comparison to the level in the other organs, and densitometric scanning of the autoradiogram revealed that the relative amount of Zfp-35 RNA in testis was increased by at least 20-fold. There were, however, small, consistent variations in amount in the other tissues, as exemplified by the raised level in thymus from both male and female mice. A second transcript in the adult testis mRNA of approximately 3.5kb was also visible. This is likely to be derived from the same gene, because at the same stringency of hybridisation to Southern blots only a single gene was detected (see Chapter 5). The 3.5kb transcript could, alternatively, be a very abundant mRNA from a weakly related gene. As expected Tcp-1 was selectively expressed in the testis, and at an approximately 10-fold higher level than Zfp-35. Actin showed a ubiquitous distribution, with cardiac actin having a smaller (approx 1.4kb) transcript. A trace of a 1.4kb actin transcript was also seen in testis (Waters et al., 1985).

It had been previously observed that some genes selectively expressed in testis also displayed elevated levels of mRNA in brain (Kilpatrick and Millette, 1986; Gizang-Ginsberg and Wolgemuth, 1987; Shackleford and Varmus, 1987;
Figure 4.7

Northern blot analysis of mouse adult organ RNA for Zfp-35 expression

10ug total RNA from a range of male and female organs was electrophoresed, transferred and hybridised sequentially with probes for Zfp-35, Tcp-1, and actin. The exposure time for the Tcp-1 hybridisation was ten fold shorter than that for Zfp-35.
<table>
<thead>
<tr>
<th>Tissue</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spleen</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Testis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lung</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thymus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heart</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kidney</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**RNA Blot Analysis**

- **ZFP-35**
  - 28S
  - 18S

- **Actin**
  - 28S
  - 18S

- **TCP1**
  - 28S
  - 18S

**Note:** The values for TCP1 are not fully visible or legible in the image.
Figure 4.8

Analysis of mouse adult brain and colon RNA for Zfp-35 expression

Northern blot with 10ug total RNA from (1), adult testis; (2), adult male brain; (3), adult male colon; (4), adult female brain; (5), adult female colon. Hybridisation was in (a) to a Zfp-35 probe, and in (b) to an actin probe.
Emoult-Lange et al., 1989) and, to a lesser extent, in colon (Wolgemuth et al., 1989). A further northern blot experiment was therefore performed to evaluate whether Zfp-35 was expressed to a significant degree in these tissues. Figure 4.8 demonstrated that there was no obvious up-regulation of Zfp-35 in samples of RNA isolated from adult male and female brain and colon, by comparison with that observed in adult testis.

4.6 Expression of Zfp-35 in testes of Sex reversed mutant mice

In order to ascertain which cell lineage Zfp-35 expression was dependent upon we examined sex reversed mutant mice. XXSxr and XXSxr' mice are phenotypically male because their genomes contain translocations of a portion of the Y chromosome bearing the testis determining gene Tdy to one of their X chromosomes (McLaren, 1988). Their testes are small and devoid of germ cells. Analysis of Zfp-35 expression in testes from these mice, compared to normal XY controls showed that an absence of germ cells correlated with an absence of the high levels of Zfp-35 RNA characteristic of the XY testes (Figure 4.9). This pattern was recapitulated with a Tcp-1 probe, whereas the expression of actin was uniform. Therefore the high level of Zfp-35 expression in the testis required the presence of germ cells, which raised the question of whether this up-regulation coincided with the presence of a particular spermatogenic stage.

4.7 Time of onset of Zfp-35 up-regulation during testis development

At around 13-14 days p.c. the germ cells in the male gonad obey a mitotic arrest signal and cease to proliferate until after birth. Post partum the primitive spermatogonia in the neonatal testis reacquire their proliferative capacity and initiate spermatogenesis. After successive mitotic divisions, accompanied by progressive changes in morphology and size, the first wave of differentiating spermatogonia
Figure 4.9
Analysis of testis RNA from adult *Sex-reversed* mice for *Zfp-35* expression

Northern blot with 10ug total RNA from (1) testes of adult XXSxr mice, (2) testes of adult XXSxr' mice, and (3) Testes of normal wild type adult mice. Hybridisation was in (a) to a *Zfp-35* probe, in (b) to a *Tcp-1* probe, and in (c) to an actin probe.
mature into resting spermatocytes at around 12 days after birth (Bellve, 1979). Developmental synchrony is maintained by differentiating germ cells throughout the first spermatogenic cycle, which takes a further two weeks to complete. Such behaviour permits the correlation of time of onset of gene expression in the testis with time of first appearance of each distinct stage of male germ cell differentiation. Morphological comparisons of germ cell populations in prepuberal testes (Nebel et al., 1961) revealed a number of important features. First, at around day 14 the first wave of differentiating germ cells have entered early prophase (i.e. leptotene and zygotene). Second, by 21 days most of the germ cells are in pachytene. Third, at day 28 a large population of terminally differentiating, haploid, round and elongating spermatids, has appeared.

Figure 4.10 is a northern analysis of the expression of Zfp-35, Tcp-1, and actin mRNAs over this period. There was a clear onset of increased Zfp-35 expression between days 14 and 21, coincident with the appearance of a large population of pachytene spermatocytes. The kinetics of appearance of Tcp-1 mRNA also showed this behaviour, implying gene activation at pachytene. Steady state levels of actin were uniform.

4.8 Expression of Zfp-35 in fractionated germ cells

To ascertain whether the increase in abundance of Zfp-35 RNA occurred within the pachytene spermatocytes, the expression of Zfp-35 in fractionated testicular cells of adult XY mice was analysed in which the full range of spermatogenic stages were represented. Testis cords were disaggregated with collagenase and the intratubular cells were fractionated according to size by unit gravity sedimentation through a gradient of bovine serum albumin solution. The separation procedure purified, on the basis of size, largely homogeneous populations of pachytene spermatocytes, round spermatids, elongating spermatids and residual bodies with mature spermatozoa. In all, sixteen fractions were collected (Figure 4.11a). The cell number per fraction was plotted in Figure 4.11b.
Figure 4.10

Analysis of testis RNA from prepuberal mice for Zfp-35 expression

Northern blot with 10µg total RNA from (1) testes of 14 day old mice, (2) testes of 21 day old mice, (3) testes of 28 day old mice, (4) and (5) testes of two adult mice. Hybridisation was in (a) to a Zfp-35 probe, in (b) to a Tcp-1 probe, and in (c) to an actin probe. Of the three bands seen in (c) the upper is Zfp-35 signal which was not removed prior to hybridisation with actin. The lower two bands are actin transcripts.
The pachytene spermatocytes came in fractions 3, 4, and 5 whereas the round spermatids were in fractions 10, 11 and 12, the elongating spermatids in fractions, 13, and 14, and fractions 15 and 16 contained residual bodies and spermatozoa. RNA isolated from each fraction was electrophoresed in a formaldehyde agarose gel and transferred to a nylon membrane. The membrane was sequentially hybridised with probes for Zfp-35, and protamine 1 (Figure 4.12). Zfp-35 RNA was found maximally in the pachytene spermatocyte fractions, although expression in haploid cells was detectable. In order to show that these fractions did indeed contain haploid-specific RNA, the blot was hybridised with a protamine 1 probe. Protamine expression was high and essentially restricted to all spermatid stages and the residual bodies, as previously described (Kleene et al., 1983; Braun et al., 1989).

4.9 In situ analysis of Zfp-35 expression in adult testis
(Collaboration with Dr Peter Koopman, National Institute for Medical Research, London)

Evaluation of gene expression by hybridisation to RNA in situ has an advantage over the cell separation procedure in that cellular interactions are unperturbed and hybridisation signals can therefore be directly interpreted within the appropriate histological context. For this reason the procedure was applied to Zfp-35. Germ cell differentiation within the testis cords involves a progressive movement of the germ cell radially, from the periphery towards the centre of the tubule (Oakberg, 1956). As it moves along a radius, so the types of cellular associations made by the germ cell changes because of the concomitant differentiation of the adjacent germ cells. Thus, in each tubule cross section there is a series of concentric rings of germ cells, each containing cells at the same stage of development. Along the length of a tubule, germ cells move into spermatogenesis in longitudinally transmitted waves as a positive signal passes down the tubule axis (Perey et al., 1961). There are several points of initiation of this signal along the
Figure 4.11

Spermatogenic cell separation by velocity sedimentation

(a) Adult mouse testis cell populations were separated into sixteen fractions by unit gravity velocity sedimentation across a gradient of BSA, using a Celsep apparatus, and examined under the Laser-sharp confocal scanning microscope. The scale bar in each frame represents a length of 25 microns. Pachytene spermatocytes are mainly in fractions 3, 4, and 5; round spermatids are in fractions 10, 11, and 12; elongating spermatids are in fractions 14 and 15; residual bodies and mature spermatozoa are in fraction 16.

(b) Graph of fraction number (abscissa) versus absolute cell number in each fraction (ordinate). Cell numbers were estimated with a Coulter Counter using an aliquot of cells from each fraction.
<table>
<thead>
<tr>
<th>Fraction No.</th>
<th>No. of Cells</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20</td>
</tr>
<tr>
<td>2</td>
<td>40</td>
</tr>
<tr>
<td>3</td>
<td>60</td>
</tr>
<tr>
<td>4</td>
<td>80</td>
</tr>
<tr>
<td>5</td>
<td>100</td>
</tr>
<tr>
<td>6</td>
<td>120</td>
</tr>
</tbody>
</table>

**Graph b:**

- **Y-axis:** No. of Cells (Millions)
- **X-axis:** Fraction No.
Figure 4.12

Analysis of gene expression in fractionated adult mouse testis cells

Northern blot analysis of Zfp-35 and protamine 1 RNA in 5μg total RNA isolated from each of the 16 fractions described in Figure 3.10. The positions of the 28S and 18S rRNA transcripts are indicated in the Zfp-35 sections. The size of the protamine 1 transcripts are indicated at approximately 0.6kb (after Kleene et al, 1983).
FRACTION NUMBER
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

28S -

18S - Zfp-35

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

0.6kb - Protamine 1

4.12
length of any one tubule, and so several waves are usually propagated simultaneously. Hence, at different positions down the tubule the germ cells are in correspondingly different states of differentiation. The consequence of superimposing the radially distributed spermatogenic cycle on top of the longitudinally extending wave is that sections of an adult testis show many juxtaposed tubule cross sections with considerable differences in the spectrum of germ cell stages that they contain and in the way that these cells are arranged within the tubules (Oakberg, 1956). This is evident in Figures 4.13a. When an antisense Zfp-35 probe was hybridised to adult testis sections, a whole range of signal strengths could be seen between tubule cross sections and between different layers of cells within the same tubule, indicative of spermatogenic stage-specific expression (Figure 4.13b). The strongest signal was over the pachytene spermatocyte layers, whereas earlier and later germ cell stages had much lower signals, and cells of non-germ cell origin showed a uniformly very low level of expression. The observed heterogeneity was clearer in the higher magnification photographs of Figures 4.13c to 4.13e (tubule classification according to Oakberg, 1956). Figure 4.13c is an example of a stage X tubule with two layers of pachytene spermatocytes (arrowed in the low power micrographs of Figure 4.13a and 4.13b). A dense localisation of silver grains around and on top of these cells was visible, by way of contrast to the low signal in the more centrally located elongating spermatid zones or in the layer of type A spermatogonia at the periphery. Figure 4.13d shows a cross-section through a Stage V tubule that contains a preponderance of early round spermatids and a single layer of spermatocytes, with intermediate or type B spermatogonia at the tubule circumference. Once again, the grain density was high only in the region of the meiotic spermatocytes. Finally figure 4.13e contains a stage XII cross section in which a few primary spermatocytes in metaphase can be seen (arrowed) surrounded by spermatogonial and spermatid stages. There are very few silver grains over these metaphase figures, suggesting that the gene is shut off at this time in the cycle. Such data are consistent with the conclusion that during spermatogenesis there is a peak of Zfp-35 expression at pachytene of meiotic prophase which declines rapidly before the end of the first reduction division. A
Figure 4.13

Localisation of Zfp-35 transcripts in adult testis by in situ hybridisation

(Collaboration with Dr Peter Koopman, National Institute for Medical Research, London)

An antisense RNA Zfp-35 probe was hybridised to all tissue sections

(a) Low power bright field view of testis cross-section. Hybridisation signal is visible as dark grains, as in all other bright field micrographs.

(b) Low power dark field view of cross section in (a) with hybridisation signal visible as white grains, as in the other dark field micrograph in (f).

(c) High power bright field view of cross section through a stage X tubule (arrowed in (a) and (b)).

(d) High power bright field view of cross section through a stage V tubule.

(e) High power bright field view of cross-section through a stage XII tubule.

In low power magnifications (a) and (b) scale bar (at bottom right) represents 0.3mm. In high power magnifications (c), (d), and (e), scale bar represents 0.03mm.
Figure 4.13
Localisation of $Zfp-35$ transcripts in adult testis by in situ hybridisation
(Collaboration with Dr Peter Koopman, National Institute for Medical Research, London)

An antisense RNA $Zfp-35$ probe was hybridised to all tissue sections

(a) Low power bright field view of testis cross-section. Hybridisation signal is visible as dark grains, as in all other bright field micrographs.

(b) Low power dark field view of cross section in (a) with hybridisation signal visible as white grains, as in the other dark field micrograph in (f).

(c) High power bright field view of cross section through a stage X tubule (arrowed in (a) and (b)).

(d) High power bright field view of cross section through a stage V tubule.

(e) High power bright field view of cross-section through a stage XII tubule.

In low power magnifications (a) and (b) scale bar (at bottom right) represents 0.3mm. In high power magnifications (c), (d), and (e), scale bar represents 0.03mm.
similar section to that in Figure 4.13a was hybridised with a sense strand Zfp-35 probe, and demonstrated that non-specific background hybridisation was negligible (data not shown).

4.10 Zfp-35 mRNA is not up-regulated in pachytene oocytes

An important question that followed from the previous conclusion was whether the observed upregulation of the Zfp-35 steady state RNA level in pachytene spermatocytes also occurred in pachytene oocytes. Since the adult ovary does not contain any oocytes at a developmental stage before diplotene of the first meiotic cell cycle the expression of Zfp-35 in the gonads of mid-gestation embryos was analysed. In XY gonads at 13-14 days p.c., T1 prospermatogonia undergo mitotic arrest in G1 of the cell cycle (McLaren, 1984). In XX embryos, however, the germ cells enter meiosis en masse at around this time. By 16.5 days p.c. the vast majority of oocytes have entered pachytene of meiotic prophase, and have experienced meiotic events similar to those occurring in the pachytene spermatocytes of the sexually mature XY mouse. Nevertheless, germ cell differentiation is ultimately controlled by the sex chromosomes, and so functional differences at the molecular level during any of the early meiotic stages might still be expected. Therefore, the expression Zfp-35 mRNA in male and female gonads from 12.5 and 16.5 day p.c. embryos, and adult testis and ovary was investigated. Figure 4.14 shows that the levels of expression of Zfp-35 mRNA in both XX and XY gonads are equivalent at 12.5 and 16.5 days p.c., and much lower than the abundance of Zfp-35 mRNA in adult testis. The apparently higher levels of Zfp-35 expression in 12.5 day p.c. gonads by comparison to 16.5 day p.c. gonads is due to loading slightly more RNA from the former, as evidenced by the higher level of actin mRNA. It could be concluded from this experiment that Zfp-35 upregulation during pachytene was male-specific, indicating that some of the regulatory networks effective during mammalian meiosis are sexually dimorphic.
Figure 4.14

Expression of Zfp-35 mRNA in mouse foetal testes and ovaries

Northern blot of 5-10ug total RNA from (1) 12.5 days p.c male gonads, (2) 12.5 days p.c. female gonads, (3) 16.5 days p.c. testes, (4) 16.5 days p.c. ovaries, (5) adult testis, and (6) adult ovary. The blot was hybridised in (a) to a Zfp35-2.0 probe, and in (b) to an actin probe.
4.11 Discussion

Zfp-35 is a unique member of the TGEKPYE zinc finger gene family with structural characteristics of nucleic acid-binding proteins.

An oligonucleotide encoding the TGEKPYE sequence was used as a probe to screen cDNA libraries. In this screen a novel mouse gene, Zfp-35, was identified that encoded a polypeptide with eighteen contiguous zinc finger repeats and that was developmentally regulated during spermatogenesis. The expression pattern was such that, at the RNA level, Zfp-35 can be considered a molecular marker for the pachytene spermatocyte. The structural significance of the TGEKPYE sequence and its close relatives remains elusive. It is, nevertheless, extremely well conserved and antibodies to a synthetic peptide containing this sequence invariably react with abundant components of the nucleus in a variety of organisms, implying a common role in gene control (Schuh et al., 1987). At the N-terminus of the open reading frame is a stretch of 79 amino acids, 25% of which have acidic side chains. A similarly high frequency of glutamic and aspartic acid residues, in the absence of a compensatory number of positively charged side chains, is a critical feature of the activating regions in a number of eukaryotic transcription factors, and may indicate a possible function of the Zfp-35-encoded protein as a transcription activator.

Computational secondary structure analysis of the N-terminal 79 residues, using a combination of eight independent methods predicted the existence of four helices within this region (Figure 4.5). Helix 4 was predicted as an amphipathic helix, and such structures confer transcription activating properties to polypeptides when attached to a DNA binding domain (Giniger and Ptashne, 1988). This region failed to match strongly any known sequences in the databases using the Intelligenetics IFIND software. However, sequence analysis using the PROSRCH program (J. Collins and A. Coulson, University of Edinburgh) revealed structural similarities to domains within desmin, apolipoprotein AIV, troponin C, and tropomyosin. These proteins contain amphipathic helices that facilitate protein-protein interactions, such as the formation of coiled coils (Steinert and
Koop, 1988; Karathansis, 1985). The similarity of these proteins to the non-finger region of ZFP-35 suggested that its putative amphipathic helix(es) could mediate analogous interactions with other proteins. Limited sequence similarity in this non-finger domain was also found with the N-terminal region of the mouse, human and rat glucocorticoid receptors (Figure 4.15). This region contains transcription activating sequences which appear to mediate protein interactions with other components of the transcription machinery (Evans, 1988). Mutagenesis of the human glucocorticoid receptor by insertion of four extra amino acids at position 120 completely abolished the ability of the protein to transcriptionally activate target genes, (Giguere et al., 1986), raising the possibility that the patch of sequence similarity in the Zfp-35 protein might be used for a similar purpose.

The remainder of the Zfp-35 open reading frame consisted of a block of eighteen zinc finger domains, each of which contained the conserved cysteine and histidine pairs, the phenylalanine at position 8, and closely related variants of the TGEKPYE motif. One point of interest was the near perfect conservation of basic residues at positions 6, 19, and 25 within each finger module, which may be important for making stable contacts with the negatively charged phosphate backbone of nucleic acid targets. The absolutely conserved serine at position 12 marks the beginning of the α-helix (Lee et al., 1989). Other residue positions within the finger domain were moderately variable, and spanned a wider region of the finger domain than the well demarcated region of clustered variability in the ZFP36-1.8 open reading frame described in Chapter Three. This may indicate that the two finger proteins have different modes of interaction with nucleic acid sequences. The current indications are that mammalian genomes contain a number of zinc finger genes with TGEKPYE-like repeats that could reach into the hundreds. Comprehensive sequence database (GenBank, EMBL) searches for similar genes revealed, at the nucleotide level, maximum overall similarity values of between 55% and 65% to the human genes HF.10 (60%), and HF.12 (64%), and to the mouse genes mKr1 (57%) and mKr2 (65%). Protein sequence searches in the PIR database revealed the same overall level of similarity to these finger proteins (HF.10, 63%; HF.12, 69%; mKr1, 65%; and mKr2, 66%). Such values indicated
**Figure 4.15**

Sequence similarity of the non-finger region in Zfp-35 to a region of the glucocorticoid receptor

Similarity of sequences within the N-terminal non-finger region of Zfp-35 (a) to sequences in N-terminal region of the human glucocorticoid receptor (b). Numbers indicate residue position within the open reading frame. Dashes indicate artificially introduced gaps to improve the alignment. Identical residues are indicated by an asterisk, conservative substitutions by a dot.
a 1 MEIQFSYESQDPHHFLSDGETKIKIGEPATEEEMTGKIGTVTEES-GSLEEDVPHDSRGKEFREFGEELNDQ 70
    * .... *.. ** *** .* . . . . . * * . . * . *** .... . .

b 98 MGNDLGFPQQQIQSLSSGETDLKLEESIAN-LN-RSTSVENPKSSASTAVSAAPTEKEFPKTHSDVSE 166

4.15
that although Zfp-35 was clearly related to these genes by structure, it had significantly different finger domains and therefore belonged to a novel subclass within the superfamily of zinc finger protein genes.

Stage-Specific Control of Zfp-35 Expression During Spermatogenesis

Based on the results of our analysis it is reasonable to postulate a transient, cell type-specific requirement for Zfp-35 during pachytene of meiotic prophase. Events occurring around this time, and found only in the germ cell lineage, include elaboration of the synaptonemal complex, assembly of the recombination nodules, as well as homologous chromatid exchange and resolution. Zfp-35 may play some role in the execution of these complex functions, and could even encode a component of synaptonemal complexes, although it is conceivable that specific post-transcriptional controls (Braun et al., 1989) might delay the biological effect of Zfp-35 until after meiosis, when the process of spermiogenesis begins. The precise arrangement of cell types within the testis, the extensive cytoplasmic processes of the Sertoli cells, and the syncytial organisation of spermatogenic cells collectively suggest that differentiating germ cells may coordinately respond to the delivery of positional information from adjacent cells, particularly since the integrity of the intratubular architecture appears to be essential for normal spermatogenesis. The stage-specific expression of Zfp-35 could reflect the involvement of this gene in mediating such responses. It is conceivable that the mechanisms governing male and female meiosis utilise a common set of core components. However the low abundance of Zfp-35 mRNA in both embryonic XX gonads and the adult ovary suggests that the upregulation observed in pachytene spermatocytes is part of a sexually dimorphic regulatory mechanism. The function(s) of Zfp-35 may, therefore, ultimately come under the control of the testis determining gene (Tdy). Potential male-specific functions for Zfp-35 include the control of spermiogenesis (Bellve, 1979), or "imprinting" of paternal gamete genomes which endows spermatozoa with male-specific epigenetic properties that are essential for
subsequent embryogenesis (Surani, 1986).

During spermatogenesis self-renewing stem cells produce a series of differentiating progeny that undergo a dramatic process of structural change, and it is therefore encouraging to find a potential regulatory gene with such a striking stage-specific pattern of expression. Indeed, a range of other mouse genes have been identified whose participation in the control of spermatogenesis is suggested based on their structure, genetics, and expression profiles. The homeobox gene Hox-1.4 has a germ cell-specific transcript that appears at pachytene and is sustained at the same level in the subsequent round and elongating spermatid stages (Rubin et al., 1986, Wolgemuth et al., 1987). Likewise, the gene DIPas1, which encodes a protein with significant structural homology to ATP-dependent RNA helicases (Leroy et al., 1989), is also activated at pachytene and its steady state levels remain high through the spermatid stages. The stability of these steady state levels is in contrast to the pattern of Zfp-35 expression, even though the time of onset is similar, suggesting that there may be qualitative regulatory differences. The data in Figure 4.10 suggest that Tcp-1, a member of the t-complex (Willison et al., 1986) and whose gene product is a component of the trans-golgi network (Willison et al., 1989), is also upregulated at the pachytene stage of prophase. The proto-oncogenes int-1 and c-abl both exhibit spermatid-specific transcription with int-1 appearing in both round and elongating spermatids (Shackleford and Varmus, 1987) whilst a novel c-abl transcript is limited to the elongating spermatids (Ponzetto and Wolgemuth, 1985). These haploid-specific events occur after the high level of Zfp-35 RNA has begun to decline.

The Zfp-35 Polypeptide as a Regulatory Protein

Our results show that Zfp-35 expression is subject to the influence of both spatially and temporally determined cues, and together with its structural analysis suggest that the protein is involved in the control of gene activity. There is already considerable evidence for regulatory roles of related genes. As detailed in Chapter
One, zinc finger protein genes have been discovered which respond *in vitro* to serum, epidermal growth factor, bradykinin, PMA, nerve growth factor, retinoic acid and synaptic activation, and some are restricted in their pattern of expression during embryogenesis (Chavrier et al., 1988; Joseph et al., 1988; Milbrandt, 1987; Chowdhury et al., 1987; Chowdhury et al., 1988; Cole et al., 1989; Wilkinson et al., 1989). These observations support the view that zinc finger proteins are important for diverse processes of development and differentiation. Moreover, the significance of the zinc finger motif continues to be confirmed in functional terms, with its recognition in an ever-widening spectrum of functionally defined eukaryotic trans-acting regulatory genes (Klug and Rhodes, 1987; Evans and Hollenberg, 1988). Notable examples include the gene encoding transcription factor Sp1, the steroid receptor gene family, *SWI5*, and *su(Hw)* (Kadonaga et al., 1988; Evans, 1988; Parkhurst et al., 1988; Spana et al., 1988; Stillman et al., 1988). Further to this, the genetically defined sex-determining region of the mouse Y-chromosome contains a gene, *Zfy-1*, which encodes a protein with a block of 13 zinc fingers (Mardon et al., 1989; Mardon and Page, 1989; Ashworth et al., 1989b). Interestingly, the expression of this gene appears to be restricted to the adult testis, and recent investigations have shown that its expression requires the presence of germ cells in the testis, since *We* homozygous mice have testes but are devoid of germ cells (Koopman et al., 1989). Although it now appears unlikely that neither *Zfy-1* nor its closely linked and highly related partner *Zfy-2* are testis-determining (Koopman et al., 1989b), there are indications that they may function in the control of spermatogenesis (Sutcliffe and Burgoyne, 1989).

Taken together, these observations suggest that the zinc finger motif is diagnostic of regulatory genes that control the flow of information to and from the nucleus, principally at the transcriptional level. The data presented in this chapter suggest that *Zfp-35* is a component of such regulatory networks, and indicate that this gene may be effective in promoting male germ cell development.
5.1 Introduction

The stage-specific expression of Zfp-35 described in Chapter Four implied that its gene product was required in the germ line during the pachytene spermatocyte and/or subsequent stages of spermatogenesis. The generation of transgenic mice harbouring mutant forms of the Zfp-35 gene is a feasible strategy in which the function of the protein can be tested in vivo. Moreover, the production of transgenic mice harbouring Zfp-35 regulatory sequences linked to a reporter gene would facilitate the identification of spermatocyte-specific regulatory elements. As an initial step towards these experiments, a set of cosmid clones containing the Zfp-35 gene were isolated and characterised. To broaden the basis for comparing the functions of this gene with those of other zinc finger genes and those of other genes known to be involved in spermatogenesis, the genomic architecture, chromosomal location, and evolutionary conservation of Zfp-35 were determined.

5.2 Isolation of cosmids containing the Zfp-35 gene

In order to elucidate the genomic organisation of the Zfp-35 gene, and to determine the sequence of its 5' flanking region, cosmids containing this gene were isolated from a cosmid library of mouse strain 129 (kindly provided by Dr Lisa Stubbs, ICRF), using the probe Zfp35-0.6 which was described in Chapter Four. Four distinct, overlapping cosmids were isolated and mapped for restriction enzyme sites using a partial restriction mapping procedure (Rackwitz et al., 1985). These clones covered 50kb of the region containing Zfp-35 (Figure 5.1). The approximate positions of the exons were located by hybridising restriction enzyme digests of cosmids with Zfp-35 cDNA probes, prior to subcloning each
Figure 5.1
Genomic organisation of the mouse Zfp-35 gene

(a) A series of overlapping cosmids that were isolated and ordered by restriction mapping. A composite map of the region of the genome from which they are derived, locating the position of Bam HI sites, B, is shown beneath.

(b) The intron - exon structure of Zfp-35 is represented by black boxes (exons) and lines (introns), and lies underneath the Bam HI restriction map to indicate the location of exons within the cosmid contig. Scale is the same as in (a). The dashed line indicates that the precise size of intron 1 is unknown but is between 2kb and 4.5kb in length.

(c) Location of structural domains in the open reading frame for the Zfp-35 transcription unit. Lines represent 5' and 3' untranscribed regions and introns, open boxes represent 5' and 3' untranslated regions, black boxes represent non-finger domains including the acidic region, and the grey box indicates the position of the zinc finger domain. The exons identified by cDNA probes Zfp35-PCR and Zfp35-2.0 are bracketed underneath.
Bam HI fragment and sequencing. The resultant hybridisation patterns were similar to the patterns obtained when Zfp-35 cDNAs were hybridised to restriction digests of genomic DNA, indicating that the cosmids had not been rearranged during cloning (data not shown).

5.3 Intron - Exon structure of the 2.4kb Zfp-35 transcription unit

The transcribed regions of the gene were sequenced in the genomic DNA by priming subclones from the cosmid with a battery of oligonucleotides encoding short stretches of the Zfp-35 cDNA sequence. In this way the positions of the intron - exon junctions that were used to generate the 2.4kb transcript were located, and the sequences at each end of the introns were determined.

The analysis revealed the presence of two introns in the transcription unit, the first located only 49 nucleotides into the 5' untranslated region, the second interrupting codon 17 of the open reading frame, in the non-finger region (Figure 5.2). Thus, exon 1 contained 49 nucleotides of 5' untranslated region, exon 2 contained 155 nucleotides of 5' untranslated region plus the first 17 codons of the open reading frame, and exon 3 contained the remainder of the non-finger region including the acidic region plus the whole of the zinc finger domain. No sequence differences in the transcribed portions of the gene to that of the cDNA were found. Intron 1 was approximately 500-700bp in length, by restriction mapping, whereas intron 2 was much larger, spanning 12kb. Comparison of the two pairs of splice donor and splice acceptor sequences showed that each corresponded to the canonical splice consensus sequences (Padgett et al., 1986).
Figure 5.2

Nucleotide sequence of the Zfp-35 gene including 5' and 3' flanking regions

The position of transcription initiation is indicated by the right angled arrow, and the locations of the intron-exon junctions are indicated as right angled lines. The positions of two putative polyadenylation signals are underlined and the point of poly(A) addition is shown by the vertical arrow. The sequence of the open reading frame is given beneath the nucleotide sequence. Numbering of the DNA sequence is on the left, and follows from the designation of the first transcribed nucleotide as +1. This numbering is arbitrary as the complete sequences of introns 1 and 2 were not determined. Numbering of the amino acid sequence of the open reading frame is on the right.
5.4 Definition of the site of transcription initiation by primer extension and identification of 5' and 3' flanking sequences

A $^{32}$P end-labelled 18-mer oligonucleotide complementary to sequences from nucleotides +175 to +192 in the cDNA sequence of Figure 4.4 was used to prime cDNA synthesis on RNA from adult male spleen, thymus, and testis, and the products were analysed on a denaturing acrylamide gel (Figure 5.3). Yeast tRNA was used as a negative control for primer extension specificity, and a dideoxy sequencing ladder was used as marker. A testis-specific product 197 nucleotides in length was generated (Figure 5.3, lane 2), and allowing for a single nucleotide cap, the site of transcription initiation was thus localised to a single position four nucleotides 5' to the first nucleotide in the cDNA sequence previously described. The position of transcription initiation thus defined is consistent with the position of the 5' ends of other cDNA clones recently isolated, which are also four nucleotides longer than the previously reported sequence. An additional extension product approximately 500 nucleotides long was also observed in all three mouse tissues, and is likely to be either an alternative transcript from this gene or an experimental artefact. A second primer complementary to the known Zfp-35 cDNA sequence detected only the testis-specific RNA (data not shown).

Primers complementary to the 5' end of the cDNA were used to sequence the region 5' to the initiation site, up to position -232 (Figure 5.2). This DNA sequence bore only one obvious homology to known transcription factor binding sites: an AP1 binding site (TGAGTC; Jones et al., 1988) starting at position -196. The site of transcription initiation as defined by primer extension was shown to be precise and limited to a single nucleotide, yet no TATA box motifs, commonly associated with directing accurate initiation to a single nucleotide, were observed. Sequencing the 3' end of the 1.3kb Bam HI fragment containing exon 3 identified two variant polyadenylation sequences (Figure 5.2), both of which were located 5' to the known position of poly (A) addition from previous cDNA cloning. Between these two sequences was the pentanucleotide CACTG, which is frequently found in close proximity to polyadenylation signals, and may interact with the U4 snRNP particle.
Figure 5.3
Definition of the site of transcription initiation by primer extension
Gel electrophoresis and autoradiography of cDNA synthesis products, using a
$^{32}$P-labelled primer complementary to positions +175 to +192 in Figures 4.4 and
5.2, with total RNA from mouse organs as template. Lanes: 1: 30ug yeast tRNA
control; 2: 30ug testis RNA; 3: 30ug spleen RNA; 4: 30ug thymus RNA; far right:
A dideoxy sequencing ladder size marker.
A substantially T-rich stretch on the sense strand was located close to the site of polyadenylation on its 3' side, and such sequences have been associated with transcription termination. The gene thus encompasses 19kb, of which only 2.4kb encode exons.

5.5 The Zfp-35 gene is conserved in mammals and is a member of a large mammalian zinc finger gene family

Many features of mammalian spermatogenesis are well conserved amongst mammals (Bellve, 1979), and it might be expected that genes required for this process would be similarly conserved. The work described in Chapter Four had established Zfp-35 as a potentially important regulatory gene that was up-regulated during the pachytene stage of spermatogenesis, and so by Southern blotting we analysed genomic DNA from a variety of eukaryotes for the presence of Zfp-35-related sequences (Figure 5.4). Using the cDNA probe Zfp35-PCR (Figure 5.1c), containing non-finger coding sequences and the 5' untranslated region, only the three cognate BamHI fragments identified by cosmid cloning were visualised in mouse genomic DNA, even under low stringency hybridisation conditions (Figure 5.4b). In addition, between one and four high molecular weight bands corresponding to the cognate Zfp-35 gene could be identified in rat, human, whale, and horse DNA. This subset of eutherians represents four mammalian orders with relatively distinct evolutionary histories and dissimilar biological adaptations. As Zfp-35 is a marker for the mouse pachytene spermatocyte, the observed sequence conservation may reflect a function during spermatogenesis that is common to mammals.

When the cDNA probe Zfp35-2.0 (Figure 5.1c), containing the acidic region and all eighteen zinc finger domains, was hybridised at low stringency to the same samples of genomic DNA, a ladder of bands could be identified in all mammals (Figure 5.4a). Much longer exposure of this blot revealed that Xenopus DNA also contained a weakly related family of genes, although chicken DNA did not produce
**Figure 5.4**

**Conservation of Zfp-35-related sequences in eukaryotes**

Low stringency hybridisation of cDNA probes from the Zfp-35 2.4kb transcription unit to Bam HI restriction digests of genomic DNA from a range of eukaryotes.

(a) Hybridisation was with probe Zfp35-2.0 (Figures 5.1c and 4.3).

(b) Hybridisation was with probe ZFP35-PCR (Figures 5.1c and 4.3).


In (a) the membrane was washed at 65°C for 20 min in 2XSSC, 0.1%SDS, and then exposed to film for 18 hours.

In (b) the membrane was washed at 65°C for 20 min in 3XSSC, 0.1%SDS, and then exposed to film for 18 hours.
a signal (data not shown). It is unlikely that these sequences are processed pseudogenes from the homologous Zfp-35 loci because none were identified by the Zfp35-PCR probe in Figure 5.4b. This indicates that a substantial number of these sequences are either functional genes with roles in vertebrate development, or nonfunctional relics of ancestral gene duplication events. The organisation of these related sequences is very similar in closely related mammals. Note, for example, the similarity in the restriction fragment patterns on the three primate DNA samples (Figure 5.4a, tracks 15, 16, and 17), and the near identical restriction fragment patterns that hybridisation to goat and sheep genomic DNA produces (Figure 5.4a, tracks 6 and 7).

5.6 Zfp-35 is located on mouse chromosome 18

Cosmid Zfp35-1 was labelled with biotin and used as a probe for in situ hybridisation to mouse metaphase chromosome spreads. Hybridisation signals were visualised with FITC-conjugated avidin (Figure 5.5). Five metaphase spreads were analysed and signals were detected on eighteen of the twenty chromatids from chromosomes 18. There were no paired signals on any chromosome other than chromosome 18. Comparison with banded karyotypes showed that 72% of all paired signals were localised to bands B3 or C of chromosome 18, indicating that Zfp-35 maps to this region. A histogram of the distribution of signals is shown in Figure 5.6.

To confirm the above assignment to chromosome 18, Southern blot analysis of genomic DNA from a pair of somatic cell hybrids was performed. R44 is a hybrid cell line with hamster background and mouse chromosomes 17 and 18, whereas MF3 contains a hamster background with chromosome 17 but definitely not chromosome 18 (kindly supplied by Dr Bernhard Herrmann, MPI, Tubingen, FRG). cDNA probe Zfp35-0.6 (Figure 4.3) was hybridised to EcoRI-cleaved DNA from these hybrids as well as from the hamster parent cell line and four different inbred mouse strains (Figure 5.5c). A 0.6kb EcoRI fragment corresponding to the
Figure 5.5

Chromosomal mapping of the *Zfp-35* gene

(Collaboration with Dr Sarah Williams, ICRF, London)

(a) Giemsa banded mouse metaphase chromosomes.

(b) In situ hybridisation of biotinylated Zfp35-1 cosmid to the same set of chromosomes visualised with FITC.

Arrows in (a) and (b) indicate position of *Zfp-35* by this method.

(c) Southern blot analysis of somatic cell hybrids for the presence of the mouse *Zfp-35* gene

Genomic DNA samples were cleaved with EcoRI. 10ug of each sample was loaded on the gel as follows: (1) *Mus spretus*, (2) C57Bl6/J, *Mus musculus*; (3) DBA, *Mus musculus*; (4) 129, *Mus musculus*; (5) R44, hybrid with *Mus musculus* chromosomes 17 and 18 only; (6) MF3, mouse-hamster hybrid with *Mus musculus* chromosome 17 but not chromosome 18; (7) V79, hamster parent cell line. Hybridisation was with probe Zfp35-0.6 (Figure 4.3), after which the membrane was washed at 65°C for 1 hour in 0.1XSSC, 0.1%SDS, and then autoradiographed for 14 days.
Figure 5.6

Histogram of the signal distribution along chromosome 18 following in situ hybridisation with cosmid Zfp35-1

Units on the vertical axis represent fractions of the total length of the chromosome and units on the horizontal axis represent the number of signals in each fraction.
mouse Zfp-35 gene was observed in R44, but not in MF3, correlating with the presence of mouse chromosome 18. No hamster-specific sequences could be detected.

Although there is a relative paucity of both genetic and molecular markers for this autosome as compared to those which exist for other chromosomes, the gene *Krox-24* has recently been localised to a region encompassing bands C and D of chromosome 18 (Jannsen-Timmen *et al.*, 1989). This gene encodes a protein with four zinc fingers and its transcription is a component of the immediate early genomic response to growth factors that serum-starved fibroblasts make in vitro. The genomic localisations of *Zfp-35* and *Krox-24* could indicate the existence of a zinc finger gene cluster in this region, a possibility which is currently under investigation. The mouse mutations *Tw, ax, pk, bc, sy*, and *shi* map to chromosome 18 and *pk, bc, and sy* are located in the central portion of this autosome. No spermatogenic defects have been associated with any of these mutations to date (although *sy* homozygotes do not breed; Lyon and Searle, 1989), which indicates that they are unlikely to reside within the *Zfp-35* gene.

### 5.7 Discussion

The genomic organisation of *Zfp-35* and that of other members of the zinc finger gene superfamily are similar.

By cosmid cloning and sequencing the organisation of *Zfp-35* has been determined. A small 47nt exon 1 in the 5' untranslated region is followed by a second small exon of 204nt containing the remaining 152nt of 5'untranslated sequence and the first 17 codons of the open reading frame. The third exon contains the remainder of the 2.4kb transcription unit. A non-coding first exon is a property conserved between several human, mouse, and *Drosophila* zinc finger genes (Tautz *et al.*, 1987; Chowdhury *et al.*, 1988; Morishita *et al.*, 1988; Parkhurst *et al.*, 1988; Chavrier *et al.*, 1989; Schneider-Gadicke *et al.*, 1989), and may indicate a
Although six of the nine zinc finger domains in Xenopus TFIIIA each reside on individual exons (Tso et al., 1986), many of the zinc finger genes that have been well characterised have zinc finger regions unperturbed by introns. To date the block of eighteen zinc finger domains in Zfp-35 constitutes the largest uninterrupted zinc finger-containing exon described for any gene. Such genomic integrity of the finger region indicates that these genes may have evolved by sequence duplication and domain shuffling by unequal crossing over. In this way recombination events between similar TGEKPYE repeats would either permit the exchange of zinc finger domains between genes, or facilitate the insertion of one or more zinc finger domains within a gene. Such mechanisms could produce genes encoding proteins with diverse and overlapping nucleic acid-binding specificities, and thereby engender the rapid evolution of new regulatory networks. The general absence of introns within the zinc finger region and the general presence of an intron in the 5' untranslated region of zinc finger genes together suggest that their genomic organisation may be conserved for a reason. The precise distribution of intron and exon sequences might confer important regulatory properties to such genes.

Given the small size of the first two exons and the large size of the introns, it is possible that there are other exons and/or promoters that could be used in conjunction with exon 3 to produce alternative mRNAs encoding a similar protein under different regulatory circumstances. In support of this, cDNA probes from the Zfp-35 zinc finger region detect both the 2.4kb mRNA, as well as a second 3.5kb mRNA of much lower abundance (Figures 4.7, 4.12, and 4.14), and this may represent an alternative transcript from the Zfp-35 gene, as discussed earlier in Chapter Four. The use of alternative promoters and exons has been documented for vertebrate homeobox genes (Cho et al., 1988; Acampora et al., 1989), as well as for the Drosophila Serendipity, Kruppel, and hunchback zinc finger genes (Vincent et al., 1985; Rosenberg et al., 1986; Tautz et al., 1987).
The 5' flanking regions of a number of genes selectively expressed in testis were compared to that of Zfp-35. Analysis of the promoters from the mouse protamine 1, human germ cell alkaline phosphatase, mouse Hox 1.4, and human PGK2 genes revealed no extensive sequence identities to Zfp-35 (Peschon et al., 1987; Millan and Manes, 1988; Galliot et al., 1989; Robinson et al., 1989). Transcription initiation is known to be precise for Hox 1.4, PGK2 and Zfp-35 even though they each lack an obvious TATA box (Bucher and Trifonov, 1986). Moreover, all three genes are co-ordinately up-regulated in the pachytene spermatocyte, and it is therefore tempting to speculate that such activation may involve a common, spermatocyte-specific, TATA-independent mechanism. A region of the PGK2 promoter that has been shown to direct pachytene spermatocyte-specific expression of a reporter transgene, contains a sequence of 12 nucleotides which matches a sequence in the 5' flanking region of Zfp-35. The sequence AGAGGTTTTTACAT, is located at positions -512 to -491 (Robinson et al., 1989), and is identical at 10 out of 12 positions with the sequence AGAAGTTTTGACAT located at positions -228 to -216 in the Zfp-35 promoter (Figure 4.2). This sequence is also similar to that spanning a site of transcription initiation in the Hox1.4 gene (Galliot et al., 1989), and therefore it will be interesting to determine whether these homologies have any functional significance.

**Zfp-35 as a common regulatory component of spermatogenesis in placental mammals**

Cell-type-specific regulatory controls have previously been shown to limit Zfp-35 up-regulation to the pachytene spermatocyte (Chapter Three), and so it is reasonable to postulate a critical role for this gene in mammalian spermatogenesis. Using a gene-specific probe, Zfp-35 was identified in a diverse group of mammals, indicating that it may indeed have a conserved biological role. To obtain further
comparisons of each Zfp-35 zinc finger domain with every one of the seventeen others were performed, and the number of nucleotide substitutions between each pair was computed (Nei, 1987). An estimate of the divergence time of the two most similar domains, based on the number of synonymous nucleotide differences, was approximately 200MY. By comparison, the mammalian radiation is believed to have occurred only 80MYago. Such estimates imply that the assembly of Zfp-35 into a gene encoding a block of eighteen zinc fingers predated the evolution of many mammalian species, and this corroborates the suggestion from the Southern analysis that Zfp-35 is required for a biologically conserved function during spermatogenesis.
When the studies presented in this thesis were initiated, the existence of large numbers of zinc finger genes in mammalian genomes had not been established. Since then a substantial body of information has accumulated which suggests that they are an important set of mammalian regulatory genes which, in the human genome, may comprise upwards of a hundred genes. The number of zinc finger domains in a protein molecule varies over a wide range. Since it has been proposed that each finger domain may bind to between 2 and 5 base pairs of DNA, the target sites for proteins with many zinc fingers may cover several turns of the double helix, although variations in the contribution of each zinc finger, in multifingered proteins, to the specificity of the interaction might be expected. Several zinc finger proteins are now known to bind to specific DNA sequences and thereby function as transcription factors. It is also conceivable that they control the structure and organisation of chromatin. Moreover, TFIIIA binds specifically to the 5S rRNA product of the gene whose transcription it controls, which suggests that other zinc finger genes might similarly interact with RNA molecules and perhaps regulate their availability to the processing or translation machinery.

Methods to recover the specific target sites to which zinc finger proteins bind from complex mixtures of genomic DNA fragments have been successfully developed (Kinzler et al., 1989). It will now be possible to determine the sequences to which the Zfp-35 and ZFP-36 proteins bind, and thereby identify the genes whose expression they control.

The studies of Zfp-35 demonstrated that its expression was up-regulated at a discrete stage of spermatogenesis, and investigations into the function of the encoded protein are underway. Polyclonal antisera have been raised to three peptides from the open reading frame of Zfp-35 (Figure 6.1), which together with transgenic mice harbouring Zfp-35 constructs will facilitate a combined biochemical and genetic approach to determining the precise role of Zfp-35 during germ cell development.
Figure 6.1
Sequences of synthetic peptides used to raise antibodies to the Zfp-35 protein.
Peptides were synthesised on an Applied Biosystems peptide synthesiser. Antisera to peptides conjugated to thyroglobulin and bovine serum albumin were raised in New Zealand White rabbits using conventional techniques (Sambrook et al., 1989).
ZFP35.1 \( \text{NH}_2\text{-MTGKIGTVEESGSLEEGC-CO}_2\text{H} \)

ZFP35.2 \( \text{NH}_2\text{-GEELNDQMLFRRRQYNGC-CO}_2\text{H} \)

ZFP35.3 \( \text{NH}_2\text{-FSQRSDELVNHQRVHADQKLQMG-CO}_2\text{H} \)


