For my father, family and friends.
ABSTRACT

Murine embryonic development relies on the precise temporal and spatial expression of genes, the protein products of which dictate cellular phenotype. Proteins controlling the transcription of genes during early embryogenesis might be expected to be regulated on the differentiation of F9 embryonal carcinoma stem cells. One such regulated activity is DRTF1 (differentiation regulated transcription factor), a transcription factor which can associate with the product of the retinoblastoma tumour suppressor gene (pRb), p107, and cyclin A. The embryonic expression pattern and biochemical characterisation of DRTF1 are the subject of this study.

DRTF1 is a sequence specific DNA binding protein that consists of at least 3 DNA binding activities referred to as la, b, and c. Complexed DRTF1, referred to as DRTF1a, has similar DNA binding specificity and DNA binding polypeptides to DRTF1b/c, which lack pRb and cyclin A. DRTF1b/c are abundant in embryonal carcinoma (EC) and embryonic stem cells (ES), and down regulated on differentiation. In contrast, DRTF1a is not abundant in EC or ES cells, and is not down regulated as EC cells start to differentiate.

The binding activity of DRTF1 was investigated during murine embryogenesis by preparing microextracts from tissues and whole embryos at various stages of development. DRTF1b was present in blastocyst stage embryos, and was abundant up to about 14 days of gestation. However, as embryogenesis progressed, the levels of DRTF1b/c decreased whereas in contrast, the binding activity of DRTF1a increased. DRTF1b/c were abundant in all tissues examined during early stages of embryogenesis, but became tissue restricted during later stages of development, and were frequently excluded from terminally differentiated tissues, for example liver and brain.

In summary, DRTF1 is a group of DNA binding activities which share common DNA binding polypeptides, and which complex with other non-DNA binding polypeptides such as pRb and cyclin A in a developmentally regulated and tissue dependent fashion.

The DNA binding polypeptides of DRTF1 were purified by sequence-specific affinity chromatography from F9 EC cells. The purification procedure was modified to incorporate a mutant site affinity column, which improved the purity of DRTF1.

A unique 360 nucleotide cDNA was isolated by using degenerate oligonucleotides against peptide sequences derived from one purified polypeptide. The distribution of its mRNA was investigated by Northern blot and RNase protection assays. These data, combined with a detailed examination of the expression during embryogenesis by in situ hybridisation, indicated that the mRNA expression was developmentally regulated, with
high levels of expression at 6.5 to approximately 13 days post coitum, and lower levels of expression at later stages of development. This expression pattern was similar to the quantitative regulation of DRTF1 DNA binding activity, and the cDNA was therefore likely to encode a polypeptide involved in the DRTF1 activity.

In conclusion, DRTF1 is a developmentally regulated and tissue-dependent transcription factor, that is subject to different levels of regulation. It is likely that DRTF1 plays an important role during embryonic development.
TABLE OF CONTENTS

ABSTRACT 2

TABLE OF CONTENTS 4

LIST OF FIGURES 10

LIST OF TABLES 12

ABBREVIATIONS 13

CHAPTER 1. INTRODUCTION 17

1.1. Murine embryonic development 18
1.2. Developmental control genes 21
1.3. Coordination of developmental control gene activity 22
1.4. Eukaryotic transcription 23
1.5. Basal polymerase II transcription 25
1.5.1. TFIID 25
1.6. Transcriptional activating proteins 30
1.7. DNA binding motif 31
1.7.1. Helix-turn-helix motif 31
1.7.2. The zinc finger motif 31
1.7.3. The basic domain 32
1.8. Protein dimerization domains 33
1.8.1. The leucine zipper 33
1.8.2. The Helix-loop-helix motif 34
1.9. Activation 35
1.9.1. Acidic activation domain 35
1.9.2. Other forms of activation domain 35
1.10. Regulation of transcription factor activity 36
1.10.1. Post-transcriptional regulation 36
1.10.2. Post-translational control 37
1.10.3. Heterodimer formation 38
1.11. Model system for early embryonic development 39
1.12. Embryonal carcinoma cells 39
1.13. Embryonic stem cells 40
1.14. Regulation of cellular genes during EC cell differentiation 40
1.15. The E1A protein 42
1.16. Role of E1A in transformation 46
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.17. Retinoblastoma</td>
<td>47</td>
</tr>
<tr>
<td>1.18. The cell cycle</td>
<td>51</td>
</tr>
<tr>
<td>1.19. Cell cycle regulation of E2F containing complexes</td>
<td>56</td>
</tr>
<tr>
<td>1.20. Characterisation of E2F</td>
<td>56</td>
</tr>
<tr>
<td>1.21. E2F-like activities in EC cells</td>
<td>57</td>
</tr>
<tr>
<td><strong>CHAPTER 2. MATERIALS AND METHODS</strong></td>
<td>59</td>
</tr>
<tr>
<td>2.1. GENERAL METHODS</td>
<td>60</td>
</tr>
<tr>
<td>2.1.1. Sterilisation</td>
<td>60</td>
</tr>
<tr>
<td>2.1.2. DNA extraction and precipitation</td>
<td>60</td>
</tr>
<tr>
<td>2.1.3. Determination of nucleic acid and protein concentration</td>
<td>60</td>
</tr>
<tr>
<td>2.1.4. Preparation of dialysis tubing</td>
<td>61</td>
</tr>
<tr>
<td>2.1.5. Autoradiography</td>
<td>61</td>
</tr>
<tr>
<td>2.2. DNA USED</td>
<td>61</td>
</tr>
<tr>
<td>2.2.1. Synthetic oligonucleotides</td>
<td>61</td>
</tr>
<tr>
<td>2.2.2. Plasmid DNA</td>
<td>63</td>
</tr>
<tr>
<td>2.3. RESTRICTION DIGESTION OF PLASMID DNA</td>
<td>63</td>
</tr>
<tr>
<td>2.3.1. Diagnostic digestion</td>
<td>63</td>
</tr>
<tr>
<td>2.3.2. Preparative digestion</td>
<td>64</td>
</tr>
<tr>
<td>2.4. RESTRICTION DIGESTION OF GENOMIC DNA</td>
<td>64</td>
</tr>
<tr>
<td>2.5. SUBCLONING</td>
<td>64</td>
</tr>
<tr>
<td>2.5.1. Dephosphorylation of DNA</td>
<td>64</td>
</tr>
<tr>
<td>2.5.2. Ligation of DNA</td>
<td>64</td>
</tr>
<tr>
<td>2.5.3. Preparation of competent cells</td>
<td>67</td>
</tr>
<tr>
<td>2.5.4. Transformation of competent cells with DNA</td>
<td>67</td>
</tr>
<tr>
<td>2.5.5. Identification of positive clones</td>
<td>67</td>
</tr>
<tr>
<td>2.6. PREPARATION OF PLASMID DNA</td>
<td>67</td>
</tr>
<tr>
<td>2.6.1. Mini-preparations (minipreps) of DNA</td>
<td>67</td>
</tr>
<tr>
<td>2.6.2. Large scale preparations of plasmid DNA (maxipreps)</td>
<td>68</td>
</tr>
<tr>
<td>2.7. PREPARATION OF GENOMIC DNA</td>
<td>68</td>
</tr>
<tr>
<td>2.8. SEQUENCING OF DNA</td>
<td>69</td>
</tr>
<tr>
<td>2.9. RADIOACTIVE LABELLING OF DNA AND RNA</td>
<td>69</td>
</tr>
<tr>
<td>2.9.1. End labelling using DNA polymerase 1 (Klenow)</td>
<td>69</td>
</tr>
<tr>
<td>2.9.2. End labelling using T4 polynucleotide kinase</td>
<td>71</td>
</tr>
<tr>
<td>2.9.3. Oligolabelling using DNA polymerase 1 (Klenow)</td>
<td>71</td>
</tr>
<tr>
<td>2.9.4. Preparation of α32P labelled riboprobes</td>
<td>71</td>
</tr>
<tr>
<td>2.9.5. Preparation of 35S labelled riboprobes</td>
<td>72</td>
</tr>
</tbody>
</table>
2.23. SIZE-EXCLUSION CHROMATOGRAPHY 82
2.24 PROTEIN PURIFICATION 82
2.24.1 General 82
2.24.2 Heparin Sepharose fractionation 82
2.24.3 Sequence specific affinity chromatography 83
2.25. PROTEIN PRECIPITATION 84
2.26. STAINING OF PROTEINS 84
2.26.1 Coomassie staining 84
2.26.2 Silver staining of proteins 86
2.27. PEPTIDE GENERATION 86
2.27.1 General 86
2.27.2 Cyanogen bromide cleavage 86
2.27.3 Proteolytic cleavage 87
2.27.4 Transfer of peptides onto PROBLOT 87
2.28. PROTEIN SEQUENCING 88
2.29. COMPUTING 88

CHAPTER 3 EMBRYONIC REGULATION OF DRTF1 DNA BINDING ACTIVITY 89
3.1 Introduction 90
3.2 Sequence requirement for binding of DRTF1 in EC extracts 90
3.3 Differentiation of EC cells 93
3.4 Regulation of DRTF1 binding activities during F9 EC cell differentiation 93
3.5 DRTF1 is present in ES cells and is regulated on differentiation 96
3.6. DRTF1 is present in blastocyst stage embryos 99
3.7 DRTF1 is regulated during murine embryogenesis 102
3.8 DRTF1 binding activity is tissue-dependent in the adult 107
3.9 Adult liver extract is not degraded 107
3.10 Tissue distribution of DRTF1 at 11.5 and 14.5 dpc 107
3.11 Tissue-dependent regulation of DRTF1 112
3.12 Discussion 120

CHAPTER 4 A BIOCHEMICAL DISSECTION OF THE DRTF1 COMPLEXES 122
4.1. Introduction 123
4.2. DNA binding components of DRTF1 123
4.3. DRTF1 complexes have very different molecular weights 126
4.3.1. Glycerol gradient centrifugation 126
4.3.2. Size exclusion chromatography of EC cell extracts 129
4.4. DRTF1a can be disrupted to produce DRTF1b/c binding activity 129
4.5. Determining the composition of DRTF1a 132
4.6. Discussion 135

CHAPTER 5 PURIFICATION OF DRTF1 139
5.1. Introduction 140
5.2. Identification of a suitable source from which to isolate DRTF1 140
5.3. Purification of DRTF1 141
5.4. Purification of F9 EC Whole cell extract by heparin Sepharose chromatography 141
5.5. Enrichment of DRTF1 binding activity by sequence specific DNA affinity chromatography 146
5.6. Removal of non-specific DNA binding polypeptides by use of a mutant site DNA affinity column 153
5.7. Determination of the molecular weight of the DRTF1 polypeptide species 153
5.8 Discussion 158

CHAPTER 6 ISOLATION OF cDNA FROM PEPTIDE SEQUENCE 166
6.1 Introduction 167
6.2. Precipitation of purified protein 167
6.3. Preparation of peptides for amino acid analysis 170
6.3.1 Chemical cleavage 170
6.3.2 Proteolytic cleavage 170
6.4. Proteolytic cleavage of DRTF1 crosslinked species 175
6.5. Endoprotease glu-C cleavage of the 98 kD polypeptide, and generation of sequence data 175
6.6. Endoprotease glu-C cleavage of DRTF1 178
6.7. Peptide sequence of DRTF1 178
6.8. Cloning of the sequenced protein 178
6.9. Predicted protein structure of the p46 360 cDNA 186
6.10. Discussion 186
### CHAPTER 7 SOUTHERN ANALYSIS AND RNA EXPRESSION PATTERN OF P46

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.1. Introduction</td>
<td>189</td>
</tr>
<tr>
<td>7.2. Southern analysis of the p46 gene</td>
<td>189</td>
</tr>
<tr>
<td>7.3. Northern analysis of p46 expression pattern</td>
<td>192</td>
</tr>
<tr>
<td>7.4. Quantitation of the p46 transcripts by RNAse protection</td>
<td>195</td>
</tr>
<tr>
<td>7.5. <em>in situ</em> hybridisation analysis of p46 gene expression during murine development</td>
<td>200</td>
</tr>
<tr>
<td>7.6. Discussion</td>
<td>208</td>
</tr>
</tbody>
</table>

### CHAPTER 8 GENERAL DISCUSSION

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.1. Summary of results</td>
<td>211</td>
</tr>
<tr>
<td>8.2. Regulation of the DRTF1 DNA binding activity during embryogenesis</td>
<td>211</td>
</tr>
<tr>
<td>8.3. Cell cycle regulation of E2F containing complexes</td>
<td>213</td>
</tr>
<tr>
<td>8.4. Regulation of DRTF1 during F9 EC cell differentiation</td>
<td>215</td>
</tr>
<tr>
<td>8.5. Purification and cloning of DRTF1</td>
<td>216</td>
</tr>
<tr>
<td>8.6. E2F-1</td>
<td>217</td>
</tr>
<tr>
<td>8.7. Comparison of p46 and E2F-1 DNA binding domains</td>
<td>217</td>
</tr>
<tr>
<td>8.8. pRb binding domains of E2F-1 and p46</td>
<td>220</td>
</tr>
<tr>
<td>8.9. Further experiments</td>
<td>220</td>
</tr>
<tr>
<td>8.9.1. Functional analyses of the recombinant DRTF1</td>
<td>220</td>
</tr>
<tr>
<td>8.9.2. A family of DRTF1 polypeptides ?</td>
<td>221</td>
</tr>
<tr>
<td>8.9.3. Regulation of complex formation during differentiation and development</td>
<td>221</td>
</tr>
<tr>
<td>8.9.4. DRTF1- a novel proto-oncogene ?</td>
<td>222</td>
</tr>
<tr>
<td>8.10. Perspectives.</td>
<td>223</td>
</tr>
</tbody>
</table>

**ACKNOWLEDGEMENTS**

**REFERENCES**

9
LIST OF FIGURES

CHAPTER 1
1.1. Early events in embryogenesis and model systems for differentiation 19
1.2. A model for the interaction of TFIID with promoter elements 26
1.3. Functional domains of E1A 43
1.4. The cell cycle 52

CHAPTER 3
3.1. Regulation of DRTF1 during F9 EC stem cell differentiation 94
3.2A. Embryonic stem cell differentiation 97
3.2B. Regulation of DNA binding activities on ES cell differentiation 100
3.3. DRTF1 and ECRE2-like binding activities are present in the blastocyst 103
3.4. DRTF1 is regulated during murine embryogenesis 105
3.5. Tissue dependent regulation of DRTF1 binding activity in the adult mouse 108
3.6. Adult liver extract is not degraded 110
3.7. Tissue distribution of DRTF1 at 11.5 dpc 113
3.8. Tissue distribution of DRTF1 at 14.5 dpc 115
3.9. Tissue dependent regulation of DRTF1 118

CHAPTER 4
4.1. DRTF1 complexes have common DNA binding subunits 124
4.2. DRTF1 species are composed of multicomponent protein complexes 127
4.3. S100 fractionation of DRTF1 complexes does not resolve DRTF1a from b/c 130
4.4. E1A can disrupt DRTF1a and a' complexes 133
4.5. Cyclin A and pRb are present in different complexes 136

CHAPTER 5
5.1. Alternative methods of growing F9 EC cells 142
5.2. The purification strategy 144
5.3. Heparin Sepharose fractionation of F9 EC cell extract 147
5.4. DRTF1 binding site affinity chromatography-gel retardation
5.5. DRTF1 affinity chromatography-silver staining analysis
5.6. Mutant site affinity chromatography
5.7. E2A promoter binding profile of proteins eluted from 4th affinity column
5.8. Comparison of polypeptide content of protein purified with or without the mutant site affinity column
5.9. Estimation of the contribution of DNA to crosslinked species
5.10. Comparison of the staining and crosslinking polypeptides in highly enriched DRTF1 preparations

CHAPTER 6
6.1. Coomassie stained PAGE of precipitated purified protein
6.2. Flow chart of protein recovery at each step of peptide production
6.3. Proteolytic cleavage of proteins by Cleveland reaction
6.4. Determination of the efficiency of endoprotease glu-C cleavage of crosslinked DRTF1 species
6.5. Endoprotease glu-C cleavage of the 98kD protein
6.6. Sequence of the p46 360 cDNA
6.7. Predictions from the amino acid sequence of p46 360

CHAPTER 7
7.1. Southern analysis of the p46 gene
7.2. p46 is encoded by a 3kb A+ RNA in F9 EC cells, and is expressed throughout embryogenesis
7.3. RNAse protection assay of the p46 transcript
7.4. in situ hybridisation analysis of p46 expression during murine embryogenesis

CHAPTER 8
8.1. Sequence comparisons of the DNA binding domains of E2F-1 and p46
## LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE</th>
<th>DESCRIPTION</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1.</td>
<td>RESTRICTION DIGESTION CONDITIONS</td>
<td>65</td>
</tr>
<tr>
<td>2.2.</td>
<td>ELECTROPHORESIS AND HYBRIDISATION BUFFERS</td>
<td>66</td>
</tr>
<tr>
<td>2.3.</td>
<td>SEQUENCING BUFFERS</td>
<td>70</td>
</tr>
<tr>
<td>2.4.</td>
<td>PROTEIN PURIFICATION COLUMN BUFFERS</td>
<td>85</td>
</tr>
<tr>
<td>3.1.</td>
<td>SEQUENCE REQUIREMENTS FOR DRTF1 DNA BINDING ACTIVITY</td>
<td>91</td>
</tr>
<tr>
<td>5.1.</td>
<td>CALCULATION OF THE ENRICHMENT OF DRTF1 POLYPEPTIDES DURING THE PURIFICATION</td>
<td>142</td>
</tr>
<tr>
<td>6.1.</td>
<td>COMPARISON OF CROSSLINKING AND COOMASSIE STAINING SPECIES IN THE LARGE PREPARATIONS OF POOLED PROTEIN</td>
<td>168</td>
</tr>
<tr>
<td>6.2.</td>
<td>PEPTIDE SEQUENCES DERIVED FROM THE 98kD POLYPEPTIDE AFTER ENDOPROTEASE-GLU-C CLEAVAGE</td>
<td>179</td>
</tr>
<tr>
<td>7.1.</td>
<td>QUANTITATION OF THE p46 RNAse PROTECTION ASSAY</td>
<td>198</td>
</tr>
</tbody>
</table>
ABBREVIATIONS

A absorbance; adenine
Ad adenovirus
AP activating protein
approx. approximately
ATF Activating transcription factor
ATP adenosine 5'-triphosphate
BHI brain heart infusion
Bis N,N' -methylenbisacrylamide
bp base-pairs
BSA bovine serum albumin
C cytosine
CaCl₂ calcium chloride
cAMP cyclic adenosine-3',5' monophosphate
cat chloramphenicol acetyl transferase, gene
CAT chloramphenicol acetyl transferase, protein
CIAP calf intestinal alkaline phosphatase
CO₂ carbon dioxide
col. column
cpm counts per minute
CR conserved region
CRE cAMP response element
CREB(-P) CRE binding (protein)
CREM modulator of CRE
CsCl caesium chloride
CTD carboxyl terminal domain (of RNA pol II)
Cys cysteine
dH₂O distilled, deionised water
dATP 2'-deoxyadenosine 5'-triphosphate
dCTP 2'-deoxycytidine 5'-triphosphate
dGTP 2'-deoxyguanosine 5'-triphosphate
dTTP 2' deoxythymidine 5'-triphosphate
dNTP 2' deoxynucleoside 5'-triphosphate
ddATP 2',3'-dideoxyadenosine 5'-triphosphate
ddCTP 2',3'-dideoxycytidine 5'-triphosphate
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ddGTP</td>
<td>2',3'-dideoxyguanosine 5'-triphosphate</td>
<td></td>
</tr>
<tr>
<td>ddTTP</td>
<td>2',3'-dideoxythymidine 5'-triphosphate</td>
<td></td>
</tr>
<tr>
<td>DEPC</td>
<td>diethyl pyrocarbonate</td>
<td></td>
</tr>
<tr>
<td>dm³</td>
<td>litre</td>
<td></td>
</tr>
<tr>
<td>DMEM</td>
<td>Dulbecco's modification of Eagle's medium</td>
<td></td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
<td></td>
</tr>
<tr>
<td>DNAse</td>
<td>deoxyribonuclease</td>
<td></td>
</tr>
<tr>
<td>DRTF1</td>
<td>differentiation regulated transcription factor</td>
<td></td>
</tr>
<tr>
<td>ds</td>
<td>double-stranded</td>
<td></td>
</tr>
<tr>
<td>DTT</td>
<td>dithiothreitol</td>
<td></td>
</tr>
<tr>
<td>E.coli</td>
<td><em>Escherichia coli</em></td>
<td></td>
</tr>
<tr>
<td>E2F</td>
<td>E2 binding factor</td>
<td></td>
</tr>
<tr>
<td>EB</td>
<td>embryoid body</td>
<td></td>
</tr>
<tr>
<td>EC</td>
<td>embryonal carcinoma</td>
<td></td>
</tr>
<tr>
<td>EDTA</td>
<td>diaminoethane tetraacetic acid</td>
<td></td>
</tr>
<tr>
<td>emc</td>
<td>extramacrochaetae protein</td>
<td></td>
</tr>
<tr>
<td>ES</td>
<td>embryonic stem</td>
<td></td>
</tr>
<tr>
<td>EtBr</td>
<td>ethidium bromide</td>
<td></td>
</tr>
<tr>
<td>EtOH</td>
<td>ethanol</td>
<td></td>
</tr>
<tr>
<td>g</td>
<td>gramme</td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>guanine</td>
<td></td>
</tr>
<tr>
<td>FCS</td>
<td>foetal calf serum</td>
<td></td>
</tr>
<tr>
<td>h</td>
<td>hours</td>
<td></td>
</tr>
<tr>
<td>HCl</td>
<td>hydrochloric acid</td>
<td></td>
</tr>
<tr>
<td>Hepes</td>
<td>N-2-hydroxyethylpiperazine-N'-2-ethanesulphonic acid</td>
<td></td>
</tr>
<tr>
<td>His</td>
<td>histidine</td>
<td></td>
</tr>
<tr>
<td>HLH</td>
<td>helix-loop-helix</td>
<td></td>
</tr>
<tr>
<td>hnRNA</td>
<td>heterogeneous nuclear RNA</td>
<td></td>
</tr>
<tr>
<td>HPLC</td>
<td>high pressure liquid chromatography</td>
<td></td>
</tr>
<tr>
<td>HPRT</td>
<td>hypoxanthine phosphoribosyl transferase</td>
<td></td>
</tr>
<tr>
<td>hsp</td>
<td>heat shock protein</td>
<td></td>
</tr>
<tr>
<td>HSV</td>
<td>herpes simplex virus</td>
<td></td>
</tr>
<tr>
<td>HTH</td>
<td>helix-turn-helix</td>
<td></td>
</tr>
<tr>
<td>ICM</td>
<td>inner cell mass</td>
<td></td>
</tr>
<tr>
<td>Id</td>
<td>inhibitor of differentiation</td>
<td></td>
</tr>
<tr>
<td>K</td>
<td>Potassium; 1000</td>
<td></td>
</tr>
<tr>
<td>K₃HPO₄</td>
<td>dipotassium hydrogen phosphate</td>
<td></td>
</tr>
</tbody>
</table>
kb  kilobase
KCl  potassium chloride
kD  kilodalton
KH2PO4  potassium dihydrogen orthophosphate
Lac Z  β galactosidase, gene
leu  leucine
M  molar, mutant
mA  milliamperes
MgCl2  Magnesium chloride
MHC  major histocompatibility class
min  minutes
ml  millilitres
mm  millimetres
mM  millimolar
MOPS  3-(N-morpholino) propanesulphonic acid
mRNA  messenger RNA
NaCl  sodium chloride
NaN3  sodium azide
NaOH  sodium hydroxide
NF  nuclear factor
ng  nanogrammes
NH2  amino
nm  nanometres
nmr  nuclear magnetic resonance
NP40  nonidet P40
OD  optical density
ORF  open reading frame
p46 360 360 bp cDNA generated from p46 peptide data
p46  DRTF1 polypeptide of 46kD which was sequenced
PAGE  polyacrylamide gel electrophoresis
PBS  phosphate buffered saline
pc  post coitum
PCNA  proliferating cell nuclear antigen
PCR  polymerase chain reaction
PE  parietal endoderm
Phe  phenylalanine
PIPIES  piperazine N,N′-bis (2-ethane-sulphonic acid)
PMSF  phenylmethylsulfonyl fluoride
pol  polymerase
Pro  proline
pRb  retinoblastoma, protein
Rb   retinoblastoma, gene
RNA  ribonucleic acid
RNase ribonuclease
rpm  revolutions per minute
rSRF related to SRF
SDS  sodium dodecyl sulphate
sec  second
Ser  serine
SRF  serum response factor
ss   single-stranded
SSEA stage specific embryonic antigen
SV40 simian virus 40
T    thymine
TAF  TBP associated factor
TBP  TATA binding protein
TCA  trichloroacetic acid
TE   10mM Tris (pH 8), 1mM EDTA
TF   transcription factor
tRNA transfer RNA
Tris tris (hydroxymethyl) methylamine
Tween 20 polyoxyethylene sorbitan monolaurate
Thr threonine
Tyr tyrosine
U    units
USF  upstream sequence factor
uv   ultra violet
V    volts
W    Watts
WT   wild-type
X    any amino acid
Δ    deleted/ truncated
µCi  microcuries
µg   microgrammes
µl   microlitres
µm  micrometres
µM  micromolar
CHAPTER 1

INTRODUCTION
Mammalian development may be defined as the process by which a single cell, the fertilised oocyte, divides, and its daughters differentiate to form the many cell types present in distinct tissues within the mature animal. Each cell, with the exception of cells of the B and T lineages, has the same genetic information and it is the differential expression of these genes that leads to selective protein production, and phenotypic differences between cells. Gene expression (of protein coding genes) is largely regulated by the frequency of binding of the RNA polymerase II complex to the control regions of genes, which is facilitated by the association of protein factors with the DNA upstream and downstream of the start site.

One pathway by which we may be able to learn more about the regulation of gene expression during early embryogenesis is by studying DNA binding proteins that are regulated during the differentiation of embryonic stem cells, because these factors potentially regulate the expression of genes whose products determine the phenotype of progeny cells.

1.1. Murine embryonic development.

The animal usually used for studying mammalian development is the mouse. Embryonic development of the mouse starts with fertilisation of the egg, and ends with the birth of the mouse at day 19 or 20 (reviewed by Hogan et al., 1986, Jackson, 1989, Rugh, 1990).

The fertilised egg divides as it passes along the oviduct towards the uterus, where implantation occurs on day 4.5. Cleavage of the egg starts approximately 24 hours after fertilisation, and it is at this two cell stage that the first zygotic transcription can be detected (Flach et al., 1982). Until this point, the embryo depends entirely on maternal mRNA for translation. Up to the four cell stage (approx. 36 hours post coitum), all cells of the morula are totipotent but, as cleavage proceeds and zygotic transcription increases, the developmental potency of the cells becomes restricted (Kelly, 1977). At the sixteen cell stage, compaction and associated polarisation of the cells occurs where the outer facing cells form trophectoderm, contributing to extraembryonic cell lineages, and the inner facing cells give rise to the inner cell mass, and eventually the embryo proper.

At the sixty four cell stage (figure 1.1), primitive endoderm cells start to appear on the surface of the inner cell mass which faces the fluid filled cavity of the blastocyst. Just prior to implantation, at about four days post coitum (pc), the primitive endoderm cells migrate onto the surface of the trophectoderm and form the parietal endoderm, or remain associated with the epiblast, to become visceral endoderm. At day 4.5, the blastocyst loses its protective zona, and implantation occurs through contact between the trophoblast and the uterine wall.
Figure 1.1. Early events in embryogenesis and model systems for differentiation.
The diagram depicts the early stages of murine development from 3 days pc to the start of gastrulation at day 7 pc. The flow chart below depicts the differentiation of embryonic stem cells which are derived from the inner cell mass, and the differentiation of the inner cell mass-like F9 EC cells to parietal or visceral endoderm-like cells.
Early events in embryogenesis and model systems for differentiation
At day 7, cells recruited from the ectoderm of the epiblast move through the primitive streak in the process of gastrulation to form an intermediate layer, the mesoderm, between the primitive ectoderm and the visceral endoderm (figure 1.1). The midline region of the mesoderm becomes notochord, and induces a thickening of the overlying ectoderm to form neuroectoderm, the precursor of the central nervous system and neural crest. Paraxial mesoderm condenses along the length of the notochord to form pairs of somites from about 8.5 days pc. The neural plate folds up and closes to form the neural tube, with closure beginning at the time of appearance of the somites, and completing by about day 9 pc. Differentiation of the somite blocks gives rise to three layers, the outer dermatome forms the connective tissue of the dermis, the myotome gives rise to the musculature of the body wall, and the sclerotome forms the vertebrae and ribs. At day 9, limb buds are apparent and by day 10 of development, organogenesis is well under way. Therefore, by 10.5 days pc, the body plan of the embryo has been established. The remainder of development is concerned with refining the structures laid down during early embryogenesis.

So how is such a complex process of coordinated tissue development controlled? Lessons have been learnt from the study of development in simpler organisms such as Drosophila, where extensive genetic knowledge and the ease of genetic manipulation have led to the elucidation of some general developmental principles.

1.2. Developmental control genes.

Developmental control genes of mammals can be isolated in several ways. One can look for conservation of genes important for body plan determination in more genetically malleable species such as Drosophila, or use the limited genetic knowledge of the chromosomal location of developmental mutations in mammals (Herrmann et al., 1990, Gubbay et al., 1990). Alternatively, one can randomly mutate genes by retroviral or transgene insertion, look for developmental defects and isolate the genes whose disruption leads to the defect. Also, the enhancer/promoter elements important in determining patterned gene expression can be isolated by ‘enhancer trapping’ (inserting LacZ constructs with a weak promoter element into the genome and looking for specific β-galactosidase staining patterns in the embryos) indicating insertion of the transgene into an actively transcribed gene.

The most widely used and possibly simplest approach has been to isolate mammalian homologues of genes which are important for morphogenesis in Drosophila. The best known class of these genes is the homeobox family, comprising two clusters in Drosophila (Antennapedia-C and Bithorax-C) encoding segmentation, maternal effect and homeotic genes. A region conserved between these gene classes consists of the homeobox, a 183 base pair (bp) sequence encoding a 61 amino acid domain which forms a helix-turn-
helix motif capable of binding DNA specifically (Desplan et al., 1985, reviewed by Scott et al., 1989).

There are four homeobox gene clusters in mice, where the position of a gene within the cluster appears to reflect the spatial expression of the gene, with the most 5' gene within the cluster having the most posterior boundary of expression within the central nervous system and mesoderm (Duboule and Dolle, 1989, Graham et al., 1989). Thus, the homeobox genes in mammals appear to play the same role in axial specification as the segmentation genes of Drosophila.

Other homeobox genes are not located in the clusters, for example, the paired box and POU gene families. These genes contain conserved motifs, in addition to the homeodomain, which are thought to play a role in the specificity of protein-DNA interactions (reviewed in Kessel and Gruss, 1990, and Gruss and Walther, 1992). The paired box motif encodes a second helix-turn-helix motif, which is found in several Drosophila segmentation genes and in vertebrates. In mice, genes of this class have been termed PAX, and the isolation of mutants of these PAX genes with developmental defects points to the conservation of these genes from arthropods to man (Balling et al., 1988, reviewed in Gruss and Walther, 1992).

The POU domain is present in a wide range of genes sharing a distinct homeodomain subclass and a conserved region upstream of the homeodomain, termed the POU box because of its conservation between the four original genes of this class (Pit-1, Oct1, 2 and Unc 86, Herr et al., 1988). The Oct1 and Oct2 genes were isolated because of their roles in transcription, and thus were the first mammalian homeodomain containing proteins shown to act as transcription factors. Since then, there has been a large expansion in the number of genes in this subclass, including members which are expressed in a very precise temporally and spatially restricted fashion during development (reviewed by Rosenfeld, 1991). For example, Pit-1 is expressed in three cell types of the anterior pituitary gland (Simmons et al., 1990), and has been shown to trans activate expression from the prolactin and growth hormone gene promoters (Ingraham et al., 1988, Nelson et al., 1988), although reported to only bind the growth hormone gene promoter by Castrillo et al. (1989). The pit-1 gene is localised on chromosome 16 with the dwarf locus (dw). Pit-1 is disrupted in the Jackson dwarf mouse, and point mutated in the Snell dwarf mouse, suggesting that mutation of pit-1 causes the dwarf phenotype, and that normal pit-1 gene expression is important for the production of somatotrophic and lactotrophic cell lineages in the development of the anterior pituitary gland (Li et al., 1990).

1.3. Coordination of developmental control gene activity.

Development is tightly controlled, with gradients of morphogens (defined as a molecule whose concentration directly specifies position) and cell-cell interactions playing

The induction of mesoderm by morphogens in amphibians is dose dependent, with differences in the levels of factors as small as 1.5 fold leading to different mesodermal cell fates (Green and Smith, 1990). Differences in the levels of morphogens also leads to a gradient of expression of homeobox genes thought to play a role in formation of the anterior-posterior axis of development (Ruiz i Altaba and Melton, 1989a,b).

Limb formation of birds can be disrupted by addition of exogenous retinoic acid, and the presence of retinoic acid and the complexity of regulation of its receptors within the limb suggest that retinoic acid plays an important role in limb development (reviewed by Tabin, 1991). This control may be partly mediated by induction of expression of homeobox genes, as within each cluster the most 3' homeobox genes are induced in vitro at a lower physiological concentration of retinoic acid than the more 5' genes (Papolopulu et al., 1990).

Mammalian developmental control genes therefore appear to encode regulators of transcription, and thus probably form hierarchical control circuits as in Drosophila, where the products of one developmental regulator serve to control the expression of subsequent regulators which may be expressed in a more restricted pattern, and interplay between these gene products results in the specification of individual cell phenotype.

Before describing specific examples of mechanisms of transcriptional control, I will review some aspects of the regulation of eukaryotic transcription.

1.4. Eukaryotic transcription

RNA polymerase II (pol II) directs the transcription of class II genes, which encode proteins. The heterogeneous nuclear RNA (hnRNA) products of transcription are processed into messenger RNA (mRNA) before translation. The start site of pol II transcription (+1) is determined by sequence elements within the promoter. There are two classes of initiation site-directing elements isolated to date, the TATA element, and the 'initiator element'. The TATA sequence is located about 30 bases upstream of the initiation site of many class II genes, and is bound by a protein, the TATA binding protein TBP, which is conserved between eukaryotes (Hahn et al., 1989, Horikoshi et al., 1989, Hoey et al., 1990, Hoffmann et al., 1990, Gasch et al., 1990). Other genes, including housekeeping genes (those which are ubiquitously expressed) usually lack a TATA sequence, and correct initiation of transcription of these genes is directed by proteins binding the 'initiator' element close to the start site of transcription (Smale and Baltimore, 1989).

The transcriptional activity of promoters is increased by enhancers which are located upstream or downstream of the promoter, and may act over long distances (up to 30 kb). Enhancer elements are functionally distinguishable from promoter elements as they
serve to increase the transcriptional activity of a promoter, whereas the promoter is critical for determining the start site of transcription. Promoter and enhancer elements are composed of modules, or enhansons (reviewed by Jones et al., 1988, Dynan, 1989), which are able to bind transcription factors. Modules are sometimes shared between promoter and enhancer elements; individually they are weak, but they may interact synergistically to promote transcription. Proteins may bind cooperatively to adjacent sites, or may promote active looping of the intervening DNA between closely spaced sites to allow protein association, or complex with each other via an intermediary protein.

RNA polymerase may have a high affinity for binding enhancer bound proteins, and then may scan along the DNA until it reaches the promoter to initiate transcription. Alternatively, proteins bound at the promoter and enhancer sites may interact to form large heteromeric complexes, with looping out of the intervening DNA. These complexes may be bound by the polymerase, thus positioning the polymerase close to the start site for transcription.

Experiments have been performed to address which of these two possibilities is the more likely. Enhancer sequences separated from the promoter by a protein bridge can still activate transcription although scanning of the DNA by the polymerase would be expected to be interrupted by the protein bridge (Müller H.-P et al., 1989). Additionally, a distant enhancer directs transcription equally from two closely spaced promoters (Heuchel et al., 1989), whereas preferential activation of the more proximal promoter would be expected if the polymerase scanned the DNA. Thus the looping model appears to be the most likely explanation for enhancer action.

The binding of pol II to initiate transcription is tightly controlled, not just by the presence of DNA binding proteins, but by the structure of the DNA itself. Eukaryotic DNA is highly packaged, with DNA wrapped around histone proteins to form nucleosomes, which are further packaged to form a 30nm fibre. The access of proteins such as RNA polymerase to the DNA is therefore restricted by the presence of these nucleosomal arrangements (as indicated by the relative insensitivity of DNA to DNAse digestion). Actively transcribed regions are often associated with a local removal of nucleosomal structure probably as initiation of pol II transcription is dependent on the unfolding of chromatin to allow progression. Translocation of the polymerase complex generates positive supercoils ahead and negative supercoils behind the assembly (Wu et al., 1988). Positive supercoiling would be expected to unwrap DNA from histone cores, and negative coiling behind the polymerase would promote reformation of the nucleosomal structure.

Chromatin structure is modified as an early response of cells to growth factors, with phosphorylation of histone H3 protein proposed to relax chromatin structure (Mahadevan et al., 1991), and rearrangement of the nucleosomal structure of hormone
responsive elements in response to steroid hormone stimulation (reviewed by Beato, 1989).

1.5. Basal polymerase II transcription

Transcription of class II genes is initiated by the binding of RNA pol II to the promoter in association with several general transcription factors, and the subsequent ATP-dependent formation of an open DNA structure at the start site. General transcription factors (TF) TFIID, A, B, pol II in association with TFIIF, and finally TFIIE, H and J bind sequentially to the TATA box of TATA containing promoters (Buratowski et al., 1989, Flores et al., 1992). All of these general transcription factors have now been cloned (TBP, Horikoshi et al., 1989, Hoffmann et al., 1990, Peterson et al., 1990, Kao et al., 1990, Gasch et al., 1990, Hoey et al., 1990, TFIIF, Sopta et al., 1989, Aso et al., 1992, TFIIE, Peterson et al., 1991, Ohkuma et al., 1991, Sumimoto et al., 1991, TFIIB, Ha et al., 1991, and a component of TFIIH, Lu et al., 1992) and use of these cloned factors to replace relatively crude preparations of purified proteins should yield more detailed information about their role in transcriptional initiation (reviewed by Lewin, 1990, Grenblatt, 1991, Sharp, 1991).

Much work has centred on the general transcription factor TBP since it was cloned from yeast in 1989 (Horikoshi et al., 1989).

1.5.1. TFIID.

Human TFIID, which is the name of the purified fraction including the TBP is 140 kD (measured by size exclusion chromatography) and is much larger than the molecular weight of the TBP expressed from the cloned gene (38 kD, Peterson et al., 1990). The most purified TFIID fraction from either human or Drosophila cells is composed of TBP tightly associated with at least six other polypeptides (TBP associated factors, TAFs), which may be dissociated in vitro by treatment with urea (Dynlacht et al., 1991, Pugh and Tjian, 1991, Tanese et al., 1991). In contrast, purified yeast TFIID is composed of a single polypeptide of 23-27 kD (Buratowski et al., 1988), with no co-purifying TBP associated factors (TAFs).

TATA binding proteins from different species are highly conserved within the carboxyl terminal domain which is required for DNA binding activity (Reddy and Hahn, 1991, Greenblatt, 1991), but have very different amino termini, in terms of size and amino acid content.
Figure 1.2. A model for the interaction of TFIID with promoter elements.

TFIID is composed of TBP, which binds the TATA element of TATA containing promoters, and TAFs which associate tightly with TBP. TAFs have been proposed to form the link between the TBP and other transcriptional activator proteins, such as Spl shown here. Different TAFs are thought to mediate interaction with different activators. In the case of the TATA-less promoter, TBP is thought to be tethered to the initiation region by interaction with a tethering factor distinct from the TAFs.
**TATA containing promoter**

- **TFIID complex**
- **TBP**
- **RNApol II and basal factors**

**TATA-less promoter**

- **TFIID complex**
- **TBP**
- **RNApol II and basal factors**

Adapted from Pugh and Tjian 1990, 1991
Recombinant TATA binding protein is sufficient to mediate basal transcription from TATA or initiator type promoters. TBPs from *Drosophila* and yeast are sufficiently conserved in the carboxyl terminal domain to functionally replace human TBP for DNA binding and basal transcription in a human transcription system reconstituted from purified proteins (Horikoshi *et al.*, 1990, Hoffmann *et al.*, 1990). Human TBP can also function efficiently in a yeast transcription system (Kelleher *et al.*, 1992).

Purified TFIID can support elevated levels of transcription in the presence of certain activating proteins *in vitro*. In contrast, recombinant TBP can only support elevated transcription by some transcriptional activators in the presence of the activating protein and purified TFIID, in which the endogenous TBP has been heat inactivated (Pugh and Tjian, 1990, Peterson *et al.*, 1990, Tanese *et al.*, 1991). This result suggests that factors which co-purify with TBP in the TFIID fraction (TAFs or coactivator proteins) are necessary for elevated transcription in the presence of some activator proteins (see figure 1.2). Deletion of the amino terminus of TBP abolishes the enhanced transcriptional activity, suggesting that TAFs promote transcriptional enhancement by certain activators through the amino terminus of the TATA binding protein (Pugh and Tjian, 1990, Peterson *et al.*, 1990, Hoey *et al.*, 1990).

Mammalian transcription reactions reconstituted with recombinant TBP from *Drosophila* could not support elevated transcription on addition of activator (Hoey *et al.*, 1990). In contrast, the same activator can promote transcription in *Drosophila* cell extracts (Pugh and Tjian, 1990). One interpretation of these experiments is that not only are there specific species differences in the structure of TATA binding proteins, but there are also species differences in the coactivator proteins with which these TBPs interact.

Recent work indicates that the carboxyl terminus of yeast TBP is sufficient to mediate transcriptional activation by the acidic GAL4 and VP16 proteins in HeLa cell transcriptional systems, indicating that the mechanism for acidic activation of transcription is not species specific, and is dependent on the carboxyl, rather than the amino terminus of TBP (Kelleher *et al.*, 1992).

RNA polymerase II is composed of two large subunits, with an assortment of smaller subunits, three of which are likely to be shared with the other polymerases (Woychik *et al.*, 1990). The largest subunits of the polymerase share homology with the large subunits of bacterial polymerases (Allison *et al.*, 1985), which form the core polymerase that binds DNA non-specifically and initiates a local DNA unwinding event. Specificity is conferred on the bacterial polymerase by the presence of another subunit, sigma, which directs the polymerase to the initiation site by reducing the affinity of the polymerase complex for other sites (reviewed by Lewin, 1987). None of the subunits of eukaryotic RNA polymerase II show homology to the sigma factor, and it has been suggested that a component of the basal transcriptional apparatus provides the DNA
binding specificity for the polymerase.

The DNA binding specificity of pol II may be provided by the TATA binding protein as there is a block of sigma homology within the carboxyl terminal domain of the protein which is important for DNA binding. Moreover, this domain is highly conserved in several species, which would explain the ability of yeast and Drosophila TFIID to support basal transcription in a reconstituted mammalian transcription system. The TATA binding proteins are also required for in vitro transcription from TATA-less promoters (Pugh and Tjian, 1990, Smale et al., 1990, Colgan and Manley, 1992). The possible interaction of the polymerase with the TATA binding protein may explain this requirement for TFIID.

Two proteins that bind the initiator element have been characterised (Roy et al., 1991, Seto et al., 1991, Shi et al., 1991), and one of these, TFII-I, has been shown to bind cooperatively with an upstream activator factor USF, which must dimerise to bind DNA. One possibility for the targeting of TFIID or the TATA binding protein to the promoter region of TATA-less promoters may be by interaction of initiator bound TFII-I with TFIID. Alternatively as suggested by Pugh and Tjian, (1990, 1991), TATA binding protein may be 'tethered' at the promoter by complexing of one of the TAF proteins associated with TBP to a DNA bound activator protein (see figure 1.2).

Any model, whereby TFIID functions to direct the polymerase to the transcriptional start site, is dependent on interaction between the polymerase and a component of the TFIID activity. Recently, evidence has increased for such an association between the TATA binding protein of TFIID and RNA polymerase II. A series of experiments by Usheva et al. (1992) used chromatography of cell extracts over columns bearing a peptide of the heptapeptide repeat, Tyr Ser Pro Thr Ser Pro Ser, which is repeated fifty two times in the carboxyl region of the murine polymerase (CTD). The proteins that bound included TFIID, and, more specifically, the TATA binding protein bound to the CTD peptide. By use of columns bearing recombinant human TBP and fractions enriched in pol II, they also showed that only the unphosphorylated form of polymerase specifically bound to TBP.

The CTD of pol II had previously been shown to be the site of phosphorylation of the large subunit of the polymerase. Phosphorylation of the CTD mediated the change in the polymerase from the non phosphorylated IIa form, associated with the preinitiation complex, to the IIb form which is associated with transcriptional elongation (Lu et al., 1991). Various kinases have been proposed to phosphorylate RNA polymerase, including a cell cycle regulatory kinase p34^cdc2 (Cisek and Corden, 1989), and a DNA dependent kinase (Peterson et al., 1992). Evidence for the role of a DNA dependent kinase in the phosphorylation of polymerase II is based on a very artificial system in which the CTD of pol II is not phosphorylated when in solution in the absence of DNA, but is phosphorylated when bound to DNA by fusion to a DNA binding protein. Better evidence for the
dependence of pol II phosphorylation on the presence of DNA, is that phosphorylation of pol II only occurs in the presence of a functional promoter, and occurs to a lesser degree with a mutant promoter bearing mutations within the TATA and initiator elements (Arias et al., 1991). These findings are consistent with a requirement for the polymerase to be bound to the DNA in association with TBP for phosphorylation to occur, and suggest that a kinase would need to be targeted to the DNA in order to achieve phosphorylation of the DNA bound polymerase. Recently it has been shown that the TFIIH component of the general transcriptional apparatus has kinase activity, and that this kinase may phosphorylate the CTD of pol II (Lu et al., 1992). The kinase activity is greatly increased by factors promoting the assembly of the transcriptional complex, and only occurs in the presence of TATA or initiator DNA elements.

These results are all consistent with a model where underphosphorylated pol II is recruited to the promoter by interaction with the TBP, and then, after phosphorylation by TFIIH, the affinity of the CTD of the polymerase for the TATA binding protein is reduced, and the polymerase is able to disengage from the initiation complex to transcribe the DNA.

It seems likely that the CTD of polymerase II binds to the carboxyl terminus of the TATA binding protein because of the non-essential nature of the amino terminal region of TBP for viability in yeast (Gill and Tjian, 1991, Reddy and Hahn, 1991, Cormack et al., 1991). The carboxyl terminus of TBPs from different species is highly conserved, and this high homology may explain the ability of the carboxyl terminal regions of TBPs from different species to function in heterologous in vitro basal transcription reconstitution experiments (Horikoshi et al., 1990, Peterson et al., 1990), and for VP16 mediated transcriptional activation (Kelleher et al., 1992). Subtle differences between the carboxyl regions of TBPs may however explain the inability of TBPs from humans or Drosophila to functionally replace yeast TBP in vivo (Gill and Tjian, 1991, Cormack et al., 1991), possibly because of an altered binding affinity for the polymerase which would not be evident in the in vitro reactions where all proteins are in excess. Alternatively, as TBP has also been shown to be required for transcription by polymerases I and III in yeast cells (Cormack and Struhl, 1992, Schultz et al., 1992), sequence variations in the TBP proteins of these other species may prevent essential polymerase I and III transcription in vivo, which would not be assayed by the in vitro pol II assays.

1.6. Transcriptional activating proteins

Proteins which bind to specific DNA sequences in the promoter or enhancer region of the DNA and promote transcription have been termed transcription factors. Many have now been cloned, and many are modular in nature, that is, they can be physically separated into domains that promote binding to a specific DNA sequence, domains responsible for
activation of transcription, and domains which are necessary for interaction with other proteins. These various properties can largely be conferred on a heterologous protein by generation of a chimaeric protein, suggesting that these functional domains are truly modular in nature.

I shall next review transcriptional regulators in terms of the properties of the various domains.

1.7. DNA binding motifs.

1.7.1. Helix-turn-helix motif.

The helix-turn-helix (HTH) motif found in the homeodomain has homology with the sequence specific DNA binding yeast proteins MATα1 and MATα2 (Shepherd et al., 1984), and the helix-turn-helix motif of many bacterial proteins (Pabo and Sauer, 1984). Indeed, many Drosophila homeodomain binding proteins have been shown to bind DNA in vitro. From studies on the structure of the engrailed homeodomain cocrystallised with its preferred DNA binding site (Kissinger et al., 1990), the general features of the structure of homeodomain-DNA complexes have been established. The 60 amino acid homeodomain is composed of three helices. The third helix is the DNA recognition helix, which has a hydrophilic region which contacts DNA in the major groove, and a hydrophobic face which packs against helices 1 and 2 lying at right angles to it. Helices 1 and 2 run antiparallel to each other, and helix 2 contacts the DNA located to the 3' side of the DNA sequence recognised by helix 3 whereas helix 1 does not contact the DNA. The amino terminal of the homeodomain forms an extended arm which fits into the minor groove of the DNA 5' to the helix 3 recognition site.

The importance of helix 3 for DNA recognition, together with the penetration of the fifth residue of the homeodomain into the minor groove, confirmed the solution nmr spectroscopic results obtained by Otting et al. (1990) with the Antennapedia homeodomain. Indeed, several residues in helix 3 are conserved between all homeodomain proteins and play a role in DNA binding (reviewed by Scott et al., 1989), while others (such as position 9 of helix 3) determine the sequence specificity of binding (Desplan et al., 1988, Hanes and Brent, 1989, Treisman et al., 1989).

POU proteins have an additional DNA binding motif situated 5' to the homeodomain which is necessary for DNA binding. All POU proteins have a conserved ninth residue of the 3rd helix, therefore the POU domain largely confers the sequence specificity of DNA binding (reviewed by Ruvkun and Finney, 1991).

1.7.2. The Zinc Finger Motif.

The zinc finger was first characterised in the polymerase III transcription factor, TFIIIA (Miller et al., 1985). TFIIIA has nine copies of a 28-30 amino acid motif of
consensus Cys X, Cys X, Phe X, Leu X, His X, His. The two cysteine and histidine residues are believed to coordinate a zinc ion to give a stable metalloprotein structure which can interact with DNA. Other proteins of this type have subsequently been cloned, and include Sp1 (Kadonaga et al., 1987), Krox 20 (Chavrier et al., 1988), PRDII-BF1 (Fan and Maniatis, 1990), and the Drosophila proteins Kruppel (Rosenberg et al., 1986) and Tramtrack (Harrison and Travers, 1990). The solution structure of a zinc finger from the Xenopus XFIN protein showed a structure rather like a right hand, with the zinc and carboxyl end of the motif located at the base of the hand, and the thumb composed of the amino terminal region of the motif (Lee et al., 1989). The thumb is an antiparallel $\beta$ sheet structure, and the rest of the hand is helical. There is a high density of basic amino acids on the exposed surface of the helix, which may, with the polar side chains on the surface of the helix, contact the phosphate backbone or make direct base contacts with the DNA. From in vitro mutagenesis studies on Krox 20 (Nardelli et al., 1991), the fingers have been proposed to contact three nucleotides, and two residues of the finger have been identified as critical determinants of DNA interaction. The fingers bind the major groove of DNA, and alternate fingers bind on opposite sides of the helix (Klug and Rhodes, 1987).

Another type of zinc finger is present in the steroid receptor DNA binding domain (Rajavashisth et al., 1989, Luisi et al., 1991). Zinc is tetrahedrally coordinated to four cysteine residues, and only two fingers are present in the DNA binding domain. The glucocorticoid and related steroid/nuclear receptors bind DNA as homodimers, and recognise palindromic sequences on one surface of the DNA, with one finger from each protein being used as a dimerization interface (reviewed by Luisi, 1992). The amino terminal finger contacts the major groove of the DNA, binding one half of the binding site in a sequence specific manner, while the second helix contacts the phosphate backbone of the flanking sequence (Chalepakis et al., 1988).

The recently solved solution structure of the yeast factor GAL4 DNA binding domain (Kralulis et al., 1992, Baleja et al., 1992) defines yet another class of zinc binding DNA binding protein. Two zinc ions are bound by six cysteine residues, and dimers of GAL4 bind target sites one and a half helical turns apart on opposite surfaces of the DNA.

1.7.3. The basic domain.

Other transcription factors use different mechanisms for binding to DNA. Many proteins which require hetero/homodimerisation to bind DNA use an ill-defined basic motif for DNA interaction. The basic domain is found adjacent to interaction domains of the helix-loop-helix (HLH) and leucine zipper types (see section 1.8). Both the oligomerization and basic domains are essential for DNA binding, and DNA binding specificity is conferred by both domains (Hope and Struhl, 1986, Kouzarides and Ziff, 1988, 1989, Sassone-Corsi et al., 1988, Landschulz et al., 1988a, Nakabeppu and Nathans, 32
Models have been proposed for the interaction of the basic domain with DNA. The basic region is hydrophilic, and α helical, and proposed to project into the major groove of the DNA (O'Neil et al., 1990, Anthony-Cahill et al., 1992). The basic region has been proposed to act as an extension of the α helical HLH or leucine zipper regions of the protein, which position the basic domain to interact with DNA.

Other transcription factors use novel motifs to bind DNA, which do not fall into the above categories. Examples include AP2 (Williams et al., 1988), CTF/NF1 (Santoro et al., 1988), NFκB (Kieran et al., 1990, Bours et al., 1990, Ghosh et al., 1990) and SRF (Norman et al., 1988).

1.8. Protein dimerization domains

As indicated in section 1.7, many transcription factors need to dimerise in order to bind DNA. The two major types of dimerisation interface used by transcription factors are the HLH and leucine zipper domains. Both of these motifs interact through α helical structures as do many of the other known dimerisation motifs (ankyrin repeats, reviewed by Blank et al., 1992, HNF1, SRF and GSRF interaction domains Nicosia et al., 1990, Norman et al., 1988, Pollock and Treisman, 1991).

1.8.1. The leucine zipper

The leucine zipper class of transcription factors were 'discovered' by Landschulz et al. (1988a), by virtue of a conserved helical region of approximately 30 residues containing a periodic repeat of leucines every seven residues. The family included the yeast transcription factor, GCN4 (Hope and Struhl, 1986), the CCAAT box/enhancer binding protein C/EBP (Landschulz et al., 1988b), and the c-Jun and c-Fos nuclear oncoproteins. Landschulz proposed a leucine zipper motif by which the proteins may dimerise, as in some cases the proteins were known to require dimerisation for DNA binding through a basic domain. The leucine rich segments were proposed to coil into α helical structures, with the leucine residues lined up along one face. Two such coiled structures were proposed to interact in an antiparallel manner, with the leucine 'teeth' projecting from the coils interdigitating to form a stable dimeric complex. Subsequent structural studies performed by O'Shea et al. (1989a) with a peptide of the leucine rich region of GCN4 showed that many of the predictions of the leucine zipper were correct. The leucine rich helices were however packed in a parallel coiled coil arrangement, with two right hand helices coiling around each other, the interaction pattern repeated every seventh (leucine) residue, and incorporating a hydrophobic (4,3) repeat within the heptad repeat. In this model, the leucines do not interdigitate to form a zipper motif, but the leucine from one
helix interacts with the hydrophobic residue of the other helix.

Some leucine zipper proteins (as they continue to be named because of the abundance of variant coiled coil interactions in biology), cannot homodimerise efficiently (e.g. c-Fos, Rauscher et al., 1988, Smeal et al., 1989) because the complex is destabilised by acidic residues positioned adjacent to the hydrophobic interaction face of the coiled coil (O'Shea et al., 1992). DNA binding of such proteins is only achieved by heterodimer formation with other members of the same class (e.g. c-Jun, reviewed by Busch and Sassone-Corsi, 1990), as such binding of the API (Jun/Fos dimer) complex is thermodynamically favoured (O'Shea et al., 1989b). Heterodimerisation of the API family of DNA binding proteins is thus regulated by the nature of the amino acids within the leucine zipper region.

Other transcription factors which bind a DNA site related to the API site are the ATF site binding activities (Lin and Green, 1988), which also dimerise through a leucine zipper arrangement and bind DNA with the associated basic motif (Hoeffler et al., 1988, Maekawa et al., 1989, Hai et al., 1989, Gonzalez et al., 1989). Selective heterodimerisation also occurs within this family of proteins (reviewed by Ziff, 1990), and selected proteins from the API family of proteins can heterodimerise with ATF family members. The DNA binding specificity of the resulting heteromeric complexes may be for either an API or ATF site, depending on the binding partners involved, for example, replacement of c-Fos by CREB as a dimerisation partner for c-Jun changes the DNA binding specificity from an API site to an ATF site (Benbrook and Jones, 1990).

1.8.2. The helix-loop-helix motif

The second major form of interaction domain employed by transcription factors is the helix-loop-helix (HLH) domain, which lies adjacent to a basic domain. Dimerisation is mediated by the HLH domains, and both basic regions are required for DNA binding of the dimeric proteins. This motif was first characterised in the ubiquitous E2 gene products, E12 and E47 (Murre et al., 1989a). The motif is formed from two α helical domains separated by a short loop. In addition to the ubiquitous proteins, the HLH domain is present in genes involved in specifying cell lineage, such as the muscle differentiation regulatory genes (Tapscott et al., 1988, Edmonson and Olson, 1989, Rhodes and Konieczny, 1989, Braun et al., 1989, Wright et al., 1989, Sassoon et al., 1989, Weintraub et al., 1991), and genes involved in Drosophila sensory neurogenesis (reviewed in Skeath and Carroll, 1991). The basic-HLH protein family also show specificity in heterodimerisation. E12 homodimers do not bind DNA because of an inhibitory domain (Sun and Baltimore, 1991), but heterodimerise efficiently with E47 to bind DNA (Murre et al., 1989b). The E47 gene product homodimerises only in mature B cells (Schlissel et al., 1991) and, in all other cell types, E47 heterodimerises with other HLH proteins.
A third class of proteins which require dimerisation for DNA binding via a basic
domain is the myc class of transcription factors, which have both leucine zipper and HLH
interaction domains (reviewed by Prendergast and Ziff, 1992). The leucine zipper and
HLH motifs of these proteins are both necessary for dimerisation specificity and
stabilisation of the interaction, conferring extra levels of regulation on dimerisation

1.9. Activation

Several types of activation domain have been identified, with perhaps the most
common being the acidic activation domain.

1.9.1. Acidic activation domain

The acidic activation domain is conserved between yeast and higher eukaryotes
1988, Preston et al., 1988). It was originally thought that the presence of a domain of
acidic residues was sufficient to confer transcriptional activation function on a protein.
However, an experiment performed by Giniger and Ptashne, (1987) dispelled this notion,
when they showed that activation (in yeast) is dependent on the amphipathic helical nature
of the acidic region, as the same acidic residues present in a scrambled order in a
chimaeric protein did not activate transcription to the same extent.

Transcriptional activation by acidic domains has been proposed to be mediated by
direct contact with the basal transcriptional apparatus. Support for this idea has come from
the change in the interaction of TFIID with the TATA region of the promoter in the
presence of the acidic yeast transcriptional activator GAL4 (Horikoshi et al., 1988a), and
in the presence of ATF (Horikoshi et al., 1988b). The acidic trans activator VP16 protein
from Herpes Simplex virus will also bind TFIID, and TFIIB in vitro (Stringer et al., 1990,
Lin Y.-S et al., 1991). Lin et al. have provided evidence that it is the binding of TFIIB to
the initiation complex which is the rate limiting step in transcriptional initiation, and it is
this step which is facilitated by acidic activator proteins.

1.9.2. Other forms of activation domain

Other DNA binding transcription factors do not have acidic activation domains,
and for these proteins the activation domain is characterised by an unusually high density
of certain amino acids, for example glutamine (Sp1, Courey and Tjian, 1988, Courey et al.,
1989, and CREB and CREM (Gonzalez et al., 1989, Foulkes et al., 1992)), or proline
(CTF, Mermod et al., 1989). It is unclear how these motifs function to increase
transcriptional activation, but from the experiments discussed in section 1.5.1, it is likely
that these factors contact the transcriptional apparatus through TBP associated factors.
1.10. Regulation of transcription factor activity

Many transcription factors are present in only a subset of cell types or during a particular period of development. This regulation may be caused by cell type specific transcription of transcription factors (see for example expression of GATA-1, Tsai et al., 1989, C/EBP, Birkenmeier et al., 1989), or by control mediated at later stages of protein synthesis.

1.10.1. Post-transcriptional regulation

There are many cases of alternative splicing of transcription factors where it is not clear what effect, if any, differences have on protein activity (Santoro et al., 1988, Schöler et al., 1990, Hatzopoulos et al., 1990). Other cases, such as in the E12 basic-HLH protein, DNA binding activity of the homodimeric protein is improved by removal of an inhibitory sequence (Sun and Baltimore, 1991), or DNA binding activity is conferred on a protein which was unable to bind DNA by addition of two amino acids to the homeodomain (Treacy et al., 1992). In both of these cases, there is no tissue dependent or developmental regulation of splice choice, and production of alternative proteins serves to increase the potential for regulation within the system.

In other cases, splicing is regulated in a tissue dependent fashion. For example, the level of the truncated TFE3 protein, which has reduced transcriptional activation potential, comprises 2 and 20% of the total TFE3 complement in lymphoid and liver cells respectively (Roman et al., 1991), which is interesting as the full length protein activates transcription of a lymphoid specific gene.

The cyclic AMP responsive CREM transcription factor is alternatively spliced (Foulkes et al., 1991), with the possibility of generating three different products, two of which, β and γ, utilise different DNA binding domains than that used by α. Both DNA binding domains however are of the basic-leucine zipper class and the proteins bind the ATF site with the same specificity as CREB. The CREM gene product resembles the positively acting transcription factor CREB, except that the CREM isoforms lack the glutamine rich activation domains of CREB, and thus may block transcriptional activation mediated by CREB. CREB is uniformly expressed, whereas CREM is differentially expressed, and is absent from the somatotroph cells of the anterior pituitary. An additional splicing product of the CREM gene is detected in the spermatid lineage of the adult testis. This form (CREMδ) has two glutamine rich activation domains which confer a transcriptional activation function (Foulkes et al., 1992). The CREM gene therefore represents a gene which gives rise to alternatively spliced transcripts with opposing functions, and this splicing is controlled in a tissue dependent and developmentally regulated manner.

Splicing of the fosB gene product produces equal quantities of ΔFosB and FosB in
serum stimulated cells, or electrically stimulated brain or regenerating liver cells of the adult mouse (Nakabeppu and Nathans, 1991). ΔFosB acts as a dominant repressor of FosB activity assayed on a reporter with a single binding site in F9 EC cells. This is probably due to the removal of part of the proposed proline rich activation domain, leaving an intact DNA binding and heterodimerisation domain for interaction with c-Jun to form inactive API site DNA binding complexes. In contrast, other workers do not report a reduction in trans activation of a reporter bearing five copies of the target sequence in 3T3 cells in the presence of ΔFosB (Mumberg et al., 1991). This discrepancy may be caused by the use of F9 cells (which essentially have no API activity) in experiments performed by Nakabeppu and Nathans. In this case, transfected ΔFosB is likely to have fewer potential antagonists than in 3T3 cells. Alternatively, several binding sites may be necessary for ΔFosB to act as a transcriptional inducer. Interestingly, ΔFosB is expressed thirty minutes after FosB on serum stimulation of 3T3 cells, and inhibits the ability of FosB to repress its own transcription. Alternative splicing therefore allows the production of two antagonistically acting proteins from the same serum stimulated gene.

1.10.2. Post-translational control.

Post translational control is exerted on many transcription factors, with phosphorylation perhaps being the most common regulatory mechanism affecting DNA binding activity (Montminy and Bilezikjian, 1987, Yamamoto et al., 1988, Raychaudhuri et al., 1989, Bagchi et al., 1989, Manak et al., 1990, Manak and Prywes, 1991, Janknecht et al., 1992) and transcriptional activity of transcription factors (Tanaka and Herr, 1990, Jackson et al., 1990). The regulation of transcription factor activity by phosphorylation allows coordination of signal pathways with the induction of gene expression (see Hanks et al., 1988 for a review of kinases and targets).

Phosphorylation may also play an important role in determining the cellular localisation of transcription factors. The yeast transcription factor SWI5 is located in the cytoplasm during most of the cell cycle, and enters the nucleus to promote transcription of its target genes only in G1 (Moll et al., 1991). Cytoplasmic retention of the protein is thought to be mediated by phosphorylation of three serine residues within the basic nuclear localising region of the protein, which directly regulates the efficiency of the nuclear localisation signal. Partial dephosphorylation of these residues during G1 allows SWI5 to enter the nucleus.

There is also evidence for glycosylation of polymerase II transcription factors: Sp1, CTF, Fos, Jun (Jackson and Tjian, 1988), the liver specific homeodomain like protein, HNF1 (Lichtsteiner and Schibler, 1989, Frain et al., 1989), and the p75 component of the pancreatic PTF1 (Sommer et al., 1991). The role of glycosylation remains unclear, and may serve to stabilise the protein, modulate kinase activity by masking potential serine/threonine recognition sites, or act as a recognition site for a component of the
transcriptional apparatus. Alternatively, glycosylation may promote nuclear localisation of the transcription factor. Experiments investigating the nuclear transport of the pancreatic PTF1 factor in the frog oocyte experimental system indicate that nuclear localisation of the DNA binding activity is dependent on the association with a 75 kD glycosylated non-DNA binding component (Sommer et al., 1991). The glycosylated residues may serve a targeting function to bind to nuclear receptors (reviewed by Silver, 1991).

1.10.3. Heterodimer formation

In addition to the regulation of DNA binding specificity by heterodimerisation, the activity of transcription factors can be modulated by dimerisation. As discussed above, complexes of transcription factors in which one partner lacks an activation domain but is still able to bind DNA reduce the expression of a reporter gene, compared with the response to a dimeric activating protein. In addition to alternative splicing of gene products, there are several genes which encode proteins that are able to dimerise, but which lack DNA binding domains. These proteins act as dominant suppressors of the activity with which they dimerise, as both partners need an intact DNA binding domain for efficient DNA binding of the complex and transcriptional activation. There are two examples of HLH type proteins which lack the basic domain required for DNA binding activity, Id (inhibitor of differentiation) in mammals (Benezra et al., 1990), and extramacrochaetae (emc) in Drosophila (Ellis et al., 1990, Garrel and Modolell, 1990). Id can heterodimerise with the ubiquitous E12 or E47 proteins, or the muscle specific MyoD protein (reviewed by Weintraub, 1991), and cotransfected Id can reduce the MyoD dependent transactivation of a muscle specific promoter in the C3H10T(1/2) cell line. Levels of MyoD and the E2 gene products are equivalent in myoblasts and differentiated myotubes, so the regulated expression of muscle specific promoters during myogenesis may be explained by the decrease in Id expression during myogenesis. Likewise emc acts as a negative regulator, but of a neurogenic gene product, achaete scute. Overexpression of emc reduces the number of sensory organs which are formed during Drosophila development.

NFkB is another transcription factor which is regulated by dimerisation, not only by the dimerisation of subunits for DNA binding, but also by the association with inhibitory proteins which retain the DNA binding subunits in the cytoplasm of the cell. NFkB is a heterodimeric complex composed of p50 and p65 subunits, both of which can bind DNA (Blank et al., 1991, Urban et al., 1991). Homodimeric p50 complexes have also been isolated, called KBF1 (Kieran et al., 1990), which have an altered DNA binding specificity compared with the heteromeric NFkB complex. KBF1 is a transcriptional inhibitor in vitro (Schmitz and Baueule, 1991), probably as the p50 subunit lacks an obvious transcriptional activating domain (Kieran et al., 1990, Ghosh et al., 1990), in contrast with the proline rich region of p65 (Nolan et al., 1991). Both p65 and p50 have
homology to the rel proto-oncogene, and the rel homology domain of the p65 subunit of NFκB is recognised by the IκB family of inhibitory proteins, which sequester the p65 subunit in the cytoplasm (Baeuerle and Baltimore, 1989, Urban and Baeuerle, 1990). This inhibitory interaction can be prevented by phosphorylation of IκB (Baeuerle and Baltimore, 1988), which may be mediated by a variety of signal transduction pathways to release p65 and allow nuclear localisation of the factor.

The p50 subunit of NFκB may also be sequestered in the cytoplasm, and prevented from DNA binding by masking of the nuclear localisation signal of the p50 subunit by the carboxyl terminal region of the precursor protein from which p50 is cleaved (Henkel et al., 1992). Alternatively, p50 nuclear translocation could be inhibited *in trans* by complexing with IκBγ, a protein which is identical to the carboxyl terminus of the p50 precursor (Inoue et al., 1992). The p50 precursor protein and cloned IκB proteins have multiple ankyrin repeat motifs within the domains required for interaction with p50 and p65 respectively (reviewed by Blank et al., 1992). Similar ankyrin repeat motifs have recently been shown to mediate protein-protein interaction in another transcription factor GABP (Thompson et al., 1991, LaMarco et al., 1991). Therefore both the p50 and p65 subunits of the NFκB activity can be prevented from DNA binding and retained in the cytoplasm by protein-protein interaction.

1.11. Model systems for early embryonic development.

The study of transcriptional control mechanisms is dependent on the availability of fairly large quantities of protein, as transcription factors are a small component of the total cellular protein. Because of the small size of embryos, the time taken to dissect them, and the heterogeneity of cell types in post-implantation embryos, use has been made of an *in vitro* culture system to investigate changes in transcription occurring during early murine embryogenesis. This culture system utilises the transformed phenotype of embryos explanted to a non-uterine location.

1.12. Embryonal carcinoma cells

The most widely used of these culture systems, the F9 embryonal carcinoma cell (EC) stem cell system, was derived from a 6 day old mouse embryo explanted to a testicular site (Bernstine et al., 1973). These cells, by virtue of various surface markers, are thought to resemble the cells of the inner cell mass of the embryo (Solter and Knowles, 1978, see figure 1.1). They can be induced to differentiate into primitive endoderm by treatment with retinoic acid (Strickland and Mahdavi, 1978), and to differentiate further into parietal endoderm (PE) by culturing as monolayers in the presence of dibutyryl-adenosine 3'5' monophosphate (cAMP) (Strickland et al., 1980). Alternatively, differentiation into visceral endoderm (VE) occurs when primitive endoderm cells are
grown as aggregates (Hogan et al., 1983). Importantly, EC cells are transformed and lose their transformed phenotype on differentiation, therefore changes occurring on differentiation cannot be ascribed to the differentiation process alone.

1.13 Embryonic stem cells.

The embryonic stem (ES) cell culture system more closely represents early differentiation events in the mouse (Evans and Kaufman, 1981, Martin, 1981). These cells are derived from the inner cell mass of the blastocyst, and can be manipulated in vitro, but retain totipotency when reintroduced into blastocysts (Bradley et al., 1984).

ES cells are cultured in the presence of differentiation inhibiting factors (reviewed by Burgess, 1989), or on STO feeder cells (Martin and Evans, 1975) which maintain the cells in an undifferentiated state. Differentiation of these cells can be achieved by aggregation to form cystic embryoid bodies after removal of the differentiation inhibiting agents. Numerous cell types characteristic of all three germ layers of the early post-implantation embryo are present in these differentiated bodies (Doetschman et al., 1985, Gossler et al., 1986) reflecting the totipotent nature of the parent cells. The differentiation of these cells cannot be controlled to the extent of the F9 EC differentiation, and, at any stage of differentiation, a wide variety of cell types are present, in contrast to the more homogeneous nature of the EC cell differentiation.

1.14. Regulation of cellular genes during EC cell differentiation

Several genes are induced during the early differentiation events in the embryo, and this regulation is mimicked during EC cell differentiation. Various structural proteins, such as laminin B, endo A, B, and type IV collagen are induced on differentiation (Wang and Gudas, 1983, Duprey et al., 1985, Singer et al., 1986), as are several surface proteins including the insulin receptor, and a component of the major histocompatibility complex (MHC) (Heath et al., 1981, Rosenthal et al., 1984). The a-feto-protein gene is induced during formation of the visceral yolk sac in the embryo, and a similar regulation of expression occurs on EC cell differentiation (Tilghman et al., 1985). In contrast, the heat shock protein, hsp70 is expressed at high levels in EC cells and in the ectoderm of 8 day mouse embryos (Bensaude and Morange, 1983), and is down regulated on differentiation (Mezger et al., 1989). The cellular proto-oncogene, c-myc is also down regulated during differentiation of EC cells, and this regulation is largely post-transcriptional (Dony et al., 1985, Griep and DeLuca, 1986).

During differentiation, gene expression is controlled partially by regulation of the abundance of transcription factors, or transcription factor-like proteins. There is an up-regulation of a murine equivalent of API, (Kryske et al., 1987), differential regulation of ATF activities (Tassios and La Thangue, 1990), differential regulation of homeobox
containing genes (Colberg-Poley et al., 1985, Schöler et al., 1989), differential regulation of zinc-finger genes (Passananti et al., 1989, Hosler et al., 1989), and down regulation of TFIIIB (White et al., 1989) during EC cell differentiation.

In addition to regulation of these cellular genes, there is also a marked difference in the ability of several viruses to productively infect EC cells or embryos, compared to their differentiated derivatives (Swartzendruber and Lehman, 1974, Jaenisch et al., 1975, Cheng and Praskier, 1982). This block to viral expression can occur at the level of viral RNA splicing in adenoviral (Cheng and Praskier, 1982) or SV40 infection (Segal and Khoury, 1979, Segal et al., 1979). Transcriptional control can be exerted on viral enhancer elements, such as Simian Virus 40 (SV40), polyoma, and Moloney murine leukaemia viruses which support higher levels of transcription on differentiation of EC cells (Linney et al., 1984, Sleigh and Lockett, 1985, Gorman et al., 1985, Cremisi and Babinett, 1986). Regulation of these viral enhancer sequences may be caused by a lack of positively acting transcription factors (Kryske et al., 1987, Speck and Baltimore, 1987, Tsukiyama et al., 1989), or the presence of repressive activities (Gorman et al., 1985, Wasylyk et al., 1988, Tsukiyama et al., 1989, Loh et al., 1990, Prince and Rigby, 1991) in EC cells, which are regulated on differentiation. Several mutant viruses which have overcome these blocks to transcription have been isolated (Fujimura et al., 1981, Hilberg et al., 1987). These have mutations which reduce binding of repressors (Hilberg et al., 1987, Prince and Rigby, 1991), or generate sites for positively acting transcription factors (Kovesdi et al., 1987a, Prince and Rigby, 1991) to promote expression in EC cells.

Adenovirus with a deletion of the E1A region (dl312) can sustain higher levels of transcription in EC cells than in differentiated cells (Imperiale et al., 1984). This finding led to the idea that EC cells possess an E1A-like activity, which is supported by correlations between the patterns of cellular gene transcription in EC and adenoviral E1A infected or transfected differentiated cells. As previously discussed, transcription of collagen type IV, MHC H2-K\(^{\alpha}\), and the insulin receptor genes, is induced on differentiation of EC cells. Young et al. (1989) showed that on infection of differentiated EC cells with a wild type adenovirus, transcription of the collagen type IV gene was repressed. Likewise, E1A could decrease the transcription of the insulin receptor (Stein and Ziff, 1987), reduce MHC H2-K\(^{\alpha}\) enhancer activity (Rochette-Egly et al., 1990), and induce transcription of hsp70 (Kao and Nevins, 1983, Imperiale et al., 1984). Also, viral E1A will repress the activity of other viral enhancers (Borrelli et al., 1984, Velcich and Ziff, 1985) which are negatively regulated in EC cells compared to their differentiated derivatives.

The E1A-like activity is not restricted to the EC tissue culture system, as an E1A-like activity has been detected by transcriptional induction of an E1A-dependent gene in pre-implantation embryos, and in oocytes of the adult ovary (Suemori et al., 1988, Dooley
1.15. The EIA protein

During early adenoviral infection, the first early gene expressed is EIA encoding two products of size 13S and 12S. The 13S product is responsible for transactivation of all other adenoviral early genes (E4, E2, E3) and the major late promoter (reviewed by Berk, 1986). The 12S and 13S products differ by the presence of 46 amino acids in 13S EIA (residues 140-185). Comparison of the sequences of various adenovirus serotypes allowed the identification of three conserved regions (Kimelman et al., 1985), two of which are shared by both proteins and the third, CR3, which is unique to the 13S region (see figure 1.3). Unique residues in CR3 are thought to mediate transcriptional activation (Moran et al., 1986, Lillie et al., 1986, 1987, Moran and Matthews, 1987), as transcription of some genes is activated only by the 13S, and not 12S EIA. Other genes, however, are activated by both the 12S and 13S products (Leff et al., 1984, Krippel et al., 1985, Zerler et al., 1987).

EIA cannot bind DNA in a sequence specific fashion to activate transcription (Ferguson et al., 1985, Chatterjee et al., 1988), and the target gene promoter/enhancer sequences have little homology. These results suggest that EIA acts indirectly to promote transcription, or may increase the abundance of a limiting factor of the transcriptional apparatus.

EIA can affect both polymerase II and III regulated transcription (Berk, 1986). In the case of polymerase III transcription, EIA increases the activity of the limiting factor TFIIIC (Yoshinaga et al., 1986). The activity of partially purified TFIID is also increased by EIA (Leong et al., 1988). Also, tethering of EIA to the promoter by fusion to a DNA binding protein activates transcription (Lillie and Green, 1989), and there is evidence that the 13S unique region of EIA binds to the basic repeat in TBP, possibly anchoring the activation domain of EIA to the promoter (Lee et al., 1991, Horikoshi et al., 1991). The proposal that TFIID tethers EIA to the promoter may explain the transcriptional induction of some polymerase II promoters through the TATA element (Simon et al., 1988). EIA appears to act through a subset of TATA elements of the TATAA class of promoters, as exemplified by E1b, hsp70, c-fos, and β-globin (Wu et al., 1987, Simon et al., 1988, Simon et al., 1990, Green et al., 1983). Replacement of the TATAA element of the adenoviral E1b promoter with the non-E1A responsive E4 TATAT element (Gilardi and Perricaudet, 1984), abolishes the E1A responsiveness of the promoter and alters the transcriptional start site (Pei and Berk, 1989).

Other viral and cellular promoters with non-E1A responsive TATA boxes, or without TATA elements are also activated by E1A (reviewed by Berk, 1986; Pei and Berk, 1989, Shi et al., 1991, Morris and Mathews, 1991). In these cases, E1A dependent transcriptional activation occurs via the initiator site. The recently cloned YY1 factor binds to two initiator sites in the adenoviral associated virus P5 gene promoter (Seto et al., 1991).
Figure 1.3. Functional domains of the E1A protein.
The upper cartoon represents the conserved regions of 13S E1A, and in the other figures
the regions of E1A important for transcriptional activation, transformation, and with
homology to other viral oncoproteins.
Functional Domains of E1A
Both binding sites, when inserted into heterologous promoters, repress transcription in vivo, but this repression is abolished in the presence of 12S E1A, whilst 13S E1A activates expression (Shi et al., 1991). Thus, E1A may activate YY1 directly by modification or association with it, or indirectly by removing a cellular inhibitor that is masking its activation domain.

12S E1A may activate transcription in an indirect fashion by sequestering the repressive Dr-1 protein from the TATA binding complex. This protein complexes with the amino terminal region of E1A (Inostroza et al., 1992), and thus both 12 and 13S E1As may sequester Dr-1 and activate basal transcription from either TATA containing, or initiator type promoters.

E1A has also been proposed to act as a coactivator linking Oct 4 with the basal transcriptional apparatus (Schöler et al., 1991) and is also required for the association of retinoic acid receptor and TFIID (Berkenstam et al., 1992).

The central portion of region 3 has similarity with metal binding domains (Culp et al., 1988), and integrity of this region is important for wild type activity of 13S E1A (Jelsma et al., 1988, Culp et al., 1988, Webster and Ricciardi, 1991). The domain is likely to form a cysteine type zinc finger (as in the steroid hormone receptor DNA binding domain, section 1.7.2), but is insufficient to confer DNA binding activity on E1A. However, this region may be important for interaction with a general transcription factor (Webster and Ricciardi, 1991), which, from the analyses by Lee et al. (1991), is likely to be TFIID.

The carboxyl terminus of the CR3 domain has been proposed as a region important for interaction with transcriptional activating proteins. Mutation of this region abolishes the ability of a CR3 domain peptide to interact with transcriptional activators (Webster and Ricciardi, 1991). 13S E1A can promote transcriptional induction through the ATF site which is present in the promoters of many E1A responsive genes (SivaRaman and Thimmappaya, 1987, Hurst and Jones, 1987, Lee et al., 1987, Morris and Mathews, 1991), by binding to the DNA bound transcription factor ATF-2 (Hai et al., 1989, CRE-BP1, Maekawa et al., 1989). ATF-2 lacks an activation domain, thus E1A may provide an activation domain in trans (Liu and Green, 1990). Interaction of E1A and ATF-2 may not be direct, and may be facilitated by other cellular proteins as binding of in vitro translated ATF-2 to an E1A column was dependent on prior incubation with cellular extracts (Maguire et al., 1991). Alternatively, ATF-2 may need to be modified by incubation with cellular proteins before interaction with E1A. The AP1 protein complex is induced by E1A in the presence of cyclic AMP (Müller, U., et al., 1989, deGroot et al., 1991) by increased transcription of the c-jun and fos-B genes. This AP1 complex can activate transcription from AP1 binding sites or from ATF sites (as AP1 can bind the site of the ATF factor, Hai et al., 1989).
E1A can also mediate increased DNA-binding by upstream activators, even in the absence of protein synthesis (Reichel et al., 1988). This suggests that E1A plays a role in the post-translational modification of such factors, possibly by activating a cellular kinase or by reducing the activity of a phosphatase. For example, in the presence of E1A, phosphorylation of both the E4F and E2F factors increases their DNA binding activity to their respective promoter binding sites (Raychaudhuri et al., 1989, Bagchi et al., 1989).

Other mechanisms by which adenovirus can stimulate the DNA binding activity of the E2F factor are reviewed in section 1.20).

1.16. Role of E1A in transformation

Transformation of cells may be mediated by 12S E1A, in cooperation with E1b or the ras oncogene (Ruley, 1983). The regions of E1A important for transformation were generally thought to be separate from those involved in trans activation (Moran et al., 1986, Lillie et al., 1986, 1987, Whyte et al., 1988), and are localised to CR1 and CR2 of the protein (Lillie et al., 1987). Transformation of primary cells is believed to be a two step process, with E1A immortalising the cells, and then acting in concert with cellular or viral oncogenes to transform them. Regions of E1A involved in in vitro immortalisation have not been separated from domains required for co-operative transformation with other oncogenes, but both features are encoded by the CR1 and 2 regions of the protein (Zerler et al., 1987, reviewed in Moran and Mathews, 1987).

E1A may immortalise cells by activation of genes required for the cell cycle. Indeed, expression of either 12 or 13S E1A in quiescent cells induces DNA synthesis (Stabel et al., 1985, Spindler et al., 1985), perhaps due to the increased levels of proliferating cell nuclear antigen (PCNA, Zerler et al., 1987) in E1A transfected cells. Also, E1A represses the transcription of many genes, and the inhibition of differentiated cellular functions by the repression of cellular enhancers may be another mechanism by which E1A immortalises cells.

Interestingly, comparison of amino acid sequence of other transforming viruses with the conserved regions of E1A, showed that conserved regions 1 and 2 of E1A are similar to domains of the papovaviral large T antigens (Stabel et al., 1985), and of the human papilloma virus type 16 E7 protein (Phelps et al., 1988, Dyson et al., 1989b). A related sequence is also present in the myc oncoprotein (Figge et al., 1988). SV40 large T antigen can transform both primary and immortalised cells (Brown et al., 1986), and the ability to transform is dependent on the intact nature of the 105-114 region (Cherington et al., 1988), which is homologous to the CR2 of E1A. This region of SV40 T can replace the E1A conserved region 2 in a functional chimaeric E1A protein, fulfilling all of the normal properties of E1A (Moran, 1988). The conserved region has an LXCXE motif at the amino terminal end, and a site for phosphorylation by the caesin kinase II (ckII) enzyme at the
carboxyl terminus. The motif encodes a helical region, and has been proposed to lie on the surface of the proteins and perhaps therefore to mediate transformation by interaction with cellular proteins.

Immunoprecipitation experiments performed using anti-E1A antibodies in adenoviral infected or E1A transformed cells specifically precipitated a range of cellular proteins of 60, 105, 107 and 300 kD (Yee and Branton, 1985, Harlow et al., 1986). Interaction of the 105 and 107 kD proteins with E1A was dependent on amino acids 121-127 of E1A, with an additional requirement for amino acids 30-60 for 105 kD binding. The binding of the 107 kD protein was destabilised by deletion of amino acids 30-85 of E1A (Whyte et al., 1989), indicating that the amino terminal domain of E1A is not essential for, but facilitates binding of the 107 kD protein. The 300 kD protein did not require residues in CR2, and associated with the amino terminal 76 amino acids of E1A (Whyte et al., 1989).

Likewise, immunoprecipitation experiments with anti-SV40 T antibodies precipitated proteins of 105 and 107 kD from SV40 T expressing CV-1 cells (DeCaprio et al., 1988, Dyson et al., 1989a, Ewen et al., 1989, Ludlow et al., 1989), and binding of these proteins was dependent on the conserved 105-114 transforming region of T.

Whyte et al. (1988) showed that the 105 kD polypeptide was the retinoblastoma gene product, pRb (Friend et al., 1986, Lee et al., 1987a, Hong et al., 1989, Bernard et al., 1989), by comparison of the peptide maps of the E1A-associated 105 kD protein with the 105 kD protein precipitated by anti-pRb antibodies. The 107 kD protein (p107) has recently been cloned (Ewen et al., 1991), and shows a region of homology with pRb in the E1A binding domain. The 60 kD polypeptide has been shown to be cyclin A (Pines and Hunter, 1989, Giordano et al., 1989), a protein which accumulates periodically during the cell cycle, and associates with and activates cell cycle regulated protein kinases (section 1.18).

1.17. Retinoblastoma

The retinoblastoma gene encodes a protein which is a regulator of cell proliferation. The loss of both alleles of the gene by somatic mutation, or by the loss of the second allele after germ-line inheritance of a mutated allele, correlates with retinoblastoma in humans. Often, in the case of familial retinoblastoma, secondary malignancies (of soft tissue, or osteosarcomas) develop later in life (reviewed by Klein, 1990).

The inactivation of both alleles of the retinoblastoma gene during tumorigenesis suggests that Rb plays an important role in the negative regulation of growth control. This notion has been supported by the reduction in cellular proliferation or tumorigenicity of cells with an inactive deletion of Rb, into which a wild type Rb gene is introduced (Huang et al., 1988, Qin et al., 1992) and the inhibition of progression into S phase of cells injected
with pRb protein during early G1 (Goodrich et al., 1991). More evidence for the importance of pRb comes from recent work performed by several groups (Lee et al., 1992, Jacks et al., 1992, Clarke et al., 1992) in which knockout of Rb in transgenic mice has shown that Rb function is crucial for embryo survival beyond 15 days. Homozygous animals showed normal development until 12.5 days, when, after the switch to hepatic haematopoiesis, there was a reduction in the production of mature enucleated blood cells, leading to anoxia in older embryos. Also, there was massive cell death in the central nervous system, and mitotic cells were present in zones where normally only cell differentiation and migration occur. These abnormalities of neural development and the anoxia were possibly the cause of the lethal phenotype of loss of Rb function.

The importance of a mutation in both Rb alleles for tumorigenesis to occur was displayed in the heterozygotic mice, which in a few cases suffered pituitary tumours. In no case did the mice develop retinoblastoma, suggesting that the pituitary of mice is more susceptible to tumorigenesis, and that in the retina, the loss of function of another tumour suppressor gene may be necessary for tumour development. Alternatively, mice simply do not suffer from retinoblastoma.

So how does Rb function to control cellular proliferation? The Rb gene has a complex structure (Hong et al., 1989), with 27 exons encoding a 105 kD polypeptide product (pRb) which is localised to the nucleus. pRb is phosphorylated in a cell cycle dependent fashion (Lee et al., 1987b, Buchkovich et al., 1989, DeCaprio et al., 1989, Chen et al., 1989, Mihara et al., 1989). The underphosphorylated form of pRb has been proposed to be the active form of pRb, as it is this form which is bound by the transforming SV40 large T antigen (Ludlow et al., 1989, 1990). The tight association of pRb with the nuclear compartment is regulated by phosphorylation (Mittnacht and Weinberg, 1991), with phosphorylated pRb forming a less strong association with the nuclear compartment. Rb might therefore be proposed to act as a regulator of transcription, but although the carboxyl terminus of pRb can bind DNA cellulose, binding is not sequence specific, and mutated non-functional versions of the protein can bind DNA with the same affinity, suggesting that DNA binding is not an important mechanism of negative regulation by Rb (Lee et al., 1987, Horowitz et al., 1989). pRb may however be targetted to the DNA by interaction with other DNA binding factors.

The region of Rb responsible for growth suppression has been identified by transfection of truncated Rb derivatives into cells with a non-functional Rb product (Qin et al., 1992). Growth suppression was associated with the 379-792 amino acid region of pRb, including that region shown to be important for interaction with the viral oncoproteins SV40 large T and adenoviral E1A (Hu, Q., et al., 1990, Kaelin et al., 1990, Huang et al., 1990).

The oncoprotein interaction domain is very large compared with the few residues
necessary in T or E1A to interact with pRb, therefore it was proposed that the interaction domain of pRb formed an enzymatic-like pocket domain (Hu, Q., et al., 1990, Kaelin et al., 1991) and that the binding of cellular proteins which normally interacted with this pRb pocket was competed by the viral proteins. The pRb pocket domain is the common site of mutation or deletion in retinoblastoma and other sarcomas (Hu, Q., et al., 1990, Huang et al., 1990, Kaelin et al., 1990, Shew et al., 1990), stressing the importance of the interaction of cellular proteins with the domain for the normal function of the Rb tumour suppressor phenotype.

Much work has been performed to isolate cellular proteins which interact with pRb. One could postulate that pRb interacts with and inactivates cellular proteins which promote the cell cycle, or alternatively, that Rb activates proteins required for growth suppression. Indeed, pRb has been shown to interact with c-myc in vitro (Rustgi et al., 1991), an interaction which occurs between the amino terminal transforming region of myc, and the carboxyl domain of pRb which is frequently deleted in tumours, suggesting that these antagonistic proteins cooperate through direct binding to modulate cell cycle control. However, this association has not been demonstrated in vivo, therefore its functional relevance is unclear.

Interestingly, pRb has also been proposed to regulate gene expression through several transcription factor binding sites, including the appropriately named retinoblastoma control element (RCE). This element was first identified in the c-fos promoter (Robbins et al., 1990), and has subsequently been identified in the promoters of several genes including transforming growth factor-β1, TGF-β1 (Kim et al., 1991), and c-myc (Pietenpol et al., 1990, 1991). The RCE sequence has similarity with an Sp1 consensus binding site, and indeed, Sp1 has been shown to bind to a subset of the RCE sites, and to be activated in the presence of cotransfected Rb (Kim et al., 1992a). Only some Sp1 sites are activated by Rb, therefore the position or context of the Sp1 site must be important for conferring regulation by Rb. TGF-β1 is also proposed to inhibit the phosphorylation of pRb, presumably by modulation of the activity of a pRb kinase (Laiho et al., 1990).

pRb has also been proposed to activate or repress transcription mediated by ATF sites. Rb stimulates transcription mediated by an ATF-2 GAL4 fusion protein, and thus pRb has been proposed to activate transcription of TGF-β2 which has a high affinity ATF-2 binding site in its promoter (Kim et al., 1992b). TGF-β2 is an inhibitory growth factor, therefore pRb can promote the expression of an inhibitory growth factor (TGF-β2), and a second induced inhibitory growth factor (TGF-β1) can maintain pRb in the active unphosphorylated state, to coordinate a reduction in the activity of target oncogenes such as c-myc.

During the course of the work presented in this thesis, it was also shown that pRb could bind E2F and DRTF1 both in vitro and in vivo (Bandara and La Thangue, 1991, 49
Bagchi et al., 1991, Chellappan et al., 1991, Chittenden et al., 1991). The binding of pRb reduces the transcriptional activity of E2F site directed transcription *in vivo* (Zamanian and La Thangue, 1992, Weintraub et al., 1992, Dalton, 1992, Hiebert et al., 1992), and mutant forms of pRb are compromised in their ability to bind E2F (Bandara et al., 1991, Hiebert et al., 1991), suggesting that the interaction of pRb with E2F is functionally important for the regulation of transcription by Rb. These results will be discussed further in chapter 8.

Other workers using a different approach have also indicated that pRb can bind to a cellular factor which has E2F DNA binding specificity (Chittenden et al., 1991, Ouellette et al., 1992). Essentially, cellular proteins which bind to pRb were incubated with pools of degenerate oligonucleotides. After multiple rounds of selection the sequences having highest affinity were amplified by the polymerase chain reaction (PCR). In addition to E2F sites, sequences resembling the Sp1 site, and the variant forms of the E2F site present in the promoters of various cell cycle regulated genes, and growth factors (for example *myc, myb* and insulin-like growth factor) were selected.

*Rb* expression vectors have been extensively used to screen cell extracts and expression libraries for proteins that interact with pRb. Several proteins were isolated which bound wild type but not mutated pRb (RBP1 and 2, Defeo-Jones et al., 1991, and RbAP46, Huang et al., 1991). RbAP46 was isolated by immunoprecipitation of pRb complexes from cell extracts. Likewise, Kaelin et al. (1991) detected the association of seven polypeptides with pRb fusion proteins which bound mutant forms of pRb with reduced affinity. The association of these proteins with pRb may be indirect, as they were isolated by immunoprecipitation. However, RBP1 and 2 were cloned by screening cDNA libraries and are therefore likely to associate directly with pRb. E2F was subsequently cloned by Helin et al. (1992), and Kaelin et al. (1992), by screening expression libraries with pRb expression vectors (see chapter 8).

If binding of the cellular proteins to pRb is necessary for Rb to control the proliferation rate, there must be extensive cell cycle regulation of the interaction. Control is not mediated at the level of Rb synthesis, as it is constitutively produced throughout the cell cycle in tissue culture cells (Mihara et al., 1989, Buchkovich et al., 1989), and is transcribed at similar levels in most tissues from about 9 days gestation in the mouse (Bernards et al., 1989). pRb is however differentially phosphorylated through the cell cycle (Mihara et al., 1989, Chen et al., 1989, Buchkovich et al., 1989, DeCaprio et al., 1989, 1992, Ludlow et al., 1990), with the underphosphorylated form of pRb thought to be the active form. There is some evidence for a correlation between the phosphorylation state of pRb and the state of cellular differentiation from experiments performed with mitogenically stimulated cells (Chen et al., 1989), where phosphorylation increased with proliferation rate, and leukaemic cell lines showed dephosphorylation of pRb as the cells were induced to differentiate.
Phosphorylation of pRb occurs on serine and threonine residues, and the sites phosphorylated in vivo can be phosphorylated in vitro by a cell cycle regulated cdc2-like kinase (Lin, B.T.-Y et al., 1991). Initially, pRb was thought to be underphosphorylated during G\textsubscript{s} and G\textsubscript{1} of the cell cycle, and phosphorylation was thought to occur during late G\textsubscript{1} phase (Mihara et al., 1989, Chen et al., 1989, DeCaprio et al., 1989, Buchkovich et al., 1989). However, recent evidence suggests that some phosphorylation occurs earlier during the G\textsubscript{1} phase of the cell cycle, and that subsequent phosphorylation events occur later, during G\textsubscript{2} and M phase (DeCaprio et al., 1992). Since pRb is presumed to be active throughout G\textsubscript{1}, these early phosphorylation events on certain sites appear not to abolish the activity of pRb.

1.18. The cell cycle.

Briefly, the cell cycle is driven by kinase activity which peaks at two positions during the cell cycle; the G\textsubscript{s}/S and G\textsubscript{2}/M transitions (see figure 1.4). Cell cycle regulation has been extensively studied in yeast, where a single catalytic subunit of kinase activity is responsible for both transitions of the cell cycle (encoded by cdc2 in S.\textit{pombe}, and CDC28 in S.\textit{cerevisiae}, reviewed by Pines and Hunter, 1990). The kinase p34\textsuperscript{cdc2}, which has functional homologues in all eukaryotic cells, is regulated by association with cyclins, proteins originally identified as undergoing cell cycle dependent accumulation in sea urchin eggs (Evans et al., 1983), and which were the only proteins whose synthesis was required to promote mitosis in \textit{Xenopus} eggs (Murray and Kirschner, 1989). Many different cyclins have now been characterised, and these are classified according to homology within the conserved cyclin box domain, thought to be important for interaction with the cell cycle regulatory kinase.

The best characterised kinase activity is the mitosis promoting activity (MPF) of higher eukaryotes (equivalent to the kinase promoting the G\textsubscript{2}/M phase transition), where activation is conferred on p34\textsuperscript{cdc2} through an association with a B type cyclin (Booher and Beach, 1987, Pines and Hunter, 1989). Cyclin B transcription is increased during late S/G\textsubscript{2} phase, and the level of protein peaks in G\textsubscript{2} and decreases rapidly after mitosis (Pines and Hunter, 1989). The p34\textsuperscript{cdc2} subunit associates with cyclin B (Draetta and Beach, 1988, Solomon et al., 1990) when levels of cyclin B have accumulated in early G\textsubscript{2} phase. Two types of phosphorylation occur on this association, with phosphorylation on Thr 167/161 (S.\textit{pombe}/human) facilitating the association of cyclin, and phosphorylation of Tyr 15/Thr 14 (S.\textit{pombe}/human) inhibiting the kinase activity until the initiation of mitosis. Dephosphorylation of Thr 14 and Tyr 15 activates the kinase, and is mediated by the cdc25 phosphatase (Dunphy and Kumagai, 1991, Kumagai and Dunphy, 1991, Strausfeld et al., 1991, Gautier et al., 1991). Kinase activity is lost on degradation of cyclin B by ubiquitin mediated degradation (Draetta et al., 1989, Glotzer et al., 1991).
Figure 1.4. The cell cycle.

Cartoon depicting the major transition points of the cell cycle. Cyclins are represented by the shaded oval figures, with differences in the shading patterns distinguishing mitotic from G1 cyclins. Interaction of G1 cyclin(s) with cdc2-related kinase(s), represented by small circle mediates activation of the kinase, necessary for Start dependent functions, and the G1/S transition. Association of mitotic cyclins synthesised during later stages of the cell cycle with the p34\(^{cd2}\) subunit causes an inhibitory phosphorylation of the kinase, which is relieved by the action of the cdc25 phosphatase. The mitotic kinase is inactivated by the destruction of the mitotic cyclin, by ubiquitin mediated proteolysis.
Mitotic Entry Functions

Active MPF

Degradation of mitotic cyclin

Mitotic Exit Functions

Mitotic cyclin synthesis

G_1 cyclin synthesis

Start Dependent Functions
The other major transition point of the cell cycle that is dependent on kinase activity was first defined in yeast, and was named START. START occurs just prior to the G₁/S transition and is similarly regulated by the CDC28/cdc2 kinase although less is known about the regulation of kinase activity by cyclins at this point. The activity of the kinase at START is regulated by a different class of cyclins which in S.cerevisiae are known as the CLN gene products. It is not known if the in vivo target specificity of the kinase is altered by association with different cyclins.

Several CLN genes have been isolated which are expressed in G₁ (CLN 1 and 2, Hadwiger et al., 1989, Richardson et al., 1989, Wittenberg et al., 1990), which activate p34\(^{CDC28}\), whereas CLN 3 has a less regulated expression pattern (Nash et al., 1988, Cross and Tinkelenberg, 1991). Expression of any member of the CLN gene family is sufficient to allow cell cycle progression, whereas functional inactivation of all three CLN genes leads to cell cycle arrest in G₁ (Richardson et al., 1989). The progression of yeast from the G₁ to the S phase of the cell cycle can be prevented by decreasing the levels of CLN transcription by starvation or in response to mating pheromones. Alternatively, stabilisation of the CLN products by removal of the PEST sequence (Rogers et al., 1986) advances the cycle into S phase, even in the absence of nutrients or in the presence of mating pheromone (Hadwiger et al., 1988, Nash et al., 1988). These experiments underline the importance of the G₁ cyclins for the control of CDC28 and START.

Cyclin genes have been isolated from mammals, and the first to be identified were required for mitosis (Pines and Hunter, 1989, Nurse, 1990). Other mammalian cyclins have now also been identified which are thought to play a role at other stages of the cell cycle, thus in both budding and fission yeasts and higher eukaryotes, different cyclins are responsible for kinase activity at different stages of the cell cycle.

Many of the human cyclins were identified by complementation of yeast deficient in all three CLN proteins: cyclins C, D, and E (Lew et al., 1991), cyclin E (Koff et al., 1991), cyclin D1 (Xiong et al., 1991), a screen which also identified the A, B1 and B2 mitotic cyclins. This finding is not so suprising when you consider that very high levels of the proteins were being expressed continuously in these yeasts. The D1 cyclin was also identified as a gene present in a rearranged parathyroid hormone locus in parathyroid tumours (Motokura et al., 1991), and as a gene whose expression is increased by colony stimulating factor-1 activation of macrophages (Matsushime et al., 1991). Two other genes were also isolated by hybridisation to the cyclin D1 cDNA; cyclins D2 and D3 (otherwise named CYL 1, 2 and 3, Matsushime et al., 1991).

These cyclins form distinct groups based on their homology with other cyclins (Lew et al., 1991). The D type cyclins are expressed in G₁ of the cell cycle, and D1 and D2 are regulated at the level of RNA transcription, message stability, and protein stability (Motokura et al., 1991, Matsushime et al., 1991), and all three D type cyclins show variant
cell specific patterns of expression (Inaba et al., 1992).

Cyclin C message accumulates two fold during the early G₁ stage of the cell cycle, whereas cyclin E RNA accumulates more dramatically during late G₁ (Lew et al., 1991). The protein levels of cyclin E and the activity of the cyclin E associated protein kinase also peak during late G₁ and early S phase, and then are reduced at later stages of the cell cycle (Dulic et al., 1992). These results, together with the observation from Koff et al. (1991) that addition of cyclin E-GST fusion proteins to G₁ extracts causes an induction of histone H1 kinase activity, suggest that cyclin E may play a role in the activation of kinase for the G₁/S progression of the cell cycle.

Cyclin A has been proposed to be important in activation of DNA synthesis immediately following G₁/S transition of the cell cycle, in addition to its role in the G₂/M phase transition (Pagano et al., 1992b). The removal of cyclin A (by injection of antisense cyclin A cDNA or anti cyclin A antibodies) from G₁ cells inhibited DNA synthesis (Girard et al., 1991, Pagano et al., 1992b). Cyclin A levels accumulate during S phase (Pagano et al., 1992b), and therefore may promote activation of kinase after the G₁/S transition.

It was originally thought that the p34cdc2 kinase activity was required for both G₁/S and G₂/M transitions in higher eukaryotes (see for example Minshull et al., 1990), as CDC28 activity is required for both points of progression in S.cerevisiae (Booher and Beach, 1987). However, the antibodies used to detect p34cdc2 in these experiments could crossreact with other subsequently identified kinases (Fang and Newport, 1991), and now several different protein kinases have been shown to regulate the cell cycle in higher eukaryotes (Lehner and O'Farrell, 1990, Elledge and Spottswood, 1991, Paris et al., 1991, Tsai et al., 1991, Meyerson et al., 1992, Matsushime et al., 1992, Okuda et al., 1992). These various kinases have the potential to regulate different substrates, for example cdk4 will not phosphorylate the histone H1 substrate generally used to measure cdc2-like kinase activity (Matsushime et al., 1992). Also, the regulation of activation of the kinases during the cell cycle, caused by the differential expression of the various interacting cyclin subunits, generates an enormous potential for regulation of the kinase activities within the cell cycle.

In addition to cdc2, which is essential for the G₂/M transition (Riabowol et al., 1989, Fang and Newport, 1991), two other vertebrate kinases, cdk2 and cdk3 could partially complement a yeast temperature sensitive cdc28 mutant (Koff et al., 1991, Elledge and Spottswood, 1991, Meyerson et al., 1992). Cdk2 had previously been identified as one of the cyclin A associated protein kinases (Elledge and Spottswood, 1991, Tsai et al., 1991, Giordano et al., 1991, Paris et al., 1991, Pagano et al., 1992b), which are necessary for entry into S phase (Fang and Newport, 1991). In addition to cyclin A, cdk2 may associate with cyclin E (Koff et al., 1991), and the cyclin E associated kinase activity is high in G₁ cells (Dulic et al., 1992).
Evidence for the role of cyclins A and E for promotion of the cell cycle has come from the experiments performed by Hinds et al. (1992). Expression of pRb in Rb deficient cells can block the cell cycle in late G1 phase, and this block can be overcome by the overexpression of certain cyclins, suggesting that these cyclins may promote cyclin dependent kinase phosphorylation of pRb (but are limiting for this activity in cells expressing large amounts of pRb). The two cyclins which are necessary for overcoming the block are cyclins A and E, both of which are found in association with cdk2 during S or G1 phases respectively. Overexpression of either of these cyclins results in phosphorylation of pRb, and the removal of the block to cell cycle progression. An alternative method of overcoming Rb suppression of the cell cycle is employed by cyclin D1, which, on overexpression, can reduce the levels of exogenous pRb and thus overcome the effect of pRb on the cell cycle without phosphorylation of pRb.

In other situations, the D1 cyclin may be involved in the phosphorylation of pRb as the D1 cyclin is found in association with the cdk4 kinase, which efficiently phosphorylates pRb and p107 in vitro (Matsushime et al., 1992).

1.19. Cell cycle regulation of E2F containing complexes

Interestingly, the transcription factor E2F, which associates with pRb, also forms complexes with both cyclins A and E and the cdk2 kinase at different stages of the cell cycle (Raychaudhuri et al., 1991, Mudryj et al., 1991, Shirodkar et al., 1992, Pagano et al., 1992a, Lees et al., 1992). The activity of these various complexes of E2F is likely to vary, therefore E2F is probably a cell cycle regulated mammalian transcription factor, and the regulation of the activity may be responsible for the cell cycle regulated expression of promoters with E2F binding sites. These findings will be discussed in more detail in chapter 8.

1.20. Characterisation of E2F

E2F was first identified as a cellular factor whose DNA binding activity on the adenoviral E1A-inducible E2A promoter was increased during viral infection (Kovesdi et al., 1986a,b). The binding sites (at -68 to -60 and -43 to -36) correlated with a region important for transcriptional activation, and for stimulation by E1A (Imperiiale and Nevins, 1984, Kovesdi et al., 1986a,b, 1987b, SivaRaman and Thimmappaya, 1987, Babiss, 1989, Loeken and Brady, 1989). Similar binding sites were subsequently found on a functionally important region of the E1A enhancer, and, on both E1A and E2A control regions, increased E2F binding activity correlated with an increase in transcription (Kovesdi et al., 1986a,b, 1987b, Reichel et al., 1987, 1988, SivaRaman and Thimmappaya, 1987, Babiss, 1989). The mechanism of this induction of binding activity was not dependent on new protein synthesis (Reichel et al., 1988), but was found to be dependent
on phosphorylation of E2F (Bagchi et al., 1989).

The stimulation of E2F binding activity by adenoviral gene products was investigated and E1A was shown to increase the binding activity of E2F (Kovesdi et al., 1986a,b, 1987b, Reichel et al., 1988, Babiss, 1989), presumably by activation of a pre-existing pool of E2F within the cell, as protein synthesis was not required (Reichel et al., 1988). The E4 gene products had also been shown to trans activate the E2A promoter (Goding et al., 1985, Reichel et al., 1989, Neill et al., 1990), and this effect was shown to be caused by a direct interaction between the E4 ORF 6/7 product and E2F (Huang and Hearing, 1989, Marton et al., 1990) promoting a stable complex formation of E2F on the E2A promoter (Hardy and Shenk, 1989, Huang and Hearing, 1989, Babiss, 1989, Hardy et al., 1989, Marton et al., 1990, Raychaudhuri et al., 1990, Neill et al., 1990, Neill and Nevins, 1991). Region 7 of the ORF 6/7 gene product was shown to be important for interaction with E2F, probably by interaction of E2F with the helix-loop-helix like domain of E4 ORF 6/7 (Neill and Nevins, 1991).

The E2F sites on the E2A promoter are close (separated by 17 base pairs, bp) and are located on the same side of the helix. Thus, cooperative binding of E2F could occur in the presence of E4 ORF 6/7 (Yee et al., 1987, Hardy and Shenk, 1989, Hardy et al., 1989, Raychaudhuri et al., 1990). Stable cooperative binding could not occur on the E1A enhancer (Hardy and Shenk, 1989, Hardy et al., 1989, Raychaudhuri et al., 1990, Neill and Nevins, 1991) probably because the E2F sites were separated by a large distance (63 bp), and a binding site for another factor, EF-1A, was present in the intervening DNA sequence (Bruder and Hearing, 1989).

1.21. E2F-like activities in EC cells.

An E1A-like activity had been reported in EC cells (Imperiale et al., 1984) which mediates transcriptional induction of the E1A dependent E2A promoter. Experiments performed to analyse E2F-like activities in F9 EC cells have shown that the complexes formed on the E2A promoter share few characteristics with those induced on adenoviral infection of fibroblast or epithelial cells (La Thangue and Rigby, 1987, Babiss, 1989, Hardy and Shenk, 1989, Hardy et al., 1989). The EC cell E2F complex is much less stable (Hardy et al., 1989), is not formed by cooperative binding of E2F (Hardy and Shenk, 1989), and has a high affinity for a single site (La Thangue et al., 1990). The EC cell E2F site binding activity (DRTF1) is abundant in embryonal carcinoma cells, and is down regulated on differentiation (Differentiation Regulated Transcription Factor, La Thangue and Rigby 1987, Reichel et al., 1987, La Thangue et al., 1990), and the reduced levels of DRTF1 on differentiation correlate with an increased requirement for viral E1A to mediate E2A transcription on EC cell differentiation (Boeuf et al., 1990).

The sequence requirements for binding of DRTF1 are very similar to those of E2F
(La Thangue et al., 1990), and semi-purified DRTF1 can activate binding site dependent transcription from the E2A promoter in vitro (Shivji and La Thangue, 1991). The regulation of the E2A promoter on differentiation of EC cells is largely due to the regulation of the E2F site binding activity, as there is little regulation of the other upstream binding activity (-82 to -70) on differentiation (La Thangue et al., 1990). This is a consensus ATF binding site, and, in EC cells, the major form of the ATF site binding activity, ECRE2 (Tassios and La Thangue, 1990) is not regulated on differentiation. Thus the down regulation of DRTF1 is likely to be involved with the regulation of the cellular E1A-like activity on differentiation of EC cells, as functional binding sites for this protein exist in a number of E1A responsive gene promoters.

The purpose of the work described in this thesis was to further characterise a protein factor which had a regulated DNA binding activity during in vitro differentiation of F9 EC cells. This factor, DRTF1, is a positively acting transcription factor, and when purified, can activate transcription of a reporter gene in a site dependent fashion.

I investigated the regulation of the DRTF1 DNA binding activity during murine embryogenesis, and showed that DRTF1 was developmentally regulated in a tissue specific manner, and that different forms of the DNA binding activity were regulated independently during development. Uncomplexed DRTF1 was abundant during early embryogenesis, but as development progressed, the levels of complexed activity increased. I demonstrated that these complexed forms of the activity were detergent sensitive, and other workers showed that the complexes were disrupted on addition of E1A. The retinoblastoma gene product, and cyclin A activities were shown to bind to DRTF1, and this complexing was regulated during differentiation and development, consistent with the proposed role for Rb in suppression of cell cycle progression, as differentiated cells show increased levels of the pRb complexed activity.

I purified the DNA binding components of DRTF1 from F9 EC cells by a combination of ion exchange and sequence specific affinity chromatography, and showed that a population of DNA binding polypeptides was enriched during the purification. I devised methods to generate peptides from these polypeptides, and successfully obtained peptide sequence data from a copurifying polypeptide. Peptide sequence from DRTF1 was successfully obtained subsequently by our collaborators from protein purified using the procedures that I developed.

The peptide sequence data was used to design degenerate oligonucleotides for PCR assisted amplification of cDNA sequence, and a short cDNA was obtained from cDNA prepared from EC cell RNA. I used this cDNA to investigate the expression pattern of the sequenced protein, by Northern, RNAse protection, and in situ hybridisation analyses through murine embryogenesis.
CHAPTER 2

MATERIALS AND METHODS
2.1 GENERAL METHODS

2.1.1. Sterilisation

Non-sterile plasticware (SW41 tubes, eppendorf tubes etc.) was sterilised by autoclaving at 121°C, 15 pounds/inch² for 30 mins. Solutions were sterilised under the same conditions, or by filter sterilisation through a 0.45 μm Nalgene disposable filter. For RNA work, no glassware was used, and all solutions were treated with DEPC (0.1% v/v, Sigma), prior to autoclaving. Protein purification columns were washed in 0.03% sodium azide (Sigma, NaN₃) in 0.1M column buffer (table 2.4) if they were to be redundant for more than 2 days.

2.1.2. DNA extraction and precipitation

DNA solutions were treated with phenol chloroform to remove contaminating proteins. Phenol (BRL) was equilibrated with 10 mM Tris (BDH) adjusted to 7.5 mM with 1 M HCl (FSA laboratory supplies), and stored at -20°C or at +4°C in the dark. All chloroform (May and Baker) used was a 24:1 mixture of chloroform: isoamyl alcohol (Sigma). DNA solutions were extracted with an equal volume of phenol chloroform isoamyl alcohol mix (25:24:1), vortexed and microfuged at 13 K rpm in a Heraeus sepatech biofuge A for 5 min. The upper aqueous layer was retained, and re-extracted with chloroform, vortexed and microfuged again to remove phenol contaminants. The chloroform extraction step was repeated, before addition of 0.1 volume of sodium acetate (Sigma, pH 5.2), and 2 volumes of 96% ethanol (Hayman Ltd.) to the aqueous phase. Precipitation was performed either on solid CO₂ for 20 min, or overnight at -20°C, and DNA collected by centrifugation for 25 min at 13K. The pellet was washed with 70% ethanol, air dried, and resuspended in H₂O.

2.1.3. Determination of nucleic acid and protein concentration

A Uvikon 860 spectrophotometer was used to determine the concentrations of nucleic acids in quartz cuvettes according to the formulae below (Sambrook et al., 1989)

\[1 \text{ A}_{260} = 50 \mu g/ml \text{ double-stranded DNA}\]
\[1 \text{ A}_{260} = 40 \mu g/ml \text{ single stranded oligonucleotides}\]
\[1 \text{ A}_{260} = 40 \mu g/ml \text{ RNA}\]
\[1 \text{ A}_{260} = 20 \mu g/ml \text{ double-stranded oligonucleotides}\]

Protein concentrations were determined by the use of the Bio-Rad protein concentration assay, based upon the method of Bradford (1976). Protein solutions were added in a 4:1 ratio to the reagent, and absorbance measured at 595 nm in disposable cuvettes. The samples were quantitated with respect to a range of BSA (Sigma) standard concentrations.
2.1.4. Preparation of dialysis tubing

Dialysis tubing (Medicell international) was prepared by boiling in 10 mM EDTA (BDH, pH 8) for 2 hours prior to sterilisation by autoclaving, and stored at 4°C in 50% ethanol, 0.03% NaN₃.

2.1.5. Autoradiography

Radioactive gels were autoradiographed by exposure to Kodak XAR 5 film with an intensifying screen at -70°C unless otherwise specified. Films were developed by a Fuji FMP 2100 x-ray automatic processor.

2.2. DNA USED

2.2.1. Synthetic Oligonucleotides

Synthetic oligonucleotides were prepared 'in-house' on an Applied Biosystems automatic synthesizer. Oligonucleotides were deprotected by incubation in 35% ammonium solution (BDH) for 6 hours at 55°C, and ethanol precipitated. Complementary single-stranded (ss) oligonucleotides were annealed by heating equimolar quantities to 90°C in ligase buffer (section 2.5.2), and cooling slowly to room temperature. The oligonucleotides used were:

71/50:-wild type binding sequence for DRTF1 from E2A promoter spanning -71 to -50 (La Thangue et al., 1990).

5' G A TC T A G T T T C G C T T A A A T T T G A 3'
3' AT C A A G C G A A T T T A A C T T C T A G 5'

64* :-as above, but with point mutation at position -64 (La Thangue et al., 1990).

5' G A TC T A G T T T T A G C T T A A A T T T G A 3'
3' AT C A A A A C G A A T T T A A C T T C T A G 5'

63* :-as above, but with point mutation at position -63 (La Thangue et al., 1990)

5' G A TC T A G T T T T C C G C T T A A A T T T G A 3'
3' AT C A A A A A G A G C G A A T T T A A C T T C T A G 5'

62* :-as above, but with point mutation at position -62.

5' G A TC T A G T T T T C G A G C T T A A A T T T G A 3'
3' AT C A A A A G A A G A C G A A T T T A A C T T C T A G 5'

61* :-as above, but with point mutation at position -61.

5' G A TC T A G T T T T C G C T C T T A A A T T T G A 3'
3' AT C A A A A G A G A G A C G A A T T T A A C T T C T A G 5'
60*: as above, but with point mutation at position -60.

5' G ATC TAG TTT TCG CGA TTA AAT TTG A
3' ATC AAA AGA GCT AAT TTA AAC TCT AG

5' G ATC TAG TTT TCG ATA TTA AAT TTG A
3' ATC AAA AGA TAT AAT TTA AAC TCT AG

60/62: as above, but with point mutation of positions -60 to -62 (La Thangue et al., 1990).

5' G ATC TAG TTT TCG ATA TTA AAT TTG A
3' ATC AAA AGA TAT AAT TTA AAC TCT AG

50/82: sequence from -82 to -50 of the E2A promoter on the non-coding strand, annealed to -82 to -70 of the coding strand (Shivji and La Thangue, 1991).

5' TGG AGA TGA CGT A
3' ACC TCT ACT ACA TCA AAA GCG CGA ATT TAA ACT

5' TGG AGA TGA TTT A
3' ACC TCT ACT AAA TCA AAA GCG CGA ATT TAA ACT

50/82 CRE-M: as above, except for 2 point mutations changing positions -72 and -73.

5' TGG AGA TGA TTT A
3' ACC TCT ACT AAA TCA AAA GCG CGA ATT TAA ACT

5' TGG AGA TGA TTT T
3' ATT GGC AAT GCA GTA AAA AAC TAG

50/82 CRE-M: as above, except for 2 point mutations changing positions -72 and -73.

5' TGG AGA TGA TTT T
3' ATT GGC AAT GCA GTA AAA AAC TAG


5' G ATC TAA CCG TTA CGT CAT TTT TT
3' ATT GGC AAT GCA GTA AAA AAC TAG

PM1: as above, but with point mutation of positions -48 and -49 (Tassios and La Thangue, 1990).

5' G ATC TAA CCG TTA AAT CAT TTT TT
3' ATT GGC AAT TTA GTA AAA AAC TAG

OCT: Immunoglobulin heavy chain enhancer position -554 to -536 (Ephrussi et al., 1985).

5' G ATC CGG TAA TTT GCA TTT CTA A
3' GCC ATT AAA CGT AAA GAT TCT AG

OCTM1: as above, except that positions -546 and -547 are mutated.

5' G ATC CGG TAA TCG GCA TTT CTA A
3' GCC ATT AGC CGT AAA GAT TCT AG

β GLOBIN: site C of β globin 3' enhancer (Wall et al., 1988). Kind gift from E.DeBoer, NIMR.

5' CGT CAG GAT GTT TAA GAT TAG CAT TCA GGA AG
3' A GTC CTA CAA ATT CTA ATC GTA AGT CCT TCG C

GATA: kind gift of S. Philipsen, NIMR.

5' TCG AGT ACT GCC TAT CTC AGC TCA A
3' CA TGA CGG ATA GAG TCG AGT TGA TC
Peptide 5 Antisense:

\[
5' \text{CGC GGA TCC AAA TTT TGA CTC TCC CGA GC} \quad 3' \\
\text{GG C G GAT T G} \\
\text{C C} \\
\text{T T}
\]

Peptide 6 Sense:

\[
5' \text{CGC GAA TCC CCA AAC ACA CAC TTC GT} \quad 3' \\
\text{G T G T T} \\
\text{C C} \\
\text{T T}
\]

2.2.2. Plasmid DNA

Plasmids used in this study are listed below.

pE2AcAt

This plasmid contains nucleotides -96 to +61 of the Adenovirus type 5 E2A transcription unit. The fragment was isolated by EcoR1 Dde1 restriction digestion.

pBluescript II SK*/KS+

Prokaryotic cloning vector derived from PUC 19. A phagemid containing the F1 phage ori and the colE1 ori. Ampicillin resistant. Stratagene.

SP64y actin

BamH1 HindIII fragment from human \( \gamma \) actin cDNA subcloned into pSP64 (Enoch et al., 1986). The plasmid was linearised with \( \text{Hin}f1 \) for SP6 polymerase reaction. Ampicillin resistant. Kind gift from E.Dzierzak, NIMR.

clone 6

p46 360 nucleotide cDNA cloned into the \( \text{BamH1} \) site of pBluescript II SK*. The plasmid was linearised with \( \text{EcoR1} \) to give sense transcription with T3 RNA polymerase, and with \( \text{Xba1} \) to transcribe antisense RNA from the T7 promoter. Ampicillin resistant.

myogenin

5' end of myogenin cDNA, \( \text{SmaI} \) to \( \text{PstI} \), subcloned into vector GSMZ from S.-P. Yee, NIMR. 150 bp Myogenin fragment released with \( \text{SmaI} \) and \( \text{PstI} \).

EF-1\( \alpha \)

Xenopus \( \text{EF-1}\alpha \) \( \text{PstI} \) \( \text{SacI} \) fragment (Krieg et al., 1989), kind gift from R.Wilson, NIMR.

2.3. RESTRICTION DIGESTION OF PLASMID DNA

2.3.1. Diagnostic digestion

Diagnostic digests used 1-10 \( \mu \text{g} \) DNA, and approximately 5 units (U) of enzyme/\( \mu \text{g} \) DNA, in the appropriate 1 X restriction buffer (table 2.1). Volumes of reactions were kept to a minimum, but the volume of enzyme never exceeded 10%. Reactions were performed at
37°C for 1 hour, DNA dye (table 2.2) was added to the reaction mixture, and the products electrophoresed through a 0.7 - 1.4 % agarose (Bio Rad) Uniscience mini-gel containing 0.5 µg/ml ethidium bromide (Sigma) at 12V/cm for 1 hour, in 1 X TBE (table 2.2). Markers (lambda DNA digested with HindIII or HindIII and EcoRI, or pBluescriptII KS+ digested with SduI) were electrophoresed in parallel. Products were examined by ultra violet (uv) illumination on a TMP 20 UVP transilluminator, and the gel photographed on Polaroid 667 film.

2.3.2. Preparative digestion

Preparative digestion of DNA was performed with upto 200 µg DNA and 2 U enzyme/µg DNA in the appropriate buffer. Reactions were incubated for 3 hours at 37°C, and an aliquot assayed by electrophoresis as described above. Once digestion was complete, the sample was phenol chloroform extracted, ethanol precipitated, and the DNA taken up in 40 µl dH2O. DNA dye was added, and the sample electrophoresed alongside size markers in a 1% low melting point (LMP, Nusieve, GTG) agarose gel in 1 X TBE. The desired fragment was excised from the gel, and melted in an equal volume of TE (10 mM Tris-HCl (pH 8), 1 mM EDTA) at 65°C for 10 min. The molten sample was extracted twice in warm phenol chloroform, with vigorous vortexing. Once the interface was clear, the sample was chloroform extracted, and ethanol precipitated.

2.4. RESTRICTION DIGESTION OF GENOMIC DNA

10 µg genomic DNA (prepared as in section 2.7) was restriction digested with approx. 7U enzyme/µg DNA in a large volume of the appropriate buffer (200 µl) for 12 hours at 37°C. Digested DNA was phenol chloroform extracted, ethanol precipitated, and resuspended in 20 µl dH2O.

2.5. SUBCLONING

2.5.1. Dephosphorylation of DNA

Dephosphorylation of the 5' ends of digested vector DNA was performed in 0.5 M Tris-HCl (pH 9), 10 mM MgCl2 with calf intestinal alkaline phosphatase (CIAP, Boehringer Mannheim). Reactions were performed using 1 µg digested DNA and 1 unit of enzyme for 30 min at 37°C, followed by a second incubation period after addition of more enzyme. DNA was phenol chloroform extracted, ethanol precipitated and resuspended in dH2O.

2.5.2. Ligation of DNA

A 10:1 molar ratio of insert:dephosphorylated vector was incubated with 10 units of T4 DNA ligase (New England Biolabs) in 66 mM Tris-HCl (pH 7.6), 10 mM MgCl2, 15 mM DTT, 0.2 mg/ml BSA and 1mM ATP at 14°C overnight. The mixture was heat treated to inactivate ligase before use.
### Table 2.1 Restriction Digestion Conditions

#### 10 X Restriction Buffers

<table>
<thead>
<tr>
<th>Component</th>
<th>High Salt</th>
<th>Very High Salt</th>
</tr>
</thead>
<tbody>
<tr>
<td>NaCl</td>
<td>1 M</td>
<td>1.5 M</td>
</tr>
<tr>
<td>Tris-HCl (pH 7.5)</td>
<td>0.1 M</td>
<td>0.1 M</td>
</tr>
<tr>
<td>MgCl₂ (Sigma)</td>
<td>0.1 M</td>
<td>0.1 M</td>
</tr>
<tr>
<td>DTT (Sigma)</td>
<td>10 mM</td>
<td>10 mM</td>
</tr>
<tr>
<td>BSA</td>
<td>1 mg/ml</td>
<td>1 mg/ml</td>
</tr>
</tbody>
</table>

**SduI Buffer**
- Tris-HCl (pH 7.5) 500 mM
- NaCl 250 mM
- MgCl₂ 100 mM

**SmaI Buffer**
- Tris-acetate 330 mM
- Mg acetate 100 mM
- DTT 5 mM

#### Digestion Conditions:

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Buffer</th>
<th>Manufacturer</th>
</tr>
</thead>
<tbody>
<tr>
<td>BamHI</td>
<td>High salt</td>
<td>Boehringer Mannheim</td>
</tr>
<tr>
<td>DdeI</td>
<td>High salt</td>
<td>Pharmacia</td>
</tr>
<tr>
<td>EcoRI</td>
<td>High salt</td>
<td>Boehringer Mannheim</td>
</tr>
<tr>
<td>EcoRV</td>
<td>Very high salt</td>
<td>Boehringer Mannheim</td>
</tr>
<tr>
<td>HindIII</td>
<td>High salt</td>
<td>Boehringer Mannheim</td>
</tr>
<tr>
<td>HinfI</td>
<td>High salt</td>
<td>New England Biolabs</td>
</tr>
<tr>
<td>PstI</td>
<td>High salt</td>
<td>Boehringer Mannheim</td>
</tr>
<tr>
<td>SduI</td>
<td>SduI buffer</td>
<td>Anglian Biotechnology</td>
</tr>
<tr>
<td>SmaI</td>
<td>SmaI buffer</td>
<td>Boehringer Mannheim</td>
</tr>
<tr>
<td>XbaI</td>
<td>High salt</td>
<td>Boehringer Mannheim</td>
</tr>
</tbody>
</table>

All digestions were performed at 37°C for a minimum of 1 hour.
### TABLE 2.2 ELECTROPHORESIS AND HYBRIDISATION BUFFERS

<table>
<thead>
<tr>
<th>Buffer</th>
<th>Components</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 X Northern running buffer</td>
<td>200 mM MOPS (Sigma) 50 mM sodium acetate 3 H₂O (GPR) 10 mM EDTA (Analar)</td>
</tr>
<tr>
<td></td>
<td>acidified to pH 7 with glacial acetic acid, autoclaved and stored in dark at room temperature.</td>
</tr>
<tr>
<td>SDS-PAGE running buffer</td>
<td>25 mM Tris- HCl (pH 8.3) 0.19 M glycine (BDH) 0.1 % (w/v) SDS</td>
</tr>
<tr>
<td>10 X TBE</td>
<td>900 mM Tris base (BDH) 900 mM boric acid (BDH) 20 mM EDTA (pH 8.0, BDH)</td>
</tr>
<tr>
<td>10 X TAE (pH 8.3)</td>
<td>400 mM Tris base 200 mM Acetic acid (FSA lab. supplies) 10 mM EDTA (pH 8)</td>
</tr>
<tr>
<td>5 X DNA Dye (for loading agarose gels)</td>
<td>0.25 % bromophenol blue (BDH) 0.25 % xylene cyanol (BDH) 25 % Ficoll 400 (Sigma)</td>
</tr>
<tr>
<td>SDS PAGE loading buffer</td>
<td>150 mM Tris (pH 6.8) 4 % SDS 20 % glycerol 5 % 2-mercaptoethanol</td>
</tr>
<tr>
<td></td>
<td>0.2 % bromophenol blue</td>
</tr>
<tr>
<td>Hybridisation buffer</td>
<td>50 % Formamide (BDH) 10 % dextran sulphate (Pharmacia) 0.5 % SDS</td>
</tr>
<tr>
<td></td>
<td>0.02 % Ficoll 400 0.02 % polyvinyl pyrolidine (Sigma) 0.02 % BSA</td>
</tr>
<tr>
<td></td>
<td>0.75 M NaCl 44 mM NaH₂PO₄ 5 mM EDTA</td>
</tr>
<tr>
<td>20 X SSC</td>
<td>3 M NaCl 0.3 M Na citrate (BDH)</td>
</tr>
<tr>
<td></td>
<td>brought to pH 7 with NaOH</td>
</tr>
</tbody>
</table>
2.5.3. Preparation of competent cells

A 50 ml overnight culture of DH5α strain of *E.coli* was prepared in L Broth (1 % tryptone (Difco), 0.5 % yeast extract (Difco), 0.5 % NaCl, adjusted to pH 7.5 with NaOH) from frozen glycerol (30%) stocks. 250 ml of L Broth was inoculated with 1 ml of the culture, and incubated at 37°C with shaking at 200 rpm, until the OD₆₀₀ was 0.9. The cells were collected by centrifugation at 3000 rpm for 5 min at 4°C in the GSA rotor of a Sorvall RT6000B centrifuge, drained, and resuspended in 50 ml of ice cold 100 mM MgCl₂. The cells were repelleted, and resuspended in 10 ml of ice cold 100 mM calcium chloride (CaCl₂, Sigma). Another 100 ml of CaCl₂ was added to the cells, and they were kept on ice for 1 hour. The cells were repelleted, resuspended in 12.5 ml 85 mM CaCl₂, 15% glycerol (FSA laboratory supplies), aliquotted and stored at -70°C.

2.5.4. Transformation of competent cells with DNA

An aliquot of the ligation mixture (approx. 50 ng DNA) was added to 200 μl of competent cells (thawed on ice), and incubated on ice for 20 min. The cells were heat shocked at 42°C for 2 min, chilled on ice for 2 min, and 800 μl of L Broth added before the cells were incubated for 1 hour at 37°C.

The cells were collected by centrifugation at 2 K rpm for 3 min, and resuspended in 100 μl of L Broth. The mixture was spread onto ampicillin (0.1 μg/ml, Sigma) agar plates (prepared by supplementing L Broth with 1.5 % agar (Difco) before autoclaving), and incubated overnight at 37°C.

2.5.5. Identification of positive clones

The presence of the correct ligation product in individual colonies was assayed by picking colonies into 10 ml of Brain heart infusion (BHI) medium (Difco) with 0.1 μg/ml ampicillin, incubating overnight at 37°C with shaking before preparing a mini-preparation of the DNA and analysing it by restriction digestion and sequencing.

2.6 PREPARATION OF PLASMID DNA

2.6.1 Mini-preparations (minipreps) of DNA

1.5 ml of the culture (section 2.5.5), was pelleted by a 5 min centrifugation at 13 K rpm in a microfuge. The pellet was resuspended in 250 μl of solution 1 (see below), 250 μl of solution 2 was added, and the solutions mixed. 250 μl of solution 3 was added, and the mixture centrifuged at 13 K rpm for 5 min to pellet the cell debris. The supernatant was retained, and phenol chloroform extracted before precipitation with an equal volume of propan-2-ol (BDH). The DNA was pelleted by a 20 min spin at 13 K rpm, and the pellet was air dried and resuspended in 50 μl TE. The RNA was precipitated out of the solution by addition of 70 μl ice cold 5 M Lithium chloride (LiCl₃) before incubating on ice for 5
min. The RNA was removed by a 20 min centrifugation, and the supernatant ethanol precipitated and dried.

Solution 1: 50 mM glucose (Sigma), 25 mM Tris-HCl (pH 8), 10 mM EDTA.
Solution 2: 20 mM NaOH (BDH), 1% SDS (BDH)
Solution 3: 5 M potassium acetate (BDH), pH 6.

2.6.2. Large scale preparation of plasmid DNA (Maxiprep)

A 250 ml overnight culture of transformed E.coli DH5α in BHI and 0.1 µg/ml ampicillin was centrifuged at 4 K rpm for 8 min at 4°C in the GSA rotor of a Sorvall RT6000B centrifuge. The cells were resuspended in 20 ml of solution 1 (2.6.1), and 40 ml of solution 2 was added and mixed by inversion. 30 ml of solution 3 was added, mixed by shaking, and centrifuged for 20 min at 10 K rpm at 4°C. The supernatant was strained through muslin and 0.6 volumes of propan-2-ol were added to precipitate the DNA at room temperature. After 20 min, the precipitate was collected by centrifugation at 12.5 K rpm for 20 min at 4°C, and resuspended in 8.8 ml TE. The solution was neutralised by the addition of 0.5 ml of 1M Tris base. 10.2 g caesium chloride (CsCl, FSA lab. supplies) was dissolved in the solution, and EtBr added to 1 mg/ml. The mixture was loaded into Ti 70.1 tubes (Beckman, polyallomer quick seal), and centrifuged for 24 hour at 48 K rpm at 20°C in a 70.1 rotor in the Beckman L8-70M ultracentrifuge. Supercoiled plasmid DNA was collected using a syringe and 19 gauge needle by side puncture of the tube under uv illumination. EtBr was removed by 4 dH₂O saturated butan-1-ol extractions, and CsCl removed by overnight dialysis against 2 dm³ TE. The DNA was ethanol precipitated, dissolved in dH₂O and checked by diagnostic restriction digest.

2.7 PREPARATION OF GENOMIC DNA

DNA was extracted from mouse liver after crushing frozen tissue in liquid nitrogen. 5 ml of proteinase K buffer (50 mM Tris, 100 mM EDTA, 100 mM NaCl, 1% SDS) was added to the powder in a 10 ml tube and mixed gently before addition of 0.5 mg/ml proteinase K (Boehringer Mannheim). The mixture was incubated overnight at 60°C with gentle shaking, 14 mg of DNAse free RNAse A (Boehringer Mannheim) was added, and incubation continued for 2 hours at 37°C on a rotating wheel. DNA was recovered after 5 phenol chloroform and 2 chloroform extractions by propan-2-ol precipitation (1:1). DNA was spooled using a heat sealed glass micropipette, and the spooled DNA washed through 70, 95, and 100% ethanol and air dried. The DNA was resuspended in 5 ml of TE by gentle agitation overnight. DNA was also extracted from F9 EC and JM tissue culture cells, using 1 ml of packed cells and 5 ml of proteinase K buffer, and resuspending the extracted DNA in 2 ml TE.
2.8. SEQUENCING OF DNA

Dideoxy sequencing was performed using Sequenase enzyme (US Biochemical Corporation), and a modification of the manufacturers protocol. 5-7 µg of plasmid DNA was alkali denatured in 25 µl 200 mM NaOH, 0.2 mM EDTA for 5 min at room temperature. The mixture was precipitated with 10 µl 5 M ammonium acetate (NH₄Ac, BDH pH 7.5) and 70 µl ethanol at -20°C for 10 min, and collected by centrifugation. 7 µl dH₂O, 2 µl sequenase buffer (table 2.3), 10 ng primer, 1 µl 100 mM DTT, 2 µl of a 1:5 dilution of labelling mix:H₂O, 5 µCi [α-³²P] dATP (Amersham) and 2 µl sequenase enzyme (1:8 dilution in TE pH 7.5) were added to the DNA pellet. After 10 min at room temperature, 3.5 µl of the mixture was added to 2.5 µl of each of the 4 termination mixes in a 60 well microtitre plate. After 20 min at 37°C, 4 µl of sequencing loading buffer (table 2.3) was added to each reaction, and the samples heated at 90°C for 2 min, prior to chilling and loading on a pre-run 6% polyacrylamide 7 M urea sequencing gel. Reactions were electrophoresed through the gel in 1 X TBE, at 40 W for 2-4 hour. After electrophoresis, the gel was soaked in 10% methanol (Hayman Ltd.), 10% acetic acid (FSA laboratory supplies) for 20 min, transferred to 3mm Whatmann paper and dried under vacuum at 80°C. The dried gel was exposed to Kodak XAR 5 film overnight at room temperature with an intensifying screen.

2.9. RADIOACTIVE LABELLING OF DNA AND RNA.

2.9.1. End labelling using DNA polymerase 1 (Klenow fragment)

300 ng of double stranded oligonucleotide or restriction fragment was incubated with 30 µCi of [α-³²P] dNTP (Amersham), and 5 Units of Klenow enzyme (Boehringer Mannheim) in 20 µl of high salt restriction endonuclease buffer (table 2.1) for 25 min at room temperature. The dNTP used was complementary to the first nucleotide of the 5' overhang, i.e. the 3' end of the DNA was labelled. The reaction volume was increased to 100 µl with dH₂O, phenol chloroform extracted and ethanol precipitated. The precipitated DNA was washed with 70% ethanol to remove unincorporated nucleotide before drying.

End labelling of the oligonucleotides for cross-linking experiments was performed by extending the coding sequence with 50 µCi of each of [α-³²P] dGTP, dATP, and dCTP; 0.1mM 5'bromo 2'deoxyuridine triphosphate (BdUR, Sigma) and dTTP, using 12 U of Klenow fragment of DNA polymerase 1 for 1.5 hours in a reaction volume of 50 µl of high salt restriction buffer, at room temperature. Synthesis was completed by addition of excess dATP, dTTP, dCTP and dGTP and an additional 5 U Klenow enzyme for 30 minutes. The reaction was phenol chloroform extracted, and ethanol precipitated as above.
### TABLE 2.3. SEQUENCING REAGENTS

**Sequenase buffer**

- 200 mM Tris-HCl (pH 7.5)
- 100 mM MgCl₂
- 250 mM NaCl

**Labelling mix**

- 7.5 µM dGTP
- 7.5 µM dCTP
- 7.5 µM dTTP

**ddN Termination mixes**

- 80 µM dGTP
- 80 µM dATP
- 80 µM dCTP
- 80 µM dTTP
- 8 µM ddNTP
- 50 mM NaCl

where N is A, G, T and C in separate mixes.

**Sequencing gel loading buffer**

- 95% deionised formamide
- 20 mM EDTA
- 0.05 % xylene cyanol
- 0.05 % bromophenol blue
2.9.2. End labelling using T4 polynucleotide kinase

Oligonucleotides were end labelled using T4 polynucleotide kinase (BRL). 300 ng of oligonucleotide was labelled in 50 mM Tris-Cl (pH 7.6), 100 mM MgCl₂, 5 mM DTT, 0.1 mM spermidine (Sigma), 0.1 mM EDTA, with 50 μCi [γ-³²P] dATP (NEN), and 10 Units T4 polynucleotide kinase at 37°C for 30 min. A second aliquot of enzyme was added, and incubation continued for a further 30 min, before phenol chloroform extraction, and ethanol precipitation. The precipitate was washed with 70% ethanol to remove unincorporated nucleotide.

2.9.3. Oligolabelling using DNA polymerase 1 (Klenow fragment)

The 360 bp fragment released by BamHI digestion of clone 6 was purified from LMP agarose (section 3.2.2). The fragment was resuspended in H₂O at a concentration of 50 ng/μl. 50 ng of DNA and 9 μl of H₂O was boiled and kept at 37°C for 10 min. 3 μl of oligolabelling buffer for dCTP (see below) was added with 13.6 μl of H₂O, 1.2 μg BSA, 30 μCi [α-³²P] dCTP and 12 U Klenow enzyme and the labelling reaction performed for 3 hours at room temperature. The reaction was terminated by phenol chloroform extraction, and ethanol precipitation of the probe. The probe was resuspended, its activity determined by scintillation (section 2.9.6), and diluted to 1 X 10⁶cpm/ml.

Oligolabelling buffer for dCTP was prepared by mixing solutions A, B, and C in the ratio 2:5:3.
Solution A : 625 μl 2M Tris-HCl (pH 8), 125 μl 1M MgCl₂, 250 μl dH₂O, 18 μl 2-mercaptoethanol (Sigma), and 5 μl of dATP, dTTP, and dGTP (each 0.1M in 3mM Tris-HCl, 0.2mM EDTA (pH 7).
Solution B : 2 M Hepes (BDH) titrated to pH 6.6 with NaOH.
Solution C : hexadeoxyribonucleotides (Sigma) resuspended in 3 mM Tris-HCl, 0.2 mM EDTA (pH 7) at 90 OD units/ml.

2.9.4. Preparation of α³²P labelled riboprobes.

Clone 6 (section 2.2.2) was linearised with EcoRI to give the sense transcript of p46 360 from the T3 promoter, and with Xbal to transcribe antisense RNA from the T7 promoter.

Human γ actin (SP64-actin, Enoch et al., 1986) was linearised with HindIII and transcribed from the SP6 promoter to give the antisense transcript.

Reactions consisting of 1 μg of linearised template DNA, 1 mM rATP, rCTP, and rGTP and 0.05 mM UTP, 40 mM Tris-Cl (pH 8.25), 6 mM MgCl₂, 2 mM spermidine, 0.02 mM DTT, 50 units RNAguard (Pharmacia) and 50 μCi [α-³²P-U TP] were incubated at 37°C for 1 hour with 12 units of phage polymerase (T3 and T7, Boehringer Mannheim), or at 40°C for SP6 polymerase (Stratagene). 12 U of DNaseI (RNAse free, BRL) was added and incubation continued at 37°C for 15 min. Reactions were phenol chloroform extracted,
and riboprobes precipitated with 20 μg yeast tRNA (Sigma) and sodium acetate in ethanol.

Full length riboprobes for RNase protection were isolated by electrophoresis of the DNAse treated samples in a 4% polyacrylamide, 7M urea, TBE sequencing gel. Samples were denatured in 80% formamide, 5% glycerol, 0.1 mM EDTA, 0.025% bromophenol blue, 0.075% xylene cyanol at 90°C for 5 min. and quenched on ice before loading alongside 32P labelled pBLCat2 cut with HinfI and EcoRI. The gel was electrophoresed at 42 W for 1.5 hours, the top plate removed, and the saran wrapped gel exposed to Kodak XAR 5 film for 1 min at room temperature. Full length probes were excised from the gel, and the probe eluted from the mashed gel slice in 400 μl of 0.5 M ammonium acetate, 1 mM EDTA, 0.1 % SDS for 2 hours at room temperature, ethanol precipitated, and resuspended to 1 X 10⁶ cpm/ml in DEPC treated dH₂O.

2.9.5. Preparation of 35S labelled riboprobes.

Riboprobes were prepared as described above, replacing the 50 μCi [32P-UTP] with 100 μCi [35S-UTP]. After DNAse digestion, RNA was separated from unincorporated nucleotide by fractionation on a 1 ml Sephadex G-50 (Pharmacia) column in 10 mM Tris-HCl (pH 7.5), 1 mM EDTA, 10 mM DTT, 0.1 % SDS. Riboprobe was ethanol precipitated, washed in 80 % ethanol, dried and resuspended to 2 X 10⁶ cpm/ml in hybridisation buffer (50% deionised formamide (BDH), 0.3 M NaCl, 20 mM Tris-HCl (pH 8), 5mM EDTA, 10 % dextran sulphate (Pharmacia), 1 X Denhardt solution, 0.5 mg/ml yeast tRNA).

2.9.6. Measurement of incorporated nucleotide.

An aliquot of the sample was taken and Cerenkov counted (32P) or added to 2ml of Econofluor scintillant (14C and 35S) in a scintillation vial. Samples were counted for 1 minute in an automatic scintillation counter.

2.10. TISSUE CULTURE.

2.10.1. General

Cells were grown at 37°C in an atmosphere of 95% humidity and 8.5% CO₂. Unless otherwise stated, cells were grown on Nunclon plasticware (Gibco, BRL). All harvested cells were snap frozen on dry ice, and stored at -70°C.

2.10.2. F9 EC cells.

F9 murine embryonal carcinoma (EC) cells (Bernstine et al., 1973) were cultured as adherent monolayers in Dulbecco’s modification of Eagle’s medium (DMEM), supplemented with 10% v/v foetal calf serum (FCS, Imperial labs, Europe), 100μg/ml streptomycin (Gibco, BRL), 500U/ml penicillin (Gibco, BRL), 1mM L-glutamine (Gibco, BRL), and 0.1% sodium pyruvate. Cells were replated at a density of 4x10⁴ /ml every 3
days after passage of trypsinised cells through a 19 gauge needle to prevent clumping.

2.10.3. Large scale culture of F9 EC cells.

Large scale culture of F9 EC cells was attempted by growing cells as adherent monolayers on Plastek microcarriers (Tekmat Co.). These beads have a small size (120 μm), and a large surface area suitable for growth of tissue culture cells. Beads were seeded at 8.3 X 10^5 cells /cm³ (1 X 10^6 cells/ 25 g of beads), in DMEM with 10% FCS and antibiotics as above, and stirred with a magnetic stirrer at 50 rpm. Cells were harvested by addition of 2 mg/ml of dispase (Boehringer Mannheim) for 20 min prior to suction of the cells from the beads under vacuum on a size 2 sintered glass funnel. Cells were washed with PBS prior to snap freezing. Beads were reusable after sterilisation in PBS. A 20 fold increase in cell number was achieved after 3 days of culture by this method. Higher yields of cells were achieved by aggregate culture. From an initial inoculum of 1 X 10^6 cells, between 60 and 120 ml packed cell volume was obtained after 10 days of culture in flasks of increasing size from 1 dm³ to 40 dm³, all stirred at 50 rpm, and at 5% CO₂. The cell suspension was concentrated by sedimentation, followed by centrifugation at 500 rpm for 20 min at 4°C.

2.10.4. Differentiation of F9 EC cells in vitro

Stem cells were induced to differentiate to parietal endoderm like cells by treating them with 0.05μM retinoic acid (Sigma), 1mM dibutyryl-adenosine 3'5' monophosphate (Sigma), and 0.1mM isobutyl methyl xanthine (Sigma), for upto 7 days, replenishing the medium every 2 days (Strickland et al., 1980). Cells were harvested at 3, 5, and 7 days after treatment.

2.10.5. Anti-SSEA1 staining

The stem cell phenotype of the EC cells was checked using immunoperoxidase staining with an α SSEA1 monoclonal antibody (kind gift of D. Solter; Solter and Knowles, 1978), using the method of Harlow and Lane (1988). Cells were fixed in 4% formaldehyde diluted in PBS, before membrane permeabilisation with 0.2% Triton X 100. The surface of F9 EC cells stained with the antibody, but the marker was lost after 1 day of differentiation, and did not reappear during the course of the differentiation.

2.10.6. Embryonic stem cells

STO G418 resistant fibroblasts were cultured on Falcon plastic ware treated with 0.1% gelatin. STO cells were plated at 3x10^5 /ml in DMEM (Flow), with 10% foetal calf serum, 1% non essential amino acids (Flow), and antibiotics and glutamine as for EC cells. Cells were cultured in 5% CO₂ at 37° C. Cells were passaged when confluent, and treated with 1mg/ml mitomycin C (Sigma) for 3 hours before use as feeder layers for the culture of embryonic stem (ES) cells, as described by Robertson, 1987.

73
CCE ES cells (Robertson et al., 1986), passage 9, were plated at 5x10^4/ml onto STO feeders, and cultured in DMEM (Flow), 20% foetal calf serum, antibiotics, amino acids, and glutamate as for STO cells, with the addition of 7x10^{-4}% 2-mercaptoethanol.

2.10.7. Differentiation of CCE ES cells in vitro

Separation of ES cells from STOs was necessary for differentiation of ES cells; cells were trypsinised, resuspended in excess medium, and cultured for 60 mins in the same plate. Fibroblasts preferentially attached to the plate during this time, and non attached ES cells were collected by centrifugation (Gossler et al., 1986). The cells were then cultured as aggregates by plating at 5 X 10^4 cells/ml onto bacteriological petri dishes (Sterilin) in STO medium. Cells were refed every 2 days and cultured for upto 12 days.

2.10.8. HeLa cells

HeLa cells (Gey et al., 1952) were grown as adherent monolayers in Eagles minimum essential medium supplemented with 10 % newborn calf serum and antibiotics as above.

2.10.9. SAOS-2 cells

SAOS-2 cells (Shew et al., 1990), were seeded at a density of 1 X 10^5/ml in DMEM supplemented with 10 % FCS, and were replated every 4 days.

2.10.10. JM cells

JM cells (Schneider et al., 1977), were grown in suspension in DMEM supplemented with 10 % FCS.

2.11. ISOLATION OF MOUSE TISSUES

2.11.1. Tissue isolation for protein analysis

Mouse development times assume mating occurs at midnight, and midday that plugs are found is taken as 0.5 days development. Tissues were isolated from adult mice (6 week old 129/ola), and from 11.5, 14.5, 17.5 days post coitum (dpc) embryos and new born mice. Whole 8.5, 11.5, 14.5 and 17.5 day embryos were also isolated (Parkes outbred, and CBA/Ca strains). All embryos and tissues were snap frozen on dry ice. Frozen whole embryos and large pieces of tissue were ground in liquid nitrogen in a pestle and mortar, and stored at -70°C.

Six week old female outbred MF1, and CBA/Ca mice were injected intraperitoneally with 5 U of pregnant mare's serum, and 48 hours later with 5 U of human chorionic gonadotrophin (HCG), to promote superovulation (Hogan et al., 1986).
After administration of HCG, one female was placed with one stud (C_{77} Bl_{10} x CBA) F1, and females checked for plugs the next morning. 3 days later, blastocysts were collected from the uterine horns of pregnant mice, by dissecting out the uterus, and flushing it with 1mg/ml BSA in PBS, using a 25 gauge needle. Blastocysts were pooled using a micropipette under a dissecting microscope, and frozen in a minimum volume of flushing solution.

2.11.2. Mouse strains for genomic DNA analysis
(CBA/Ca X C_{77}Bl_{10} ScSn) F1 mice were used.

2.11.3. Mouse strains for RNA isolation
(CBA/Ca X C_{77}Bl_{10} ScSn) F1 mice were used.

2.11.4 Mice for in situ hybridisation
CBA/Ca mice were used.

2.12. SOUTHERN ANALYSES
0.5 µg ethidium bromide and 5 µl of DNA dye was added to 10 µg restriction digested DNA samples (section 2.4), and the samples electrophoresed through a 0.6% agarose, 1 X TBE, 0.3 mg/ml ethidium bromide gel at 70v for 4 hours in 1 X TBE, 0.3 mg/ml ethidium bromide. Markers were run in parallel (lambda DNA digested with HindIII and EcoRI). After electrophoresis, the gel was placed on a TMP-20 UVP transilluminator, and DNA visualised by uv illumination. An acetate copy was made of the position of migration of the markers, and the wells of the gel, and the gel photographed. The gel was acidified in 0.25 M HCl twice for 15 min each, rinsed in H_{2}O, and denatured for 30 min in 0.5M NaOH, 1.5 M NaCl, with two changes of solution. The DNA was transferred to a nylon membrane (Amersham Hybond N) by capillary blotting as described for RNA transfer (section 2.14), except that Hybond N and 3mm paper were soaked in 2 X SSC (table 2.2), and the wicks dipped into a reservoir of 20 x SSC. After 12 hours of transfer, the Hybond N was rinsed in 6 X SSC, uv crosslinked for 3 min, and baked for 2 hours at 80°C. Filters were prehybridised in hybridisation buffer (table 2.2), at 42°C in heat sealed bags in a shaking water bath for 1 hour, before addition of 1 X 10^6cpm/ml of oligolabelled probe (section 2.9.3) and 200 µg/ml sonicated salmon sperm DNA which had been heat treated at 95°C for 5 min, and quenched on ice. Filters were hybridised for 12 hours at 42°C, prior to washing at 65°C in 2 X SSC (table 2.2), 0.1% SDS for 30 min, and in 0.1 X SSC, 0.1% SDS for 30 min, also at 65°C. Damp filters were autoradiographed for the times indicated in the figure legends.

Filters were stripped prior to reprobing by covering the filter in 500ml of boiling 0.1 % SDS solution, and allowing to cool to room temperature. Filters were
autoradiographed for 48 hours to ensure that no probe remained bound to the filters before their re-use.

2.13. RNA EXTRACTION

2.13.1. Total cellular RNA
Total cellular RNA was prepared from tissues or whole embryos ground in liquid nitrogen, or from harvested tissue culture cells, by a modification of the method of Chirgwin et al. (1979). 1g of powdered sample or 5x 10⁶ cells was lysed with 8ml of 4.25M guanidine isothiocyanate (Fluka, deionised with Bio-Rad AG-501-X8(D) resin), in 0.5% sodium lauroylsarcosine (Sigma), 5mM sodium citrate (pH7, BDH), 0.1M 2-mercaptoethanol. After removal of cell debris by a 5 min spin at 3K (Sorvall RT6000B centrifuge), the homogenate was layered onto 3.5 ml of 5.7M CsCl, 50 mM EDTA, 5 mM Na citrate (pH 7), in a Beckman polyallomer SW41 centrifuge tube, and centrifuged at 30 Krpm for 20 hours at 22°C in an SW41 rotor in a Beckman L8-70M centrifuge. After centrifugation, the supernatant was aspirated, and the tube inverted and wiped dry with tissue paper. The RNA pellet was resuspended in 400 µl of DEPC treated H₂O, phenol chloroform extracted twice, ethanol precipitated and resuspended to 5 mg/ml in DEPC H₂O, and stored at -70°C.

2.13.2. Poly A+ selected RNA
Poly A+ RNA was selected by use of a Poly (A) Quik mRNA purification kit (Stratagene) from 0.5 mg of total RNA from F9 EC cells. The RNA was heated for 5 min at 65°C, quenched on ice, and adjusted to 1mM Tris-HCl (pH 7.5), 0.1 mM EDTA, 500 mM NaCl, before application to a 1ml oligo dT column pre-equilibrated in high salt buffer (10 mM Tris-HCl (pH 7.5), 1 mM EDTA, 500 mM NaCl). The sample was reapplied twice before two washes with the high salt buffer, and 3 washes with reduced salt buffer (100 mM NaCl). The A+ RNA was eluted with buffer containing no salt (10 mM Tris-HCl (pH 7.5), 1 mM EDTA) which had been preheated to 65°C. The concentration of the A+ RNA was determined by spectrophotometry, and stored at -70°C. 0.5 mg of total RNA yielded approximately 0.015 mg of A+ enriched RNA.

2.14. NORTHERN ANALYSES
A 1.5 % agarose gel was prepared (14 x 11 x 0.7 cm), in 20 mM MOPS (3-[N-morpholino]-propane-sulfonic acid, Sigma), 5 mM NaAc 3H₂O, 1 mM EDTA (pH 7), 18% formaldehyde (BDH). 10 or 20 µg samples of total cellular RNA or A+ RNA prepared from 33.3 µg cellular RNA were heated at 65 ºC in 14 µl northern dye mix (10% 10X northern running buffer (table 2.2), 62.5% deionised formamide, 18% formaldehyde,
0.075% xylene cyanol, 0.025% bromophenol blue, 5% glycerol, 0.1 mM EDTA). After 5 min, the samples were quenched on ice, 0.5 μg of ethidium bromide added, and samples loaded alongside 5 μg of 0.24-9.5 Kb RNA ladder (BRL). The gel was electrophoresed in 1 X northern running buffer, 10 % formaldehyde at 30V overnight.

The gel was photographed on Polaroid 667 film while illuminated on a TM20 UVP 220 uv transilluminator. Prior to denaturation of the RNA in 50 mM NaOH, 10 mM NaCl, the gel was soaked in several changes of water. After 20 min of denaturation, the gel was rinsed in H2O and neutralised for 45 min in 500 mM Tris-HCl (pH 7), 1.5 M NaCl. Transfer of the RNA to nylon membranes (Amersham Hybond N) was achieved by capillary blotting for 12 hours. The gel was lain on top of a 3mm Whatman filter soaked in 10 X SSC, with wicks dipping into a tank of 10 X SSC. The gel was overlain with Hybond N and 2 pieces of 3mm also soaked in 10 X SSC, followed by a stack of dry paper towels, a glass plate and a 500g weight. Evaporation was reduced by covering the tank with Saranwrap.

Efficiency of transfer was checked by uv illumination of the ethidium bromide stained samples on the Hybond N, and the position of migration of the RNA ladder and ribosomal bands was marked on the filter. The filter was uv crosslinked for a total of 3 min and baked at 80°C for 2 hours.

Pre-hybridisation was performed in 10 ml of hybridisation buffer (table 2.2 ) in heat sealed plastic bags in a thermostatically controlled shaking H2O bath at 65°C. After 1 hour, 2.5 X 106cpm/ml hybridisation buffer of 32P labelled riboprobe, 200 μg/ml sonicated salmon sperm and 100 μg/ml yeast tRNA were heated to 95°C for 5 min and quenched on ice before addition to the hybridisation mixture. Hybridisation was performed for 12 hours at 65°C. Filters were washed in 1 X SSC, 1% SDS for 30 min at 65°C followed by 30 min in 0.1 X SSC, 1% SDS at 65°C, followed by an additional 15 min at 70°C for the p46 RNA tissue distribution northern (Chapter 7, figure 7.2C).

Damp filters were autoradiographed for 96 hours (p46 360), and 48 hours (human γ actin), and 72 hours for the A+ northern.

2.15. RNAse PROTECTION ANALYSES

20 μg RNA samples were mixed with human γ actin and p46 360 gel purified riboprobes in 40mM Pipes (pH 6.4, BDH), 1 mM EDTA (pH 8), 400 mM NaCl, 54 % deionised formamide. Samples were denatured for 5 min at 90°C, and annealed for 12 hours at 55°C. Samples were treated with 8 μg/ml RNAse A (Boehringer Mannheim), and 40 U/ml RNAse T1 (Boehringer Mannheim) in 300 mM NaCl, 10 mM Tris-HCl, 5 mM EDTA (pH 8) for 1 hour at 37°C. RNAse was destroyed by addition of 5 μg proteinase K in 0.6% SDS for 15 min at 37°C, and RNA was phenol chloroform extracted and precipitated with 20 μg yeast tRNA and ethanol on dry ice for 20 min. RNA was collected by centrifugation (20 min at 13 K rpm in a Sartorius microfuge), and dried under vacuum.
for 3 min before addition of 5 µl of 80% deionised formamide, 5% glycerol, 0.1 mM EDTA, 0.025% bromophenol blue and 0.075% xylene cyanol. Samples were loaded onto a pre-run 4% polyacrylamide, 7M urea, 1X TBE sequencing gel, and electrophoresed in 1X TBE at 40 W for 2 hours with 3P labelled pBLcat2 HindIII, EcoRI as markers. The gel was dried onto 3mm Whatman paper, and autoradiographed for 1 hour (γ actin signal) and 12 hours (p46 360 signal).

2.16. IN SITU HYBRIDISATION

The protocol for the preparation and sectioning of mouse embryos (CBA/Ca) and the in situ hybridisation details were taken from Wilkinson and Green (1990). Probes used were 35S [UTP] labelled sense and antisense p46 360 riboprobes (section 2.9.5), hybridisation was performed at 55°C, and washes were performed under high stringency conditions at 65°C. Ilford K5 emulsion was used to dip the slides for autoradiography under safelight conditions (sepia filter no. 914 from Ilford), and slides were developed in Kodak D19 developer after 17 days exposure at 4°C.

Sections were photographed using Kodak Ektachrome 5018 EPY 64T film, using a Zeiss photomicroscope III with illumination set at 12V. Bright field photographs were automatically exposed at ASA 25, and dark field photographs were taken after 2 sec exposure (6.5, 8.5 day embryos), 6 sec exposure (10.5 day embryos) and 8 sec exposure (12.5 and 14.5 day embryos) at ASA 100.

2.17. PROTEIN EXTRACTS

2.17.1. Preparation of Manley whole cell extracts

Whole cell extracts were prepared from 5mls of packed cells, exactly as described by Manley et al., 1980.

2.17.2. Preparation of large scale Manley extracts.

Large scale whole cell extracts for protein purification were prepared from 47 ml of packed cells, as above, except that solution C was replaced by 0.1 M col.buffer (table 2.4).

2.17.3. Preparation of Schöler micro whole cell extracts

Whole cell extracts from small numbers of cells or small quantities of tissue were prepared using the buffer described by Schöler et al., 1989, except that 0.5 mg/ml protease inhibitor (Sigma), and 1mg/ml trypsin inhibitor (Sigma) were used in place of pepstatin as protease inhibitors.

Extracts were prepared by freeze thawing 3 times in place of sonication, and were
either used immediately in gel retardation assays, or were frozen on dry ice, and stored at -70°C.

2.17.4. Preparation of micro nuclear extracts

Small nuclear extracts of tissues and cells were prepared by resuspending material in 400μl of buffer A described by Schreiber et al., 1989, with added protease inhibitors: 0.5mg/ml leupeptin (Boehringer Mannheim), 0.5mg/ml protease inhibitor (Sigma), 1mg/ml trypsin inhibitor (Sigma), 0.5mg/ml aprotinin (Sigma), and 40mg/ml bestatin (Boehringer Mannheim). After leaving on ice for 15 minutes, 25μl of Nonidet P40 (NP40, BDH) was added, the solution vortexed for 10 seconds, and microfuged for 30 seconds. An aliquot of the pellet was resuspended in PBS, and mixed with an equal volume of Unna stain (kind gift of E. De Boer, NIMR), and viewed under the light microscope to check the integrity of the nuclei. The nuclear pellet was resuspended in 50μl of the modified whole cell microextraction buffer, freeze thawed 3 times, and the supernatant stored at -70°C.

2.18. ANTIBODIES

2.18.1. Polyclonal antibodies

Anti-cyclin A antibody was raised against bovine cyclin A fusion protein. A cyclin A cDNA clone (residue 26 to the 3' non-translated end) was inserted into the T7 inducible open reading frame of pET56, protein expressed, and recovered from inclusion bodies by SDS polyacrylamide electrophoresis and electroelution. Rabbits were injected with 200 μg of protein at intervals of 4 weeks. This antibody was a kind gift from J. Adamczewski and T. Hunt, ICRF, South Mimms.

2.18.2 Monoclonal antibodies

XZ55 is a monoclonal antibody (subclass IgG1) raised against the carboxyl terminal 60 kD polypeptide of human pRb. This antibody will immunoprecipitate and Western blot pRb from a variety of species, and preferentially interacts with the underphosphorylated form of pRb (Hu et al., 1991). Kind gift from Q. Hu and E. Harlow, Massachusetts General Hospital.

IG4E was used as a control antibody for experiments incorporating XZ55, as it is of the same subclass (La Thangue et al., 1984). Kind gift from N. La Thangue, NIMR.

Anti-SSEA1 recognises a stem cell marker protein present on undifferentiated F9 EC cells, and lost on their differentiation. The antibody is described in Solter and Knowles, (1978) and was a kind gift from D. Solter.
2.19. IN VITRO TRANSCRIPTION AND TRANSLATION

4 μg template DNA was linearised (3' to the insert), ethanol precipitated, and resuspended in 5 μl H2O. 2μl 0.1M DTT, 2.5 μg BSA, 1μl each of 10 mM ATP, CTP, UTP, and 1mM GTP, 2 μl 5mM GpppG cap (Pharmacia), and 1 μl RNA guard were added to the DNA in a reaction volume of 20 μl of 1x transcription buffer (Pharmacia). 40 units of phage polymerase were added, and the reaction incubated at 37°C (T3 and T7 polymerases) or at 40°C (SP6) for 2 hours. Products were phenol extracted and ethanol precipitated, before resuspension in 30 μl DEPC treated dH2O.

6 μl RNA was heated, quenched on ice, and added to 10 μl of rabbit reticulocyte lysate (Promega), 0.4 μl RNA guard, 0.4 μl of amino acid mixture lacking methionine (Promega), and 3.2 μl of 35S-methionine (Amersham, 15 μCi/μl). The mixture was incubated for 1 hour at 30°C, and efficiency of translation assessed by SDS polyacrylamide electrophoresis of 2 μl of sample alongside 14C methylated protein molecular weight markers, followed by autoradiography.

For use in gel retardation assays, the translation was performed with 3.2 μl 150 μM methionine (Sigma) replacing the 35S labelled methionine.

2.20. GEL RETARDATION ASSAYS

Binding reactions were set up in a total volume of 20 μl, using 2 μg of sonicated salmon sperm DNA as non specific competitor DNA, with 6 or 12 μg of protein extract. Binding solutions were adjusted to reduce the total salt concentration of the mixture to 135 mM NaCl, in 50 mM Tris, 0.2mM EDTA, 1mM DTT, 15% glycerol. After preincubation at 30 °C for 10 minutes, approximately 3ng of Klenow end labelled probe was added, and incubation continued for a further 10 minutes. The reaction mixture was chilled before loading on a 4% (20:1 acrylamide: bis acrylamide) polyacrylamide gel and electrophoresed in Tris acetate EDTA buffer (pH 8.3). Electrophoresis was at 500 volts for 2 minutes, followed by 2.25 hours at 150 volts at 4°C, with recirculating buffer. Gels were vacuum dried, and exposed to Fuji RX 100 film at -70°C, with an intensifying screen.

In competition experiments, competitor DNA in molar excess of the probe was included at the preincubation step. Experiments incorporating polyclonal antibodies included 1 or 2 μl of preimmune or immune sera in the binding reaction. Experiments including monoclonal antibodies used 7 or 10 μl of antibody with an equivalent volume of unrelated antibody of the same isotype as control.

In experiments using blastocysts, the amount of salmon sperm DNA was reduced to 0.5 μg, and the extraction buffer was supplemented with 1 mg/ml BSA. Upto 10 μl of extract was used in the binding reaction, therefore salt concentration was raised to 200 mM NaCl.
2.21. UV CROSSLINKING ANALYSES

2.21.1 Identification of DNA binding proteins in complexes

Binding reactions were set up essentially as described for gel retardation assays using the 82/50 or 82/50 CRE-M probe (2.9.1). Just prior to electrophoresis, samples were uv irradiated for 20 minutes on a TM 20 UVP 220v transilluminator. After electrophoresis, the wet gel was exposed to Kodak XAR film for 1 hour at room temperature. Retarded complexes were excised, boiled in SDS loading buffer, and placed in the wells of a 7.5 % SDS polyacrylamide gel. The gel was electrophoresed at 60 volts overnight in SDS-PAGE running buffer. The gel was dried and autoradiographed.

2.21.2 Identification of purified DNA binding proteins

Double volume gel retardation reactions were set up with purified protein, 500 ng of salmon sperm DNA and competitor oligonucleotides. After 10 min incubation at 30°C, end labelled probe was added (50/82 CRE-M), and incubation continued for 10 min. Samples were uv illuminated for 30 min, and 3 µl of 25 mM CaCl₂, 50 mM MgCl₂ was added with 5 µg DNase1. After 10 min at 37°C, the reaction was stopped by precipitation with 8% trichloroacetic acid (TCA) for 20 min at room temperature. The precipitate was collected by a 13 K rpm 5 min centrifugation, and the pellet dried and resuspended in 30 µl of SDS PAGE loading buffer (table 2.2). The mixture was vortexed, heated at 90°C for 5 min, and loaded alongside ³C methylated proteins onto a 7.5 or 10 % SDS polyacrylamide gel, and electrophoresed at 20 mA for 6 h. The gel was transferred to 3mm Whatman paper, dried and autoradiographed.

2.22. GLYCEROL GRADIENTS

0.5 ml of 10, 20, 30 and 40 % glycerol solutions in 450mM NaCl, 20mM Hepes pH 7.9,1 mM DTT, 0.2mM EDTA were layered into 11x 34 mm polyallomer tubes (Beckman). Gradients were equilibrated overnight at 4°C, and 200 µl of sample (1 mg of whole cell microextract) in 10 % glycerol solution was layered on top of the tube. Gradients were centrifuged at 55K rpm for 8 hours at 4°C in the TLS-55 rotor of the TL-100 ultracentrifuge at acceleration and deceleration setting 5. After centrifugation, 100 µl aliquots were carefully taken from the top of the gradient, and stored at -70°C. 6 µl of each fraction was assayed by gel retardation using the 71/50 probe. In each experiment, 200 µl of molecular weight markers diluted to 0.5 mg/ml for each protein in 10 % glycerol solution were sedimented in parallel, and 100 µl aliquots stored at 4°C. 10 µl samples of the fractions were electrophoresed through 7.5 % SDS polyacrylamide gels, alongside 5 µg
samples of the individual protein species, and gels stained with coomassie blue. The protein size markers used were carbonic anhydrase (29 kD), BSA (66 kD), alcohol dehydrogenase (150 kD) and amylase (200 kD).

2.23. SIZE-EXCLUSION CHROMATOGRAPHY

A 30 ml column (30 cm X 1 cm diameter) and 20 ml column were packed with Sephadex S-100 (Pharmacia) and equilibrated in 0.1 M col.buffer. Bromophenol blue dye was added to molecular weight marker proteins (used in section 2.22.) dissolved in 0.1 M col.buffer (0.5 mg/ml), and the 200 µl sample applied at 1 ml/hour, and washed through with 0.1 M col.buffer. Fractions were collected, and the migration of the proteins used to calibrate the columns. The protein samples of interest were then applied to the columns under exactly the same conditions described above, and fractions assayed by gel retardation.

2.24. PROTEIN PURIFICATION

2.24.1. General

Eleven preparations of purified DRTF1 were undertaken. Each preparation was from whole cell extracts prepared from 47 ml F9 EC cells (1.2–1.5 g of protein, section 2.17.2). Pharmacia columns with adaptors were used, flow was generated by LKB pumps, and fractions were collected by Pharmacia and Gilson fraction collectors. Following loading of protein onto the columns, non binding proteins were washed off the column by use of 4 column volumes of 0.1M col.buffer, and bound proteins eluted with a linear gradient of increasing molarity of KCl generated by a gradient maker from 2.5 column volumes of both 0.1 M and 1 M col.buffers (table 2.4). All fractions were assayed for binding the E2A promoter probe in gel retardation assays (section 2.20), and for specificity of this binding by competition with 71/50 and 60/62 oligonucleotides.

2.24.2. Heparin Sepharose fractionation

A 30 ml heparin Sepharose (Pharmacia) column was packed and equilibrated with 0.1 M col.buffer (table 2.4). Approx. 100 ml of whole cell extract (section 2.17.2) was loaded onto the column at a rate of 5ml/hour, and the column washed at 10 ml/hour before elution of the bound proteins at 3 ml/hour. 5 ml fractions were collected during loading washing, and elution. The column was washed with 1 column volume of 1 M col.buffer and 3 volumes of 0.1 M col.buffer before re-use. Heparin Sepharose columns were re-used a maximum of 4 times. The protein concentration of fractions was
determined. Fractions with specific DRTF1 binding activity were pooled, dialysed for 8 hours against 2 changes of 0.1 M col.buffer, and clarified by a 10 min 10 K rpm centrifugation at 4°C in a Sorvall SS34 rotor.

2.2.4.3. Sequence specific affinity chromatography

Affinity matrix was prepared by coupling double stranded concatemerised binding site oligonucleotides to CNBr activated Sepharose 4B (Pharmacia). 2 mg of each complementary single stranded oligonucleotide was mixed and kinased, using a scaled-up version of the method described in section 2.9.2. Only 20 μCi [γ-32P] dATP was used, with 10 mM rATP, and 100 units of T4 polynucleotide kinase. The complementary single stranded oligonucleotides were annealed (section 2.2.1), and ligated in a volume of 100 μl of ligation buffer described in section 2.5.2, with 50 units of T4 DNA ligase. 1 μl of this reaction mixture was taken after overnight incubation at 14°C, and end labelled with klenow fragment as described in section 2.9.1. The labelled DNA was resuspended in 20 μl H₂O, 2 μl taken and mixed with 8 μl of sequencing loading buffer (table 2.3), heat denatured at 90°C, quenched on ice and loaded onto a 10% polyacrylamide, 7M urea, TBE sequencing gel. Kinased monomer oligonucleotide (section 2.9.2) was used as a monomer marker, and the gel was electrophoresed for 1.5 hours at 40 W. A 1 hour autoradiographic exposure of the gel was taken, and successful ligation of the oligonucleotides was indicated by a ladder of increasing length, up to 30 times that of the monomeric species.

The remainder of the ligated oligonucleotides were phenol extracted and precipitated, and resuspended in 100 μl H₂O. 1 μl of this solution was retained for scintillation counting. 5 g of CNBr activated Sepharose 4B was resuspended in 1mM HCl, washed with 600 ml of 1 mM HCl in a sintered glass funnel, followed by 600 ml of 1 mM KPO₄ (pH 8). 10 ml KPO₄ was added to the resin in a 50 ml tube, with the oligomerised oligonucleotides, and coupled for 16 hours at 4°C on a rotating wheel. The resin was collected on a sintered glass funnel, washed twice with 100 ml H₂O, and once with 100 ml 1M ethanolamine-HCl (pH 8, Sigma). The filtrate and washes were retained, their volume measured and an aliquot scintillation counted. The resin was inactivated by resuspension in 20 ml 1M ethanolamine-HCl (pH 8) in a 50 ml tube which was mixed by rotation for 5 hours at room temperature. The resin was once again washed on the sintered glass funnel with 100 ml 10 mM potassium phosphate buffer (pH 8, prepared from KH₂PO₄ and K₂HPO₄), 100 ml 1M potassium phosphate buffer (pH 8), 100 ml 0.1 M KCl, and 100 ml of 10 mM Tris-HCl (pH 7.6), 0.3 M NaCl, 1mM EDTA. The resin was resuspended in this final wash buffer, and 15 ml used to pack the first affinity column, and 10 ml for all subsequent columns (both wild type and mutant sequences). Efficiency of coupling of the DNA to Sepharose was calculated from the results of the scintillation counting. Coupling efficiencies varied from 60 to 90%, so resin carried 80 μg to 120 μg DNA /ml resin.
Packed columns were washed with 10 column volumes of 0.1 M col.buffer, before application of the pooled dialysed DRTF1 activity from the heparin Sepharose fractionation. Sonicated salmon sperm DNA was added to 0.1 mg/ml before the protein was applied to the first affinity column.

Protein was loaded at 5 ml/hour onto the first wild type site affinity column (15 ml col.volume), and at 3 ml/hour on the subsequent mutant site column (10 ml col.volume), and wild type site columns (10 ml). Fraction sizes during elution were of 1.5-3 ml. Fractions with DRTF1 binding activity in column gradient elutions were pooled and dialysed to 0.1 M KCl against 0.1 M Col.buffer, before addition of 200 μg/ml salmon sperm DNA and application to the next affinity column. DRTF1 binding activity was present in the flow through fractions from the mutant site column, and this activity was loaded directly onto the 3rd (WT) affinity column. Affinity columns were used a maximum of 6 times before replacement.

The enrichment of specific polypeptides during the fractionation was assessed by silver staining of proteins after electrophoresis through 7.5 or 10 % SDS polyacrylamide gels (section 2.26.2.), and by uv crosslinking of the DNA binding species (section 2.21.2). The active fractions from the 11 preparations were pooled and precipitated (section 2.25) prior to electrophoresis and coomassie staining (section 2.26.1).

2.25. PROTEIN PRECIPITATION

The efficiency of protein precipitation by various methods was assessed. Yields were quantitated by determining recovery of ¹⁴C methylated proteins after TCA precipitation (various concentrations of TCA) or acetone precipitation, in the presence or absence of carrier DNA (salmon sperm DNA). Highest yields of protein were obtained after TCA precipitation (at a concentration of 20%) in the presence of 20 μg/ml carrier DNA.

Precipitation was performed for 14 hour at -20°C in polypropylene centrifuge tubes (Nalgene, 50 ml), precipitate collected by 13 K rpm centrifugation for 30 min at +4°C in a Sorvall SS34 rotor. The precipitate was washed 5 times with 1 ml acetone to remove acid, air dried, and resuspended in SDS PAGE loading buffer prior to electrophoresis through a 10 % polyacrylamide, SDS gel.

2.26. STAINING OF PROTEINS

2.26.1. Coomassie staining

After electrophoresis, SDS polyacrylamide gels were fixed in 40% methanol, 10 % acetic acid for 20 min. Gels were coomassie stained in 0.25% coomassie R (Sigma), 10% acetic acid and 50% methanol for 30 min, and destained for 1 hour in 5% methanol, 10 % acetic acid.
**TABLE 2.4 PROTEIN PURIFICATION COLUMN BUFFERS**

**0.1 M Column Buffer**

- 20 mM Hepes pH 7.9
- 0.2 mM EDTA
- 1 mM DTT
- 20 % glycerol
- 100 mM KCl
- 0.5 µg/ml aprotinin
- 0.1 µg/ml trypsin inhibitor
- 0.1 µg/ml protease inhibitor
- 1 µM leupeptin
- 0.1 µM pepstatin
- 0.1 % NP40
- 6 mM MgCl₂

**1 M Column Buffer**

- 20 mM Hepes pH 7.9
- 0.2 mM EDTA
- 1 mM DTT
- 20 % glycerol
- 1 M KCl
- 0.5 µg/ml aprotinin
- 0.1 µg/ml trypsin inhibitor
- 0.1 µg/ml protease inhibitor
- 1 µM leupeptin
- 0.1 µM pepstatin
- 0.1 % NP40
- 6 mM MgCl₂
2.26.2. Silver staining of proteins

Silver staining of gels was performed using the Bio-Rad silver staining kit, according to manufacturers instructions. Briefly, small gels were fixed for 10 min in 40 % methanol, 15 min in 10 % ethanol, 5 % acetic acid. After draining the gel, 100 ml of oxidiser solution was added for 5 min, and the gel rinsed extensively in dH$_2$O for 10 min. 100 ml of silver staining reagent was added for 20 min, the gel rinsed quickly in dH$_2$O, and 100 ml developer added for 30 sec. The developer was changed two more times, until the desired band intensity was achieved. The reaction was stopped with 5% acetic acid.

2.27. PEPTIDE GENERATION

2.27.1. General.

Efficiency of peptide generation was checked by determining the quantitative recovery of peptides produced from $^{14}$C methylated proteins, or by using visual estimation of recovery of peptides by coomassie staining.

2.27.2. Cyanogen Bromide cleavage

Cyanogen bromide (CNBr) cleavage of protein is not efficient in acrylamide (I observed very low yields of peptides from BSA when I used the method of Nikoderm et al., 1979). Proteins were therefore resolved by electrophoresis, and recovered by blotting and elution. Protein was electrophoresed through 7.5 % polyacrylamide SDS gels, and blotted to nitrocellulose (Schleir and Schull, 0.45 μm) using a Biorad miniprotean II gel blotting apparatus. Three transfer buffers were tested: A) 10 mM CAPS (pH 11, 3-[cyclohexylamino]-1-propanesulfonic acid, Aldrich), 10 % methanol, B) 25 mM Tris, 192 mM glycine, C) 25 mM Tris, 192 mM glycine, 20% methanol. Blotting time was optimised to 1.5 hour on ice at 50V using ice cold buffer C, which gave 79 % overall recovery of protein (molecular weight range 14-200 kD) with higher recovery of smaller proteins (upto 94% for 14 kD protein).

The efficiency of protein elution from nitrocellulose was tested again using a variety of buffers: A) 2% SDS, 1% Triton X 100, 50 mM Tris–HCl (pH 9), B) 20 % acetonitrile, 1 % Triton X 100, 50 mM Tris–HCl (pH 9), C) 2 % SDS, 20 % acetonitrile, 50 mM Tris–HCl (pH 9), D) 40 % acetonitrile in dH$_2$O. Buffer C was the most efficient eluant, with 96.5% of 14-200 kD proteins eluted. The eluted protein was acetone precipitated (5:1 ratio of acetone : protein) at -20ºC overnight. The protein was recovered by centrifugation, and dried under vacuum. The pellet was resuspended as a 1% solution in 70 % acetic acid, and CNBr was added to 1 mg/ml. The reaction was degassed, and cleavage allowed for 24 hour at room temperature in the dark. Products were dried with
care, taken up in SDS loading buffer, and acetone precipitated before electrophoresis through a 15\% polyacrylamide, SDS gel. The initial volume of protein was run alongside the digested products to allow quantitation of the protein recovery by densitometry after autoradiography of the gel. Average yield of proteins was 31.5\%.

### 2.27.3 Proteolytic cleavage

Multiple samples of BSA (30 μg) were electrophoresed in a 10 \% polyacrylamide, SDS gel (Bio-Rad mini apparatus, 0.75 mm spacers), and excised after coomassie staining of the gel. Gel slices were rinsed in dH\(2\)O, and soaked for 20 min in 0.125 M Tris–HCl (pH 6.8), 0.1 \% SDS, 1 mM EDTA. These gel slices were inserted into the wells of a 15 \% polyacrylamide SDS gel poured with a large stacking gel (ratio of 2:3 compared with normal ratio of 1:4 stacking:separating gel). The Cleveland reaction was performed (Cleveland et al., 1977), by overlaying the gel slices with 10 μl of buffer containing 20 \% glycerol, 0.125 M Tris–HCl (pH 6.8), 0.1 \% SDS and 1 mM EDTA, and finally 10 μl of the same buffer but at 10 \% glycerol and including the proteolytic enzyme. Marker proteins were also loaded, and the gel electrophoresed at 50V in SDS PAGE running buffer until the dye front (from the coomassie stained protein) was close to the separating gel. The power was turned off for 2 hours to allow cleavage to occur, and peptides were electrophoresed at 100 volt through the separating gel. The gel was coomassie stained to visualise the peptides.

For cleavage of the purified polypeptide species, protein bands were excised from the coomassie stained gel. After soaking, the gel slices were inserted into the wells of a 15 \% polyacrylamide SDS gel (1.5 mm thick).

Samples of the protein alone (if available), and of the protease alone were electrophoresed in parallel. Large gels were electrophoresed at 50 mA, turned off for 2 hours, and products electrophoresed through the separating gel at 40 mA. After electrophoresis, the gel was cut to allow coomassie staining of the control lanes, and direct transfer of the purified protein cleavage products onto PROBLOT (ABI).

### 2.27.4 Transfer of peptides onto PROBLOT

The gel was soaked in cold transfer buffer (25 mM Tris, 192 mM glycine, 20 \% methanol). A piece of PROBLOT of the same dimensions as the gel was soaked in 100 \% methanol, and in transfer buffer. 3 mm paper (Whatman) was also soaked in transfer buffer, and blotting of peptides to PROBLOT was performed in a Bio Rad mini trans blot apparatus assembled according to manufacturers instructions. Transfer was carried out for 8 hour at 4°C at 200 mA. PROBLOT membranes were rinsed in dH\(2\)O, and methanol, and stained with PROBLOT stain (0.1 \% coomassie blue R–250, 1% acetic acid, 40 \% methanol, filtered through 0.45 μm filter) for 1 min, and destained with many changes of 50 \% methanol. The filters were rinsed with dH\(2\)O, before excision of the protein bands and
sequencing.

2.28. PROTEIN SEQUENCING

Pieces of PROBLOT were inserted into the cartridge of an Applied Biosystems automatic protein sequencer, and sequence analysis performed by A. Aitken and A. Harris, Lab of protein structure, NIMR.

2.29. COMPUTING

Protein and DNA sequence searches were performed using the GCG package (University of Wisconsin genetics computer group), searching EMBL and Leeds databases (Devereux et al., 1984). Secondary structure predictions were made using the Macvector protein toolbox facility, Macintosh (classic II). Protein homology between E2F-1 and p46 was assessed by BESTFIT (GCG), and displayed with a Pustell matrix (Pustell and Kafatos, 1984), with a window size of 8, hash value of 2, and minimum % score set at 40 %.
CHAPTER 3

EMBRYONIC REGULATION OF DRTF1 DNA BINDING ACTIVITY
3.1 Introduction

As discussed in section 1.21, DRTF1 is a sequence-specific DNA binding protein, first identified in F9 EC cells by binding to the adenovirus type 5 E2A promoter (La Thangue and Rigby, 1987). The binding activity of this protein decreased on differentiation of the cells to parietal endoderm (La Thangue et al., 1990), mimicking the regulation of the transcriptional activity of the E2A promoter measured in vitro, in protein extracts derived from these cells (La Thangue et al., 1990). DRTF1 is composed of three activities in F9 EC cells, the abundant b and c forms of the protein, and the less abundant DRTF1a which is b/c complexed with cyclin A (Bandara et al., 1991) and the retinoblastoma gene product (pRb) (Bandara and La Thangue, 1991, Bandara et al., 1991), or p107 (L. Bandara, unpublished observation). Affinity purified DRTF1b/c can activate transcription from the E2A promoter in vitro (Shivji and La Thangue, 1991), but the ability of the pRb complexed DRTF1a to activate transcription is reduced in vivo (Zamanian and La Thangue, 1992). The adenoviral 12S E1A protein can bind pRb, p107, (Whyte et al., 1989) and cyclin A (Giordano et al., 1989) proteins, and when added to DRTF1a can sequester these proteins from DRTF1a, producing the DRTF1b/c binding activity (Bandara and La Thangue, 1991). The abundance of DRTF1b/c and low levels of DRTF1a in EC cells correlates with the presence of a cellular E1A-like activity, suggested previously by the ability of E1A dependent adenovirus genes to be expressed in these cells (Imperiale et al., 1984). A cellular E1A equivalent has also been proposed to mediate transcriptional activation of E1A dependent promoters in preimplantation embryos (Suemori et al., 1988, Dooley et al., 1989), but the activity is down regulated after implantation.

I have examined the developmental regulation and tissue distribution of DRTF1 DNA binding activities during murine embryogenesis, and also during the differentiation of embryonal carcinoma and embryonic stem cell lines. Initially, I investigated the binding requirements of DRTF1 more closely to determine if there was any difference in the binding specificity of the DRTF1 DNA binding activities.

3.2 Sequence requirement for binding of DRTF1 in EC extracts.

DRTF1 in crude embryonal carcinoma (EC) cell extracts produced a DNAse I insensitive footprint extending from -71 to -60 on the non-coding strand of the adenovirus 5 E2A promoter (La Thangue et al., 1990). Gel retardation analysis of EC proteins on the 71/50 oligonucleotide probe (table 3.1), showed that DRTF1 was composed of three complexes of different mobility. Analysis of the exact sequence requirements for these interactions, by use of mutated sites in gel retardation competition analyses, indicated that
Table 3.1. Sequence requirements for DRTF1 DNA binding activity.

The inverted repeat of DRTF1 sites within the adenovirus type 5 E2A promoter are highlighted (−67 to −60, and −43 to −36). The upstream site has the higher affinity for DRTF1 binding, and therefore oligonucleotides used for gel retardation analyses encompass this site (71/50). Mutations within the 71/50 oligonucleotide are denoted by the position of mutation marked by an asterisk, and the mutated nucleotide(s) is highlighted.
E2A Promoter: TGGAGATGACGTAGTTTCGCGCTTAAATTGAGAAAGGGCGCGAAACTAG

<table>
<thead>
<tr>
<th>Variant</th>
<th>Sequence (with changes)</th>
<th>DRTF1 binding</th>
</tr>
</thead>
<tbody>
<tr>
<td>50/82</td>
<td>TGGAGATGACGTAGTTTCGCGCTTAAATTGAGAAAGGGCGCGAAACTAG</td>
<td>++++</td>
</tr>
<tr>
<td>50/82 CRE-M</td>
<td>TGGAGATGACGTAGTTTCGCGCTTAAATTGAGAAAGGGCGCGAAACTAG</td>
<td>++++</td>
</tr>
<tr>
<td>71/50</td>
<td>TAGTTTTTTCGCGCTTAAATTGAGAAAGGGCGCGAAACTAG</td>
<td>++++</td>
</tr>
<tr>
<td>64*</td>
<td>TAGTTTTTTCGCGCTTAAATTGAGAAAGGGCGCGAAACTAG</td>
<td>++++</td>
</tr>
<tr>
<td>63*</td>
<td>TAGTTTTTTCGCGCTTAAATTGAGAAAGGGCGCGAAACTAG</td>
<td>+</td>
</tr>
<tr>
<td>62*</td>
<td>TAGTTTTTTCGCGCTTAAATTGAGAAAGGGCGCGAAACTAG</td>
<td>+++</td>
</tr>
<tr>
<td>61*</td>
<td>TAGTTTTTTCGCGCTTAAATTGAGAAAGGGCGCGAAACTAG</td>
<td>++++</td>
</tr>
<tr>
<td>60*</td>
<td>TAGTTTTTTCGCGCTTAAATTGAGAAAGGGCGCGAAACTAG</td>
<td>++</td>
</tr>
<tr>
<td>60/62</td>
<td>TAGTTTTTTCGCGCTTAAATTGAGAAAGGGCGCGAAACTAG</td>
<td>-</td>
</tr>
</tbody>
</table>

**TABLE 3.1**

Sequence requirements for DRTF1 binding activity
the binding activity had a 5' border at -67, and a 3' border downstream of -58, and that some nucleotides, but not all within the G/C rich region -60 to -64 were necessary for binding (La Thangue et al., 1990).

Closer examination of the binding requirements within the -60 to -64 region indicated that mutation of nucleotides -60 or -62 reduced binding, whereas mutation of -61 had little effect on binding activity. Mutation of -63 showed a reduction in binding activity to levels marginally higher than the -60/-62 mutation. These data are summarised in table 3.1. The binding activity of all species of DRTF1 to the mutant sequences was tested by gel retardation assay. All DRTF1 activities shared identical sequence requirements for binding.

3.3 Differentiation of EC cells

A series of whole cell extracts was prepared from EC cells over a 7 day course of differentiation (Strickland et al., 1980). The state of differentiation of these cells was assessed by visual inspection and by using an antibody which recognises the embryonic stage specific stem cell marker SSEAl (Solter and Knowles, 1978). After 2 days of differentiation, most (approx. 90%) of the cells no longer expressed SSEAl, and after 5 days of differentiation, the stellate, flattened morphology of the differentiated parietal endoderm-like cells was clearly visible.

3.4 Regulation of DRTF1 binding activities during F9 EC cell differentiation.

The oligonucleotide probe 71/50, encompassing the upstream DRTF1 binding site of the E2A promoter (table 3.1), was used to investigate the regulation of DRTF1 during EC cell differentiation. After 5 days of differentiation, DRTF1 binding activity was reduced (La Thangue et al., 1990). The regulation of the three individual species of DRTF1 binding activity was determined during a time-course of differentiation (figure 3.1A, lanes 1-10). DRTF1b/c was abundant in F9 EC cells whereas the slower migrating la complex was not. As EC cells started to differentiate, there was a rapid reduction in the binding activity of DRTF1 b/c, whereas the a complex was not down regulated, and appeared to be slightly induced (compare track 3 with track 2). This trend of reduction in b/c binding activity continued throughout differentiation, when DRTF1b was most strongly down regulated. After 7 days of differentiation, the binding activity of DRTF1a and c were comparable, and DRTF1b was the least abundant activity (tracks 5 and 10). Therefore, during EC cell differentiation, DRTF1b and c were both down regulated, and DRTF1a was initially upregulated, and henceforth maintained above the level of DRTF1a in EC cells. This analysis of DRTF1 during differentiation of F9 EC cells suggested that the activities were subject to different regulatory cues.

The integrity of the cellular extracts was checked by investigating the binding
Figure 3.1. Regulation of DRTF1 during F9 EC stem cell differentiation.

A. DRTF1 is regulated on EC cell differentiation.

F9 EC stem cells were grown as stem cells, or differentiated (section 2.10.4) for the indicated times. Manley extracts were prepared (section 2.17.1), and assayed for DRTF1 binding activity (lanes 1-10), ECRE2 (lanes 11-15), or octamer binding activities (lanes 16-20) in gel retardation assays (section 2.21) using the binding site probes as listed in section 2.2.1. Tracks 1, 6, 11, and 16 show probe alone. Note that DRTF1b and c activities were down regulated whereas DRTF1a was induced by 3 days of differentiation (compare tracks 2 and 3). ECRE2 was constitutively active (tracks 12-15), but Oct 4 was detected only in F9 EC stem cells (track 17). Tracks 1-5 show an increased exposure of tracks 6-10.

B. Specificity of the complexes

Specificity of binding to the 71/50 binding site was tested (lanes 2-7) in EC cell extracts (lanes 2-4), and in extracts prepared from cells after 3 days of differentiation (lanes 5-7). Wild type (WT) and mutant (M) competitors were 71/50 and 63* respectively. ECRE2 (tracks 8-11), and octamer binding proteins (tracks 12-14) were also tested for specificity of binding, with P and PM1, and Oct and OctM1 competitors respectively. All competitors were used at a 50 fold molar excess, and * indicates a non specific activity. 12 μg of protein was used for binding to DRTF1 and octamer sites, and 6 μg of protein for P motif binding activities.
activity ECRE2 with the oligonucleotide probe P, which consisted of a consensus ATF site (lanes 11-15). This ATF activity is not regulated during differentiation of F9 EC cells (Tassios and La Thangue, 1990), and was relatively constant during this analysis.

Binding to a consensus octamer site was also investigated, as in addition to the ubiquitous Oct 1 protein, the Oct 4 protein is present in EC cells (Schöler et al., 1989; Oct 3 in Rosner et al., 1990). As expected, in EC cells Oct 4 and Oct 1 activities were present (lane 17). Oct 1 was present throughout the differentiation, whereas Oct 4 was rapidly down regulated, so that no binding activity was detectable after three days of differentiation (lane 18). Oct 4 and DRTF1b therefore showed a similar regulation during the differentiation of EC cells.

The specificity of binding of the various DRTF1 binding activities was assessed by competition (figure 3.1B). Competition with the wild type site abolished binding of DRTF1b/c to the probe (lane 3), whereas competition with a fifty fold excess of the mutant site marginally reduced the binding activity (lane 4). The specificity of DRTF1a was tested using extracts prepared from cells differentiated for three days (where it was most abundant, PE 3, lanes 5-7). DRTF1a was specifically competed only by the wild type site (compare lanes 6 and 7).

ECRE2 binding was specific for the P motif as competition with the homologous probe led to a loss of binding activity (lane 10), whereas competition with the mutant site PM1 (M) did not (lane 11).

The specificity of binding of the various octamer species was also assessed by competition (lanes 12 to 14), both Oct 1 and Oct 4 were found to be specifically competed only by the wild type and not by the mutated site.

3.5 DRTF1 is present in ES cells and is regulated on differentiation

Embryonal carcinoma cells are transformed cells, and lose their transformed phenotype on differentiation (Strickland and Mahdavi, 1978). The loss of DRTF1 binding activity could therefore be attributed to changes occurring during the differentiation process, or as a result of the loss of transformed phenotype.

Earlier studies within the laboratory (La Thangue and Rigby, 1987) showed that DRTF1 was present in a characteristic three complex form only in undifferentiated cells; differentiated SV40 transformed 3T3 cells (c138, Rigby et al., 1980) did not show this pattern. It was therefore of interest to determine if DRTF1 was present in totipotent embryonic stem cells, which are undifferentiated and non-transformed cells, and whether DRTF1 binding activity was regulated upon differentiation of these cells.

Embryonic stem cells (figure 3.2A, 1) were allowed to differentiate by culturing in the absence of STO cells (Gossler et al., 1986). The cells form embryoid bodies, which are
Figure 3.2.A. Embryonic stem cell differentiation

1) ES cells grown on STO cells. 2) Embryoid body produced after 8 days of differentiation. Note the cystic nature of the body, and the variety of cell types present. 3) Embryoid body after 10 days of differentiation, with a blood island (BI) present. 4) The blood island at higher magnification (see section 2.10.7).
composed of a variety of different cell types, representing all three embryonic germ layers (Doetschman et al., 1985). After two days of differentiation, the bodies formed an outer rind-like layer, the Reichert membrane, and by the seventh day, they had become cystic bodies. Differentiation into cardiac muscle often occurred by eight days (figure 3.2A, 2), and by ten days (figure 3.2A, 3 and 4), blood islands were formed in some cultures.

Extracts were prepared from these embryoid bodies during differentiation, and assayed for the presence of binding activity in the presence of competing mutant binding site (figure 3.2B). As in EC cells, the predominant forms of DRTF1 in ES cells were the b/c complexes, with little DRTF1a binding activity (lane 1 and 6). Differentiation of the cells resulted in a marked decrease in the binding of b/c complexes by 2 days of differentiation (compare lanes 2 and 1). Levels of DRTF1 binding activity varied at later stages of differentiation and between different experiments, despite consistent culture conditions, but DRTF1 binding activity at any stage of differentiation was consistently less than in ES cells. The reason for the variability in DRTF1 binding activity during differentiation, and between experiments was probably due to differences in the final phenotype of the cells produced. Visual examination of the cells showed beating cardiac muscle, striated muscle, and epithelial and neuronal like cells, which were present in different proportions in different cell cultures.

The binding activity of other transcription factors was also assessed during ES cell differentiation. ECRE2 was present in ES cells, and its binding activity was slightly increased over the time course (lanes 11-15), in contrast to Oct 1 which showed a reduction in binding activity during ES cell differentiation (figure 3.2C, lanes 2-5). Interestingly, Oct 4 was not as rapidly down regulated on differentiation of ES cells as on differentiation of EC cells, with binding activity still present after two days of ES cell differentiation, compared to loss after one day of EC cell differentiation (data not shown). GATA binding activity was induced (Tsai et al., 1989) within these cultures after eight days of differentiation. GATA-1 is a factor enriched in the haematopoietic cell lineage, and the timing of the appearance of the GATA binding activity preceded the appearance of blood islands in these bodies at ten days of differentiation (figure 3.2C lane 5, figure 3.2A, 3 and 4).

The regulation of DRTF1 binding activity during ES cell differentiation suggested that the reduction of DRTF1 during EC cell differentiation was due to a loss of undifferentiated phenotype, and not a result of reversion of the transformed phenotype.

3.6 DRTF1 is present in blastocyst stage embryos

DRTF1b/c is present at high levels in EC cells and in ES cells, which are thought to resemble (Solter and Knowles, 1978) or are derived from (Evans and Kaufman, 1981) the inner cell mass of the blastocyst. It was therefore of interest to assess directly if
ES cells were grown and differentiated for the indicated times, extracted, and assayed for DRTF1 (tracks 1-10, probe 71/50) or ECRE2 (tracks 11-15, probe P) binding activity, in the presence of 50 fold excess of 60/62 or PM1 respectively. DRTF1 was down regulated within 2 days of differentiation, although the activity fluctuated in differentiating cells (for example, tracks 4 and 9). Tracks 6-10 show an increased exposure of tracks 1-5. 12 μg of protein was assayed for DRTF1 and 6 μg was assayed for ECRE2 binding activity.

C. GATA binding factor is induced as ES cells differentiate

β Globin probe (which encompasses GATA and octamer binding sites) was used to assess binding activities on ES cell differentiation. Tracks 2-5 show the regulation of binding activities after 2-8 days of differentiation. GATA binding activity was induced by 8 days, and Oct 1 binding activity decreased during differentiation. Oct 4 binding activity was lost between 2 and 4 days of differentiation. Tracks 6-11 show the specificity of binding activities in the 8 day extract. Oct 1 was competed by titration of the Oct competitor (tracks 6-8), and GATA binding was abolished by titration of GATA site competitor (tracks 9-11). 12 μg of protein was used in each track.
DRTFl binding activity was present in blastocyst stage embryos. Whole cell extracts were prepared from 3.5 dpc blastocysts collected from superovulated mice, and assayed for DRTFl and other transcription factor binding activities. As predicted from the stem cell studies, DRTFlb did appear to be present in blastocysts (figure 3.3, compare track 2 and 3). The specificity of the binding activity was tested by assaying binding to the mutant site 63* (table 3.1), which confirmed that the activity binding the 71/50 probe had DRTFl specificity (track 4). Another slower migrating complex also bound the 71/50 probe specifically, but the nature of this species is unknown.

ECRE2 binding activity was present in blastocysts, binding specifically to the P but not PM1 probe (figure 3.3, lanes 5 and 6). In addition, as published by Schöler et al., 1990, Oct 4 binding activity was present in the blastocyst.

Therefore, DRTFlb, Oct 4 and ECRE2-like activities are present in the blastocyst. The presence of DRTFlb at this stage of embryogenesis is consistent with the abundance of DRTFlb in stem cells. The predominant form of DRTFl in blastocysts is the b complex, but because of the very small amounts of protein assayed, it is possible that other less abundant forms of DRTFl also exist in the 3.5 dpc embryo.

3.7 DRTFl is regulated during murine embryogenesis

The EC and ES cell differentiation studies indicated that the different forms of DRTFl were independently regulated, as b/c forms of DRTFl which were abundant in stem cells decreased on differentiation, whereas DRTFla binding activity increased on differentiation of EC cells.

To assess the regulation of DRTFl activities during murine embryogenesis, extracts were prepared from whole embryos at 8.5, 11.5, 14.5, and 17.5 dpc, and assayed for DRTFl binding activity in the presence of either a wild-type or mutant competing DRTFl binding site (figure 3.4a). In 8.5 dpc embryos, DRTFlb/c were the abundant forms (lane 3), and as embryogenesis progressed, levels of binding activity decreased. In contrast, levels of DRTFla increased during development, and between 8.5 and 14.5 dpc (lanes 3-6), the mobility of the a complex increased to a' (a switch that occurred at different times in different tissues, see later). The regulation of DRTFlb/c was therefore inversely correlated with the regulation of the complexed forms of DRTFl during this window of development. The DNA binding activity of ECRE2 was also assayed during this time period, and specific binding activity was upregulated (figure 3.4b).

DRTFl is therefore not only a differentiation regulated transcription factor, but is also developmentally regulated. The increase in DRTFla/a' binding activities during development suggested that there was also a developmental regulation of protein complexing with DRTFlb/c.
Figure 3.3. DRTF1 and ECRE2-like binding activities are present in the blastocyst.

Approximately 40 blastocysts were isolated from superovulated mice, extracted, and assayed for specific DRTF1 binding activity on the 71/50 (track 2) and 63* (track 4) mutant site probe. Track 3 shows F9 EC cell extract, and 71/50 probe alone is shown in track 1. The arrow indicates DRTF1b in blastocyst and EC extracts, and * indicates non specific DNA binding activity. The more slowly migrating complex seen only on the 71/50 probe (track 2) was specific for the DRTF1 binding site, but it was not always observed in blastocyst extracts.

3.5 dpc blastocyst extract was also assayed for binding to the P and PM1 probes (tracks 5 and 6). A specific ECRE2-like activity was abundant in blastocysts, and a shift of ECRE1 mobility was also reproducibly observed.
Figure 3.4. DRTF1 is regulated during murine embryogenesis.

Extracts were prepared from whole embryos at the indicated times, and assayed for DRTF1 (a), and ECRE2 (b) binding activities, in the presence of either mutant or wild type sequences (tracks 1-6, and 7-11 respectively). For comparison, an EC cell extract was assayed in lane 2, and lane 1 shows probe alone. DRTF1 and ECRE2 specific complexes were abolished by competition with wild type but not mutant competitors. Note that the activity of DRTF1b/c decreases during development, whereas the binding of the a and a' forms increases. 1a' appears between 8.5 and 14.5 days, and replaces a in older embryos. Approximately 12 µg, and 6 µg extract were assayed for DRTF1 and ECRE2 respectively.
(a) DRTF1

(b) ECRE2

Legend:
- **a**: Probe
- **a'**: Competitor
- **b/c**: d.p.c.

<table>
<thead>
<tr>
<th></th>
<th>Probe</th>
<th>Competitor</th>
<th>d.p.c.</th>
</tr>
</thead>
<tbody>
<tr>
<td>EC</td>
<td>8.5</td>
<td>11.5</td>
<td>14.5</td>
</tr>
<tr>
<td>F9</td>
<td>11.5</td>
<td>14.5</td>
<td>17.5</td>
</tr>
<tr>
<td>63*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>71/50</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Legend:
- **PM1**: Probe
- **P**: Competitor
- **d.p.c.**: d.p.c.
3.8 DRTF1 binding activity is tissue-dependent in the adult.

The down regulation of DRTF1 binding activity during embryogenesis, and the low levels of DRTF1 in 17.5 dpc embryos suggested that DRTF1 might not be present in the adult mouse. To test this possibility, extracts were prepared from various tissues of 6 week old mice, and assayed on the 71/50 binding site (figure 3.5A).

There was a striking difference in the levels of binding activity in the different tissue extracts, with no binding activity present in adult liver or brain, and large amounts of a' present in adult thymus. Spleen had DRTF1b and a' binding activity, and testis and ovary had predominantly a' and c, and a', b and c activities respectively.

The lack of binding activity in some tissues could have been caused by degradation of the activity during extract preparation. To control for this possibility, the binding activity of ECRE2 and octamer proteins was also assayed (figure 3.5B and C). These proteins had very different tissue dependent patterns of binding activity. However, large amounts of ECRE2 were present in testis, ovary and brain, confirming the integrity of these extracts.

DRTF1 therefore showed a tissue-dependent pattern of binding in the adult, with DRTF1a forms being the predominant binding activities.

The low level of all three DNA binding activities in the liver extract was a cause for concern. Further experiments were performed to ensure that the extract had not been degraded by proteolytic action.

3.9 Adult liver extract is not degraded

Addition of adult liver extract to EC cell extract did not result in a reduction of DRTF1 binding activity (figure 3.6A), or of ECRE2 or octamer binding activities (figure 3.6B and C). Therefore proteases capable of reducing binding activity were not present in the liver extract. This result contrasts with the extensive degradation of EC binding activities that was observed when adult bone marrow extract was added to EC extract. However, this analysis did not address the possible problem of proteolysis occurring during the preparation of the liver extract. Proteolysis did not occur during extract preparation as coextraction of EC cells both with and without liver, yielded equivalent binding activity. Therefore the lack of DRTF1 and low levels of Oct 1 and ECRE2 binding activities are probably a true reflection of the levels of binding activities in adult liver.

3.10 Tissue distribution of DRTF1 at 11.5 and 14.5 dpc

The strong regulation of DRTF1 binding activities between different tissues of the adult could be a reflection of absolute tissue dependent DRTF1 binding activity, or could be caused by stronger developmental down-regulation of the binding activities in some tissues (for example brain), than in thymus. The tissue dependence of DRTF1 was
Figure 3.5. Tissue dependent regulation of DRTF1 binding activity in the adult mouse.

Tissues from adult mice were dissected, whole cell microextracts were prepared, and analysed for DRTF1, ECRE2, and octamer binding activities (A, B, and C respectively). Binding activity in F9 EC cell extracts was used as a control (lane 2), and activities in thymus, spleen, testis, ovary, liver and brain extracts (lanes 3-8 respectively) were examined. DRTF1a was not present in adult tissues, only the 1a' form was seen. This activity was abundant in thymus (lane 3), and was present along with other forms of the activity in spleen, testis and ovary. DRTF1 was undetectable in liver and brain, although the brain extract had ECRE2 binding activity. ECRE2 and octamer binding activities were also strongly tissue regulated, although the regulation of each factor was different. Additional octamer binding proteins were present in some tissues (eg brain, lane 8, figure C), as shown by others (Schöler et al., 1989). 12 μg of extract was assayed for DRTF1 and octamer binding activities, and 6 μg for ECRE2. * represents non-specific activity.
Figure 3.6. Adult liver extract is not degraded.

12 μg of adult liver extract was added to a titration of F9 EC cell extract (2-8 μg, lanes 6-10), and DRTF1, ECRE2 and octamer binding activities compared with the titration of EC extract alone (lanes 1-5, figures A, B, and C respectively). There was no reduction in any of the EC binding activities in the presence of liver extract. (The ECRE2 binding activities were assayed in a titration of 1-4 μg of protein, with 6 μg liver extract added).
therefore examined at earlier time points during embryogenesis, at 11.5 and 14.5 dpc (figures 3.7 and 3.8).

Extracts were prepared from embryonic and extraembryonic tissues at 11.5 days, and were assayed for DRTF1 binding activity. DRTF1 is abundant in whole embryos prepared from this stage, and this abundance was reflected in liver and brain tissue, with binding activity also present in somites, limb bud and gut. DRTF1b/c were the major forms of the binding activity, with high levels of DRTF1b and both b and c present in liver and brain respectively (figure 3.7, lanes 8 and 9). None of the tissues examined showed much of the DRTF1a activity. Extraembryonic tissues had very low levels of DRTF1 binding activity (lanes 5 and 6). Genital ridge, which at this stage is being populated by primordial germ cells, showed little DRTF1 binding activity (lane 7).

DRTF1 binding activity is therefore expressed in a tissue dependent fashion during embryogenesis, but the binding activity is not restricted to certain tissues throughout development. For example, DRTF1 is abundant in embryonic liver and brain, but is not present in those tissues in the adult.

DRTF1 binding activity was also assayed at 14.5 dpc. DRTF1b and c forms were widespread at this stage, with the highest level of binding activity in the liver. DRTF1 levels in the liver increased between 11.5 and 14.5 dpc, whereas binding activity in the brain decreased during this time period. However, DRTF1 binding activity was present in the spinal cord at 14.5 dpc, whereas it was absent at 11.5 dpc. Both DRTF1b and c forms were present in the testis, whereas only the c form is maintained with a' in the adult. Binding activity in the limb was also reduced at 14.5 days compared to 11.5 days.

There was no correlation of the levels of DRTF1 binding activity with differentiation of tissues from different germ layers of the embryo. The regulation of DRTF1 binding activity was therefore next examined in four individual tissues whose developmental biology is well established.

3.11 Tissue-dependent regulation of DRTF1

Each tissue examined had its own characteristic profile. In liver, DRTF1b and c were abundant between 11.5 and 14.5 days, when liver starts to establish hepatic function, undergoes maximum growth and differentiation and is the major site of haematopoiesis (reviewed by Shafritz and Panduro, 1988). Later, as the liver became more differentiated, there was a decline in DRTF1 binding activity, and DRTF1 was undetectable in the adult tissue. Although b and c were more abundant than a and a', (a' appeared by 14.5 dpc in liver), all activities were quantitatively down regulated in a similar fashion after 14.5 dpc. This contrasts with the earlier observations from EC cell differentiation where regulation was non coordinate. ECRE2 binding activity was also examined during liver differentiation, it was abundant at 11.5 days, but down regulated to a low level of binding.
Figure 3.7. Tissue distribution of DRTF1 at 11.5 dpc.

Tissues were dissected from 11.5 dpc embryos, and assayed for DRTF1, ECRE2 and octamer binding activities (A, B, and C respectively). Lane 1 shows the probe alone, and lanes 2 and 3, EC and adult thymus extracts to serve as controls. DRTF1 was abundant in whole embryo extracts of this stage (lane 4), and DRTF1 binding activity was present in liver, brain, gut, somites and limb bud (lanes 8,9,11,12,13). Octamer and ECRE2 activities showed different tissue dependent regulation of binding. 12 μg of extract was assayed for DRTF1 and octamer activities, and 6 μg for ECRE2.
Figure 3.8. Tissue distribution of DRTF1 at 14.5 dpc.

Binding activities of 14.5 dpc embryonic tissues were assayed on 71/50, P, and octamer probes. DRTF1 binding activity was widespread at this stage. Lane 1 shows probe alone, and 2 and 3 EC and adult thymus binding activity respectively. * represents non-specific binding activity. 12 μg of extract was assayed for DRTF1 and octamer binding activities, and 6 μg was assayed on the P motif for ECRE2.
activity in the adult.

The binding activity in embryonic brain was also examined. Again, all forms of DRTF1 were quantitatively down regulated during embryogenesis, and were virtually undetectable in the adult tissue (figure 3.9b, tracks 3-6, probe 71/50). As with liver, the regulation of DRTF1 binding activity correlated with maximal growth and differentiation (Rugh, 1990). However, in embryonic brain, the predominant form of DRTF1 was c, in contrast to the situation in liver, EC and ES cells, where b was most abundant. ECRE2 was present throughout brain development, but was reduced in adult tissue.

The strong down regulation of DRTF1 during brain and liver development and low levels of binding activity of DRTF1 and ECRE2 in adult tissue could be attributable to cytosolic dilution of the binding activity in adult tissue. There was however a more marked down regulation in the binding activity between 11.5 and new born mouse tissue when nuclear, rather than whole cell extracts were prepared, indicating that cytosolic dilution was not responsible for the observed regulation. This experiment, along with the data showing that DRTF1 is expressed at high levels in some adult tissues allowed me to conclude with reasonable confidence that the regulation of DRTF1 in these extracts reflected the regulation of DRTF1 in vivo.

DRTF1 binding activity was next assayed in testis, which at 14.5 dpc contained DRTF1a, b and c, but only a' and c in the adult tissue. DRTF1 binding activity was down regulated during testis development, although still detectable in adult tissue (figure 3.9c, lanes 3-5, probe 71/50). Again DRTF1c was more abundant than DRTF1b, and in this respect testis and brain clearly resemble each other. In testis, ECRE2 binding activity was constitutive over this time period (figure 3.9c, lanes 3-5, probe P).

Finally, the level of binding activity in thymus was investigated. At 17.5 dpc, DRTF1a' and b were present, yet in contrast to other tissues (except adult spleen figure 3.5A lane 4) DRTF1c was absent (figure 3.9d, lane 3, probe 71/50). Adult thymus extracts had much higher levels of DRTF1a' binding activity than any other tissue examined, and also had DRTF1b (figure 3.9d, lane 4, probe 71/50).

In summary, DRTF1 binding activity is not restricted to embryonic stages when pluripotent stem cells are present, since it was present in tissues that contain stem cells with limited potency such as liver and brain. Secondly, DRTF1 is regulated as tissues undergo maximal proliferation and differentiation, and thirdly, different forms of DRTF1 were present in different tissues. DRTF1 is therefore a tissue dependent, developmentally regulated binding activity.
Figure 3.9. Tissue dependent regulation of DRTF1.

Whole cell extracts were prepared from liver (a), brain (b), testis (c), and thymus (d) at the indicated times, and assayed for DRTF1 (probe 71/50), or ECRE2 (probe P) binding activities. For comparison, an F9 EC cell extract was assayed in lane 2, and the probe alone was shown in lane 1. DRTF1b was down regulated in all of the tissues studied, whereas DRTF1c was abundant in adult testis and present at 17.5 dpc in brain. Testis, thymus and liver (after 14.5 dpc) contained DRTF1a*. Non specific complexes are shown by *.

12 µg of extracts were assayed for DRTF1 and octamer activities, whereas 6 µg were assayed for ECRE2.
3.12 Discussion

During the course of this work, much care was taken to ensure that the developmental regulation of the DRTF1 DNA binding activities was a true reflection of the regulation, and not an artefact caused by degradation of the extract, differential extraction, or cytosolic dilution of the activity. The characterisation of several other DNA binding activities with different developmental regulation, the combination of techniques used to check the integrity and efficiency of extraction, and the fact that some adult tissues possessed large amounts of DRTF1 binding activity, convinced me that true down regulation of DNA binding activity occurred during development of some tissues.

During differentiation of F9 EC cells, the abundant b/c forms of DRTF1 were down regulated. The same result was obtained during ES cell differentiation, although at stages during the differentiation process, there was an increase in the binding activity of DRTF1. ES cells are pluripotent, therefore numerous cell types are produced on their differentiation, and presumably levels of DRTF1 binding activity varied between these cell types, resulting in the non-linear regulation of DRTF1. Different cultures gave slightly variant results, possibly correlating with differences in the cell types produced; for example, not all cultures produced blood islands, or gave rise to myogenic cells.

ES cells are derived from the inner cell mass of the blastocyst, therefore it was reassuring to observe DRTF1b binding activity in extracts prepared from blastocysts, in addition to the abundant ECRE2-like activity. Other forms of DRTF1 binding activity may be present in blastocysts, but b was seen as it is the predominant form of the activity in stem cells, and small amounts of material were assayed.

DRTF1 was therefore abundant in all stem cells examined, and levels of binding activity decreased on differentiation of most tissues. However, complexing of DRTF1 with non-DNA binding proteins increased during the development of some tissues. Some complexing was seen at 8.5 dpc, the earliest stage during post-implantation development examined. It is therefore possible that the proposed cellular E1A-like activity involved in the maintenance of DRTF1b/c in the uncomplexed form in EC cells (Bandara and La Thangue, 1991), also acts in a similar manner during early preimplantation development, and is absent or reduced in the post-implantation embryo (as suggested by Suemori et al., 1988, and Dooley et al., 1989). Unfortunately there was not enough material available from embryos at earlier stages of post-implantation development to test for the presence of complexed DRTF1, as it is likely to be scarce.

DRTF1 binding activity was greatly reduced in tissues which undergo terminal differentiation, such as liver and brain, but higher levels of the complexed activity were retained in adult thymus and testis. Adult thymus contains many maturing thymocytes, which undergo an average of eight mitotic divisions with very short duration (8 hours); that is they are a population of very actively dividing, differentiating cells (Kendall, 1981). The
other adult tissue with high levels of DRTF1 binding activity is testis. Spermatocyte production involves coordinate waves of mitosis and meiosis from progenitor cells along the seminiferous tubules, and this division and differentiation of spermatocytes is maintained throughout adulthood. Therefore two of the tissues of the adult mouse which possess DRTF1 binding activity, contain cells which are undergoing division and differentiation at faster rates than in tissues such as liver and brain which lack DRTF1 activity. Further characterisation of the role played by DRTF1 in these cellular populations requires the isolation of cDNAs encoding DRTF1. The nature of the various DRTF1 complexes is investigated in chapter 4, and a fuller discussion of the results of this chapter in the light of those results is presented in the general discussion, chapter 8.
CHAPTER 4

A BIOCHEMICAL DISSECTION OF THE DRTF1 COMPLEXES
4.1. Introduction

During murine embryogenesis, four forms of DRTF1 were identified. These were regulated independently during the development of different tissues, although during embryogenesis there was an enrichment of DRTF1a binding activities.

At the time that this work was performed, initial purification studies from F9 EC cell extracts had shown that DRTF1a was lost during purification, and that b/c complexes were composed of two DNA binding polypeptides of 50 and 30 kD (Shivji and La Thangue, 1991). DRTF1a could be disrupted by detergent treatment to produce a faster migrating shift (similar to a'), and DRTF1a binding was prevented by phosphatase treatment (Shivji and La Thangue, 1991).

I was interested to determine the relationship between the DRTF1a/a' and b/c activities. Two models could be proposed for the different forms of DRTF1. The different complexes could be formed by distinct polypeptides with a related DNA binding motif (allowing identical sequence specificity of binding) as shown by the octamer family of proteins (Sturm and Herr, 1988, Scheidereit et al., 1988, Schöler et al., 1990). Alternatively, a common DNA binding polypeptide(s) associates with other non-DNA binding proteins to form multicomponent DNA binding complexes which recognise the same sequence.

DNA crosslinking assays were performed to address these possibilities, and the relative molecular weights of the complexes were assessed by glycerol gradient centrifugation and size exclusion chromatography.

4.2. DNA binding components of DRTF1

The size of the DNA binding components in each of the DRTF1 complexes was determined by crosslinking the polypeptides to a bromodeoxyuridine substituted probe by uv irradiation. The probe used previously for determining sizes of purified DRTF1 DNA binding proteins spanned -82 to -50 of the E2A promoter, encompassing both the CRE and DRTF1 binding sites (Shivji and La Thangue, 1991). This probe bound CRE proteins efficiently in whole cell extracts (figure 4.1A, compare tracks 4 and 5). To ensure that only proteins of DRTF1 specificity crosslinked to the probe, a double point mutation was incorporated to disrupt binding of CRE proteins. Only complexes of DRTF1 specificity bound to the 50/82 CRE-M probe in F9 EC and thymus extracts (figure 4.1A, tracks 6-15).

Binding reactions were performed with both F9 EC cell and adult thymus extracts, uv crosslinked, and DRTF1 complexes separated in native gels by gel retardation. Discrete complexes were excised, and the DNA bound crosslinked species resolved by electrophoresis through an SDS denaturing gel alongside 14C methylated protein markers. I
Figure 4.1. DRTF1 complexes have common DNA binding subunits

A. A probe spanning -82 to -50 of the adenovirus E2A promoter was synthesised incorporating BdUR and $^{35}$P nucleotides across the binding region (-70 to -50, Shivji and La Thangue, 1991). Binding reactions were set up with F9 EC extract (lanes 1-5) in the presence of oligonucleotide competitors (table 3.1, chapter 3), to determine the specificity of binding complexes. CRE binding activities bound the probe, and were competed by P but not by PM1 (lanes 4 and 5), in addition to DRTF1 activities (competed by 71/50 but not 63*, lanes 2 and 3). A double point mutation was therefore incorporated into the probe to abolish binding of CRE activities, but maintain specific binding of DRTF1a, b, and c, and a' in F9 EC and adult thymus extracts respectively (lanes 6-15). 12 µg of whole cell extract, and 300ng of competitor were used for these assays.

B. Binding reactions were set up as for a gel retardation assay with F9 EC and adult thymus extracts with the 50/82 CRE-M bromodeoxyuridine substituted probe, in the presence of 60/62 competitor. After uv irradiation, complexes were fractionated through a native gel, located, excised, and crosslinked polypeptides resolved by SDS gel electrophoresis. The crosslinked polypeptides in a and b/c from EC extract are shown in lanes 3 and 2, and those in adult thymus a' in track 4. Molecular weight markers are shown in track 1. Note that the predominant crosslinking polypeptide in each complex is p50, and that p30 (*) crosslinks inefficiently under these assay conditions.
could be confident therefore that the polypeptides in a, a' and b/c were present in discrete complexes, although it was not possible to separate DRTF1b from c by this technique.

Both DRTF1a and b/c from F9 EC cells contained similar DNA binding polypeptides, since a common polypeptide of corrected molecular weight 47 kD (see chapter 5, figure 5.9 for an explanation) (p50) was crosslinked (figure 4.1B, lanes 2 and 3), and likewise, DRTF1a' from thymus contained this p50 species (lane 4). A polypeptide of molecular weight 27 kD (p30) also crosslinked in these experiments, but the crosslinking efficiency was weaker, and more variable than that of the p50 species. This was also the case when affinity purified preparations of DRTF1 were crosslinked (see chapter 5). These studies however clearly demonstrated that DRTF1 complexes shared a DNA binding polypeptide of 50 kD, and since all complexes have identical sequence specificity (chapter 3), it is likely that this polypeptide is shared by all DRTF1 complexes. Differences in migration of the complexes in gel retardation assays are likely to be caused by modification of this DNA binding polypeptide, or by the association of other proteins with the common DNA binding subunit(s). To differentiate between these possibilities, the native molecular weight of the individual complexes was determined.

4.3. DRTF1 complexes have very different molecular weights

4.3.1. Glycerol Gradient centrifugation

Whole cell extracts of EC cells or adult thymus or testis were centrifuged through glycerol gradients, and the position of individual complexes within the gradient determined by gel retardation of the fractions, and compared to the position of molecular weight markers centrifuged in parallel (figure 4.2). The specificity of the complexes seen in these experiments was tested by competition with wild type and mutant DRTF1 binding sites (figure 4.2d).

DRTF1a and b/c from EC cell extracts (figure 4.2a) were present in distinct fractions, DRTF1a had a molecular weight of approximately 200 kD, and DRTF1b/c activity peaked in the 50-75 kD range. I was unable to resolve DRTF1b from c under these assay conditions, but based on their mobility, and the results of the SI00 assays (figure 4.3), it was reasonable to suggest that c had the smaller molecular weight. Indeed, analysis of adult testis extract showed DRTF1c equilibrating with an apparent molecular weight of 50 kD (figure 4.2c), whereas DRTF1b assayed in adult thymus extract (figure 4.2b) had an approximate molecular weight of 60 kD, similar to that of DRTF1b/c in EC extracts. In both adult thymus and testis, DRTF1a' binding activity peaked in the 165 kD range.

Therefore the molecular weights of the DRTF1 complexes reflected their migration in native gels, DRTF1a was larger than a' by about 35 kD, and a' larger than b/c by about 80 kD, and b was larger than c. The large differences in molecular weight between the a forms and other DRTF1 complexes were probably caused by assembly of multiprotein
Figure 4.2. DRTF1 species are composed of multicomponent protein complexes

Whole cell extracts from either F9 EC cells (a), adult thymus (b), or testis (c) were fractionated through glycerol gradients, and fractions assayed by gel retardation with the DRTF1 binding site probe (71/50). The positions of standard molecular weights were determined in parallel, and the DRTF1 complexes are shown. Note that the size of DRTF1c in testis is smaller than DRTF1b in thymus, and that both are considerably smaller than a' in thymus. DRTF1a in EC extracts has the largest molecular weight of all the complexes.
a) F9 EC:
Probe 71/50 (DRTF1)
Load 29 68 150 200 M.W. (kD)

b) Adult thymus
Probe 71/50 (DRTF1)
Load 29 68 150 200 M.W (kD)

C) Adult Testis
Probe 71/50 (DRTF1)
Load 29 68 150 200 MW (kD)
complexes which could form in the absence of DNA.

The molecular weight of the b/c complexes seen in this assay was much smaller than determined previously by S200 size exclusion chromatography (Shivji and La Thangue, 1991). It was possible that very weak interactions with the b/c complex could occur which were assayable by size exclusion chromatography, but disrupted by the less gentle glycerol gradient centrifugation. Alternatively, the difference in apparent molecular weight could be accounted for by the complexes adopting different conformations when assayed by the two techniques (as seen with NFκB, Urban et al., 1991).

4.3.2. Size exclusion chromatography of EC cell extracts.

F9 EC whole cell extract was subject to chromatography through an S100 column under the same conditions as marker proteins, and fractions assayed by gel retardation. DRTF1b binding activity peaked at 200 kD, and DRTF1c activity peaked at 150 kD (figure 4.3a). DRTF1a showed the same molecular weight of 200 kD as seen by glycerol gradient centrifugation. The difference in molecular weight for DRTF1b and c obtained by the two methods indicated that b/c complexes were free to change conformation or form "supercomplexes", whereas DRTF1a was not.

However, when a glycerol gradient fraction of molecular weight 50 kD containing DRTF1b/c binding activity was loaded onto the S100 column, eluted DRTF1 binding activity peaked at an apparent molecular weight of 150 kD (figure 4.3b). DRTF1b/c could therefore form complexes during chromatography which were dissociated during gel retardation (as there was no difference in the migration of the complexes in the loaded and eluted fractions). Alternatively, DRTF1b/c complexes could adopt different conformations and therefore show different molecular weights when assayed by the two techniques.

The molecular weight analysis of the complexes indicated that DRTF1a is composed of an activity complexed to DRTF1b/c, and that the complex is unable to adopt conformations of larger apparent molecular weight on chromatography. The b/c complexes are smaller than the a complex, and can either form "supercomplexes" or adopt a much less compact conformation on chromatography than during glycerol gradient centrifugation.

4.4. DRTF1a can be disrupted to produce DRTF1b/c binding activity

DRTF1a can be dissociated into component DRTF1b/c DNA binding subunits by treatment with detergent (Partridge and La Thangue, 1991). This effect is also mediated in F9 EC cell extract by the adenoviral E1A oncoprotein (Bandara and La Thangue, 1991). 12S E1A is sufficient to promote dissociation, and dissociation is dependent on the presence of intact conserved regions 1 and 2 (Bandara and La Thangue, 1991). These regions are conserved between different adenoviruses and several other viral oncogenes (adenoviruses, Kimelman et al., 1985, SV40, and polyoma, Stabel et al., 1985, HPV E7, 129
Figure 4.2d. The specificity of binding of each of the complexes in EC and thymus extracts was assayed. Complex * of fast migration in all fractionations (lanes 8-12 in a and b, and lanes 7-11 in c) was non-specific as it was competed by both 71/50 and 63* (lanes 2-10 and 13-15). Also the slower migrating complex * in EC fractions 7-11 was non specific (see lanes 2-7).

Figure 4.3. S100 fractionation of DRTF1 complexes does not resolve DRTF1a from b/c

(a). F9 EC whole cell extract was applied to an S100 column, and fractions assayed for DRTF1 binding activity. The position of the peak of each of the standard marker proteins is indicated. Note that under these conditions, DRTF1b/c is in the same fractions as DRTF1a, although the peak of a binding activity is approx. 200 kD, and that of b/c about 150 kD. Note also, that b is larger than c in this assay.

(b). DRTF1b/c complexes migrate differently in S100 and glycerol gradient assays

Fraction 5 and 6 from an EC cell glycerol gradient fractionation were pooled and loaded onto an S100 column. Fractions were assayed by gel retardation, and the position of the peak of DRTF1 b/c binding activity was compared to that of molecular weight standards. DRTF1b/c binding activity peaked at 150 kD in this assay, although the apparent molecular weight of applied protein was only 50kD as measured in the glycerol gradient assay.
(a) F9 EC  
Probe 71/50  

(b) EC fraction 5/6  
Probe 71/50
Phelps et al., 1988), and Bandara and La Thangue (unpublished observation) showed that DRTF1a can also be dissociated to DRTF1b/c by addition of SV40 large T antigen, or HPV E7 protein.

I found that 12S E1A could dissociate DRTF1a in fractions of F9 EC cell extract enriched for this activity (figure 4.4A, lane 5) and could also dissociate DRTF1a’ in adult thymus extract to produce DRTF1b (figure 4.4A, lane 9). The dissociation of the DRTF1a forms by E1A was specific, as addition of reticulocyte lysate to these binding reactions had no effect on DRTF1a complex formation (lanes 3, 6, 8, 11), and E1A had no effect on binding of Oct 1 (figure 4.4B) although the Oct 4 complex was abolished by the addition of reticulocyte lysate alone (lanes 2 and 3).

4.5. Determining the composition of DRTF1a

Several cellular proteins are known to bind E1A (Yee and Branton, 1985, Harlow et al., 1986) by interaction with the conserved regions 1 and 2 (Whyte et al., 1989). It was likely that one or more of these proteins were present in the DRTF1a complex, and sequestered from that complex by binding to E1A in the dissociation experiments. The presence of these proteins in the complex was tested (where reagents were available), and the retinoblastoma tumour suppressor gene product (pRb) and E1A associated p60 (cyclin A) were shown to be components of the a complex by antibody mediated perturbation of the a shift in JM and EC cell extracts respectively (Bandara and La Thangue, 1991, Bandara et al., 1991).

In the presence of the anti-cyclin A antibody (figure 4.5A), the a shift in EC cells was entirely lost. A faster migrating shift of a’ mobility was produced, presumably caused by antibody mediated dissociation of cyclin A from the DRTF1a complex (figure 4.5A, lane 3). Incubation of the anti-cyclin A antibody with thymus extract (lane 6) had no effect on the a’ shift, indicating that cyclin A was not involved in the a’ shift.

Monoclonal antibody XZ55 (Hu et al., 1991) crossreacted with murine pRb in immunoprecipitation and Western analyses, and so was tested for reactivity with the murine DRTF1a complex in EC cells and a’ in mouse thymus extract (figure 4.5B). Some of the a’ complex was supershifted by the antibody but unaffected by a control antibody of the same isotype (IG4E, compare lanes 5 and 6), suggesting that pRb was present in the a’ complex of adult thymus. However, no supershifting of the EC DRTF1a activity was observed on incubation with XZ55 (lane 3). The a complex of EC cells either did not contain pRb, or the XZ55 reactive epitope on pRb was masked within the a complex.

The recognition of pRb in the thymus DRTF1a’ complex was inefficient compared with the supershifting of DRTF1a obtained with XZ55 in JM cell extracts (compare with JM shifts in Bandara and La Thangue, 1991), which could be for a number of reasons. The XZ55 antibody may recognise human pRb more efficiently than murine pRb in gel.
Figure 4.4 E1A can disrupt DRTF1a and a' complexes

A. 12S E1A was in vitro transcribed and translated, and added to binding reactions before gel retardation, or an equivalent amount of reticulocyte lysate was used as a control. Addition of E1A to F9 EC extract (lane 2), or to a fraction enriched for DRTF1a (by affinity purification of DRTF1), lane 5, resulted in a disruption of the a complex, and production of increased b/c binding activity. This result was not observed on addition of the reticulocyte lysate alone (lanes 3, 6, 8 and 11). Addition of E1A to the thymus extract also caused disruption of the a' complex (lane 9). E1A also disrupted the a complex in SAOS-2 cells (lane 12). This effect was specifically mediated by E1A action on DRTF1a and a' complexes, no effect was seen on DRTF1b/c, or on addition of E1A to binding reactions with the Oct probe (figure 4.4B). Figure 4.4C shows the molecular weight of 35S methionine labelled in vitro transcribed and translated 12S E1A gene product.
A  71/50

<table>
<thead>
<tr>
<th></th>
<th>F9 EC</th>
<th>'a' EC</th>
<th>THY</th>
<th>SAOS-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probe</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extract</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lysate added</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- a
- b/c

Oct-1

Oct-4

B

<table>
<thead>
<tr>
<th>F9 EC</th>
<th>THY</th>
<th>SAOS-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>ELA RL</td>
<td>ELA RL</td>
<td>ELA RL</td>
</tr>
</tbody>
</table>

- Probe
- Extract
- Lysate added

C

MW (kD)

- 205
- 116
- 97.4
- 66
- 45
- 30
- 12S
retardation assays. The antibody may also preferentially detect the underphosphorylated form of pRb in gel retardation assays (it preferentially associates with underphosphorylated pRb in immunoprecipitation and Western assays, Hu et al., 1991), and the pRb present in the murine complexes may be of a different phosphorylation status to that in the JM complex. Alternatively, XZ55 may crossreact with a related polypeptide eg p107 (Ewen et al., 1991) and this cross reactivity might be weaker than a bona-fide recognition of pRb.

The status of Rb in various cell lines from sarcomas has been characterised, and one of these, SAOS-2, is known to have a truncated pRb gene product, lacking exons 21-27, a region necessary for complexing SV40 large T antigen (Shew et al., 1990) and DRTF1 (Bandara et al., 1991). SAOS-2 cells have a DRTF1a-like complex, which can be dissociated by 12S E1A (figure 4.4A, lane 12). The ability of the XZ55 antibody to cause a supershift of the a complex in the presence or absence of cyclin A antibody was tested, and results from the single experiment performed, indicated that cyclin A was a component of the a complex, but that XZ55 could not cross react with non-pRb proteins within this DRTF1a complex. Another protein in addition to cyclin A must complex DRTF1b/c in SAOS-2 cells, as removal of cyclin A resulted in a shift of similar migration to a'. Based on the observation that a p107 GST fusion protein will complex with DRTF1b/c from EC cells (L. Bandara, unpubl.obs.), it is likely that p107 is this other protein in the a'-like complex in SAOS-2 cells, (and possibly in the a complex in EC cells) and that it is not detected by XZ55.

4.6 Discussion

DRTF1 is therefore a multicomponent family of proteins, composed of a common DNA binding polypeptide, which associates with other non-DNA binding proteins such as pRb, cyclin A and p107 to form the various a complexes in different cell types and during development. The DRTF1a' form of DRTF1 found at later stages of murine development, and characterised here in adult thymus extract, is composed of the DRTF1 DNA binding proteins complexed with pRb in the absence of cyclin A. In contrast, the complexed DRTF1 in the Rb mutant SAOS-2 cells is composed of DRTF1b/c, cyclin A and another E1A dissociable protein, possibly p107. In F9 EC cells, the complexed DRTF1 activity also contains cyclin A, and an E1A dissociable protein which is unlikely to be pRb. Bandara et al. (1991) have shown that the EC DRTF1a complex could be reconstituted in vitro from purified DRTF1b/c by incubation with a glutathione-S-transferase pRb (amino acids 379-928) and cyclin A (amino acids 77-432) protein A fusion proteins. Neither fusion protein alone complexed efficiently with purified DRTF1b/c. In contrast, results obtained from immunoprecipitation studies from various cell lines have not shown complexes of pRb and cyclin A with E2F in vivo. Complexes of E2F and pRb have been observed (Bandara and
Figure 4.5. Cyclin A and pRb are present in different a complexes.

A. Cyclin A is present in the F9 EC DRTF1α complex

A rabbit polyclonal antiserum raised against purified bovine cyclin A was added to F9 EC and adult thymus whole cell extracts, and the effect assayed by gel retardation on the DRTF1 binding site probe. No effect was seen on the thymus shifts compared with the effect of the preimmune sera (compare lanes 5 and 6), but the addition of the anti cyclin A antibody specifically caused an increase in the mobility of the a complex in EC extract to a shift of comparable migration to a' (lane 3). This experiment therefore indicated that cyclin A was present in the EC a complex, but was not present in the thymus a' complex.

B. The Rb protein is involved in the thymus a' complex

Monoclonal antibody XZ55 against pRb was added to F9 EC or adult thymus whole cell extracts (lanes 3 and 6), and the effect on DRTF1 assayed by gel retardation. Some supershifting of the thymus a' complex to a' occurred (lane 6), whereas addition of control antibody IG4E had no effect on any of the complexes assayed (lanes 2 and 5). XZ55 caused a slight reduction in the intensity of the a complex in EC extracts, but no supershifted complex was resolved (lane 3). This experiment suggested that pRb was present in the thymus a' complex, but not in the EC cell a complex.
A

71/50

<table>
<thead>
<tr>
<th>F9 EC</th>
<th>THY</th>
</tr>
</thead>
<tbody>
<tr>
<td>PI</td>
<td>α32P</td>
</tr>
<tr>
<td>α32P</td>
<td>PI</td>
</tr>
</tbody>
</table>

: Probe
: Extract
: Sera added

a

a'

b/c

B

71/50

<table>
<thead>
<tr>
<th>F9 EC</th>
<th>THY</th>
</tr>
</thead>
<tbody>
<tr>
<td>IG4E</td>
<td>X255</td>
</tr>
<tr>
<td>IG4E</td>
<td>X255</td>
</tr>
</tbody>
</table>

: Probe
: Extract
: Sera added

a

a'

a^s

1 2 3 4 5 6
La Thangue, 1991, Bagchi et al, 1991, Chellappan et al., 1991), as have complexes of E2F, p107, cyclin A and associated kinases (Bandara et al., 1991, Raychaudhuri et al., 1991, Mudryj et al., 1991, Cao et al., 1992, Shirodkar et al., 1992, Devoto et al., 1992), but complexes of cyclin A, pRb and E2F formed in vivo have not been detected. Such complexes may exist, but are not detected because they are of low abundance compared with other E2F complexes, or they may not be recognised by the antibodies used in these studies.

Cyclin A interacts with the 'spacer' region of p107 (Ewen et al., 1992) and is thought to bind indirectly to E2F via this complexing with p107 (Faha et al., 1992). Cyclin A cannot bind pRb directly (Ewen et al., 1992), but it is possible that cyclin A interacts directly with E2F, and that pRb may also bind E2F through a different region of the protein, and that such complexes have not been identified by immunoprecipitation because of masking of the immunoreactive epitopes.

The various combinations of the proteins which may complex with a common DRTF1 DNA binding polypeptide allows much scope for control of the transcriptional activity of the protein, and the changes that occur, for example during development, indicate that the different forms may have different transcriptional effects. Indeed, it has been shown in vivo, that DRTF1 dependent transcriptional activation in SAOS-2 cells may be abrogated by cotransfection of Rb, and that this repression of DRTF1 activity may be relieved by transfection of 12S E1a (Zamanian and La Thangue, 1992).

Cloning of the DNA binding subunits of DRTF1 would allow a more detailed analysis of the interactions of DRTF1 with other proteins and the transcriptional activity of these complexes. Chapter 5 describes the purification of DRTF1 for sequence analysis.
5.1 Introduction

Having characterised the DNA binding polypeptides of DRTF1 and determined the relationship between the various DRTF1 binding activities, it became necessary to clone the DNA binding subunits for further molecular characterisation.

Various cloning methods had been attempted previously, such as screening cDNA expression libraries with an oligomerised DNA binding site (Singh et al., 1988), preparing monoclonal antibodies against purified protein, and amino terminal sequencing of the purified protein.

Amino terminal sequencing of purified protein was unsuccessful, possibly because the polypeptides were blocked by acetylation or because there was insufficient protein for sequence analysis (A. Aitken, pers. comm.). Semi-purified protein was also used to generate monoclonal antibodies, but the injected mice did not respond efficiently to the immunogen (P. Tassios, J. Gannon, M. Shivji, and N. La Thangue, pers. comm.), possibly because of the low levels of DRTF1 in the immunogen, and the small quantities of protein used in these experiments. Additionally, expression library screening for DRTF1 site binding polypeptides was unsuccessful (R. Girling, A. Whyte and N. La Thangue, pers. comm.)

Two other approaches to cloning DRTF1 were considered. Firstly, polyclonal antibodies could be raised against purified protein, and these could be used to screen cDNA expression libraries for DRTF1 encoding clones. Since generation of polyclonal antibodies requires large amounts of purified protein, and as success was not guaranteed, I decided to take the alternative approach of generating peptides from purified protein to obtain amino acid sequence.

DRTF1 had been purified previously (Shivji and La Thangue, 1991), by a combination of ion exchange and sequence specific affinity chromatography (Kadonaga and Tjian, 1986). I improved the purification procedure by incorporating a mutant binding sequence affinity column, which selectively removed non-specific DNA binding proteins, thereby reducing the number of cycles of affinity chromatography required to obtain highly purified protein.

It was necessary to undertake the purification on a large scale, and therefore that the source of protein, F9 EC cells be grown in large enough quantities to meet the demand. I therefore investigated alternative methods of growing the cells, and other possible sources of DRTF1.

5.2. Identification of a suitable source of material from which to isolate DRTF1.

Purification of DRTF1 was previously performed on a relatively small scale (Shivji and La Thangue, 1991). The source of material for these experiments was F9 EC cells
grown as monolayers. Large scale purification of DRTF1 required protein from approximately $5 \times 10^{11}$ cells, (calculated from the yield of purified protein from the first preparation, and assuming 15 $\mu$g of protein would be required for peptide generation). Growth of this number of cells by monolayer culture would require much time, effort and expense. Therefore alternative sources of protein were sought.

From the tissue analysis experiments, (chapter 3), it was clear that thymus isolated from adult mice contained fairly large amounts of DRTF1, albeit mainly in the form of the $\alpha$ complex. Early fractionation experiments indicated that although the $\alpha$ complex was retained during heparin Sepharose fractionation, it dissociated to form DRTF1b/c during the first affinity chromatography step. Therefore it was theoretically possible to isolate DRTF1b/c from mouse thymus. However, since mouse thymus is a small tissue it was not a practically viable exercise. Use of thymus from larger animals would also be inappropriate as murine cDNAs would eventually be required, and their isolation might prove difficult if DRTF1 was not well conserved at the nucleotide level. Therefore, alternative methods of growing F9 EC cells were tested.

Growth of cells on plastic beads (Plastek microcarriers, Tekmat Co.) was tested (section 2.10.3, figure 5.1B). DRTF1 binding activity was maintained (lane 2), but yields of cells obtained by this culture method were low. Suspension culture of F9 EC cells was also attempted. DRTF1 binding activity was maintained over 5 days of culture (figure 5.1A), and there was little loss of binding activity over a 10 day culture period. As suspension cultures could be maintained for 10 days, high yields of cells were obtained. DRTF1 was therefore purified from F9 EC cells grown in suspension culture.

5.3 Purification of DRTF1

The overall purification strategy for DRTF1 is shown in figure 5.2. Table 5.1 gives details of the yields and enrichment of DRTF1 during a typical purification.

5.4 Purification of F9 EC whole cell extract by heparin Sepharose chromatography.

Eleven preparations of purified DRTF1 were carried out, each from whole cell extracts (chapter 2, section 2.17.2) derived from 47ml of F9 EC cells. A typical extract (1.2 to 1.5g of protein) was loaded onto a heparin Sepharose column, and bound proteins eluted with a linear gradient of KCl (section 2.24.). Fractions were assayed for DRTF1 binding activity by gel retardation assay (figure 5.3A). The E2A promoter was used as a probe to assay fractions, as less protein was required than on the 71/50 probe to produce a shift (increasing the sensitivity of the assay), and it was easier to distinguish DRTF1b from c on the promoter.

Some E2A binding activity eluted from the column in the flow through (fractions I, II, and 1, figure 5.3A), but this activity did not have DRTF1 binding specificity. No
Figure 5.1. Alternative methods of growing F9 EC cells.

A. DRTF1 binding activity is maintained in cells grown in suspension.

Whole cell extracts were prepared from cells grown in suspension over a 5 day period, and DRTF1 binding activity in 12 μg extract was assayed by gel retardation on probe 71/50. DRTF1 binding activity was maintained over this period of culture in suspension (compare lanes 2 and 5), and levels of binding activity were equivalent from cells extracted after 10 days of culture. Yields of cells from these cultures were high, ranging from 60-120 fold increase in cell number.

B. Growth of cells on Plastek microcarriers.

Lane 1 shows the DRTF1 binding activity extracted from cells grown as monolayer cultures, and lane 2, the binding activity from cells grown on Plastek microcarriers on probe 71/50. 12 μg of whole cell extract was used for each experiment, and yields of DRTF1 obtained indicated that microcarrier culture was as efficient as monolayer culture. Disappointingly, only a 20 fold increase in cell number could be obtained by this method.

Table 5.1. Calculation of the enrichment of DRTF1 polypeptides during purification.

* binding activity is shown as arbitrary units. During purification, there was no apparent loss in DRTF1 DNA binding activity. Enrichment is defined as the units of binding activity/mg/ml compared with the units binding activity/mg/ml loaded onto the heparin Sepharose column. Purification (increase in specific activity) is defined as the units of binding activity/mg protein.
Table:

<table>
<thead>
<tr>
<th>Protein source</th>
<th>Amount (mg)</th>
<th>Volume (ml)</th>
<th>Binding act.*</th>
<th>DRTF1 enrichment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heparin Sepharose load</td>
<td>1300</td>
<td>92</td>
<td>100</td>
<td>1</td>
</tr>
<tr>
<td>Binding Heparin Sepharose</td>
<td>520</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1st Affinity load</td>
<td>148</td>
<td>60</td>
<td>100</td>
<td>5.7</td>
</tr>
<tr>
<td>2nd Affinity load (mut).</td>
<td>5</td>
<td>15</td>
<td>100</td>
<td>42.4</td>
</tr>
<tr>
<td>4th round purified protein</td>
<td>0.013</td>
<td>13</td>
<td>100</td>
<td>14130</td>
</tr>
</tbody>
</table>
Figure 5.2. The purification strategy.

The final purification strategy involved preparation of whole cell extract from F9 EC cells, which was fractionated on a heparin Sepharose matrix. Protein was eluted with increasing salt concentration, and fractions with DRTF1 binding activity were pooled, dialysed and applied to an affinity column bearing ligated wild type DRTF1 binding sites. Binding activity was eluted, and applied to a mutant site affinity column. DRTF1 binding activity could not bind to this column, and was present in the 'flow through'. This activity was subject to two further rounds of affinity chromatography on the wild type site column. The final preparation of DRTF1 was enriched $1.4 \times 10^4$ fold for DRTF1 binding activity. The numbers in brackets indicate the enrichment of DRTF1 binding activity at each stage of the preparation. See table 5.1 for details.
The Purification Strategy

F9 EC cells

Whole cell extract

heparin Sepharose fractionation

(5.7)

DRTF1 binding site

Affinity chromatography

(42.4)

Mutant site

Affinity chromatography

2x

(14130)

Highly enriched DRTF1 DNA binding polypeptides
activity that bound to the E2A promoter was eluted during washing of the column, but E2A binding species were eluted by concentrations of KCl exceeding 0.1M (figure 5.3B). A promoter binding activity which did not have DRTF1 binding specificity (*) was seen in fractions 45–47 (figure 5.3A). At higher salt concentrations, (of about 0.35 M KCl), DRTF1 binding activity eluted, with a and b eluting first (fraction 47, track 19; figure 5.3A), followed by co-elution of b and c (fractions 48–53).

Approximately 40% of the loaded protein bound to the heparin Sepharose column (figure 5.3B), and fractions containing DRTF1 binding activity constituted only 33% of that protein. Thus the heparin Sepharose fractionation yielded an approximately 5.7 fold enrichment of DRTF1 binding activity, and 8.8 fold increase in specific activity (See table 5.1). Fractions with DRTF1a, b or c activity were pooled and dialysed against 0.1M KCl col. buffer before use in the subsequent stages of purification.

5.5 Enrichment of DRTF1 binding activity by sequence specific DNA affinity chromatography

An effective method of purifying sequence specific DNA binding proteins is to utilise the affinity of interaction with their cognate binding site. The 71/50 DRTF1 binding site was multimerised and coupled to cyanogen bromide activated sepharose (Kadonaga and Tjian, 1986). Dialysed protein from the heparin Sepharose fractionation was applied to the column. The column was eluted and fractions were assayed by gel retardation (figure 5.4A).

Some E2A binding proteins were present in the flow through and washes from the column, (tracks 2–9), but these were not specific for the DRTF1 binding site. Protein in fractions 33–36 contained DRTF1b/c binding activity (tracks 16–19), although fractions 36 and 37 also contained a non specific DNA binding activity (figure 5.4B, tracks 13–18). The a form of DRTF1 was not seen in the elution profile, and most probably dissociated into DRTF1b/c during dialysis and fractionation, as low levels of a were present in the load onto the column (figure 5.4A, track 1). This first affinity purification, in conjunction with the heparin Sepharose fractionation yielded a 42 fold enrichment of DRTF1 binding activity. Multiple application of protein to this column allowed further enrichment of DRTF1.

Silver staining analysis of the proteins present throughout a preparation showed that although DRTF1 binding activity was maintained, there was a reduction in the total protein concentration, i.e., DRTF1 was enriched (figure 5.5A, compare lanes 1 and 5). It was not possible from this analysis to pinpoint a particular protein species that was enriched during the fractionation, as the protein profile of the purest fractions (after 4 rounds of affinity purification) was still complex (figure 5.5B). The fractionation was therefore modified to reduce the number of contaminating polypeptides without loss of
Figure 5.3. Heparin Sepharose fractionation of F9 EC whole cell extract

A. Gel retardation column profile of E2A promoter binding activity.

Lanes 3-5 show material present in the flow through from the column, lanes 6-14 the material in the wash, and lane 15 onwards the material eluted from the column. * indicates protein that was not specific for the DRTF1 binding site eluting before DRTF1 which eluted from fraction 47. Tracks 1-16 show shift produced by 1 μl of fraction, whereas tracks 17-28 show binding activity in 0.1 μl of fraction.

B. Protein elution profile of heparin Sepharose fractionation.

Approximately 40% of the applied protein bound to the column but DRTF1 was present in fractions constituting only 12% of the applied protein. Heparin Sepharose fractionation therefore yielded a 5.7 fold enrichment and 8 fold purification of DRTF1 binding activity, (see table 5.1).
A

E2A

Probe

Fraction No.

B

Fraction Number

Protein Conc. (rel. units)

KCl Conc. (M)
Figure 5.4. DRTF1 binding site affinity chromatography- gel retardation analysis.

A. Gel retardation profile of DRTF1 affinity chromatography fractionation.

Tracks 2-8 show E2A promoter binding activity in flow through fractions, tracks 9-13 show the binding activity in wash fractions, and tracks 14-20 the binding activity eluted from the column. DRTF1b/c eluted at a lower salt concentration than the non specific DNA binding activity * in tracks 18 and 19. 0.33 μl of each fraction and of the load (L), was assayed.

B. Specificity assay of the proteins binding the DRTF1 affinity column.

Specific DRTF1 binding activity was present in fractions 33, 34, and 35 (tracks 4-12), but non-specific activity (*), was present in fractions 35-37 (tracks 10-18). 0.2 μl of each fraction was assayed, and 300 ng of competitor oligonucleotides were used.
Figure 5.5. DRTF1 affinity chromatography—silver staining analysis of purified polypeptides.

A. Silver stained analysis of the polypeptides purified during a preparation of DRTF1.

Equivalent units of DRTF1 binding activity (as measured by gel retardation), from the protein loaded onto each column after the heparin Sepharose fractionation, were electrophoresed through a 7.5% SDS polyacrylamide gel, and assayed by silver staining. This purification was performed with wild type binding site affinity columns only. During the purification, there was a marked reduction in the complexity of the polypeptide species.

B. Silver staining profile of the DRTF1 binding fractions present after 4 rounds of affinity chromatography.

A sample of the fractions with peak DRTF1 DNA binding activity from the 4th affinity column were loaded in tracks 1-3. These samples had 10 times the number of units of binding activity than the protein shown in figure 5.5A, track 1. Despite enrichment for DRTF1 polypeptides, there were still many polypeptides present in these fractions.
5.6 Removal of non-specific DNA binding polypeptides by use of a mutant site DNA affinity column

The purification was modified by using a DNA affinity column prepared in the same way as the wild-type column, but using the mutated binding site 60/62, which does not bind DRTF1. This technique had been successfully used in the purification of Sp1 and NFκB (Kadonaga et al., 1987, Kawakami et al., 1988). Application of a mixture of proteins to this column should result in a separation of DRTF1 and non-specific DNA binding proteins. DRTF1 did not bind to the mutant site, and was therefore present in the flowthrough from the column (figure 5.6A, tracks 5-10), whereas the majority of the protein that binds DNA non-specifically bound to the column and was eluted in fractions separate from the DRTF1 binding activity (lanes 15-19). Binding specificity assays were performed on these proteins, and are shown in figure 5.6B. Comparison of the protein that bound the column with that excluded from it (by silver staining, data not shown), indicated that much of the applied protein bound to the column, and therefore inclusion of this step in the purification became routine, since it increased the purity of DRTF1.

The DRTF1 containing fractions from this column were loaded directly onto a wild type column and subjected to two further rounds of affinity chromatography yielding a final preparation of protein enriched $1.4 \times 10^4$ fold for DRTF1. The gel retardation profile from the final column (4th round purified protein) is shown in figure 5.7. This protein was pooled, and the purity of a unit of binding activity was compared by silver staining to the purity of an equivalent unit of activity from 4th round affinity purified protein prepared without the use of the mutant site column (figure 5.8). Track 1 shows the complex series of polypeptides present in preparations purified without the mutant column, whereas it is difficult to see any polypeptide species in track 2 prepared with the mutant column. This analysis demonstrates the efficiency of the mutant column in removing non-specific contaminants.

5.7 Determination of the molecular weight of the DRTF1 polypeptide species

As seen in figure 5.8, track 2, very pure preparations of DRTF1 did not display a prominent silver staining band (although by using more protein, several species could be detected; figure 5.10A, lane 2). A more sensitive technique was therefore required to determine the approximate molecular weight of DNA binding components. A method was used that relied on the covalent crosslinking of DRTF1 polypeptides to the $^{32}$P labelled 50/82 CRE-M binding site (described in chapter 4, section 4.2). Samples were treated with DNase to remove DNA not protected by binding of DRTF1, and the molecular weight of the DNA bound polypeptides determined by electrophoresis through a denaturing gel,
Figure 5.6. Mutant site affinity chromatography.

A. Gel retardation profile of E2A promoter binding activity in fractions from the mutant site affinity column.

Tracks 2–8 show promoter binding activity of the flow through fractions from the column, and tracks 9–13 the activity washed off the column. Bound material was eluted in fractions 13–18. Absolute amounts of the protein binding to the column could not be calculated, as fractions possessed low protein concentration. However, silver staining analysis suggested that the majority of the protein bound the column. 1 μl of fraction or load (L) was assayed.

B. Specificity of binding activity in fractions from the mutant column.

DRTF1 binding activity was present in fractions that did not bind the column (specificity of fractions 6 and 8 shown here, lanes 1–6). Material that bound the column was largely not of DRTF1 specificity (lanes 7–15). 1 μl of each fraction was assayed, with 300 ng of competitors.
Figure 5.7. E2A promoter binding profile of proteins eluted from the 4th affinity column.

Specific DRTF1 binding activity was eluted, and the activity of 1 μl of each fraction and of the load is assayed in this gel retardation. Specific DRTF1 activity is present in lanes 8-12. Non-specific DNA binding proteins were not present in these fractions, and there was no promoter binding activity in wash or flowthrough fractions (data not shown).

Figure 5.8. Comparison of polypeptide content of protein purified with or without the mutant site affinity column.

Equal units of DRTF1 binding activity from the final stage of affinity purification of DRTF1 prepared either with or without the mutant site affinity column was loaded onto a 7.5 % SDS polyacrylamide gel, and the polypeptide complexity was assayed by silver staining. Track 1 shows that DRTF1 purified without the use of the mutant column was much less pure than DRTF1 prepared with one mutant site chromatography step incorporated into the purification (track 2).
alongside protein markers. The amount of DNA remaining bound to the polypeptides after DNAse treatment was not known, therefore the contribution to the molecular weight by the DNA remaining bound to the DRTF1 polypeptides was estimated by determining the apparent molecular weight of the polypeptides in the absence of DNAse treatment (see figure 5.9 for explanation). This led to a correction value of 4.6 kD for DNA remaining bound after DNAse digestion, allowing a more accurate determination of the molecular weight of DRTF1 polypeptides by cross linking assays.

Individual cross linking experiments, and results obtained with different preparations of DRTF1 varied in the proportions of p50 and p30 species that could crosslink to the probe. Presumably this was because of differences in probe synthesis and possibly DRTF1 polypeptide composition differences in cells with varying growth characteristics. Therefore, DRTF1 fractionations were pooled, and crosslinking experiments, and silver and coomassie staining carried out on a sample of the protein (figure 5.10). In essence, DRTF1 species that crosslinked were of corrected molecular weights 43, 46, and 53 kD, and a doublet of 27 and 30 kD. A species of 38 kD was also observed in several other experiments. These values correlated with the molecular weights of polypeptides visualised by coomassie staining (31 and 32 kD, and 39, 46, 48, 53, 54 and 56–68 kD).

The complexity of the crosslinking species, and difficulties in extrapolation of this data to the polypeptide profiles obtained by staining underline the importance of obtaining DRTF1 in as pure a form as possible, and stress the importance of the mutant column in the purification.

5.8 Discussion

In summary, the purification of DRTF1 yielded protein that was enriched 1.5x10^4 fold for DRTF1. This protein mixture consisted of at least six species of DRTF1 binding activity, and several contaminating polypeptides.

The DNA binding polypeptides appeared to cover a wide range of molecular weights from 27 to 53kD. Crosslinking species in whole cell extracts were of approximately 30 and 50kD. These polypeptides could therefore homodimerise or heterodimerise to produce DNA binding activity, as is common amongst API family members (Kouzarides and Ziff, 1988, Rauscher et al., 1988, Sassone-Corsi et al., 1988, O'Shea et al., 1989b), ATF family members (Hai et al., 1989), and the helix loop helix family of transcription factors (Murre et al., 1989a,b), or each could bind to the same sequence independently of one another as exemplified by the octamer family of proteins (Sturm and Herr, 1988, Scheidereit et al., 1988, Schöler et al., 1990, Rosner et al., 1990, Suzuki et al., 1990). Alternatively, the smaller species are DNA binding proteolytic products of the larger
Figure 5.9. Estimation of the contribution of DNA covalently bound to DRTF1 to the molecular weight of crosslinked DRTF1 species.

Lanes 2-4 show the specificity of species crosslinked to 50/82 CRE-M for the DRTF1 binding site. Crosslinking species of 31, 53 and 55kD were specifically abolished by inclusion of 300ng of 71/50, but not by 300ng of the 60/62 mutant site (tracks 3 and 4 respectively). Lane 1 shows the crosslinked species present in the absence of DNAse treatment of the probe after crosslinking, the 31kD species has an apparent molecular weight 5kD larger than after DNAse treatment. Note that in this experiment, crosslinking of the 31 kD species was much more efficient than the crosslinking of the 53 and 55 kD species. A 10 % SDS polyacrylamide gel was used to resolve the crosslinked species. The accompanying determination of the weight of DNA remained bound to DRTF1 after DNAse 1 treatment is based on the assumption that each nucleotide contributes a weight of 300 D, therefore the full length probe contributes 32 x 300 = 9.6 kD.
Total weight of the protein + complete probe = 36 KD
Total weight of the protein + DNAsed probe = 31 KD
Total weight of complete probe = 9.6 KD
Total weight of probe remaining bound to protein after DNAse treatment = 9.6 - 5 = 4.6 KD

Apparent cross-linked weight = weight of polypeptides + 4.6 KD

31 = 26.4 + 4.6

Corrected molecular weight of the polypeptide = 26.4KD
Figure 5.10. Comparison of the (A) silver staining, coomassie staining and (B) crosslinking polypeptides in highly enriched DRTF1 preparations.

A small sample of pooled purified DRTF1 binding activity was electrophoresed through a 10% SDS polyacrylamide gel, and silver and coomassie stained. The molecular weights of the stained polypeptides were compared to the corrected molecular weights of proteins specifically crosslinking to the DRTF1 binding site, to determine which of the polypeptide species were DRTF1.

A. Track 2 shows the silver staining species present in 50 μl of purified protein, and track 3 shows the coomassie staining polypeptides TCA precipitated from 3 ml of protein. Tracks 4 and 5 show coomassie stained BSA (2 and 1 μg respectively). Lanes 1 and 6 show molecular weight markers. Notice the negative staining species of 33 and 63 kD in the silver stained track 2, and the correlation of polypeptides stained in both tracks 2 and 3 of 33, 39, 48, and 54 kD.

B. DNAse crosslinking was performed on pooled purified protein. Specificity of binding was checked by inclusion of 300ng of 71/50 or 60/62 competitor in lanes 3, 6, 9, and 12; and lanes 4, 7, 10, and 13 respectively. Two separate pools were tested in lanes 2-7 and 8-13. In track 4, a specific doublet of DRTF1 binding activity of corrected molecular weight 27 and 30 kD was seen, in addition to a species of 53kD. The pooled protein in tracks 8-13 showed different crosslinking proteins, two abundant species of 46 and 43 kD were present as was a low abundance binding species of 30kD (tracks 11 and 13). Tracks 5-7 and 11-13 were duplicates of tracks 2-4 and 8-10 respectively.
proteins. This is intrinsically unlikely as the levels of the smaller species did not increase relative to the larger species during the purification, as would be expected for degradation products.

If multiple proteins were required for DNA binding activity, one would expect the crosslinking assays to show a constant ratio of their abundance, and not the observed variation of crosslinking activity between experiments. Crosslinking of a protein depends on the incorporation of bromodeoxyuridine into the probe in place of dTTP, in a position close enough to the protein to allow covalent attachment of the protein to the probe. Differences in the incorporation of bromodeoxyuridine between different probe syntheses could account for some of these differences in efficiency of crosslinking. However, comparison of different pools of protein crosslinking to the same batch of probe (as seen in figure 5.10B) can only be explained by differences in the polypeptide composition of different preparations of DRTF1, perhaps dependent on the different growth characteristics of the cells from which the protein was extracted.

There are apparently multiple DNA binding forms of DRTF1 proteins. These different forms could be the products of separate genes (as for the ATF family of transcription factors, with a range of different molecular weight polypeptides encoded by different genes capable of binding the same site), or may be the product of one gene with the differences in molecular weight attributable to post-translational modifications, or alternative splicing.

Many polymerase II transcription factors are glycosylated (Jackson and Tjian, 1988; Lichtsteiner and Schibler, 1989; Sommer et al., 1991), with eight possible glycosylation sites on Sp1, the best characterised of these factors (Jackson and Tjian, 1988). However, only a very small proportion of DRTF1 was glycosylated (as determined by affinity for a wheat germ agglutinin column, Shivji and La Thangue, 1991), therefore a glycosylated form of DRTF1 would be a minor component of the mixture.

Phosphorylation is a common modification of transcription factors (Montminy and Bilezikjian, 1987, Hoeffler et al., 1988, Yamamoto et al., 1988, Gonzalez et al., 1989, Tanaka and Herr, 1990) and can contribute a 10 kD molecular weight difference to the 95 kD Sp1 protein (Jackson et al., 1990). Phosphorylation may affect DNA binding (Raychaudhuri et al., 1989, Bagchi et al., 1989, Manak et al., 1990), and transcriptional activation (Tanaka and Herr 1990). Dephosphorylation of heparin Sepharose fractions with calf intestinal alkaline phosphatase (CIAP) had no effect on the DNA binding activity of DRTF1b/c (Shivji and La Thangue, 1991), therefore some of the different DNA binding species of DRTF1 may be produced by differential phosphorylation as the dephosphorylated species could still bind DNA. Alternative splicing is another possibility, with the production of multiple forms of protein which may bind the same site (Santoro et al., 1988, Schöler et al., 1990, Hatzopoulos et al., 1990).
Because of the complexity of the profile of the DRTF1 DNA binding species, and the copurifying contaminants, it was critical to determine which of these species was DRTF1. Other groups faced with similar problems have tested individual polypeptides (separated by SDS polyacrylamide gel electrophoresis) for specific DNA binding after renaturation (Hager and Burgess, 1980, Hurst and Jones, 1987, Landschulz et al., 1988b, Scheidereit et al., 1988), or performed SouthWestern assays (Miskimins et al., 1985). I attempted a SouthWestern assay using purified protein, and although I detected DNA binding activity of renatured protein expressed from the L20 clone (Landschulz et al., 1988b), I could not renature DRTF1 binding activity. This could be explained by poor renaturation of DRTF1 into a structure capable of DNA binding, or because two polypeptide components were needed together either for DNA binding or for successful renaturation. Similarly, attempts to clone DRTF1 using a binding site λgt11 cDNA library screen (also dependent on renaturation of the protein to confer DNA binding activity) proved unsuccessful (R. Girling, A. Whyte, N. La Thangue, pers. comm.).

Antibodies could also have been raised against the various polypeptides, and used in the gel retardation assay to detect either a 'shifted shift', or an abolition of the shift indicating that the immunogenic protein was DRTF1. However, efforts to prepare both polyclonal and monoclonal antibodies against purified DRTF1 failed (P. Tassios, M. Shivji, N. La Thangue), indicating that the protein is not particularly immunogenic, or that insufficient immunogen was used for these experiments.

Therefore, the assignment of polypeptides constituting the DRTF1 binding activity rested heavily on the crosslinking data, hence the importance of defining molecular weights of crosslinking species as accurately as possible, and the requirement for very highly purified DRTF1.

The level of DRTF1 binding activity is comparatively low, even in cells such as F9 EC cells, with DRTF1 showing approximately 25% of the levels of Oct 1 binding activity. Large numbers of cells were therefore required to purify DRTF1 for sequence analysis.

The purification of DRTF1 was relatively successful (compared with the purification of other transcription factors), with 14000 fold enrichment (1 X 10^4 fold purification). The most efficient steps in the purification were the first and second affinity chromatography steps, with a 30 fold purification at the first affinity column, and an estimated 20 fold purification on the mutant site column. Subsequent purification was difficult to quantitate, but application of DRTF1 to more than four rounds of affinity chromatography resulted in a loss of DNA binding activity, and presumably of protein, so there was no advantage in trying to prepare purer DRTF1 by that method. The yield of DRTF1 binding activity after four rounds of affinity chromatography was approximately 14 %, which is significantly greater than the yield of 3.4 % obtained after only two affinity steps using the previous purification strategy (Shivji and La Thangue, 1991). Also, 14 % is
probably an underestimate of the true efficiency of the purification as DRTF1 DNA binding activity decreases in solutions of reduced total protein concentration.

The next chapter describes the preparation of peptides from the purified protein, and the results of the amino acid sequence analysis.
CHAPTER 6

ISOLATION OF cDNA FROM PEPTIDE SEQUENCE.
6.1 INTRODUCTION

Purification of DRTF1 yielded a large volume of dilute heterogeneous protein containing several polypeptides. DRTF1 polypeptides needed to be isolated from copurifying contaminants, and cleaved to generate internal amino termini suitable for sequence analysis, as direct sequencing from the amino terminus was likely to be prevented by acetylation.

The method chosen to cleave DRTF1 was largely determined by the small amount of material available; good recovery of DRTF1 peptides was required, and in a form suitable for sequence analysis. Various methods were tested and optimised and a proteolytic approach was taken to cleave DRTF1.

A trial cleavage was performed on a copurifying contaminant 98 kD polypeptide, from which peptides were produced and amino acid sequence was obtained. Peptide production from DRTF1 was unsuccessful using the same methods as used to generate peptides from the 98 kD polypeptide. Peptides were later prepared and sequenced by J. Hsuan and N. Totty, (Ludwig Institute, Riding House Street, London, W1) from protein purified by R. Girling, in a collaborative venture. cDNA encoding a portion of the sequenced protein was successfully isolated by R. Girling and myself.

6.2 Precipitation of purified protein

The purification of DRTF1 yielded dilute, highly purified protein. UV crosslinking and small scale precipitation experiments (chapter 5, figure 5.10) suggested that the purified protein solution consisted of several species of DNA binding polypeptides with DRTF1 specificity of 38, 43, 46, 53 and 27-30 kD, and that there were several other contaminating polypeptides in these preparations. Separation of these polypeptides was necessary to obtain amino acid sequence. However, the dilute protein needed to be concentrated approximately 660 fold before the polypeptide species could be separated by polyacrylamide gel electrophoresis.

The purified protein was in a buffer consisting of 20% glycerol and so could not be evaporated to less than 20% of the original volume. The protein was therefore concentrated by precipitation with trichloroacetic acid (TCA) (chapter 2, section 2.25).

The precipitated proteins were separated by electrophoresis through a 10% polyacrylamide gel, and yields quantitated by comparison with BSA standards on coomassie staining (figure 6.1). Note that the profile of track 2 is much more complex than that in tracks 3 and 4; this is because protein in track 2 was from a mixture of protein purified with and without the use of the mutant site column, whereas all of the protein in tracks 3 and 4 was prepared with the aid of the mutant column. The polypeptides seen in tracks 3
Figure 6.1. Coomassie stained polyacrylamide gel of precipitated purified protein.

Purified protein was precipitated with 20% TCA in the presence of 20 μg/ml of salmon sperm DNA, and electrophoresed through a 10% gel (lanes 2-4). Protein in lane 2 was pooled from preparations prepared with and without the use of the mutant site column, whereas protein in lanes 3-4 was prepared with the use of the mutant column. 5 μg of BSA was also precipitated (lane 5). Lane 1 shows marker proteins. Note the common 52, 54, 46, 43, 40, and 35 kD proteins present in all pooled preparations of protein.

Table 6.1. Comparison of crosslinking and coomassie staining species in the large preparations of pooled protein.

Compilation of the polypeptide molecular weights of coomassie staining and crosslinking species present in the purified DRTF1 preparations. Bracketed number represents a species that was only present in some preparations.
Purified DRTF1

MW (KD)

Cross linked species KD

<table>
<thead>
<tr>
<th>Cross linked species KD</th>
<th>27-30</th>
<th>30</th>
<th>38</th>
<th>43</th>
<th>46</th>
<th>53</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coomassie stained species KD</td>
<td>(30)</td>
<td>34, 35</td>
<td>40</td>
<td>43</td>
<td>46</td>
<td>50</td>
</tr>
</tbody>
</table>
and 4 show some differences with the polypeptides observed on coomassie staining of small scale precipitated protein experiments, for example the 43 kD species was not seen previously (Chapter 5, figure 5.10). Protein bands that are common between these large-scale precipitations, and those performed subsequently by R.Girling are compared with the molecular weights of the crosslinking species in table 6.1.

6.3. Preparation of peptides for amino acid analysis

The major limitation to the techniques available for peptide generation was the small amount of protein available. Therefore, trial experiments were performed using ¹⁴C methylated proteins for a quantitative analysis, and BSA and other 'standard' proteins for a qualitative assessment of the efficiency of cleavage and recovery of peptides. Two major methods of peptide generation were available, using an enzymic or a chemical cleavage approach. Figure 6.2 shows a flow chart of the steps involved in each process, and a value for the yield of protein at each stage using optimised conditions.

6.3.1 Chemical cleavage.

Cleavage was tested using cyanogen bromide, which cleaves at methionine residues. Cyanogen bromide cannot efficiently cleave proteins in the presence of acrylamide, therefore proteins were blotted onto nitrocellulose, and eluted before cleavage in acidic solution with cyanogen bromide (section 2.27.2). Cleavage products were reprecipitated before electrophoresis. Complete cleavage was not attained using this method, and efficiency of recovery of peptides was poor (25%) compared to the recoveries attainable by proteolytic cleavage.

6.3.2 Proteolytic cleavage.

Proteolytic cleavage of isolated proteins, in polyacrylamide gel slices, was carried out using the Cleveland reaction (Cleveland et al., 1977, section 2.27.3). Peptides were electrophoresed through the separating gel, prior to blotting and staining of the peptides. Trial reactions were performed to assess the activity of a variety of proteases, and to determine optimal conditions for complete cleavage of protein.

Figure 6.3 shows the peptides produced from BSA (A, B, C) using a range of enzymes. Trypsin digestion, using up to a 2:3 ratio of enzyme : protein was highly inefficient (figure 6.3A), as was thermolysin digestion (figure 6.3B). Endoprotease glu-C from Staphylococcus aureus V8 however cleaved BSA with much greater efficiency, digesting almost to completion at a similar ratio of enzyme : protein (see figure 6.3C, track 3). The activity of this enzyme (which cleaves at glutamate and aspartate residues) was tested on another protein, carbonic anhydrase (figure 6.3D). Apparently complete cleavage of carbonic anhydrase was achieved by a 3:5 ratio of enzyme to protein. Although endoprotease glu-C was capable of complete digestion of proteins, it was necessary to check that the enzyme could cleave DRTF1.

170
Figure 6.2. Flow chart of protein recovery at each step of peptide preparation.

The efficiency of each stage of peptide preparation was shown by the % value, and cumulative yield of peptides was shown by the unit value (U), where the initial protein is represented by 100 U. Enzymic cleavage yielded 43 % overall recovery of peptides, whereas chemical cleavage yielded only 25 % recovery using optimised conditions.
Dilute Protein (100U) → TCA Precipitation

60% 60U → Separation of Protein Species by SDS PAGE

Chemical Cleavage

Gel blotted onto nitrocellulose

77% 46.2U

Protein of interest eluted from nitrocellulose, and dried

95% 45U

Protein cleaved with CNBr

70% 31.5U

Products acetone precipitated and separated by SDS PAGE

Enzymic Cleavage

Protein Bands excised and Cleveland Reaction performed in SDS PAGE

90% 54U

Amino Acid Sequence Analysis
Figure 6.3. Proteolytic cleavage of proteins by Cleveland reaction

BSA was cleaved with trypsin (A), thermolysin (B), and endoprotease glu-C (C) in Cleveland reactions, and peptide products were separated through a 15% PAGE before coomassie staining. Endoprotease glu-C gave the most efficient cleavage, and was therefore tested for ability to cleave carbonic anhydrase also (D). Gel slices containing 30 μg BSA, and 10 μg of carbonic anhydrase were used in this experiment with the enzyme quantities shown in the figures.
6.4. Proteolytic cleavage of DRTF1 crosslinked species

DRTF1 does not silver stain efficiently, and much protein would be required to obtain coomassie stainable amounts of peptides. Therefore, the approach taken to test whether endoprotease glu-C was capable of cleaving DRTF1 relied upon the use of DRTF1 crosslinked to a radioactive probe. Crosslinking analyses of DRTF1 showed three polypeptides of molecular weight 38, 43, and 53 kD that crosslinked efficiently to XLCRE-M (figure 6.4A). These polypeptides were excised from the gel, and used in Cleveland assays with the endoprotease glu-C enzyme. The results of such cleavage reactions were visualised by autoradiography where the observed polypeptides had retained the ability to bind the probe. Figure 6.4B shows the results of such an experiment. The endoprotease glu-C cleaved each of the crosslinked DRTF1 polypeptides to generate peptides which retained crosslinked DNA binding activity. The concentration of endoprotease glu-C could be increased without any further loss of DNA bound peptide, suggesting that digestion of the crosslinked species was complete and that the peptides produced were stable to further endoprotease glu-C cleavage. Interestingly, cleavage of the 38 and 43 kD species (tracks 8-10 and 5-7 respectively) generated a common size DNA binding peptide of 15.5 kD, whereas the 53 kD species generated a peptide 4.5 kD larger (compare tracks 5 and 8 with 2). The molecular weights of the generated peptides do not necessarily reflect the molecular weights that would be obtained if the uncrosslinked proteins were cleaved, as crosslinking of the protein to DNA could protect other potential cleavage sites.

*Staphylococcus aureus* V8 endoprotease glu-C could be used to generate peptides from DRTF1 with much higher efficiency than any of the other methods tested, and was therefore the enzyme chosen for the large scale cleavage experiment. However, it was necessary to establish that the cleavage reaction, and subsequent blotting and staining of PROBLOT did not result in a modification of the generated amino termini of the peptides that was incompatible with amino acid sequencing. Endoprotease glu-C cleavage was therefore performed on the 98 kD contaminating polypeptide present in the profile of track 2, figure 6.1.

6.5. Endoprotease glu-C cleavage of the 98 kD polypeptide, and generation of sequence data.

The region of the polyacrylamide gel containing the 98 kD species was excised, and Cleveland digests performed using endoprotease glu-C. Peptides were fractionated prior to blotting onto PROBLOT (using optimised conditions). Transferred peptides were stained with a modified coomassie stain (figure 6.5, section 2.27.4). Two major species of peptide were generated from the 98 kD protein, of 9.2 and 12.4 kD (lane 3). The peptides were generated from the 98 kD protein and were not degradation products of the protease, as an
Figure 6.4. Determination of the efficiency of endoprotease glu-C cleavage of crosslinked DRTF1 species.

A. Autoradiograph of a crosslinking assay in which DRTF1 crosslinking species were treated with DNAse 1 before separation in a 10% SDS PAGE. The protein used in the crosslinking assay was purified DRTF1. The crosslinked species were excised from the gel, and a Cleveland reaction performed on the individual species (B), in a 15% gel. Crosslinked species were cleaved with 100 ng (lanes 3, 5, 8) and 200 ng (lanes 2, 6, 9) of endoprotease glu-C, and peptides remaining bound to the DNA probe were visualised by autoradiography. Lane 1 shows position of migration of "C labelled protein markers.
A

Pool 1

Pool 2

'Corrected' M.W. of X-linking Species (KD)

M (KD)

92.5 -
69 -
46 -
30 -
14 -

1 2 3 4

B

M (KD)

53
43
38

200 100 -
100 200 -
100 200 -

'Corrected' M.W. of X-linking Species (KD)

ng endo glu-C

peptides

'Corrected' M.W. (KD)

20 -
15.5 -

1 2 3 4 5 6 7 8 9 10
adjacent track (lane 2) was loaded with an equal concentration of enzyme, and no peptides were produced.

The two peptide species were subject to Edman degradation (by A. Aitken and A. Harris, Lab. of protein structure, NIMR), and novel amino acid sequence was generated (table 6.2). The peptide generation protocol devised could therefore be successfully used to obtain peptide sequence information.

6.6 Endoprotease glu-C cleavage of DRTF1

Cleavage and sequencing of the coomassie staining species of 43, 46 and 53 kD was attempted using exactly the methodology outlined above, except that twice the amount of enzyme was used as the molarity of these species was apparently twice that of the 98 kD species. Peptides were generated and sequenced, but all sequence obtained was of the enzyme. The endoprotease had autocleaved only in the presence of the DRTF1 species, as in adjacent tracks with enzyme alone, there were no equivalent peptides. Some sequence was present which could not be accounted for by the enzyme, but this sequence was heavily masked by the endoprotease sequence, and was therefore not useful.

6.7 Peptide sequence of DRTF1

After the unsuccessful attempt to generate peptide sequence data from DRTF1, more protein was purified using exactly the procedure outlined in chapter 5, by R. Girling. This protein was used to generate and sequence peptides by J. Hsuan and N. Totty (Ludwig Institute, London). The 46 kD region of the gel was isolated without staining the protein track, and the protein was electroeluted. Cleavage of the protein was performed in solution using an enzyme which cleaves carboxy terminal of lysine residues, and peptides were separated by high performance liquid chromatography (HPLC), before amino acid sequencing.

Sequence was successfully obtained for several peptides, and I used the amino acid and degenerate nucleotide sequences of these peptides to screen both EMBL and Leeds databanks, by FASTA and Wordsearch programmes (see section 2.29). There was no homology between any of these sequences and sequences present on the databases (higher than 70%) over these very short sequences, suggesting that the sequenced protein was novel.

6.8 Cloning of the sequenced protein

Fully degenerate oligonucleotides were designed against regions of peptides showing the least degeneracy. These oligonucleotides were used in Polymerase chain reactions (PCR), with template cDNA from a λ gt11 11.5 day mouse library by R. Girling. A BamH1 site was incorporated into the 5' termini of the oligonucleotides, to facilitate
Figure 6.5. Endoprotease glu-C cleavage of the 98 kD protein

A Cleveland reaction was performed with 25 μg of endoprotease glu-C on the 98 kD protein band excised from figure 6.1 lane 2. After blotting of the fractionated peptides from the 15% gel onto PROBLOT, staining of the blot revealed two peptides of 9.2 and 12.4 kD (lane 3). Sequences derived from these peptides are shown in table 6.2. 25 μg of endoprotease glu-C alone was loaded into track 2, but no staining was visible on the PROBLOT.

Table 6.2. Peptide sequences derived from the 98 kD polypeptide after endoprotease glu-C cleavage.

Two peptide sequences were derived from each of the peptide bands shown in figure 6.5. These are unique peptide sequences, and are represented here by single letter code.
12.4 KD PEP1  L E G P S S G R R
12.4 KD PEP2  V D K A K D K Y A
9.2 KD PEP1   G A N S N F K D
9.2 KD PEP2   L D F Q V S P
subcloning of products into pBluescript II SK- vectors for sequence analysis. Multiple PCR products were generated for each combination of oligonucleotide pools used, and I subcloned these and sequenced them from both T3 and T7 primer sites. PCR products should contain oligonucleotide sequence derived from each pool used in the reaction, at either end. Alternatively, if an internal BamHI site is present in the PCR product, cleavage of the product prior to subcloning should result in loss of one oligonucleotide sequence, replaced by a BamHI site alone. If the PCR product is derived from the sequenced protein, nucleotide sequence adjacent to the oligonucleotide sequence should encode the remainder of the peptide sequence.

Sequence generated from ten subcloned PCR products showed that the PCR was specifically amplifying DNA with the oligonucleotides incorporated, but in none of these cases did sequence adjacent to the oligonucleotide sequence match that encoding the rest of the peptide. I screened the PCR product sequences against EMBL and Leeds databanks, and in all cases except one, there was no significant homology to any cloned DNA. The exception encoded a 500 bp region of DNA from phage lambda. This product was amplified because the sequence of the selected oligonucleotides showed 69% and 53% homology to lambda sequences separated by 500 nucleotides. The stringency of hybridisation was therefore low enough to allow such mismatches to occur under the PCR conditions employed.

Several options were available to maximise the chance of amplifying DNA encoding the sequenced protein. We could reduce the number of mismatches between primers and template DNA by increasing the stringency of hybridisation of the PCR, or we could enrich the template DNA for the required species. The second alternative was chosen, as increasing the stringency of the PCR by raising the temperature of annealing may prevent the desired hybridisation.

cDNA synthesis was primed from antisense pools of oligonucleotides by reverse transcription of F9 EC cell RNA by R. Girling. This oligonucleotide primed cDNA was used as the template in PCR with other pools of sense oligonucleotides. Products were obtained from five such reactions, and were subcloned and sequenced. One of these products (clone 6), from peptide 5 antisense primed cDNA and peptide 6 sense oligonucleotides, generated a 360 nucleotide product. Sequence adjacent to the peptide 5 antisense oligonucleotide sequence correlated with adjacent peptide sequence, but the sequence adjacent to peptide 6 sense oligonucleotide gave no further information because the oligonucleotide was derived from the carboxyl terminal end of peptide 6. Translation of the 360 nucleotide product showed that two other peptide sequences (peptides 21 and 3) were included in the sequence, confirming that 360 was a portion of cDNA encoding the sequenced protein. The nucleotide and amino acid sequences encoded by 360 are shown in figure 6.6.
Figure 6.6. Sequence of the p46 360 cDNA.

Amino acid and nucleotide sequence of the p46 360 bp cDNA, represented by single letter codes. Peptide sequences are shown in bold type and are underlined, and are in the order peptide 6, peptide 21, peptide 3, and peptide 5.
Figure 6.6. SEQUENCE OF THE p46 360 cDNA.
Figure 6.7. Predictions from the amino acid sequence of p46 360.

The Mackintosh Macvector toolbox was used to generate secondary structural predictions, hydrophilicity predictions (Kyle and Doolittle, 1982), and flexibility predictions (Karplus and Schulz, 1985) for the 360 segment of p46. The amino acid numbering in these diagrams relates to the position of 360 within the full length cDNA, encompassing residues 84-204 (R. Girling, pers.comm.).

CF indicates Chou–Fasman prediction (Chou and Fasman, 1974), and RG, Robson–Garnier prediction (Garnier et al., 1978).
6.9 Predicted protein structure of the p46 360 cDNA

The p46 360 sequence showed no homology to sequences in Leeds and EMBL databases, and did not show any characteristic domains of transcription factors. Chou-Fasman protein structure predictions indicated that the amino terminal portion of the region formed a coiled struture, followed by two helical regions interrupted by a short region of coils, and another helical region at the carboxyl end of the region encoded by p46 360. This region of the protein therefore seemed to possess a large degree of flexibility and was mainly hydrophilic (see figure 6.7).

6.10 Discussion

Other workers who have purified transcription factors for sequence analysis have either purified a single polypeptide species, or a highly enriched population of protein, where the species of interest is by far the predominant polypeptide (CTF, Jones et al., 1987, Sp1, Briggs et al., 1986, ATF, Hurst and Jones, 1987, CREB, Montminy and Bilezikjian, 1987, Yamamoto et al., 1988, Oct 2, Scheidereit et al., 1987). These polypeptides have been sequenced directly (AP2, Williams et al., 1988, C/EBP, Landschulz et al., 1988b), or peptides sequenced after cleavage in solution by trypsin (AP2, Williams et al., 1988, Sp1, Kadonaga et al., 1987, CTF/NF1, Santoro et al., 1988) or by CNBr (AP2, Williams et al., 1988, Oct 2, Scheidereit et al., 1988). These approaches were not feasible for DRTF1 as the purified protein was a complex mixture of DRTF1 site binding polypeptides and contaminants. Purified DRTF1 was therefore precipitated, and polypeptides species separated by electrophoresis. Cleavage might have been performed by in situ digestion with trypsin, after blotting to nitrocellulose (Aebersold et al., 1987, CREB, Gonzalez et al., 1989, NFkB (p50), Kieran et al., 1990, Ghosh et al., 1990, NFkB (p65), Nolan et al., 1991) but the recovery of such low abundance peptides by HPLC could not be guaranteed. I therefore assessed the recovery of peptides after enzymic digestion in polyacrylamide gels by the Cleveland method, and successfully obtained peptide sequence data after S. aureus V8 endoproteinase glu-C cleavage of a 98 kD copurifying polypeptide. This approach was however unsuccessful with DRTF1, as the enzyme autocleaved in the presence of DRTF1, and DRTF1 peptide sequence was masked by the sequence of these cleavage products.

Fortunately, peptide sequence data were subsequently generated, which enabled the isolation of the 360 bp cDNA by use of degenerate oligonucleotides in PCR reactions, with oligonucleotide primed cDNA template. The p46 360 cDNA showed little homology to other proteins at the nucleotide or amino acid level, and did not have any obvious structural features common to transcription factors. The p46 360 cDNA was used to screen libraries for full length clones, at much higher stringency than would have been possible using degenerate oligonucleotides derived from peptide sequences, and full length cDNAs
were isolated by R. Girling.

The isolation of this p46 360 cDNA was an important step in the cloning of the sequenced protein. This cDNA product was also used to determine the expression pattern of the gene. The results of expression pattern analyses, and Southern hybridisation using this cDNA as probe are presented in the next chapter.
CHAPTER 7

SOUTHERN ANALYSIS AND RNA EXPRESSION PATTERN OF p46
7.1 Introduction

We were able to isolate a short cDNA (p46 360) by using the p46 peptide sequences to design degenerate primers for PCR. It was, however essential to determine if the sequenced protein was authentic DRTFl. Consequently, I used the cDNA to investigate the expression pattern of the p46 360 defined RNA, and to determine if there was a correlation with the DRTFl DNA binding activity.

7.2. Southern analysis of the p46 gene

The genomic structure of p46 was analysed by hybridisation of the p46 360 fragment to murine DNA digested with various restriction enzymes. Each enzyme released several fragments of up to 7 kb which hybridised to p46 360 (figure 7.1). These multiple bands were not caused by an incomplete digestion of the genomic DNA, as pBluescript seeded at genomic copy number was digested to completion, and reprobing of the experiment with a myogenin cDNA probe (after stripping the blot) showed hybridisation to fragments of the expected size (S.-P.Yee, pers. comm.).

The p46 360 cDNA could, therefore, either hybridise to several related genes, or hybridise to a region of a single gene encompassing several introns with restriction sites. The first possibility is unlikely, given the stringency of the conditions used for hybridisation and washing, as blots were washed at 65°C in 0.1% SSC and 1% SDS. When similar conditions were used by others (Hara et al., 1992), the brain-2 POU domain hybridised much more strongly to the brain-2 genomic DNA than to the highly related brain-1,-4 or SCIP-1 genomic POU regions.

The p46 360 cDNA was known to span an intron (from isolation of a cDNA clone with an intron within the 360 sequence, R.Girling, pers. comm.). However the hybridisation of p46 360 to four fragments of genomic DNA (for example, the hybridisation pattern of BamHI digested DNA, lane 1) suggested that if p46 was unique, at least three introns each bearing a BamHI restriction site (in addition to other enzyme sites) must interrupt the genomic DNA which hybridised to the 360 cDNA sequence. There was some variation in the hybridisation intensity of the 360 probe to different fragments of the digested genomic DNA (see for example the variation in signal intensity with PstI digested DNA, lane 2, figure 7.1), but this variation was much less marked than the hybridisation intensity differences seen with the brain-2 hybridisation to brain-2 and related genomic DNAs, which was performed under less stringent conditions (Hara et al., 1992). Therefore, p46 appears to be a unique gene, and the more intensely hybridising fragments probably include more of the 360 region than the weaker bands, possibly reflecting the extent of hybridisation of the digested genomic DNA fragments to p46 360
Figure 7.1. Southern analysis of the p46 gene.

Genomic DNA from CBAxC57 mice was digested with *BamH*I (lane 1) and *Pst*I (lane 2), and probed with the p46 360 cDNA. A complex series of fragments produced by each enzyme could hybridise to the p46 360 cDNA. The blot was autoradiographed for 3 days.
under stringent hybridisation conditions. If however the sequence encoding p46 does belong to a family of related genes the variation in signal intensity may be caused by different levels of homology between the 360 probe and genomic fragments.

I next investigated the expression pattern of p46 by Northern analysis.

7.3. Northern analysis of p46 expression pattern.

Since p46 protein was purified from F9 EC cells, a Northern assay was performed to determine the size of the p46 RNA transcript in these cells (figure 7.2A). The p46 360 probe hybridised to a 3kb transcript in total cellular RNA from EC cells. Poly A+ messenger RNA (mRNA) was prepared from F9 EC cellular RNA, and when hybridised with the 360 probe, a 3kb transcript was detected, indicating that p46 is predominantly expressed as a 3 kb poly A+ message in F9 EC cells.

The expression pattern of the p46 RNA was investigated in a range of murine tissues, and from embryos at various stages of development. Approximately equal amounts of RNA (quantitated by spectrophotometry and stained by ethidium bromide, figure 7.2B) from tissues of adult mice, from embryos, and from tissue culture cells were assayed for the presence of p46 RNA. Hybridisation was performed under conditions of high temperature (65°C), with yeast transfer RNA (tRNA) and salmon sperm DNA included in the hybridisation buffer to prevent non-specific hybridisation of the riboprobes (see section 2.14). The major hybridising transcript in murine tissue (with the exception of brain) was a 3 kb species. The 3 kb message was the only species present in EC cells (lane 1, figure 7.2C), and its abundance was marginally reduced in F9 PE cell RNA (lane 2). The 3kb species was also down regulated during murine development, with only low levels of hybridisation present at 17.5 days pc (figure 7.2C, lane 5). The hybridisation of a human γ actin riboprobe served as a control for the integrity of the RNA used in these experiments. A 2 kb γ actin transcript was detected in F9 EC cells, but was also regulated during differentiation and development. The γ actin RNA was upregulated on differentiation of EC cells to parietal endoderm like cells in contrast to the down regulation of the p46 transcript on differentiation, whereas both p46 and γ actin transcripts were down regulated during murine embryogenesis (lanes 3–5, figure 7.2C and D).

Various adult tissues were examined for the presence of p46 RNA. Adult thymus and testis had high levels of the 3kb species (figure 7.2C, lane 6 and 8), whereas brain had low levels (lane 7), and the message was undetectable in liver (lane 10). In contrast, levels of hybridisation of γ actin in thymus and brain were equivalent (figure 7.2D, lanes 6 and 7), and in common with p46, there were high levels of γ actin in testis (lane 8), and no hybridisation of γ actin in liver RNA. Another probe was also used to ensure that the liver RNA was intact. The Xenopus EF1α probe (kind gift from R.Wilson, NIMR, Krieg et al., 1989) hybridised to a transcript of approximately 1.7 kb in liver RNA and in other tissues,
Figure 7.2. p46 is encoded by a 3 kb A+ RNA in F9 EC cells and is expressed throughout embryogenesis

A. 20 μg of total RNA from EC cells, and A+ RNA isolated from 33.3 μg EC cell RNA was blotted and hybridised to an antisense RNA probe transcribed from the 360 cDNA. A 3 kb A+ species hybridised, and was the only major transcript present in F9 cells.

B. 20 μg of total RNA from EC cells and their differentiated derivatives (lanes 1 and 2), from mouse embryos of various stages (lanes 3-5), and from adult mouse tissues and human tissue culture cells were electrophoresed in the presence of ethidium bromide prior to blotting. C. The blot was hybridised with the 360 antisense RNA probe. D. 10 μg of the same RNA samples were assayed for the abundance of γ actin RNA as a control for the integrity of the samples. The γ actin probe was antisense RNA transcribed from a partial human γ actin cDNA, and equivalent activity probes were used in the two experiments. Hybridisation was performed in the presence of 100 μg/ml yeast tRNA, and 200μg/ml sheared salmon sperm DNA at 65°C, and washed at 65°C (actin) and 70°C (360).
indicating that the lack of hybridisation of the γ actin and p46 360 probes to liver RNA was caused by a low abundance of p46 and γ actin transcripts in liver, and possibly lower levels of RNA loaded for this sample, as there was less ethidium staining of ribosomal RNA in the liver than in other RNA samples (figure 7.2B, lane 10).

In contrast to all other murine tissues examined, in which the 3kb message was the most abundant form of p46 RNA, the major transcript of p46 in murine brain was of 5 kb (figure 7.2C, lane 7). This 5 kb species was also abundant in human JM and HeLa cell RNA, as was the 3 kb species, and a 2 kb hybridising species (lanes 11 and 9 respectively). Very low levels of the 2 and 5 kb p46 hybridising species were present in most of the tissues and, because of the similarity in size of these transcripts to the size of ribosomal RNAs (rRNA), it was necessary to consider that the 2 and 5 kb hybridisation signals may be caused by non-specific hybridisation to ribosomal RNAs. However, several observations argued that this was not the case, and that the hybridisation of the 2 and 5 kb species reflected hybridisation of p46 RNAs. Firstly, the hybridisation was performed under stringent conditions, with yeast tRNA as a non-specific competitor and secondly, the abundance of ribosomal RNA in all samples was equivalent and certainly not enriched five-fold in brain (see figure 7.2B).

In summary, p46 is expressed as a 3 kb poly A+ transcript in most tissues, with the exception of adult liver and brain. The expression pattern of the RNA appears to be regulated in a tissue specific pattern, with higher levels of the 3 kb RNA in thymus and testis than in brain, where a 5 kb transcript is enriched. Three forms of p46 RNA (of 2, 3 and 5 kb) are enriched in human tissue culture cells, and the abundance of these RNAs appears to be greater than in EC cells, where the 3 kb species of p46 is predominant.

7.4. Quantitation of the p46 transcripts by RNAse protection

RNAse protection was used to quantitate the abundance of the p46 transcript in the murine tissues examined in the Northern assays. RNAse protection can be used as a very stringent test of homology between a probe and the RNA of interest. Nucleotide differences between the probe and target RNA which do not hybridise are degraded by the RNAses, which generate shorter hybridisation products that can be resolved on high resolution polyacrylamide gels.

The p46 probe (total length 460 nt) protected a doublet of RNA species of approximately 330 bp in all murine samples examined (figure 7.3A). These were shorter than the expected 360 bp fragment and were probably caused by digestion of the RNA transcribed from the slightly mismatching nucleotides incorporated at the 3' and 5' boundaries of the cDNA. The doublet could be caused by alternative splicing generating an additional RNA species with a slightly shorter region homologous to the 360 sequence, or by incomplete digestion of G/C regions within the mismatching termini of the cDNA by the RNAses. There were some additional low abundance hybridisation species which were
Figure 7.3. RNase protection assay of the p46 transcript.

RNase protection assays were set up incorporating both γ actin and 360 antisense RNA probes, and hybridised to 10 µg of yeast t, EC, PE, 12.5-17.5 dpc embryonic RNAs (lanes 3-8), in addition to adult tissue RNA (lanes 9-12). The γ actin transcripts shown are from a 1 hour exposure of the gel (B), and the 360 transcripts from 12 hour exposure (A).
Table 7.1 Quantitation of p46 RNAse protection assay

Mean results from two experiments scanned by laser densitometry are shown. The abundance of γ actin and p46 RNA is not directly comparable, as different amounts of radioactivity were incorporated into the probes. However, variation of RNA levels between tissues measured by each probe are quantitative.
<table>
<thead>
<tr>
<th>RNA Source</th>
<th>p46 360 hybridisation</th>
<th>γ Actin hybridisation</th>
</tr>
</thead>
<tbody>
<tr>
<td>F9 EC</td>
<td>57.6</td>
<td>4721</td>
</tr>
<tr>
<td>F9 PE</td>
<td>49.2</td>
<td>9428</td>
</tr>
<tr>
<td>12.5 dpc</td>
<td>45.0</td>
<td>7072</td>
</tr>
<tr>
<td>14.5 dpc</td>
<td>38.9</td>
<td>3286</td>
</tr>
<tr>
<td>17.5 dpc</td>
<td>24.3</td>
<td>2349</td>
</tr>
<tr>
<td>Adult thymus</td>
<td>48.5</td>
<td>2161</td>
</tr>
<tr>
<td>Adult brain</td>
<td>24.0</td>
<td>2817</td>
</tr>
<tr>
<td>Adult testis</td>
<td>56.0</td>
<td>5416</td>
</tr>
<tr>
<td>Adult Liver</td>
<td>7.3</td>
<td>712</td>
</tr>
</tbody>
</table>

**TABLE 7.1. QUANTITATION OF RNase PROTECTION ASSAY.**
shorter than the 330 bp major protected fragment, but these showed the same tissue dependent regulation as the major protected species, and therefore were likely to be degradation products of the 360 transcripts.

The protected RNAs show similar regulation to the 3 kb message in the Northern analyses. For example, the protected fragments were marginally down regulated from the levels in EC cell RNA to PE cell RNA, and were of low abundance in brain RNA. The low level of protected fragment in brain RNA may result from insufficient homology between p46 360 and the 5 kb RNA to withstand RNAse cleavage. The \( \gamma \) actin probe (137 nt) was also present in the hybridisation mixture for these RNAse protection analyses (figure 7.3B), and protected a fragment of 77 bp in murine tissue, which served as an internal control for degradation.

In summary, the p46 RNA species is marginally down regulated on differentiation of EC cells to parietal endoderm cells, whereas \( \gamma \) actin RNA levels increased on differentiation. Since \( \gamma \) actin RNA levels are thought to remain constant during differentiation of EC cells (P.Tassios, pers.comm.), the levels of p46 RNA may be more dramatically down-regulated during the differentiation of EC cells than suggested by these experiments. The levels of p46 RNA decreased during embryonic development, and p46 RNA was more abundant in adult thymus and testis than in brain. A 5 kb p46 360 hybridising species was abundant in Northern analyses of brain RNA, whereas only low levels of transcript were detected in RNAse protection analyses. The quantitation of the RNAse protection assays is shown in table 7.1.

A more detailed examination of the expression pattern of p46 RNA during embryogenesis was performed by \textit{in situ} hybridisation of sectioned mouse embryos.

7.5. \textit{In situ} hybridisation analysis of p46 gene expression during murine development

p46 RNA expression was examined during murine embryogenesis by \textit{in situ} hybridisation. The levels of RNA hybridising to the 360 antisense probe were compared with the levels of hybridisation to the sense probe in embryos at various stages of development.

The hybridisation signal detected from a particular region of the embryo is dependent not only on the level of expression of the gene in those cells, but is also dependent on the cell density within that region. Therefore, in order to compare the levels of hybridisation of the p46 RNA probe in different cell types, it was important to consider the cell density of the region by examination of the bright field photographs of the sections that accompany the dark field photographs depicting hybridisation intensity. Also, it was important to note that some areas of brightness in the dark field photographs are due to the refractile properties of certain tissues (most notably the lens of the eye and the blood cells) and do not correspond to autoradiographic silver grains. The distinction between these refractile areas and areas of genuine hybridisation signal was made by
Figure 7.4. *In situ* hybridisation analysis of p46 expression during murine embryogenesis

Figures A,C,E,G,I,L,P and P show bright field images of embryonic sections hybridised with control (C,I,L,P), or p46 probes (A,E,G,N). These pictures are used to determine the relative cell density when comparing dark field photographs of p46 expression (F,H,K,O,R), or where hybridisation is likely to be non-specific (probed with control probe D,J,M,Q). Sections shown in figures C,D,E,F,L to R were sagittal sections, and those in G and H were frontal sections.

p46 RNA was expressed at 4.5 days pc (A), at 8.5 dpc (C-H), at 10.5 dpc (I-K), 12.5 dpc (L-Q) and at 14.5 days (P-R). The hybridisation pattern of p46 to the 4.5 dpc embryo (shown in figure 4A) is shown in cartoon version in figure 4B. The boxed region shows the cell density of deciduum and endometrium, and dots represent hybridisation intensity. The increased hybridisation signal of the inner cell mass compared with the trophectoderm may reflect differences in cell density.

Various structures and tissues are labelled: A Aorta, AC Amniotic cavity, Am Amnion, C Coelom, D Intervertebral disc, De Deciduum, E Eye, F Foregut, FB Forebrain, LB limb bud, G Gut, GC Giant cells, H heart, HF Headfold, HL Hindlimb, HM Head mesenchyme, ICM Inner cell mass, K Kidney, Li Liver, Lu lung, M mesocoel, Ma mandible, Me mesenchyme, MG midgut, My myelocele, N Nasal cavity, Ne Neural epithelium, NF neural fold, NG neural groove, NT neural tube, PV prevertebral condensation, R ribs, S stomach, SC spinal cord, So Somite, T telocoel, Th thymus, To tongue, Tr trophectoderm, YS yolk sac.

Probes used were antisense RNA transcribed from the 360 cDNA encoding a portion of p46, and the control probe was the sense transcript from the same cDNA. Equal activity probe was used for these experiments, and exposure was for 17 days. Dark field photographs comparing p46 and control hybridisation efficiency were taken using exactly the same conditions of illumination, and exposure time. Bars indicate the magnification, and in A and B, bar represents 50 \( \mu M \), C-H, bar represents 100 \( \mu M \), I-K bar represents 500 \( \mu M \), and in L-R it represents 1000\( \mu M \).
comparing the sections probed with sense and antisense probes (see figure 7.4Q for example).

At 4.5 dpc (figure 7.4A), the blastocyst has hatched from the zona pellucida, and attaches to the uterine wall prior to implantation. P46 RNA was present in the embryo at this stage, both in cells of the inner cell mass which give rise to the embryo proper, and in the trophectodermal cells which generate extraembryonic membranes (refer to cartoon in figure 7.4B). The endometrial cells of the uterine wall also expressed p46 RNA, whereas cells undergoing the decidual reaction around the site of implantation of the blastocyst in the uterine endometrium showed less hybridisation to the p46 RNA. The abundance of p46 RNA in the inner cell mass cells appeared to be higher than in the trophectodermal cells of the embryo, although the differences in hybridisation intensity could reflect the relative density of cells in these regions. The reduced RNA levels seen in the decidual cells compared with the endometrial and embryonic cells cannot however be explained by differences in cell density, and therefore probably reflect real differences in p46 gene expression.

p46 therefore appears to be expressed in the 4.5 day embryo. By 8.5 days (figure 7.4C-H), gastrulation has occurred, and there is further differentiation of the body layers of the embryo. Expression of the p46 RNA was widespread at this stage, and hybridisation can be seen clearly within mesenchyme and extraembryonic membranes (figure 7.4F). In the dark field photographs 7.4F and 7.4H, the silver grains overlying the neural ectoderm appear to be clustered along one edge of this sheet of cells. This does not reflect a heterogenous distribution of p46 transcripts within these cells, but is a consequence of the organisation of this columnar epithelium. In this tissue, the cell nuclei are not only darkly stained by the toluidine blue, but are also closely spaced within the tissue. This effectively excludes illumination in the dark field situation, so that silver grains lying directly over the nuclei are not visualised. Comparison of the hybridisation signal obtained with the p46 probe and the control probe (compare D to F and H) showed that the p46 expression detected in the mesenchyme was specific.

At later stages of development (day 10.5, figure I-K) there was a fairly uniform expression of the p46 RNA throughout the embryo when cell density differences were also taken into account. The apparent exception was neural tissue, and cells lining the gut (figure K) in which the real hybridisation intensity could not be visualised because of the intensity of the blue staining in the epithelium.

At 12.5 dpc (figures 7.4L-O), there was apparently more p46 RNA expression in the snout, mandible (Ma) and tongue (To) of the animal than in other areas. This apparent expression difference may be due to a higher density of cells in these tissues (see figure N). Also, increased levels of p46 RNA were evident in and around the intervertebral discs (labelled D in figure N). This could be a reflection of the increased cell density in this
region, compared with the prevertebral condensations (Pv).

By 14.5 days (figures P-R), there was still widespread hybridisation to the p46 RNA probe (figure R). The hybridisation signal obtained with the p46 RNA probe over the whole embryo was much greater than in the more lateral section probed with the control probe (figure Q). This difference in antisense to sense probe hybridisation intensity was maintained in more medial sections of the embryo.

As in the 12.5 day embryo sections, there were regions such as the mandible and tongue of the 14.5 day embryo which appeared to show a stronger hybridisation of p46 RNA than other regions of the embryo such as the neural tissue. These variations were probably largely attributable to cell density differences rather than differences in the level of p46 expression in individual cells. One place where this was particularly obvious was in the ribs of 14.5 day embryos (figure R), where there was apparently less hybridisation signal than in the adjacent intercostal regions.

Expression of p46 RNA is therefore widespread throughout embryogenesis, with cell density and cell staining effects accounting for most of the regional distribution differences seen up to 14.5 days. P46 RNA was detected at 4.5 dpc in both the cells of the inner cell mass and the trophectoderm, and expression of p46 continues in both the extraembryonic and the embryonic cell lineages through development. There is apparently little tissue specific regulation of the expression of p46 RNA during the examined period of embryogenesis. Possibly there is a more marked tissue specific expression pattern of p46 at later stages of embryogenesis, but embryos of over 14.5 dpc were not examined in this study.

7.6. Discussion

The results from the Southern analyses indicate that a single gene is likely to encode p46, and that this gene has a complex intronic structure encompassing at least three introns within the 360 cDNA region. A alternative is that a family of genes are highly conserved with the 360 region of p46, and can hybridise to 360 under high stringency conditions.

Both of these alternatives are very interesting, as the 360 region of the p46 cDNA is sufficient for DNA binding of a 360 cDNA glutathione-S-transferase fusion protein to the adenoviral E2A promoter (L.Bandara and N.La Thangue, pers. comm.). If the 360 region is indeed shared by a number of related genes, these could constitute a novel family of E2F-site DNA binding proteins. Alternatively, if a single gene encodes p46, it is likely that the DNA binding region of p46 is encoded by several exons, as experiments to further define the region of 360 responsible for DNA binding activity have so far proved unsuccessful (L.B. and N.L.T. pers.comm.). Several transcription factors have recently been shown to have alternatively spliced transcripts, and the DNA binding activity of one of
these, twin of I-POU, (Treacy et al., 1992) is generated by alternative splicing incorporating six nucleotides into the homeodomain of an otherwise non-DNA binding protein I-POU. Thus if the DNA binding activity of p46 is dependent on the splicing of multiple exons within the 360 region, there is a possibility that alternative splicing within the 360 region of p46 may produce a different protein with altered properties. Such alternative splicing of the 360 region of p46 was possibly suggested from the doublet nature of the protected species in RNAse protection assays.

The p46 gene encodes a 3 kb poly A+ RNA. Northern analyses of the p46 transcript indicate that the RNA is not strongly regulated on differentiation of F9 EC cells, but is down regulated during development, with low levels of the RNA present at 17.5 dpc. There are marked tissue specific differences in the abundance of p46 RNA in the adult mouse, with low levels of RNA expressed in liver, and high levels of expression in thymus and testis. The level of the 3 kb message in brain RNA is low, and, instead, a 5 kb transcript was detected in brain tissue. A 5 kb transcript was also abundant in the human tissue culture cell RNAs examined, as was the 3 kb transcript, and also a 2 kb transcript which was present only at very low levels in the murine tissues. The in situ hybridisation experiments indicated that p46 is widely expressed, and that the gene is expressed at 4.5 dpc. The last time point examined in the in situ hybridisation analysis was 14.5 dpc, and at this stage, the expression of p46 was widespread, and did not show any marked tissue dependent variation.

The purpose of these experiments was to investigate the expression pattern of the p46 gene and to determine if there were any tissues in which DRTF1 binding activity was present, but in which the p46 gene was not expressed. All tissues with DRTF1 binding activity showed expression of the p46 gene, although in some cases the regulation of DRTF1 DNA binding activity between cell types must be at other levels in addition to the regulation of the steady state levels of p46 RNA expression. For example, the regulation of the DRTF1 DNA binding activity on differentiation of EC cells to PE-like cells (chapter 3, section 3.4) is much more marked than the slight down regulation of the p46 RNA levels on differentiation of EC cells. Thus DRTF1 DNA binding activity may be regulated by both changes in the steady state level of p46 RNA expression and by other mechanisms such as post-translational modification.
CHAPTER 8

GENERAL DISCUSSION
8.1. Summary of results.

In this study, I have shown that the DRTF1 DNA binding activity is down-regulated during stem cell differentiation, and generally, during murine development. As development progresses and the proportion of differentiated cells increases, there are increased levels of complexed DRTF1 activity. DRTF1 can complex with various cell cycle regulatory polypeptides, such as pRb and cyclin A, and I have shown that these interactions are likely to vary between different cell types, and at different stages of embryogenesis. As development progresses, higher levels of the complexed activity are detected, consistent with a role for Rb mediating its growth suppressive effects through transcriptional control.

The DRTF1 DNA binding polypeptides were purified, and amino acid sequence determined from peptides of one purified protein (p46). Nucleotide sequence was then deduced, and used to isolate a cDNA fragment (p46 360). This fragment was used as a probe to study the regulation of steady state RNA levels during differentiation and murine development. Both Northern and RNAse protection analyses indicated that the pattern of RNA expression was consistent with the earlier studies on DRTF1 DNA binding activity, and experiments performed subsequently by others in the laboratory have indicated that the cloned cDNA does indeed encode a component of the DRTF1 binding activity.

8.2. Regulation of the DRTF1 DNA binding activity during embryogenesis.

The regulation of DRTF1 DNA binding activity during development was assessed by preparing whole cell microextracts from tissues at various stages of embryogenesis. A similar technique had been used by Schöler et al. (1989) to demonstrate the wide variety of octamer DNA binding activities which were present during development. Once a portion of cDNA encoding p46 was isolated, the tissue specific patterns of gene expression could be compared with the regulation of the DRTF1 DNA binding activity during embryogenesis.

In situ hybridisation and gel retardation analyses indicated that DRTF1 is expressed early during embryogenesis. Indeed, hybridisation to the p46 360 cDNA probe was detected in the 4.5 dpc blastocyst, with hybridisation to the inner cell mass cells which subsequently give rise to the embryo. The expression of p46 RNA at this early stage of embryogenesis was expected as DRTF1 DNA binding activity was present in the 3.5 day embryo. At this stage, the only form of DRTF1 detected was DRTF1b. This result was not surprising, as DRTF1b is the most abundant form of the activity in stem cells (EC and ES), and thus was the most likely form of activity to be detected in early embryonic cells. Complexing of proteins with DRTF1 in preimplantation embryos may be prevented by the
presence of the cellular E1A like activity proposed to be present in embryos of this stage (Imperiale et al., 1984, Suemori et al., 1988, Dooley et al., 1989).

Another alternative is that protein complexing with DRTF1b does not occur at this stage of development because the associated proteins (eg pRb) are not yet expressed within the embryo. pRb complexes with DRTF1 during later stages of embryogenesis, but expression of \( Rb \) is not detected (by Northern analysis) until about nine days of development (Bernards et al., 1989). The pRb-like protein, p107, may also associate with DRTF1, but it is not yet known at what stage of development this protein is expressed. Therefore, uncomplexed DRTF1b activity is present at the earliest time of development examined, when cells are undergoing rapid cell division.

Since DRTF1 is a transcriptional activator (Shivji and La Thangue, 1991), and binding sites are present in a number of cellular genes whose products are involved in promoting cellular growth (Blake and Azizkhan, 1989, Hiebert et al., 1989, 1991, Mudryj et al., 1990, Thalmeier et al., 1991), it is possible that DRTF1 influences cell cycle progression. Accordingly, the levels of uncomplexed DRTF1 DNA binding activity remain high through early development, and levels decrease in some tissues including brain between 11.5 and 14.5 dpc, after the peak of proliferation of the tissue. By 17.5 days, the levels of DRTF1 DNA binding activity are low in tissues which are terminally differentiated (for example brain and liver), and are maintained only in tissues which retain a stem cell component, or in cells which are undergoing more cell division than in other tissues of the embryo (for example, thymus and testis).

During the late stages of development (from 14.5 dpc), the DRTF1 DNA binding activity is probably complexed to pRb. This association is widespread, suggesting that DRTF1 is bound by pRb in most cells. Indeed, comparison of the expression patterns of \( Rb \) (Bernards et al., 1989) with p46 indicate that both genes are expressed at higher levels in thymus and testis than in the liver of adult animals.

Complexing of DRTF1 by pRb is thought to reduce the transcriptional activity of DRTF1 as the coexpression of pRb in cells transfected with DRTF1 reporter constructs leads to a reduction in reporter activity (Zamanian and La Thangue, 1992). Therefore it is possible that as development progresses, and the rate of cellular proliferation decreases, there is a reduced requirement for DRTF1 activity. Expression of the p46 gene is reduced during development, and although not rigorously quantitated, the in situ hybridisation results suggest that the expression of p46 message is high in most tissues during the foetal period, and the RNAse protection analyses indicate that p46 expression is retained only in some tissues of the adult (eg thymus and testis). The tissues which retain DRTF1 binding activity have a higher proportion of complexed activity than seen at earlier stages of development, therefore it is likely that in these tissues, the activity of DRTF1 is reduced not by decreased expression of p46, but by complexing with pRb.
Such a mechanism of regulation allows a rapid generation of the transcriptionally active uncomplexed form of the protein in response to the appropriate stimulus, as opposed to the relatively slow steps of transcription and translation to produce the active protein. This is a novel mechanism for the reduction in transcriptional activity of a protein, and differs from that of retention in the cytoplasm (as shown by IkB inhibition of NFkB activity, Baueuerle and Baltimore, 1989) or heterodimerisation with a non-DNA binding partner (e.g. Id, Benezra et al., 1990) to prevent a transcriptional activator from binding DNA. The mechanism also contrasts with the reduction in activity of a protein by heterodimerisation with a less active partner (e.g. Δ fos B, Nakabeppu and Nathans, 1991).

This model whereby transcriptionally active DRTF1 is produced by dissociation of pRb from the complexed form of the activity assumes that there is a mechanism within the cell whereby pRb may be released from DRTF1. The release of pRb may be mediated by phosphorylation of pRb as Western assays have shown that the pRb protein copurifying with E2F is a fast migrating form of pRb, suggesting that it is underphosphorylated (Chellappan et al., 1991). Since the underphosphorylated form of pRb (from bacterial expression of a pRb fusion protein) can complex DRTF1 (Bandara and La Thangue, 1991) these studies suggest that it is the underphosphorylated form of pRb that binds E2F or DRTF1, and that these complexes might be disrupted by phosphorylation of pRb.

8.3. Cell cycle regulation of E2F containing complexes

Extensive studies of the cell cycle regulation of DRTF1 have not yet been performed, but from the data collected to date (L. Bandara and N. La Thangue, pers.comm.), there is no evidence to suggest that the regulation of association of proteins with DRTF1 is different to that of E2F during the cell cycle. This discussion is therefore based on the assumption that the cell cycle regulation of the DRTF1 activity is indistinguishable from that of E2F.

As discussed in section 1.17, pRb is phosphorylated in a cell cycle dependent fashion. Because pRb was thought to be underphosphorylated and active during the G1 stage of the cell cycle (Chen et al., 1989, Mihara et al., 1989, Buchkovich et al., 1989, DeCaprio et al., 1989), the presence of the underphosphorylated form in association with E2F suggested a model whereby Rb may regulate the cell cycle through E2F.

pRb/E2F complexes were shown to be present during the G1 phase of the cell cycle in cells which were cell cycle coordinated by serum starvation (Mudryj et al., 1991) and in cells synchronised by elutriation, the pRb/E2F persisted into S phase (Shirodkar et al., 1992).

Recent data suggests that there are earlier phosphorylation events that occur during G1 (DeCaprio et al., 1992), and the underphosphorylated form of pRb does not persist until the G1/S boundary. Therefore, it is possible that E2F is complexed with
underphosphorylated pRb during early G₁, and that this complex may persist after pRb has undergone the first steps of phosphorylation, or even until after the major phosphorylation events occurring at the G₁/S transition. Alternatively, E2F may only bind unphosphorylated pRb, in which case, pRb containing complexes found in S phase enriched cells may be 'contamination' of the S phase cells with G₁ phase cells, or caused by binding of E2F to newly synthesised pRb. If the underphosphorylated form of pRb is the only active form, the finding that pRb is phosphorylated at a point in G₁ may explain the observation by Goodrich et al. 1991, that injected pRb prevented cell cycle progression only if injected during early G₁, and not at later times close to the G₁/S boundary. It is possible that some critical phosphorylation event occurs at the G₁/S transition which prevents pRb binding to E2F or DRTF1, possibly by changing the conformation of pRb. The kinase that phosphorylates pRb is likely to be the cdc2-related kinase, cdk2, which consists of p33 MAP and either cyclin A or E (dependent on the stage of the cell cycle, Dulic et al., 1992). At the stage of the cell cycle when the initial phosphorylation of pRb is thought to occur, p33 MAP is probably complexed with cyclin E (as cyclin A is not synthesised until the S phase of the cell cycle).

During S phase, a novel E2F containing complex is detected, which accumulates and peaks during S phase, and disappears by the G₂/M phase transition. This complex has been shown by antibody recognition experiments to contain cyclin A, the pRb-like p107 protein, and the cdk2 kinase activity (Mudryj et al., 1991, Pagano et al., 1992a, Devoto et al., 1992, Shirodkar et al., 1992, Cao et al., 1992). It is not yet clear whether the association of cyclin A with E2F is dependent on the presence of p107, but the characterisation of complexes in which cyclin A is absent and p107 is complexed to E2F (Pagano et al., 1992a) together with the observation that cyclin A binds p107 via the spacer region separating the domains of the pocket region of the protein (Ewen et al., 1992) suggests that cyclin A interaction with E2F is mediated by p107.

Recently, another form of the E2F complex has been characterised, which is also present during the G₁ phase of the cell cycle, and consists of cyclin E, p107 and cdk2 complexed with E2F (Lees et al., 1992). Thus at the same stage of the cell cycle, E2F, and presumably DRTF1 can associate with pRb and also the kinase complex which may effect the release of DRTF1 from pRb.

The finding that cyclins A and E associate with E2F suggests several possible mechanisms by which E2F may play a role in cell cycle progression, as both cyclins A and E are capable of overcoming an Rb mediated block to the cell cycle (Hinds et al., 1992). The kinase activity of p33 MAP may be targetted to the promoter elements of genes via E2F site binding proteins in order to phosphorylate other DNA binding proteins. Alternatively, the cyclin A kinase complex may be required for the initiation of DNA synthesis, as cyclin A has been shown to be important for DNA replication (Girard et al., 1991). Another
possible explanation, is that the kinase is incorporated into the E2F complex to phosphorylate either E2F, or the E2F complexing activities possibly to modulate the transcriptional activity of E2F.

The predominance of the various forms of DRTF1 activity at different stages of development correlates well with what is known about the regulation of E2F during the cell cycle and with the developmental regulation of the cell cycle. For example, at early stages of development, the cell cycle has a very short $G_1$ phase (reviewed by Pardee, 1989), correlating with low levels of pRb complexed DRTF1 activity in the early embryo. As development progresses, and as the rate of proliferation slows, the cells spend longer in the $G_1$ phase of the cell cycle, and hence increased levels of the pRb complexed DRTF1 activity can be detected. As the cells become terminally differentiated, they exit the cell cycle. Experiments performed with serum starved cells have shown that there is no detectable E2F binding activity during $G_0$ (Mudryj et al., 1990), and the absence of DRTF1 DNA binding activity in the adult liver and brain suggests that these cells have exited from the cell cycle.

Several E2F-site containing promoters (\textit{myc, myb} and dihydrofolate reductase, \textit{DHFR}) are coordinately expressed on serum stimulation (reviewed in Mudryj et al., 1990), and the cell cycle regulation of E2F site binding activities may be important for this control. The E2F sites in the \textit{myc} and \textit{DHFR} promoters have been shown to be functionally important, and \textit{myc} is \textit{trans} activated by E1A in an E2F site-dependent fashion (Blake and Azizkhan, 1989, Hiebert et al., 1989, 1991, Thalmeier et al., 1989), suggesting that activation is mediated by removal of pRb or other proteins from E2F. The \textit{DHFR} promoter has two E2F sites situated 3' to its transcriptional start site, and is first transcribed in late $G_1$, possibly correlating with the release of E2F from pRb repression as pRb is phosphorylated, or with the formation of other activatory forms of complexed E2F activity. Similarly, the \textit{cdc2} promoter is repressed during $G_1$, and activation occurs at the $G_1$/S transition of the cell cycle, with expression maintained until $G_2$. Transcription of \textit{cdc2} is also repressed during $G_0$, and can be stimulated in non dividing cells by 12S E1A (Dalton, 1992), as reported for endogenous \textit{cdc2} (Wang et al., 1991). Therefore the cell cycle regulation of E2F transcriptional activity is likely to be important for regulated expression of genes associated with proliferation control.

8.4. Regulation of DRTF1 during F9 EC cell differentiation.

In addition to the regulation of DRTF1 transcription factor activity by complexing with pRb, there is regulation of the DNA binding activity during differentiation of EC cells to parietal endoderm. DRTF1 DNA binding activity is strongly down regulated upon differentiation of EC cells. However, there is comparatively little change in the abundance of the p46 RNA transcript, and experiments could be performed to determine whether the
level of the p46 polypeptide is regulated or if the polypeptide is modified during differentiation.

8.5. Purification and cloning of DRTF1.

DRTF1 was purified to generate peptides for sequence analysis. The purification was improved by inclusion of a mutant site affinity column in the procedure to eliminate non-specific DNA binding activity. Purified DRTF1 yielded several DNA binding polypeptides in crosslinking analyses, together with a range of coomassie staining polypeptides. Experiments were therefore designed to identify the major polypeptides of DRTF1 and to cleave these for peptide sequencing. It was difficult to determine which of the coomassie staining proteins were components of DRTF1, for the reasons outlined in chapter 5. The crosslinking of polypeptides to the DRTF1 DNA binding site was the best method to identify the DRTF1 polypeptides, but it suffered the disadvantage of the unknown contribution of the DNAse treated probe to the molecular weight of the crosslinking polypeptides. A crude correction value was obtained, and the molecular weight of the crosslinked polypeptides was modified accordingly to obtain a more accurate determination of their molecular weight.

The variety of crosslinking species may be different forms of the p46 gene product produced by alternative splicing, by use of alternative polyadenylation sites, or by post-translational processing. Furthermore, the faster migrating species may be produced by proteolytic degradation of the p46 protein. Alternatively, several different gene products may crosslink to the DRTF1 binding site, and these gene products may use similar or different DNA binding domains to p46.

Other polypeptides in the preparation did not crosslink to DNA. These polypeptides co-purified with DRTF1 through a stringent purification protocol, with non-specific DNA included at all times to prevent non-specific binding of polypeptides to the affinity matrices. Also, the mutant site affinity step was specifically included to prevent the co-purification of non-specific DNA binding activities. Despite these measures, some non-specific DNA binding activities might have co-purified with DRTF1. Also, polypeptides might have purified with DRTF1 because they are tightly associated with the DRTF1 activity, as for example coactivators are purified with TBP (Pugh and Tjian, 1991). These polypeptides are tightly associated with TBP, and can only be dissociated in vitro by treatment with urea (Tanese et al., 1991). Possibly therefore, some of the 'contaminants' of the DRTF1 preparations are coactivators of DRTF1. Further experiments comparing the nature of the recombinant p46 polypeptide with the affinity purified activity will resolve these issues.

The estimation of molecular weight of the DRTF1 polypeptides by crosslinking allowed the subsequent isolation and preparation of peptides from a component of DRTF1.
Once amino acid sequence of the peptides was obtained, it was important to determine whether the purified p46 protein was a component of DRTF1. Therefore, I attempted in vitro translation of both EC cellular RNA and poly A+ selected RNA to assess if DRTF1 activity could be produced, and whether this activity could be abolished by inclusion of p46 specific antisense oligonucleotides which should prevent the translation of p46 RNA. Unfortunately, both wheatgerm and rabbit reticulocyte translation systems included large amounts of DRTF1 or E2F binding activity which was sufficient to mask any production of activity from the translated products. Subsequent experiments performed in the laboratory have indicated that p46 is indeed a component of DRTF1.

8.6. E2F-1.

Recently another component of the E2F or DRTF1 site binding activity has been cloned and characterised (Helin et al., 1992, Kaelin et al., 1992). The cDNA was isolated from a human pre B leukaemic cell line expression library screened with a segment of pRb encompassing the pocket domain. The gene product possesses many of the characteristics expected for E2F, but it probably is not involved in all E2F complexes (Kaelin et al., 1992), as the RNA does not appear to be present at all times of the cell cycle when E2F protein is present (Kaelin et al., 1992). Kaelin has therefore suggested that E2F-1 is a component of the E2F activity, but that there could be other genes encoding further components of the activity. p46 is a possible candidate, as an anti-p46 peptide antibody showed that p46 was involved in all DRTF1 complexes in EC cell extracts (N. La Thangue, pers.comm.)

8.7. Comparison of p46 and E2F-1 DNA binding domains

The DNA binding domain of E2F-1 was crudely mapped to the amino terminal 284 amino acids of E2F-1 (Kaelin et al., 1992). This portion of the protein has been proposed to encompass a HLH motif and a leucine zipper-like domain. These motifs may be important for mediating DNA binding and/or protein dimerisation. However, the DNA binding domain of E2F-1 was more finely mapped by Helin et al., 1992, and amino acids 89-191 were shown to be sufficient for DNA binding (see figure 8.1A), indicating that the leucine zipper like domain was not essential for DNA binding activity. Analysis of the sequence of E2F-1 with GCG cube search has not detected significant homology with other proteins of the helix-loop-helix families.

The 360 DNA binding domain of p46 showed little homology with any known DNA binding motifs suggesting that p46 has a novel DNA binding domain. Comparison of the amino acid sequence of E2F-1 and p46 using the GCG v7 BESTFIT program indicated a low overall similarity (44%). This homology increased to 46% over the DNA binding domains of the proteins, which, interestingly, lie at very similar positions within the two
Figure 8.1. Sequence comparison of E2F-1 with p46 360

A. Amino acid sequence of E2F-1 from amino acids 90 to 191, which encompasses the DNA binding domain. Brackets indicate the basic region, and and the potential helices of the protein (Kaelin et al., 1992).

B. Bestfit sequence comparison of the DNA binding domain of E2F-1 with the DNA binding region of p46 (360). Sequence identities are indicated by vertical lines, and conserved amino acid differences represented by dots. The sequences were 46% similar within the DNA binding regions, and showed 25% identity. The similarity between the two domains was increased to 65% (46% identity) in the region within the solid bars (165-191 of E2F-1, and 167-192 of p46).

C. Matrix plot of sequence comparison between p46 360 amino acid sequence, and the sequence of the DNA binding domain of E2F-1. Diagonal lines represent regions with at least 40% homology over a window of eight residues. Data prepared by the Macvector Pustell matrix programme (Macintosh, IBI).
Sequence comparisons of the DNA binding domains of E2F-1 and p46.
gene products. The alignment shown in figure 8.1B and C demonstrates that the region of highest homology between the two proteins lies within the carboxyl terminus of the proposed basic helix loop helix zipper region of E2F-1. The proteins are 65% similar, and 46% identical within this region, which is predicted to form an α helical region, with the helix forming an amphipathic structure. The helical region may play a part in determining the specificity of protein-DNA interaction, as it is conserved between these proteins that recognise a common sequence.

8.8. pRb binding domains of E2F-1 and p46

E2F-1 binds pRb directly through a short region of 18 amino acids close to the carboxyl terminus of E2F-1 (Helin et al., 1992). These residues are sufficient to bind pRb when they are expressed alone in a fusion protein. This pRb binding motif differs from the LXCXE motif present in viral oncoproteins which bind pRb, and represents a novel class of pRb binding domain. The homology search between E2F-1 and p46 did not identify a region in p46 with homology to either the pRb binding domain of E2F-1 (amino acids 409-426, Helin et al., 1992), or the LXCXE motif conserved between viral oncoproteins and RBP1 and 2 (Defeo-Jones et al., 1991). C-myc has also been proposed to bind pRb (Rustgi et al., 1991), but this protein also lacks the novel E2F-1 or the LXCXE motif. It is possible therefore that additional motifs can confer pRb binding activity, or that p46 and c-myc do not bind pRb directly but complex with proteins which bind pRb.

8.9. Further experiments.

The isolation of cDNA clones encoding a component of DRTF1 has allowed the generation of many useful reagents for DRTF1. Polyclonal antibodies have been raised which can be used for analysis of the regulation of the protein. These tools, in combination with probes for the E2F-1 activity should help resolve questions concerning the relationship of the two activities, and how they regulate E2F-site dependent transcription.

8.9.1. Functional analyses of the recombinant DRTF1.

The DNA binding domain of p46 has been mapped to the 360 region of the cDNA (residues 84-192 of the protein). It is possible that a smaller region is sufficient for DNA binding activity, and this region could be mapped by exonuclease digestion of the 360 cDNA fragment.

The trans activation domain of DRTF1 may be mapped by generation of GAL4 fusion proteins with fragments of p46, and assaying the activation potential of these proteins by cotransfection with a reporter bearing GAL4 binding sites. The pRb binding domain of p46 could be mapped, and it could be determined if, as in E2F-1, the pRb binding domain overlaps the activation domain of the protein (Kaelin et al., 1992), and
whether binding of pRb to the protein masks the activation domain, and prevents its interaction with the general transcriptional apparatus.

It would be interesting to compare the properties of the recombinant p46 protein with purified DRTF1 DNA binding activity to determine if both proteins can function as transcriptional activators, or whether purified DRTF1 has accessory proteins or coactivators which modulate the transcriptional activity of the protein. Additionally, antibodies raised against both p46 and E2F-1 could be used to determine if the purified DRTF1 binding activity is the product of p46 gene expression alone, or whether the E2F-1 DNA binding activity contributes to the EC cell DNA binding activity.

8.9.2. A family of DRTF1 polypeptides?

Several cellular genes possess more than one DRTF1 binding site (see Chittenden et al., 1991, Mudryj et al., 1990), and it is possible that DRTF1 binds cooperatively to these sites and forms more stable interactions with repeated binding sites than on binding to a single site. Experiments could be performed to assess this possibility, and also whether there is a cellular equivalent of the adenoviral E4 ORF6/7 gene product, which mediates stable cooperative binding of E2F to the E2A promoter during adenoviral infection. It would also be interesting to determine if p46 could heterodimerise with other proteins, for example E2F-1, and whether these complexes show differential DNA binding specificities or activation potential compared with the homodimeric proteins. If p46 could homo or heterodimerise, it would be useful to define the domains necessary for this interaction.

The Southern analyses used a probe encompassing the DNA binding domain of p46, and showed hybridisation to several DNA fragments, some of which may have been caused by cross-hybridisation to related genes. Potential homologues of p46 (and E2F-1) could be cloned, and experiments could be designed to determine whether there is a degree of functional redundancy within the E2F site binding proteins, or whether each gene product plays a specific role in transcriptional activation and in cell cycle progression.

8.9.3. Regulation of complex formation during differentiation and development.

The cell cycle dependence of DRTF1 complex formation could be analysed to see if indeed the p46 gene product is complexed by the same polypeptides as E2F-1 at the same stages of the cell cycle. One possibility is that if p46 and the murine homologue of E2F-1 are both expressed in the same cell at the same time, each polypeptide is complexed by a different set of regulatory polypeptides. For example, E2F-1 may complex pRb during the G1 phase of the cell cycle, whereas p46 complexes p107, cyclin E and cdk2. The complexes formed at each stage of the cell cycle can be isolated and their constituents analysed now that it is possible to generate probes for many of the component activities. This type of analysis will clarify the data presented in this thesis regarding the regulation
of the complexes formed during differentiation and development. For example, such an
analysis may explain the incomplete supershifting by anti-pRb antibodies of the DRTF1a'
complex enriched during embryogenesis, identify the other components of the a' complex,
and clarify whether p107 is involved with cyclin A in the EC cell DRTF1a complex.

Experiments could be designed to determine whether E2F or DRTF1 can bind pRb
that has been phosphorylated, and to ask why the major phosphorylation events of the
G_{1}/S transition apparently prevent interaction of pRb with DRTF1 or E2F. The kinases
responsible for the various phosphorylation steps could be identified; can p33^{adt} complexed
with cyclin E phosphorylate all of the sites of pRb which are phosphorylated during the G_{1}
phase of the cell cycle, or are additional kinase activities involved? Also, other targets of
the regulatory kinases could be identified, such as other transcriptional activators, or
components of the basal transcriptional apparatus, which may yield information on the
coordination of the cell cycle progression.

8.9.4. DRTF1- a novel proto-oncogene?

Although p46 gene expression is reduced in some tissues of the adult, there is
expression of p46 in tissues such as the thymus and testis, where the activity of the protein
is apparently regulated by complexing with the pRb polypeptide. If one component of this
control system was perturbed, for example, if pRb was mutated so that it could no longer
complex efficiently with DRTF1, active DRTF1 would be present in several tissues,
presumably activating transcription of E2F site containing genes such as c-myc at stages of
the cell cycle when these genes are normally transcriptionally silent. Also, if the region of
p46 or E2F-1 responsible for interaction with pRb was mutated, the E2F-1 site binding
activity would no longer be under cell cycle regulation, and target genes would be
inappropriately activated. Additionally, the constitutive overexpression of either p46 or
E2F-1 by chromosomal translocation to a constitutively active promoter may be
tumorigenic. The chromosomal location of these genes has not yet been mapped, but it
would be interesting to see if the mapping correlated with regions associated with mutation
or translocation in tumours. However, there is the possibility of degeneracy within the
control system, for instance some cell cycle regulated transcription might be controlled by
p107 in the absence of Rb, and E2F-1 may substitute for DRTF1. The Rb knockout mice
die by 15 days of embryogenesis, showing that Rb is an essential gene. It is possible that
some of the effects of loss of pRb were caused by loss of cell cycle control of the E2F site
binding activities, but not enough is known about other roles for pRb to determine if any
of the developmental defects were caused by an inability of other gene products to
compensate for Rb in control of the E2F site binding activities.
DRTF1 is therefore a developmentally regulated transcription factor, which is likely to play an important role in the control of cell cycle progression. The abundance of the activity at early stages of embryogenesis, the increased levels of pRb association in tissues retaining DRTF1 binding activity in the adult, and the presence of DRTF1 binding sites in the promoters of genes whose products promote cell cycle progression suggests that the activity of the protein is modulated by the cell cycle.

DRTF1 was the first transcription factor characterised whose activity was regulated by pRb, and thus established the first link between a tumour suppressor gene product and transcriptional control. Additionally, the identification of complexes of DRTF1 associated with cyclins provided a further link between transcription and the control of the cell cycle.

The identification of multiple cDNA products which can bind the DRTF1/E2F site suggests there may be a degree of degeneracy within this transcriptional control system. Alternatively, the individual DNA binding polypeptides may play distinct roles in the control of cell cycle progression. The isolation of such molecular probes will facilitate experiments to determine how the pRb tumour suppressor and other cell cycle regulatory proteins control transcriptional activity.

It will be interesting to determine whether the regulation of transcription factors by tumour suppressor or cell cycle regulatory polypeptides is a common mechanism for control, and to determine whether perturbations of such regulatory systems lead to cancer.
ACKNOWLEDGEMENTS

I would like to thank Nick La Thangue for his expert and enthusiastic supervision of this project, and for his support and encouragement over the last four years. I am grateful for the opportunity he gave me to join his lab at NIMR, and for the scientific training I received. Thanks also to David Latchman for his enthusiasm and interest in the octamer project, and for support during my time at NIMR.

I am grateful to Ed Harlow, Nic Jones, Tim Hunt and David Solter for gifts of reagents. Warm thanks to Robin Lovell-Badge and Blanche Capel for their expert help with embryo dissections and to SPY and Jo Chan for superovulating many mice. Many thanks also to Margaret Goggin and Colin Young for their willingness to try out new methods of bulk culture, Rossi for useful suggestions and ES cell cultures and Paula Timmons and Wendy Hatton for the embryo sections for in situ hybridisation. Thanks to Alistair Aitken for lengthy discussions on peptide production, and Allan Harris for successfully sequencing the 98 kD peptides.

I am also grateful to the members of the photographic and graphic units for their efficient and good humoured artistry, and indebted to Sarah Carrick for her art work in chapter 1.

I greatly appreciate the support and useful technical advice from past and present members of both the La Thangue and Rigby labs, who are too numerous to list, but individually remembered. Special thanks to SPY for his technical tips and great cooking, and to Paula for help with the in situ. Thanks also to Panayotis Tassios, Lan Bandara, Rowena Girling, Maryam Zamanian and Mahmud Shivji for sharing of reagents and ideas, and Paula, Dave Cox and Chandi Patel for the smooth running of the lab.

Many thanks also to Vicky Buck, Lan and Nick for careful reading of the manuscript and useful suggestions on its improvement, and to Debbie Duthie for Wordperfect tips. Lastly, thanks to the EEC and SGM for subsidising meetings and SERC for providing financial support for the first three years.

Cell 42, 599-610.

Science 255, 979-983.

J. Biol. Chem. 266, 8055-8061.


Cell 65, 1063-1072.

Cell 55, 531-535.


Nature 334, 37-42.

Bioessays 10, 166-169.

Trends Genet. 6, 36-40.


Science 243, 814-817.

Nucleic Acids Res. 16, 10237-10247.

EMBO J. 7, 835-841.

EMBO J. 7, 29-35.

Cell 65, 1053-1061.

Cell 58, 1193-1198.


Biochemistry 18, 5294-5299.

Cell 65, 1073-1082.

Biochemistry 13, 222-245.


Edmonson, D.G., Olson, E.N. (1989). A gene with homology to the myc similarity region of MyoD1 is expressed during myogenesis and is sufficient to activate the muscle differentiation program. *Genes Dev.* 3, 628-640.


Cell 61, 27-38.

Molec. Cell. Biol. 6, 801-810.

Science 227, 134-140.

Nature 292, 154-156.

Cell 33, 389-396.

Science 255, 85-87.

Cell 58, 257-267.

Cell 66, 1155-1164.

Science 255, 87-90.

Genes Dev. 4, 29-42.

Cell 66, 731-742.

Ferguson,B., Kripp,B., Andrisani,O., Jones,N., Westphal,H., Rosenberg,M. (1985). E1A 13S and 12S mRNA products made in E.coli both function as nucleus-localized transcription activators but do not directly bind DNA. 


EMBO J. 1, 681-686.

233
*J. Biol. Chem.* 267, 2786-2793.

*Cell* 64, 739-749.

*Nature* 355, 80-84.


*Nature* 323, 643-646.

*Cell* 23, 809-814.

*Science* 250, 805-808.

*J. Mol. Biol.* 120, 97-120.


*Cell* 67, 197-211.

*Cancer Res.* 12, 264-265.

*Cell* 62, 1019-1029.

*Nucleic Acids Res.* 12, 7877-7888.


Genes Dev. 3, 1062-1074.


structure-function relationships of yeast TATA box binding factor TFIID. 
*Cell* 61, 1171-1178.


*Science* 243, 937-940.


Hu, Q., Dyson, N., Harlow, E. (1990). The regions of the retinoblastoma protein needed for binding to adenovirus E1A or SV40 large T antigen are common sites for mutations. 
*EMBO J.* 9, 1147-1155.

*Genes Dev.* 4, 1741-1752.

*Science* 242, 1563-1566.

*Genes Dev.* 3, 1699-1710.


Hurst, H.C., Jones, N.C. (1987). Identification of factors that interact with the E1A-inducible adenovirus E3 promoter. 
*Genes Dev.* 1, 1132-1146.


Klein,G. (1990). Multiple emancipation of tumors from growth control: can it be curbed in a single step? 


determine transcriptional activity and inducibility.  


Li, S., Crenshaw, E.B., Rawson, E.J., Simmons, D.M., Swanson, L.W., Rosenfeld, M.G. (1990). Dwarf locus mutants lacking three pituitary cell types result from mutations in the POU-domain gene *pit-1*.  


*Cell* **50**, 1091-1100.


Liu, F., Green, M.R. (1990). A specific member of the ATF transcription factor family can mediate transcription activation by the adenovirus E1a protein.  
*J. Biol. Chem.* 264, 6572-6579.


*Cell* 56, 57-65.

*Cell* 60, 387-396.


*Cell* 51, 113-119.

*Cell* 55, 443-446.

*EMBO J.* 8, 2023-2028.

*Oncogene* 6, 1417-1422.

*Cell* 65, 775-783.

Genes Dev. 4, 955-967.


Nikoderm, V., Fresco, J.R. (1979). Protein fingerprinting by SDS-gel electrophoresis after partial fragmentation with CNBr.


at two points in the human cell cycle.

*EMBO J.* 11, 961-971.


and protein regulation in the cell cycle and for interaction with p34cdc2.
Cell 58, 833-846.


Genes Dev. 5, 2327-2341.

Trends Genet. 8, 91-96.

J. Virol. 65, 1803-1811.


Cell 61, 1187-1197.

Nucl. Acids Res. 12, 643-655.

Genes Dev. 6, 953-964.

Science 245, 640-643.

Genes Dev. 2, 1687-1699.

Genes Dev. 5, 1200-1211.

J. Virol. 64, 2702-2710.

Genes Dev. 3, 620-627.


Structural homology of the product of the Drosophila Kruppel gene with Xenopus transcription factor IIIA.

Genes Dev. 5, 897-907.

Nature 310, 415-418.

Nature 345, 686-692.


Cell 57, 317-326.

Nature 304, 602-607.

Bioessays 14, 89-95.

Nature 352, 541-544.

Cell 64, 475-478.

Nature 335, 563-564.


multiple cDNAs.


Simmons, D.M., Voss, J.W., Ingraham, H.A., Holloway, J.M., Broide, R.S., Rosenfeld, M.G., Swanson, L.R. (1990). Pituitary cell phenotypes involve cell-specific Pit-1 mRNA.
translation and synergistic interactions with other classes of transcription factors.

*Genes Dev.* 4, 695-711.


*Cell* 52, 723-729.


*J. Biol. Chem.* 261, 538-547.


*Cell* 52, 415-423.


*Genes Dev.* 5, 984-995.


*EMBO J.* 4, 3831-3837.


*Cell* 57, 103-113.


*Genes Dev.* 3, 2091-2100.


*Cell* 63, 1013-1024.


Sommer, L., Hagenbuchle, O., Wellauer, P.K., Strubin, M. (1991). Nuclear targeting of the transcription factor PTF1 is mediated by a protein subunit that does not bind to the PTF1 cognate sequence.

*Cell* 67, 987-994.


*J. Cell physiol.* **85**, 179-188.

*Cell** **66**, 199-217.

*Cell** **60**, 375-386.

*Genes Dev.* **5**, 2212-2224.


*New Biol.* **2**, 1123-1134.


*Science* 243, 1689-1694.


*EMBO J.* 10, 1817-1825.

*Cell* 69, 871-881.

*Cell* 40, 705-716.


*Science* 251, 761-766.


in response to F9 embryonal carcinoma stem cell differentiation.

_Cell_ 59, 1081-1092.


_Cell_ 56, 67-75.


_Nature_ 334, 124-129.


_Oncogene_ 7, 423-432.


_Genes Dev._ 2, 1557-1569.


_Genes Dev._ 4, 313-323.


_Cell_ 56, 607-617.


_Cell_ 53, 433-440.


_Cell_ 65, 691-699.


_Nature_ 334, 494-498.


