STRUCTURAL AND FUNCTIONAL ANALYSIS OF THE RABBIT ERYTHROID-SPECIFIC LIPOXYGENASE GENE PROMOTER.

John David Chester.

A thesis submitted for the degree of
Doctor of Philosophy
in the University of Glasgow.

The Beatson Institute for Cancer Research, Glasgow.

September, 1988

A.M.D.G.

And to make an end is to make a beginning
The end is where we start from.

T. S. Eliot (1888-1965)

"Little Gidding"

Summary of Contents.

Acknowledgements	i
Abbreviations	ii
List of Figures	iii
List of Tables	iv
List of Text Headings	v
Abstract	xiii
INTRODUCTION	1
MATERIALS AND METHODS	55
RESULTS	83
DISCUSSION	128
REFERENCES	150

Acknowledgements.

I am extremely grateful to my supervisor, Dr. Paul Harrison for his constant encouragement, for innumerable discussions on strategy, and for his ability to frequently make me feel more confident in my work. I also owe a great deal to Dr. Jon Frampton for his early guidance, constant enthusiasm, perpetual optimism and excellent scientific example. Drs. Mark Plumb and David Llewellyn have provoked numerous helpful discussions, and I am also endebted to all the other members of R4 for their help over the last three years, particularly to Janis Fleming and to Jim O'Prey as fellow members of the LOX sub-group.

None of this work would have been possible without the previous hard work and continued collaboration of others. I would like to thank Dr. Jon Frampton for so much trail-blazing, especially for establishing the transient expression system used here and for his experience of mRNA analysis, but also for the construction of pHSV-βgal, for the gift of pαCAT1, and for help in the construction of pCAT-40E and pCAT-150E plasmids. Jim O'Prey has also been of immense help in providing the genomic sub-clones, pLOX6/1.5R and pLOX6/2.5R, and CAT recombinants pCAT19, pCAT-40, pCAT-150 and pCAT-2700. Thanks are also due to Janis Fleming for providing the sequence of plasmid pGF6 and for masses of technical advice, and to Dr. Bernd Thiele for communication of cDNA sequences prior to publication. I am also grateful to Dr. Mark Plumb, to Kay Macleod and to Helen Wainwright for nuclear protein extracts and their footprinting expertise.

I am personally endebted to Rosey Proctor, Jon and Helen Frampton, Mark Rigby, and Richie and Sadie Vile for their lasting friendship, for providing most of the good bits of the last three years, for innumerable excellent meals, and for invaluable assistance and moral support during the final preparation of this thesis. I gratefully acknowledge the blessings of a supportive family and a marvellous bunch of friends, to all of whom this work is affectionately dedicated.

This work has been supported by a Research Studentship from the Cancer Research Campaign.

Abbreviations.

DNA	deoxyribonucleic acid
cDNA	complementary DNA
RNA	ribonucleic acid
tRNA	transfer RNA
mRNA	messenger RNA
mRNP	messenger ribonucleoprotein particle
A	adenine
C	cytosine
G	guanine
T	thymine
N	A, C, G, or T as above
U	uracil
Pu	purine (A or G)
Py	pyrimidine (C or T)
dNTP	deoxyribonucleoside triphosphate
dATP	deoxyadenosine triphosphate
dCTP	deoxycytidine triphosphate
dGTP	deoxyguanosine triphosphate
dTTP	thymidine triphosphate
ddNTP	dideoxyribonucleoside triphosphate
LOX	lipoxygenase
DNase	deoxyribonuclease
RNase	ribonuclease
CAT	chloramphenicol acetyltransferase
nt	nucleotides
bp	base pairs
kb	kilobase pairs
kD	kiloDaltons
ml	millilitres
µg	micrograms
nm	nanometres
pmol	picomoles
V	Volts
mA	milliAmperes
OD	optical density (absorbance)
% w/v	weight in grammes per 100ml water
% v/v	volume in ml per 100ml water
SDS	sodium dodecyl sulphate
Amp	ampicillin
Tris	Tris(hydroxymethyl)methylamine
EDTA	Ethylenediamine tetra-acetic acid
YFG	Your Favourite Gene

List of Figures.

		Following page
Figure 1.	Differentiaton of multiple cell lineages during haemopoiesis.	33
Figure 2.	Catalytic activity and substrate specificity of the lipoxygenase enzyme family.	42
Figure 3.	Strategy for sequencing genomic sequences flanking the translation initiation site.	84
Figure 4.	Comparison of rabbit RBC 15-LOX genomic and cDNA sequences.	84
Figure 5.	S1 nuclease analysis using a 72bp Sau3A probe derived from pGF6 cDNA insert.	87
Figure 6.	Summary of strategy and results in mapping of the 5′ end of the rabbit RBC 15-LOX mRNA.	89
Figure 7.	S1 nuclease mapping of the 5′ end of the first protein-coding exon using a single-stranded 276nt SstI genomic probe.	89
Figure 8.	S1 nuclease analysis using a single-stranded 540nt EcoRI/BalI genomic probe.	89
Figure 9.	Optimum hybridisation temperature for primer extensions using a single-stranded 58nt Sau3A/SstI genomic primer.	91
Figure 10.	Primer extension and mRNA sequencing using a single-stranded 58nt Sau3A/SstI genomic primer.	91
Figure 11	Optimum hybridisation temperature for primer extensions using the 166nt BalI cDNA-derived primer; preliminary mRNA sequencing; and detection of minor transcripts.	94
Figure 12.	Primer extension on a 166nt BalI cDNA-derived primer: mRNA sequence and minor transcripts.	94
Figure 13.	Northern blot analysis of RBC 15-LOX transcripts.	97
Figure 14.	Position of the transcription initiation site and 5′ flanking sequences of the RBC 15-LOX gene.	104
Figure 15.	*In vitro* DNaseI footprinting of the non-coding strand of the rabbit RBC 15-LOX promoter using murine and human nuclear proteins.	108

Figure 16. Schematic diagram of murine nuclear protein binding to the non-coding strand of the rabbit RBC 15-LOX promoter. 108

Figure 17. Plasmids used in the construction of RBC 15-LOX/CAT recombinants or as transfection controls. 112

Figure 18. RBC 15-LOX/CAT recombinants used in the investigation of RBC 15-LOX promoter function. 112

Figure 19. RBC 15-LOX/CAT recombinants used in the investigation of the response of the RBC 15-LOX promoter to a heterologous enhancer. 112

Figure 20. CAT expression driven by the rabbit RBC 15-LOX promoter in murine cell lines, and response to a heterologous enhancer. 117

Figure 21. Basic RBC 15-LOX promoter function and response to a heterologous enhancer. 117

Figure 22. CAT expression in murine cell lines from various progressive deletions of RBC 15-LOX 5′ flanking sequences. 121

Figure 23. Effect of sequences upstream of -150nt on CAT expression from the basic RBC 15-LOX promoter. 121

Figure 24. Comparison of available LOX protein sequences from various species. 147

Figure 25. Conserved regions of LOX proteins from various species. 147

List of Tables.

Table I. CAT activities of various RBC 15-LOX/CAT recombinants, corrected for β-galactosidase activity. 117

Table II. Occurrence of the sequence motif GGGAG in the 5′ flanking sequences of various genes. 135

List of Text Headings.

Chapter 1. INTRODUCTION. 1

1.1 MECHANISMS OF GENE REGULATION IN EUKARYOTES. 2

1.1.1. A variety of mechanisms for regulation of gene expression. 2

1.1.2. Regulation of transcription initiation. 4

1.1.3. Alterations in chromatin structure and transcriptional activation. 6

1.1.4. DNaseI hypersensitive sites associated with regions important for transcriptional activation. 7

1.1.5. Mechanisms of open chromatin structure formation. 9

1.1.6. Maintenance of an open chromatin structure: DNA methylation. 10

1.2. Cis-ACTING SEQUENCES INVOLVED IN THE REGULATION OF TRANSCRIPTION INITIATION. 11

1.2.1. Modular organisation of eukaryotic promoters. 11

1.2.2. TATA-like boxes. 13

1.2.3. Upstream promoter elements. 14

1.2.4. Enhancers. 16
1.2.4.1. 3′ enhancers. 17
1.2.4.2. Mechanism of action of enhacers. 17
1.2.4.3. Modular structure of enhancers. 18

1.3 TRANSCRIPTION FACTORS. 19

1.3.1. Positively-acting transcription factors. 20

1.3.2. Negative regulation and negatively-acting transcription factors. 22

1.3.3. Structure of transcription factors. 25
1.3.3.1. Activation domains. 25
1.3.3.2. DNA-binding domains. 27

1.3.4. An evolutionary relationship of components of the transcription pre-initiation complex. 29

1.4 GENE REGULATION IN ERYTHROID CELLS. 31

1.4.1. Erythropoiesis as a model system for the study of cell type-specific co-expression during differentiation. 32
1.4.1.1. Ontogeny of erythroid cells. 32
1.4.1.2. Murine erythroleukaemia (MEL) cells. 35

1.4.2. Regulation of globin gene expression 36
1.4.2.1. Promoter sequences of globin genes. 36
1.4.2.2. Upstream sequences involved in the modulation of globin gene expression. 37
1.4.2.3. Sequences downstream of the transcription initiation site involved in regulation of globin gene expression. 38

1.4.3. Co-expression of genes in erythroid cells. 39

1.5. THE LIPOXYGENASE FAMILY OF ENZYMES. 42

1.5.1. Diversity in the positional specificity of various lipoxygenases. 43

1.5.2. Diversity in the function of various lipoxygenases. 44

1.5.3. Common features of the lipoxygenase enzyme family. 45

1.5.4. Structure-function relationships of the lipoxygenases. 46

1.6. THE ERYTHROID-SPECIFIC RBC 15-LOX OF RABBIT RETICULOCYTES. 46

1.6.1. Biological function of the rabbit RBC 15-LOX. 47

1.6.2. Characterisation of the rabbit RBC 15-LOX protein and its mRNA. 48

1.6.3. Erythroid-specific expression of the RBC 15-LOX. 50

1.6.4. Cloning of the RBC 15-LOX mRNA. 52

1.6.5. Cloning of the rabbit RBC 15-LOX gene. 53

1.6.6. Aims of this project. 53

Chapter 2. MATERIALS AND METHODS. 55

2.1. MATERIALS. 56

2.2. GENERAL PROCEDURES INVOLVING NUCLEIC ACIDS. 57

2.2.1. Handling of nucleic acids. 57

2.2.2. Organic extraction of nucleic acid solutions. 58

2.2.3. Concentration of nucleic acids. 58

2.2.4. Spectrophotometric determination of nucleic acid concentrations. 58

2.3. TRANSFORMATION OF BACTERIAL CELLS WITH DNA. 59

2.4. PREPARATION OF NUCLEIC ACIDS. 60

2.4.1. Eukaryotic RNA. 60
2.4.1.1. From rabbit reticulocytes. 60
2.4.1.2. From other rabbit tissues. 60
2.4.1.3. From cultured cell lines. 61
2.4.1.4. Selection of poly(A)$^+$ RNA. 62

2.4.2. Prokaryotic plasmid DNA. 62
2.4.2.1. Small scale preparation. 62
2.4.2.2. Bulk preparation. 63

2.4.3. Single-stranded M13 viral DNA. 64

2.5 GEL ELECTROPHORESIS OF NUCLEIC ACIDS. 65

2.5.1. Agarose gel electrophoresis of DNA. 65

2.5.2. Agarose gel electrophoresis of RNA. 65

2.5.3. Native polyacrylamide gel electrophoresis. 66

2.5.4. Denaturing polyacrylamide gel electrophoresis. 66

2.6. PURIFICATION OF DNA RESTRICTION FRAGMENTS. 67

2.6.1. From low melting-point agarose. 67
2.6.1.1. Using silica beads. 67
2.6.1.2. Using Elutip-d mini-columns. 67

2.6.2. From polyacrylamide. 68

2.7. MANIPULATION OF DNA. 68

2.7.1. Digestion of DNA with restriction endonucleases. 68

2.7.2. Conversion of 5′ single-stranded regions of duplex DNA to double-stranded DNA. 69

2.7.3. Ligation of DNA fragments into vectors. 69

2.7.4. Removal of 5′ terminal phosphate groups. 69

2.7.5. Labelling of DNA with radio-isotopes. 70
2.7.5.1. Random-priming using Klenow polymerase. 70
2.7.5.2. Uniform labelling of S1 probes. 70
2.7.5.3. Kinase 5′ end-labelling of DNA. 71

2.8. S1 NUCLEASE PROTECTION ANALYSIS. 72

2.9. PRIMER EXTENSION ANALYSIS. 72

2.10. SEQUENCING OF NUCLEIC ACIDS. 73

2.10.1. mRNA sequencing. 73

2.10.2. DNA sequencing. 73
2.10.2.1. Sub-cloning of fragments for sequencing. 73
2.10.2.2. Dideoxy chain-termination sequencing. 74
2.10.2.3. Computer analysis of sequences. 75

2.11. DNA:DNA and DNA:RNA HYBRIDISATION TECHNIQUES. 75

2.11.1. Capillary transfer from agarose gels by Southern blotting. 75

2.11.2. Electro-blotting. 75

2.11.3. Northern blotting. 76

2.11.4. Hybridisation of nucleic acids immobilised on filters. 76

2.12. In vitro DNaseI DIGESTION. 77

2.13. EUKARYOTIC CELL CULTURE. 78

2.13.1. Culture of cells. 78

2.13.2. Storage of cells in liquid nitrogen. 78

2.14. ASSAY OF TRANSIENT EXPRESSION OF EXOGENOUS DNA IN CULTURED CELLS. 79

2.14.1. DNA-mediated gene transfer. 79

2.14.2. Harvesting transfected cells and preparation of soluble protein extracts. 79

2.14.3. Determination of protein concentration of extracts. 80

2.14.4. β-galactosidase assay. 80

2.14.5. Measurement of chloramphenicol acetyltransferase (CAT) activity. 80

2.14.6. Correction of CAT activities for β-galactosidase activity. 81

2.15. PREPARATION OF NUCLEAR PROTEIN EXTRACTS FROM EUKARYOTIC CELLS. 82

Chapter 3. RESULTS. 83

3.1. SEQUENCE SURROUNDING THE ATG TRANSLATION INITIATION CODON OF THE RABBIT RBC 15-LOX GENE. 84

3.1.1. The genomic clone, λ8, is a genuine RBC 15-LOX recombinant, accurately predicting the N-terminal amino acid sequence of the RBC 15-LOX protein. 85

3.1.2. The 5′ end of the pGF6 cDNA insert is not derived from a RBC 15-LOX mRNA. 86

3.2. MAPPING OF THE TRANSCRIPTION INITIATION SITE OF THE RABBIT RBC 15-LOX GENE. 88

3.2.1. Strategy for mapping the 5′ end of the RBC 15-LOX mRNA. 88

3.2.2. S1 nuclease analysis using a single-stranded 276nt SstI probe: the 5′ end of the first coding exon lies 28nt upstream of the ATG translation initiation codon. 90

3.2.3. S1 nuclease analysis using a single-stranded 540nt EcoRI/BalI probe confirms the localisation of the 5′ end of the first coding exon, and reveals that micro-heterogeneity of protected fragments is not a function of digestion temperature or enzyme concentration. 91

3.2.4. Primer extension using a single-stranded 58nt genomic DNA primer: transcription initiation occurs 28nt upstream of the ATG translation initiation codon. 91

3.2.5. mRNA sequencing using the 58nt genomic primer: priming is from a unique mRNA sequence which matches that of the RBC 15-LOX gene. 93

3.2.6. Primer extension using a single-stranded 166nt cDNA-derived BalI primer: confirmation that the 5′ end of the RBC 15-LOX mRNA lies 28nt upstream of the ATG translation initiation codon. 94

3.2.7. mRNA sequencing using the 166nt BalI cDNA-derived primer. 95

3.2.8. A genomic primer which extends beyond the predicted 5′ end of the RBC 15-LOX mRNA does not primer extend on rabbit reticulocyte RNAs. 96

3.2.9. None of the available upstream sequences show detectable levels of hybridisation with Northern blots of rabbit reticulocyte RNAs. 97

3.2.10. Possible multiple minor sites of upstream transcription initiation detected by primer extension and S1 protection analysis. 101

3.2.10.1. Primer extension using the 166nt cDNA primer identifies multiple minor transcripts. 102

3.2.10.2. S1 nuclease analysis using the 540nt EcoRI/BalI genomic S1 probe also produces multiple minor products. 103

3.3. ORGANISATION OF RABBIT RBC 15-LOX GENE 5′ FLANKING SEQUENCES AND BINDING OF NUCLEAR PROTEINS. 104

3.3.1. Sequence elements upstream of the RBC 15-LOX transcription initiation site. 104

3.3.2. DNaseI *in vitro* footprinting of 5′ flanking sequences of the RBC 15-LOX gene. 107

3.3.2.1. Footprinting using murine nuclear protein extracts. 108

3.3.2.2. Footprinting using HeLa nuclear protein extract. 108

3.3.3 Proximal RBC 15-LOX upstream sequences may be involved in constitutive promoter function. 109

3.4. FUNCTIONAL ANALYSIS OF RABBIT RBC 15-LOX 5′ FLANKING SEQUENCES IN MURINE CELLS. 110

3.4.1. Strategy for investigating the function of the RBC 15-LOX promoter. 110

3.4.2. Variability in assays of transient expression of CAT enzyme activity. 112

3.4.3. β-galactosidase enzyme assay as internal standard for transfection efficiency. 114

3.4.4. Strategy to define the minimal functional promoter. 116

3.4.5. Sequences between 40 and 150bp upstream of the rabbit RBC 15-LOX transcription initiation site are sufficient to elevate expression of a linked CAT gene in both erythroid and non-erythroid murine cells. 118

3.4.6. Sequences present in the RBC 15-LOX promoter can respond to a heterologous enhancer. 119

3.4.7. Sequences between 1000 and 2700bp upstream of the rabbit RBC 15-LOX transcription initiation site have a cell type-specific effect upon expression of a linked CAT gene in murine cells. 121

3.4.8. Analysis of CAT mRNAs from murine cells transfected with RBC 15-LOX/CAT recombinants. 123

3.4.8.1. Analysis of CAT mRNAs by primer extension. 124

3.4.8.2. Analysis of CAT mRNAs by S1 protection. 126

Chapter 4. DISCUSSION. 128

4.1. OVERVIEW OF THE PROJECT. 129

4.2. FUNCTIONAL ANALYSIS OF THE RBC 15-LOX PROMOTER. 129

4.2.1. Two regions of RBC 15-LOX upstream sequences important for expression of a reporter gene from the RBC 15-LOX promoter in murine cell lines. 130

4.2.2. Potential cis-regulatory sequences in the RBC 15-LOX promoter. 131

4.2.3. DNA-protein interactions in the RBC 15-LOX promoter. 132

4.2.4. GGGAG: a sequence motif present in the RBC 15-LOX promoter and in upstream regions of a variety of other genes. 134

4.2.5. A distal element which confers cell type-specificity upon expression driven by the proximal element. 136

4.2.6. Is the cell type-specific effect seen with the -2700 to -1000 region a positive or a negative effect? 137

4.2.7. Future analysis of the cell type-specific regulatory element. 139

4.2.8. Other cis-regulatory elements in the RBC 15-LOX gene? 140

4.2.9. Regulation of the rabbit RBC 15-LOX gene in the intact animal. 141

4.3. POSSIBLE ROLES FOR NON-CODING mRNA SEQUENCES IN CONTROL OF RBC 15-LOX EXPRESSION. 143

4.4. AN EVOLUTIONARY RELATIONSHIP BETWEEN THE LIPOXYGENASES. 146

4.5. FUTURE DIRECTIONS. 148

ABSTRACT.

The rabbit erythroid-specific lipoxygenase (RBC 15-LOX) is a member of a family of enzymes which catalyse the addition of molecular oxygen to certain polyunsaturated fatty acids. It functions in the inactivation and degradation of mitochondria during the final stages of erythroid maturation, following extrusion of the nucleus. The protein has been purified to homogeneity and extensively characterised (Rapoport et al, 1979), and both cDNA and genomic clones have recently been obtained (Thiele et al, 1987). RBC 15-LOX expression shows several interesting features, including an erythroid-specific accumulation of its mRNA (Thiele et al, 1987), and translational inactivation of the messenger (Thiele et al, 1982).

Data presented in this thesis includes the sequencing of the recently-cloned RBC 15-LOX gene in the region surrounding the ATG translation initiation codon, and mapping of the transcription initiation site to a position 28nt upstream of the ATG codon by a combination of primer extension and S1 nuclease protection analyses. No transcripts arising from the available 2.7kb of sequences upstream of this point have been detected by Northern blot, primer extension or S1 nuclease protection analyses. A unique mRNA sequence, matching the genomic sequence immediately downstream of the putative transcription initiation site, has been obtained by primer extension mRNA sequencing.

Functional analysis of 5′ flanking sequences reveals two regions which affect transient expression of a cis-linked bacterial chloramphenicol acetyltransferase (CAT) reporter gene in established murine erythroleukaemia (MEL) and murine fibroblast cell lines. A proximal region within 150nt of the transcription initiation site functions as a promoter in both erythroid and non-erythroid cell lines, elevating CAT enzyme activity 7- to 8-fold above background, and can respond to a heterologous Friend Murine Leukaemia Virus enhancer. A distal region between 1.0 and 2.7kb upstream of the transcription initiation site confers a 6- to 8-fold cell type-specific effect upon CAT expression from the RBC 15-LOX promoter in the same assay system. The

proximal region contains several sequence motifs which are also seen in the 5′ flanking sequences of a variety of other genes, including: a TATA-like sequence; two CACCC motifs; a GGGCGG element; and a sequence resembling the cognate binding site for the CTF/NF-1 family of transcription factors. In vitro DNaseI footprinting of this proximal region using nuclear protein extracts from erythroid and non-erythroid murine cells shows protection of both the proximal CACCC sequence and the CTF/NF-1-like element.

Future experimental strategies for investigating a possible role for the regulation of transcription in the erythroid-specific accumulation of RBC 15-LOX mRNA are discussed. Also considered are aspects of the data presented here which, in conjunction with the work of others, may have implications for the translational inactivation of RBC 15-LOX transcripts, and for the evolution and structure-function relationships of the lipoxygenase enzyme family.

Chapter 1.

INTRODUCTION

1.1 MECHANISMS OF GENE REGULATION IN EUKARYOTES.

During the development of eukaryotic cells, a wide variety of differentiated cell types is produced. The phenotype of each cell is the product of the genes which have been, or are being, expressed. Several lines of evidence, including nuclear transplantation experiments, cytogenetic analysis of chromosomal banding patterns and determination of gene copy number using molecular probes, indicate that, for most genes, somatic cell differentiation is not accompanied by any alteration in the genome (reviewed in Alberts *et al*, 1983). The characteristic phenotypes of differentiated cell types are thus largely the result of differential gene expression, qualitative and quantitative differences in gene products being seen in each cell type, at different stages during differentiation, and in response to intra- and extra-cellular stimuli. Much attention has been devoted in recent years to establishing the mechanisms involved in such cell type-specific, stage-specific and inducible regulation of gene expression.

1.1.1 A variety of mechanisms for regulation of gene expression.

The additional complexity in the cellular structure of eukaryotes offers additional possibilities for regulation of gene expression over those seen in prokaryotes. Every stage in the pathway of gene expression is a potential point for regulation (reviewed in Darnell, 1982). Thus, as will be discussed further below, examples are known where protein-coding genes are regulated at the level of transcription; processing, transport or stability of mRNA; translation of mature mRNA transcripts; and post-translational modification of proteins. Studies of a large number of genes have revealed that control at the level of transcription is an important mechnism for the regulation of gene expression in eukaryotes. As transcriptional control has been the main focus of this project, the evidence for, and mechanisms of, transcriptional regulation will be dealt with in greater detail in later sections. However, as outlined briefly below, an increasing number of examples are now known of control at almost

every other level of gene expression.

Quantitative regulation of gene expression at the level of transcription can be achieved either by increasing the number of templates available for transcription, as is seen in the 30-fold amplification of chorion protein genes during Drosophila oogenesis (Spradling and Mahowald, 1980) or by increasing the rate of initiation of transcription per DNA template, as occurs in response to a wide variety of intra- and extra-cellular signals, and as will be discussed in detail below (Section 1.1.2). Quantitative regulation need not just involve increases in the amount of gene product, however. Regulation of the stability of mRNAs can involve both stabilisation and destabilisation of mRNAs (reviewed in Raghow, 1987). Stabilisation of mRNA involves a decrease in its rate of turnover, as seen in the 30-fold increase in half-life seen with the chicken vitellogenin gene in response to oestrogen (Brock and Shapiro, 1983) and for the 17- to 25-fold increase in casein mRNA in response to prolactin (Guyette et al, 1979). Conversely, Shaw and Kamen (1986) have demonstrated that an A/U-rich sequence from the 3' end of the mouse and human granulocyte/macrophage colony-stimulating factor (GM-CSF) mRNA, and which is similar to A/U-rich sequences seen in a variety of lymphokine, cytokine and proto-oncogene mRNAs, destabilises the rabbit β-globin mRNA when incorporated into the 3' untranslated region of this mRNA. Regulation of translation can involve the presence of upstream AUG codons, as in the translational control of GCN4 mRNA (Fink, 1986), or the sequestering of mRNA as ribonucleoprotein particles, as occurs for a variety of maternal mRNAs in sea urchin and Xenopus eggs (reviewed in Hunt, 1985).

Differential gene expression in eukaryotes also involves a number of qualitative regulatory mechanisms, many producing different products from transcription of the same gene. For example, comparisons of cDNA and genomic sequences have revealed that the initiation of transcription from alternative promoters is used to produce alternative transcripts in different cell types for mouse α-amylase (Hagenbuchle et al, 1981), the haem biosynthetic enzyme, porphobilinogen deaminase (PBG-D) (Grandchamp et al, 1987; Chretien et

al, 1988) and for human α1-antitrypsin (Perlino et al, 1987) genes. Termination of transcription seems to be rarely used as a means of genearating alternative mRNAs in eukaryotes; rather, alternative forms of 3′ processing seem to be used (reviewed in Birnstiel et al, 1985). Thus, during B lymphocyte differentiation, alternative 3′ cleavage and polyadenylation site selection events are involved in the production of membrane-bound and secreted forms of the immunoglobulin μ heavy chain (Alt et al, 1980; Rogers et al, 1980; Early et al, 1980). Alternative splicing can also be used to generate different mature mRNAs, and, therefore, different protein products, from the same primary transcript. This mechanism for differential expression has been known for some time in both DNA tumour viruses and in retroviruses (reviewed in Darnell, 1982), and has been recently shown to have a role in cell type-specific expression of the calcitonin/calcitonin gene-related peptide (calcitonin/CGRP) gene in neurons and thyroid cells (Leff et al, 1987). Splicing is also used as a means of quantitative regulation of Drosophila P elements, and suppressor-of-white-apricot and transformer loci (Bingham et al, 1988). In these cases, transcription initiation is constitutive, and it is splicing which determines whether or not an RNA is translated. Another important qualitative control of gene expression is seen in the post-translational modification of enzymes of the glycogen metabolism cascade by phosphorylation (reviewed in Cohen, 1982).

Of course, control of the expression of a particular gene may occur at more than one of these potential sites. For example, the expression of the mouse α-amylase gene from alternative promoters in different cell types also involves differential splicing of the alternative 5′ sequences onto a common body of downstream sequences (Hagenbuchle et al, 1981), while cell type-specific expression of the calcitonin/CGRP gene involves alternative splice site selection which then determines alternative polyadenylation site utilisation (Leff et al, 1987).

1.1.2 Regulation of transcription initiation.

Since much of the information concerning the mechanism of

regulation of transcription initiation has been obtained for transcription by RNA Polymerase II (PolII), which transcribes protein-coding genes, "transcription" will be considered here to refer to transcription by RNA Polymerase II, unless otherwise stated.

As the first step in gene expression, regulation of the rate of transcription initiation provides important possibilities for quantitative control of levels of gene products in eukaryotes, even when subsequent steps may result in qualitatively different products from the same primary transcript. Evidence has been obtained in favour of a role for regulation of transcription in the expression of a variety of eukaryotic genes. So-called "nuclear run-on" transcription experiments, an *in vitro* technique in which previously-initiated RNA molecules are elongated in the presence of radio-labelled RNA precursors in isolated nuclei and the products hybridised with cloned DNA fragments, have been shown to accurately reflect transcription *in vivo*, as determined by pulse-labelling of RNAs in intact cells (reviewed in Darnell *et al*, 1982). Nuclear run-on transcription has therefore been widely used in establishing transcriptional control both in cell type-specific expression and in induced alterations in gene expression. For example, McKnight and Palmiter (1979) demonstrated a 20-fold increase in ovalbumin mRNA, and a 2- to 3-fold increase in conalbumin mRNA synthesis following oestrogen and progesterone treatment of chickens, while Derman *et al* (1981) showed cell type-specific synthesis of a variety of RNAs in isolated nuclei from mouse liver, but not from mouse brain cells. Similar experiments, in which the DNA fragment hybridised with the *in vitro* labelled RNA comes from the very 5′ end of the transcription unit indicate that regulation of transcription occurs at the level of initiation (Wilson *et al*, 1979).

Recent evidence suggests that the initiation of transcription is the final stage in a multi-step process of transcriptional activation. The salient points, discussed in greater depth in following sections, can be summarised as follows: an altered chromatin structure is believed to represent a pre-activation stage (Section 1.1.3) necessary for the following step, the binding of so-called "trans-acting"

sequence-specific DNA-binding proteins (Section 1.3) to multiple, short cis-regulatory DNA sequences (Section 1.2). The final step in transcriptional activation is then the association of various transcription factors to form a pre-initiation transcription complex which has a dual function in positioning RNA Polymerase on the DNA template ready for transcription initiation, and in maximising the efficiency of the initiation process.

1.1.3 Alterations in chromatin structure and transcriptional activation.

One of the principal differences between gene regulation in prokaryotes and in eukaryotes concerns the conformational state of the genetic material. The much larger genome size of eukaryotes is present in a more condensed state as chromatin within the nucleus. An association of actively transcribed genes with alterations in chromatin structure has been appreciated for some time. Pulse-labelling of newly-synthesised transcripts in giant salivary gland cells of Drosophila larvae using ^{3}H-uridine has revealed that active transcription in these cells is associated with decondensed "chromosome puff" regions of chromatin (Bonner and Pardue, 1977). This relationship between transcription and chromatin structure was first observed at the molecular level by Weintraub and Groudine (1976), who observed an association of the chromatin of actively transcribed genes with increased sensitivity to the enzyme DNaseI, presumed to reflect increased accessiblity of the enzyme to the DNA. In chick erythroblasts, the untranscribed ovalbumin gene was relatively insensitive to single-stranded cutting by DNaseI, while a domain of increased DNaseI sensitivity surrounded actively-transcribed globin genes. Within this domain, gene flanking sequences were intermediate in sensitivity between inactive genes and the high sensitivity of transcribed regions. However, within the flanking sequences were regions of even greater sensitivity than those of the transcibed region, the so-called DNaseI hypersensitive sites (HSS), within which cuts are introduced into both strands (Stalder et al, 1980; Wu et al, 1979).

1.1.4 DNaseI hypersensitive sites associated with regions important for transcriptional activation.

DNaseI HSS have now been found in association with a wide variety of transcriptionally active eukaryotic cellular genes, usually in either 5′ or 3′ flanking regions (see Section 1.2). Evidence that such HSS are associated with regions important for regulation of gene expression has been provided from studies on the Sgs4 gene of Drosophila, which encodes one of the pupal glue proteins. Shermoen and Beckendorf (1982) discovered a hierarchy of DNaseI HSS upstream of Sgs4. Deletions of the HSS regions resulted in greatly reduced or completely abolished Sgs4 expression, depending upon which of the HSS had been deleted. Further evidence for the importance of DNaseI HSS comes from the observation that they are frequently associated with regions which have been shown by deletion or mutation analysis to be important for regulation of transcription. For example, four of the nine DNaseI HSS regions within 8kb of 5′ flanking sequences of the chicken lysozyme gene have been shown by transient expression experiments to contain regulatory sequences having either positive or negative effects upon gene expression (Theisen *et al*, 1986; Steiner *et al*, 1987). Similarly, cell type-specific DNaseI HSS are associated with regions which enhance transcription of genes, such as chicken β-globin and immunoglobulin light chains, which are expressed in a cell type-specific fashion (Emerson *et al*, 1987; Parslow and Granner, 1982). Furthermore, DNaseI HSS have been shown to include binding sites for proteins likely to be involved in transcription. For example, Plumb *et al* (1985 and 1986) have observed the *in vitro* binding of nuclear proteins to the DNaseI HSS regions upstream of the chicken β^A- and β^H-globin genes, while Kemper *et al* (1987) have demonstrated protein binding to sequences within a DNAseI HSS immediately 5′ of the chicken α^D-globin gene. The same α^D sequences are protected in both *in vivo* and *in vitro* "footprinting" assays for nuclear protein binding.

The altered chromatin structure indicated by increased DNaseI sensitivity seems to be necessary but not sufficient for transcription. Weintraub *et al* (1982) demonstrated that, following

the release of differentiation arrest upon switching to a higher temperature, the globin genes of chicken erythroblasts infected with temperature-sensitive Avian Erythroblastosis Virus showed DNaseI HSS before the initiation of globin mRNA expression, while Groudine and Weintraub (1982) showed that DNaseI HSS persisted for multiple cell divisions following the cessation of (inappropriate) transcription of globin genes in temperature-sensitive Rous Sarcoma Virus-infected chick embryo fibroblasts. This observation that altered chromatin conformation is not sufficient to ensure transcriptional activation can be ascribed to the requirement for subsequent steps, including the binding of transcription factors (see Section 1.3), for formation of an active transcription complex.

Sharp boundaries between domains of increased DNaseI sensitivity and surrounding insensitive regions observed by Weintraub et al (1981) prompted the suggestion that the boundaries are defined by specific sequences. Evidence in favour of this has been presented by McGinnis et al (1983) who demonstrated that a single point mutation in one of the DNaseI HSS regions of a naturally-occurring under-producer mutant resulted in a 50% reduction in expression of the Sgs4 glue protein gene of Drosophila. These domain boundary sequences may include so-called scaffold attachment regions (SARs) involved in attachment of DNA to the nuclear matrix and containing recognition sites for topoisomerase II cleavage. Gasser and Laemmli (1986) found SARs both 5′ and 3′ of Drosophila aldehyde dehydrogenase, Sgs4 and fushi tarazu genes, and that the upstream SARs co-map with sequences required for enhanced expression of these genes. Furthermore, five of these SARs contained sequences closely related to the consensus sequence for topoisomerase II cleavage. Recently, the domain boundaries of the human β-globin locus, some 70kb apart, have been analysed by Grosveld et al (1987), who have demonstrated that these regions can confer cell-type specific, high-level, position-independent expression of the human β-globin gene in transgenic mice. Whether the sequences of such domain boundary sequences show any similarity with those of the SARs is not yet known, and it remains to be seen whether the β-globin boundary sequences are conserved and can have similar effects when flanking other genes, or whether other genes contain distinct

sequences at their domain boundaries.

1.1.5 Mechanism of open chromatin structure formation.

The mechanism of alteration of chromatin structure as a pre-activation event in the regulation of transcription is uncertain at present. Weintraub and Groudine (1976) suggested that, as each newly-synthesised DNA strand is packaged into nucleosomes, altered chromatin structure might arise at replication. The behaviour of looped chromatin domains, anchored to the nuclear matrix and each containing one or more transcription units, as independent units of supercoiling and replication (reviewed in Goldman, 1988) may provide a basis for such indirect effects of replication upon transcription. A model which views these independently-replicating chromatin domains as being units of gene regulation suggests that an open, pre-activated chromatin structure is determined by the time of replication of a chromatin domain (Goldman, 1988; Gottesfeld and Bloomer, 1982). This model is supported by studies of the replication times of various transcribed genes (reviewed in Goldman, 1988) which show that transcriptionally active genes are generally replicated early in S phase, while transcriptionally inactive genes are replicated later (e.g. Xenopus somatic and oocyte 5S rRNA genes - Guinta and Korn, 1986). Goldman (1988) has suggested that determination of late-replicated genes as being inactive is not the result of late replication per se, but may be due to limiting amounts late in S phase of chromatin assembly factors required for replication in a transcriptionally active open chromatin conformation. However, if this is the case, it is still unclear what determines replication of a particular chromatin domain as being early or late in S phase in a particular cell type.

As explained above, the formation of an open chromatin conformation, whatever the mechanism by which it is achieved, does not seem to be sufficient to ensure transcription. This depends upon the next step in transcription activation - the formation of a pre-initiation transcription complex by the binding of transcription factors to multiple cis-acting regulatory sequences.

1.1.6 Maintenance of an open chromatin structure: DNA methylation.

DNaseI HSS can persist through multiple cell divisions (see Section 1.1.4 above). One mechanism proposed to explain this stable maintenance of open chromatin structure involves DNA methylation, a modification of DNA known to be reproduced during DNA replication (Razin and Riggs, 1980). In general, DNA methylation is associated with inactive genes, while expressed genes are frequently, but not always, hypomethylated (reviewed in Cedar, 1988; Bird, 1984). Thus, unmethylated copies of the human γ-globin gene introduced into mouse fibroblast cells, in which the endogenous γ-globin genes are methylated and transcriptionally inactive, can be transcribed, whereas methylated exogenous γ-globins are transcriptionally inactive. Hypomethylated CpG-rich "islands" are associated with constitutively-expressed "housekeeping genes" (Bird, 1986). Conversely, genes expressed in a cell type-specific fashion are heavily methylated in non-expressing tissues, but are usually unmethylated in expressing tissues (reviewed in Cedar, 1988). Expressing tissues seem to be capable of demethylating sequences required for transcription, since *in vitro* methylated muscle-specific α-actin is demethylated and expressed when introduced into myoblasts, but not in fibroblasts (Yisraeli *et al*, 1986). It is unclear at present, however, whether the demethylation which accompanies transcriptional activation in such cases is a cause or an effect of transcriptional activation (reviewed in Bird, 1986). Whether demethylation is involved in causing activation of transcription, or whether it is a "passive" effect of the binding of factors required for transcription which also conincidentally hinder methylation (Bird, 1986), the emerging picture seems to be that genes are "marked" as being transcriptionally inactive by DNA methylation, until, for example, transcription is activated in a cell type-specific or developmental stage-specific fashion, whereafter the hypomethylated DNA methylation pattern is maintained through subsequent cycles of replication.

1.2 Cis-ACTING SEQUENCES INVOLVED IN THE REGULATION OF TRANSCRIPTION INITIATION.

Regions of open chromatin structure likely to be important for the regulation of gene expression have been found in both 5′ and 3′ flanking sequences and in internal regions of genes. For example, in the β-globin gene family, the human β-globin gene locus shows DNaseI HSS in both 5′ and 3′ regions, the upstream HSS at least being erythroid-specific (Grosveld et al, 1987). Similarly, the developmentally regulated expression of the chicken adult β^A-globin gene correlates with a developmental stage-specific DNaseI HSS (Emerson et al, 1985), and erythroid-specific DNaseI HSS are located in both 5′ flanking sequences and within the second intron of mouse β-globin (Galson and Housman, 1988). Indeed, as will be discussed further below, sequences important for the regulation of transcription, referred to as cis-acting regulatory sequences, have been found either in 5′ flanking, 3′ flanking or in intervening sequences. In several cases, most notably in immunoglobulin and globin genes, transcriptional control involves cis-acting sequences both 5′ and 3′ of the transcription initiation site (Sections 1.2.4.1 and 1.4.2.3).

The formation of a transcription pre-initiation complex, following the extablishment of an open chromatin structure, can be summarised as involving the interaction of various transcription factors with multiple cis acting-sequences (reviewed in Dynan and Tjian, 1985; Maniatis et al, 1987). Characteristic combinations of cis-acting sequences thus provide a "blue-print" for the assembly of the particular set of transcription factors required for initiation of transcription of a given gene, as will be discussed in greater depth in the sections which follow.

1.2.1 Modular organisation of eukaryotic promoters.

For several reasons, much of the early work in the investigation of regulation of transcription initiation in eukaryotes was directed towards the analysis of sequences immediately upstream of

the transcription site, the so-called 5′ flanking sequences. Firstly, intuition suggested that sequences involved in the control of transcription initiation are likely to affect the association of RNA Polymerase with DNA near the initiation site itself. Secondly, transcriptional control in prokaryotes provided a precedent for transcription regulatory signals immediately upstream of the transcription unit.

By analogy with prokaryotic transcription, these 5′ flanking sequences immediately upstream of the eukaryotic transcription initiation site, and containing sequences involved in the control of transcription initiation, are referred to as "promoter" regions. Initially, sequence elements within the promoter region likely to be of importance for the regulation of transcription were revealed by sequence comparisons between genes (Goldberg, 1979; Breathnach and Chambon, 1981; Efstratiadis et al, 1980). The importance of these sequences has since been confirmed, and additional cis-regulatory sequences revealed, by a variety of methods involving the generation in vitro of point mutations, deletions and re-arrangements of 5′ flanking sequences followed by analysis of expression levels after the re-introduction of the modified genes into living cells (reviewed in Spandidos and Wilkie, 1984a), or by analysis using in vitro transcription systems (Manley et al, 1980).

As will be detailed in following sections, the emerging picture is that the promoter regions of eukaryotic protein-coding genes are modular (Serfling et al, 1985). There are no extended regions of sequence homology between promoters of unrelated genes. Rather, different promoters contain various combinations of multiple cis-acting sequences, often no more than a few nucleotides in length, and sometimes separated by considerable distances. Evidence obtained by introducing additional sequences corresponding in length to either full or half-turns of a DNA double helix into the repressor-binding operator sequences of bacteriophage lambda suggest that the spacing of such elements is important for efficient transcription initiation (Ptashne, 1986).

In principle, a large number of regulatory possibilities can be achieved by the combination of a small number of such regulatory sequence elements, as recognised by Britten and Davidson (1969). Indeed, many sequence elements have been found to be associated with regulation of a wide variety of genes (Section 1.2.3). None of these sequence elements, however, have been found to be ubiquitous, and there would appear to be no such thing as a "typical" eukaryotic promoter. Nevertheless, elements of the organisation and function of regulatory sequences in various genes show many similarities, not just between eukaryotic protein-coding genes, but over an evolutionary range from prokaryotes, via lower eukaryotes such as yeasts, to primates including man (Guarente, 1988; Struhl, 1987a - see Section 1,3,4).

1.2.2 TATA-like boxes.

In bacteria, the promoter is the site for binding of RNA Polymerase, whereas, in eukaryotes it appears that RNA Polymerase does not itself bind DNA. Rather it is positioned relative to specific promoter sequences. In many genes, an A/T-rich region lying 25-30bp upstream of the transcription initiation site, and typified by the so-called "TATA" or Goldberg-Hogness box first identified in Drosophila histone genes (Goldberg, 1979; Breathnach and Chambon, 1981), appears to serve the positioning or "selector" function for RNA Polymerase II. Experiments in which the TATA boxes of SV40 early and sea urchin histone H2A genes were deleted and re-introduced into living cells, in contrast to results obtained using *in vitro* transcription systems, did not abolish transcription, but resulted in lower levels of "ragged" initiation of transcription at multiple sites (Grosschedl and Birnstiel, 1980; Benoist and Chambon, 1981; Ghosh *et al*, 1981). Point mutations of the 5′ flanking regions of β-globin genes have since confirmed that mutation of a similar TATA-like region results in a 2- to 5-fold decrease in levels of transcription, as well as affecting the site of initiation (Dierks *et al*, 1983; Myers *et al*, 1986). In this respect, the function of the TATA box differs from that of the bacterial Pribnow box, of similar sequence, in that deletion of the Pribnow box completely abolishes transcription

(Grosschedl and Birnstiel, 1980; Miller, 1978).

Not all eukaryotic genes contain the "canonical" TATA sequence; rather, for many genes, an A/T-rich region seems to serve the positioning function. For example, many mammalian globin promoter regions contain a conserved A/T-rich sequence (Efstratiadis, 1980) at a similar distance from the transcription initiaition site to that of the TATA box in other genes, suggesting a similar function for this region in determining the position of transcription initiation. Indeed, *in vitro* point mutagenesis of this region of the rabbit and mouse β-globin genes have revealed that this region is of importance for accurate and efficient transcription (Dierks *et al*, 1983; Myers *et al*, 1986). The promoter regions of some other genes completely lack any similarity to a TATA-like sequence. These genes have G/C-rich promoters, and often encode gene products expressed in all cell types at all times - so-called "housekeeping" genes (Dynan, 1986). In these genes, other elements, or combinations of elements, are presumed to govern the positioning of RNA Polymerase.

1.2.3 Upstream promoter elements.

The quantitative efficiency of transcription initiation is now known to be regulated by a wide variety of cis-acting sequence elements (Wingender, 1988). Such sequences are analogous to the operator sequences of prokaryotic promoters, being involved in up- or down-regulation of the rate of transcription initiation, and have been referred to as "modulator" sequences (Grosschedl and Birnstiel, 1980). These sequences are frequently found close to the transcription initiation site, but further upstream than the A/T-rich region, and are consequently frequently referred to as upstream promoter elements.

Some upstream promoter elements are found in a wide variety of genes. For example, the CCAAT sequence, first noticed upstream of ovalbumin genes (Benoist *et al*, 1980), is also found in the promoters of many eukaryotic genes, including mammalian globin genes, sea urchin histone genes and the Herpes Simplex viral thymidine kinase gene (Efstratiadis *et al*, 1980; Barberis *et al*, 1987; McKnight and Tjian,

1986). Similarly the so-called G/C-box, GGGCGG, which was originally identified in the 21bp repeats of the SV40 viral Early promoter, is present in a variety of eukaryotic genes (reviewed in Dynan and Tjian, 1985; Kadonaga et al, 1986). These sequences are consequently often referred to as "general" or "constitutive" cis-acting elements and have important roles in the formation of the transcription pre-initiation complex, but not necessarily in the modulation of transcription in different cell types, at different stages of differentiation or development, or in response to cellular or external stimuli. Such functions seem to be due to other "specific" or "regulatory" sequences restricted to the promoter regions of single genes or families of genes whose gene expression shows characteristic regulatory features (Serfling et al, 1985). For example, the octamer motif of immunoglobulin genes, when linked to a minimal mouse renin-1 promoter retaining only sequences downstream of the TATA box, is sufficient to confer lymphoid cell-specific expression when introduced into lymphoid and fibroblast cell lines (Dreyfus et al, 1987).

Both general and specific cis-acting elements are frequently found in multiple copies within promoter regions. For example, the SV40 early promoter contains 6 copies of the G/C-box GGGCGG sequence within its three 21bp repeats, while the H2B-1 sperm histone gene of sea urchins contains two CCAAT motifs (Barberis et al, 1987), and most of the heat shock gene promoters studied in a variety of species contain multiple copies of the so-called heat shock element (HSE) unique to heat shock genes (Pelham, 1985).

In addition to genes possessing multiple copies of the same regulatory sequence, the organisation of many promoters is "modular", containing combinations of cis-acting sequences: some "specific" to a single gene or small sub-set of genes, others "general" elements of widespread occurrence (reviewed in Serfling et al, 1985). 5' flanking sequences of the human metallothionein II_A gene, for example, contain several different types of cis-acting elements, including TATA and G/C boxes, a glucocorticoid response element (GRE) also found in steroid hormone-responsive genes, an AP-1 protein-binding site also seen in a variety of genes, including SV40 and human collagenase regulatory

sequences, and several copies of a so-called metal-response element (MRE) (Lee et al, 1987a and b).

1.2.4 Enhancers.

A third class of cis-acting sequences, known as enhancers, serve an important role in quantitative regulation of transcription initiation. Enhancers are capable of greatly elevating transcription from a linked gene, including genes transcribed from a heterologous promoter, in an orientation- and position-independent fashion, from either 5′ or 3′ of the gene body and often over large distances (Banerji et al, 1981; Serfling et al, 1985). Originally identified in viruses, many sequences from cellular genes which show similar properties have now been identified (see Harrison, 1988 for a recent review) and sequence homologies between certain cellular and viral enhancers have been observed. For example, Goodbourn et al (1986) observed homology between portions of the human β-interferon enhancer, revealed by deletion studies to be capable of acting as a strong constitutive transcription element, and sequences within Adenovirus E1a, Polyoma, Cauliflower Mosaic Virus and SV40 viral enhancers. The homology with the SV40 enhancer "core" sequence is also seen in the immunoglobulin heavy chain enhancer.

Cellular genes containing enhancers include those encoding enzymes such as rat elastase and chicken lysozyme (Hammer et al, 1987; Theisen et al, 1986; Steiner et al, 1987); hormones and growth factors such as insulin and interleukin-2 (Walker et al, 1983; Fujita et al, 1986); structural molecules such as mouse α2 collagen (Rossi and de Crombrugghe, 1987); nuclear histone proteins (Trainor et al, 1986); α- and β-globin genes (Choi and Engel, 1986; Antoniou et al, 1988); and components of the immune system such as complement Factor B, immunoglobulins and the T cell receptor β-chain (Wu et al, 1987; Banerji et al, 1983; Queen and Baltimore, 1983; Gillies et al, 1983; Krimpenfort et al, 1988).

Like upstream promoter elements, enhancers have been classified according to their function, as "general" and "specific"

(or "regulated"), depending upon the range of circumstances in which they exert their effects upon transcription. Thus, some enhancers, such as the archetypal SV40 enhancer, function equally well in a variety of cell types; others, such as that of polyoma virus or the cellular immunoglobulin enhancer, produce cell type-specific effects on transcription (see Serfling et al, 1985; Maniatis et al, 1987 for recent reviews).

1.2.4.1 3′ enhancers.

Enhancers are capable of functioning in a largely position-independent fashion. For example, the SV40 72bp repeats enhance transcription of a linked rabbit β-globin gene in HeLa cells when artificially placed 3.3kb downstream of the globin transcription initiation site as well as from 1.4kb upstream (Banerji et al, 1981). Enhancers positioned downstream of the transcription initiation site have now been identified in naturally-ocurring situations: in immunoglobulin heavy chain genes (Queen and Baltimore, 1983; Gillies et al, 1983; Banerji et al, 1983) and mouse α2 (type I) collagen (Rossi and deCrombrugghe, 1987) genes, for example, where they are located in an intron; and in globin (Choi and Engel, 1986; Hesse et al, 1986; Emerson et al, 1987; Kollias et al, 1987; Bodine and Ley, 1987), chicken histone H5 (Trainor et al, 1987) and human T cell receptor β-chain (Krimpenfort et al, 1988) genes, where the enhancers lie beyond the 3′ end of the transcription unit.

1.2.4.2 Mechanism of action of enhancers.

Enhancers function by increasing the frequencey of correct initiation of transcripts, increasing the density of RNA Polymerase molecules as demonstrated by nuclear run-on transcription experiments in which transient expression of globin genes in HeLa cell nuclei resulted in increased levels of newly-initiated globin transcripts in the presence of the SV40 viral enhancer (Weber and Schaffner, 1985; Treisman and Maniatis, 1985). The mechanism by which enhancers function is still unclear, however. Various models of enhancer function have been suggested, proposing a role either in the

pre-activation stage or in the formation of the transcription complex itself. Enhancers have been seen as entry sites for the RNA Polymerase molecule or other components of the transcription apparatus on the basis that the enhancer's effect is often greatest on the most proximal promoter (Wasylyk et al, 1983). This model envisages that RNA Polymerase would then be positioned at the transcription initiation site either by sliding along the DNA template until encountering a positioning factor such as a TATA-binding protein, or by looping out of DNA (Ptashne, 1986).

Other models suggest a role for enhancers in establishing an active, potentially-transcribable chromatin structure. Enhancers might be thought of as acting as "organiser" sites for alteration of chromatin topology or assignment of a given gene to a particular "nuclear address". These ideas are supported by the observation that enhancers are frequently associated with cell type-specific DNaseI HSS, as in the cases of the chicken adult β-globin, chicken lysozyme and mouse immunoglobulin kappa chain enhancers (Emerson et al, 1987; Theisen et al, 1986; Parslow and Granner, 1982). The observation that immunoglobulin genes can retain transcriptional activity even following the loss of their enhancer during the gene rearrangements which accompany B lymphocyte differentiation suggest that enhancers are required for establishment, but not maintenance, of an active transcription complex (Serfling et al, 1985).

1.2.4.3 Modular structure of enhancers.

Enhancers, like promoters, have modular structures. For example, in vivo studies of deletions of the SV40 enhancer have shown that it has a domain structure, containing multiple sequence elements (Zenke et al, 1986; Herr and Clarke, 1986). Furthermore, in vitro binding of nuclear proteins to enhancers shows protection of several sequence motifs against DNaseI digestion. For example, multiple cell type-specific or developmental stage-specific factors protect at least five discrete regions in the chicken adult β-globin 3' enhancer (Emerson et al, 1987). It has been suggested that the elements of an enhancer are themselves bi-partite, being composed of so-called

enhansons (Ondek et al, 1988).

The modular structure of both promoters and enhancers serves to emphasise that the classification of cis-regulaory sequences using terminology such as upstream promoter elements and enhancers is somewhat artificial, and serves mainly to provide a conceptual framework for model-building. There is an increasing blurring of functional distinctions between upstream promoter elements and enhancers as envisaged by this type of classification (Serfling et al, 1985). For example, a single copy of the Drosophila hsp70 heat shock element, regarded as a "typical" upstream promoter element, has little effect upon transcription of a linked human β-globin gene, but two copies behave as a strong heat-inducible enhancer (Bienz and Pelham, 1986). Further evidence of a functional homology between both promoter and enhancer elements comes from immunoglobulin genes, where the same octamer sequence is found in both promoter and enhancer regions of the immunoglobulin heavy chain gene (Falkner and Zachau, 1984; Parslow et al, 1984). This may well be the basis for the synergism of function observed between immunoglobulin promoters and enhancers (Garcia et al, 1986) where the immunoglobulin enhancer has a greater effect upon the homologous immunoglobulin promoter than on heterologous cis-linked promoters. The functional homology of modules of promoters and enhancers is the basis for the suggestion made by Ondek et al (1988) that modules of regulatory sequences ("enhansons") can be used in novel combinations to produce new enhancers.

The essential point to be made here is that modular sequence motifs, whether components of promoters or enhancers, act as binding sites for sequence-specific DNA-binding transcription factors and that characteristic combinations of cis-acting sequences and transcription factors are employed in the formation of an activated transcription complex for a particular gene.

1.3 TRANSCRIPTION FACTORS.

All the transcription factors identified to date are proteins (including the products of several nuclear proto-oncogenes such as

c-jun and c-fos, which are now being discovered to have probable functions in the initiation of transcription - Bohmann et al, 1987; Chiu et al, 1988; Sassone-Corsi et al, 1988); but there is no reason, in principle, to expect that all will prove to be so.

Most of the transcription factor proteins identified thus far have DNA-binding activity, as has been demonstrated by "footprinting" and electrophoretic mobility shift assays using nuclear protein extracts containing mixtures of transcription factors. Exceptions, however, include the GAL80 protein of yeast, which inhibits the transcriptional activation function of another transcription factor, GAL4, by interacting with its C-terminal region (Ma and Ptashne, 1987b) and, probably, the product of the c-fos proto-oncogene, which interacts with the c-jun product, the transcription factor AP-1 in stimulating transcription (Chiu et al, 1988; Sassone-Corsi et al, 1988). Such transcription factors illustrate the principle that protein-protein, as well as DNA-protein interactions, play a vital role in the formation of a pre-initiation complex (Sigler, 1988). The role of some transcription factors may thus be to mediate indirect interactions between other components of the transcription complex (between two DNA-binding factors, or between a DNA-binding protein and RNA Polymerase, for example).

In contrast to these heterologous protein-protein interactions, other transcription factor combinations consist of homologous sub-units. Various eukaryotic DNA-binding transcription factors, like those in prokaryotic systems, bind to their cognate sequences as dimers - the yeast transcription factor GCN4 (Hope and Struhl, 1987), and a cAMP-response element-binding (CREB) protein which binds to the CAMP response element of the rat somatostatin gene, for example (Yamomoto et al, 1988).

1.3.1 Positively-acting transcription factors.

Most of the transcription factors characterised to date function in the activation of transcription, as has been demonstrated by analysis of expression from reporter genes driven by promoters

which include transcription factor recognition sequences, either by addition of purified proteins to *in vitro* transcription systems, or by expression of cDNA clones in cells *in vivo*.

Many DNA-binding factors show precise sequence recognition. For example, the transcription factor AP-1 protects the same sequence (TGACTCA) in *in vitro* DNaseI footprinting experiments on SV40, metallothionein and collagenase promoters (Lee *et al*, 1987b). Yet the adenovirus transcription factor, ATF, recognises a sequence which differs by only one nucleotide from that recognised by AP-1 (Hurst and Jones, 1987). Such subtle differences in sequence specificity are reminiscent of the sequence specificity of DNA-binding proteins in prokaryotes, the restriction endonucleases. Not all transcription factors show absolute specificity for the sequences to which they bind, however. For example, a single yeast trans-activator, HAP1, appears to bind to different sequences in the CYC1 and CYC7 promoter regions (Pfeifer *et al*, 1987b). Conversely, not all cis-acting sequences act as recognition sites for only a single transcription factor. The same yeast CYC1 gene promoter sequence appears to be bound by two different transcription factors, HAP1 and RC2 (Pfeifer *et al*, 1987a).

A similar situation is seen in higher eukaryotes, where there appear to be a family of transcription factors capable of binding to sequences which include the CCAAT sequence found in a variety of eukaryotic promoters (Santoro *et al*, 1988; Chodosh *et al*, 1988a, b; Barberis *et al*, 1987; Dorn *et al*, 1987). Within this family, cDNA cloning has revealed that several different forms of the CTF/NF-1 transcription factor are generated by alternative splicing of a single gene (Santoro *et al*, 1988). The interaction of various members of the family with their appropriate recognition sequences seem to be subtly different, methylation interference analyses indicating that different residues are important for sequence recognition by the transcription factors CP1, CP2 and CTF/NF-1 (Chodosh *et al*, 1988b). Furthermore, these three show widely different affinities for each other's extended recognition sequences, as determined by mobility shift assays in the presence of competitor oligonucleotides, suggesting that DNA sequences

on either side of the common CCAAT recognition sequence may also be involved in protein binding (Chodosh et al, 1988b).

Regulation of the activity of transcription factors is thought to be involved in cell type-specific, stage-specific and inducible regulation of gene expression. For example, recognition sequences in certain genes are bound in vitro by nuclear proteins only from cells of a particular type or at a particular stage of differentiation. Such differential DNA-binding activities might result from differences in the expression of the transcription factor itself, or to a modulation of its binding activity which allows it to bind to its recognition sequence in certain situations or cell types, but not in others. The appearance of DNA-binding activity of several transcription factors including the differentiation stage-specific induction of binding of transcription factor NF-kB to immunoglobulin gene regulatory sequences (Sen and Baltimore, 1986), has been found to be unaffected by inhibitors of new protein synthesis such as cycloheximide. This suggests that binding activity arises by modification of a pre-existing protein. The post-translational modification event might involve a covalent modification or a non-covalent interaction, either with a regulating ligand or with other essential components of the transcription complex. A situation in which transcription is altered by a covalent post-translational modification of an inactive, pre-existing factor is seen with the cAMP response element-binding (CREB) protein mentioned earlier, which binds its cognate sequence as a dimer. Phosphorylation of CREB in response to cAMP appears to be necessary for the activation of transcription in vivo (Montminy and Belizikjian, 1987). An explanation of this may be provided by the observation that dimerisation and DNA binding activity in vitro are reduced by dephosphorylation (Yamomoto et al, 1988).

1.3.2 Negative regulation and negatively-acting transcription factors.

An increasing number of cis-acting sequences involved in negative regulation of transcription have recently been identified in a wide variety of eukaryotic genes (reviewed in Harrison, 1988). One

of the most intensively studied examples is the negative regulation of the yeast mating type HMR and HML loci. These loci contain transcriptionally inactive copies of the α and a mating type alleles, respectively. Either allele, when moved to the MAT locus by a recombination event which occurs almost every cell division, becomes transcriptionally active, resulting in a switch from a to α mating phenotypes, or vice versa (reviewed in Miller et al, 1985; Brent, 1985). In studies of the mechanism by which the mating type alleles at HML and HMR are maintained in a transcriptionally "silent" state, Brand et al (1985) discovered a cis-acting "silencer" sequence element, HMRE, which shows similar properties to those of an enhancer, but acting to inhibit rather than to enhance transcription i.e. the silencer can function at a distance, in an orientation- and position-independent fashion. A "dehancer" region upstream of the P1 promoter of the c-myc gene appears to act in an analogous fashion, being capable of exerting a negative regulatory effect at a distance, from either 5′ or 3′ of the transcription initiation site, in either orientation, and can also exert its negative effect upon a heterologous promoter (Remmers et al, 1986).

Negative regulation of this sort may be involved in the regulation of gene expression during differentiation. Somatic cell fusion experiments involving fusion of differentiated with undifferentiated cell types can result in the "extinction" of genes expresssed in a cell type-specific fashion which are characteristic of the differentiated state. Although the mechanisms by which extinction occurs are still unclear, the use of segregating inter-species hybrids has allowed the identification of chromosomes bearing genes encoding the factors involved in extinction. For example, genes involved in extinction of liver-specific tyrosine aminotransferase and albumin genes have been localised to mouse chromosomes 11 and 3 respectively (Killary and Fournier, 1984; Petit et al, 1986).

Transcription factors involved in negative regulation have long been recognised in prokaryotic systems. The most extensively studied have been the lac repressor of Eschericia coli (Miller, 1978) and the lambda (cI) and cro repressors of bacteriophage lambda (Ptashne,

1986). These act by binding to DNA in a sequence-specific fashion, preventing access of RNA Polymerase, and thereby inhibiting the initiation of transcription. (Lambda repressor, however, shows the unusual property of being able to act as both a repressor and as an activator of transcription, having different effects upon transcription of genes on either side of its binding site - reviewed in Ptashne, 1986)).

In some cases of negative regulation in eukaryotes, DNA-binding proteins analogous to prokaryotic repressors appear to be involved. One example is that of the human β-interferon gene. Goodbourn et al (1986) demonstrated that the inducible enhancer located between 77 and 36nt upstream of the transcription initiation site consists of two regions. Deletion of the more 3′ region resulted in constitutive high-level expression of β-interferon mRNA in the absence of the viral or double-stranded RNA inducers which are normally required for activation of transcription. It thus seems that the 3′ half of the element of the enhancer negatively regulates transcription from the constitutive 5′ portion, except in the presence of inducers. In vivo footprinting of this region of the β-interferon gene (Zinn and Maniatis, 1986) show that a factor bound to the negative regulatory region of the enhancer dissociates upon induction, followed by binding of another factor to the constitutive region - a situation reminiscent of repressor action in prokaryotes. Similarly, binding of a so-called CCAAT displacement protein (CDP) to sequences overlapping the more proximal of two CCAAT motifs in the 5′ flanking region of the sea urchin sperm histone H2B-1 gene prevents binding of a CCAAT-binding protein in the embryo, where the H2B gene is not expressed (Barberis et al, 1987).

Intriguingly, recent studies of protein-binding to the HMRE element involved in silencing of the yeast mating type HMR locus have resulted in the cloning of the gene encoding a transcription factor, RAP1, which appears to be involved in the regulation of a variety of genes in addition to the HMR locus, functioning as either a repressor or as an activator of transcription depending upon the context of its binding site (Shore and Nasmyth, 1987).

1.3.3 Structure of transcription factors.

A substantial number of eukaryotic transcription factors have been purified (Rosenfeld and Kelly, 1986; Jones et al, 1987; Briggs et al, 1986; Kadonaga and Tjian, 1986; Lee et al, 1987b) and an increasing number are being cloned, including a factor, Sp1, which binds to the GGGCGG box (Kadonaga et al, 1987), a family of CCAAT-binding proteins (Santoro et al, 1988) and various steroid hormone receptors (reviewed in Maniatis et al, 1987). The availability of structural information concerning transcription factors, and especially those involved in transcriptional regulation in yeast, are beginning to provide insights into the interactions involved in the transcription pre-initiation complex. Furthermore, comparisons of the structures of these factors further suggest an evolutionary conservation of elements of the transcriptional apparatus (see Section 1.3.4 below).

Many of the transcription factor proteins characterised show similar structural characteristics, consisting of multiple domains which can function independently of one another. This structural feature probably has a functional basis. Unlike bacterial repressors, where the function of the protein is merely to exclude the binding of another protein, eukaryotic transcriptional activators must interact with other components of the transcription complex in order to activate transcription. As discussed earlier, protein-protein interactions, as well as sequence-specific DNA-binding, are important in the formation of the transcription pre-initiation complex. Recent studies have begun to provide information concerning the structures both of regions of transcription factors involved in DNA-binding and of regions important for protein-protein interactions.

1.3.3.1 Activation domains.

The yeast transcription factors GAL4 and GCN4, each involved in the co-ordinate regulation of a number of metabolic enzymes (Giniger et al, 1985; Hill et al, 1986 and references therein), have

both been shown to contain functionally-independent DNA-binding and activation domains (Brent and Ptashne, 1985; Hope and Struhl, 1986). Functional dissection experiments, involving progressive deletions of coding portions of the respective GAL4 and GCN4 genes, have revealed that the analogous functional regions are located in different portions of the two proteins - the activation domain of GCN4 lies in the middle portion of the protein (Hope and Struhl, 1986), while that of GAL4 is near the C-terminus (Brent and Ptashne, 1985). In the absence of the activation domain, DNA-binding activity is retained, but activation of transcription cannot occur. Furthermore, the expression of artificially-constructed fusion proteins in which the activation domains of either trans-activator are fused to the DNA-binding region of the bacterial lexA repressor protein, results in the activation of transcription from reporter genes whose promoters contain the recognition sequence for the lexA protein (Brent and Ptashne, 1985; Hope and Struhl, 1986), demonstrating that the DNA-binding and activation domains are functionally independent. This functional independence of DNA-binding and activation domains would seem to suggest that there is no conformational signalling process between the two domains necessary to activate the function of the activation domain following DNA binding.

Characterisation of the activation domains of GAL4 and GCN4 has not revealed any similarities in the primary sequence other than a preponderance of acidic residues. Mutations of the activating region of GAL4 which increased the activation of transcription all involved increases in the acidity of this region (Gill and Ptashne, 1987). It seems that almost any acidic region will suffice - Ma and Ptashne (1987b) observed that the expression of fusion proteins containing the DNA-binding domain of GAL4 and random polypeptides encoded by E. coli DNA fragments resulted in transcription activation. Each of the activating bacterial sequences were found to contain an excess of acidic residues. Hence, it would seem that the negative charge of such regions, rather than any particular protein conformation, as defined by primary sequence, is important for the activation function. It has been suggested that these activator domains, referred to as "acid blobs" or "negative noodles", function by making non-covalent

protein-protein interactions with basic or hydrophilic regions of other components of the transcription complexes, including perhaps the C-terminal repeat domain of RNA Polymerase II (reviewed in Sigler, 1988). This kind of "molecular imprecision" may allow the interaction of RNA Polymerase II with the wide variety of transcription complex components necessary for transcription of all the genes expressed in a cell.

Functionally-distinct domains are also seen in the case of the steroid receptor family of transcription factors in which DNA-binding and transcription activation domains within the N-terminal half of the hormone receptor molecule are separated from the hormone-binding domain by a "molecular hinge" (reviewed in Maniatis et al, 1987). This modular design of transcription activator proteins and the ability to form functional trans-activators by fusing regions of transcription factors from widely-diverged species, may reflect the evolution of transcription factors by the assembly of multi-functional proteins from various combinations of protein domains by "exon-shuffling", as has been suggested to be the case for the evolution of other proteins (Blake, 1978).

1.3.3.2 DNA-binding domains.

In contrast to the activator domains discussed above, DNA-binding domains of eukaryotic transcription factors contain relatively well-defined structural motifs. Thus far, only a very few such motifs have been identified. The same motif is frequently seen in transcription factors from widely-diverged species, and involved in the regulation of widely different classes of genes. For example, the DNA-binding domains of the steroid receptor family of transcription factors contain a peptide sequence similar to a so-called "zinc finger" motif first identified in the repeating units of the RNA Polymerase III transcription factor, TFIIIA (Miller et al, 1985; Brown et al, 1985). This motif has been suggested to exist as a loop projecting from the protein suface, held in place by co-ordination of a zinc ion at the base of the loop, and capable of interacting directly with DNA. Similar protein sequence motifs, including several

containing two pairs of cysteine residues rather than the pair of cysteines and pair of histidines seen in TFIIIA, have now been recognised in protein sequences predicted by cloned cDNAs of a wide variety of transcription factors, including Sp1 (Kadonaga et al, 1987), the mammalian testis-determining factor (Page et al, 1987), Xfin of Xenopus (Ruiz i Altaba et al, 1987), and the product of a developmental regulatory gene, Kruppel, in Drosophila (Preiss et al, 1985; Rosenberg et al, 1986). Not all of these, however, have been proven to contain co-ordinated zinc ions (reviewed in Klug and Rhodes, 1987).

Prior to the discovery of this large family of zinc-finger proteins, the only motif identified as being involved in DNA binding was the so-called "helix-loop-helix" motif first recognised in bacterial repressors (reviewed in Pabo and Sauer, 1984). This motif, too, has since been found in DNA-binding proteins from widely-diverged organisms, including the homoeobox domain of insect and mammalian developmental genes (Laughon and Scott, 1984), and a factor involved in the determination of yeast mating type (Shepherd et al, 1984).

Recently, Landschulz et al (1988) have suggested that a periodic repetition of leucine residues, present in transcription factors CCAAT/enhancer-binding protein (C/EBP) and GCN4, and the products of c-jun, c-fos and c-myc proto-oncogenes, may be of importance in protein-protein interactions essential for DNA-binding. Such a "leucine zipper" is suggested to be involved in the formation of dimers between pairs of transcription factors by interdigitation of leucine residues projecting from one face of an α-helical region of each transcription factor protein. Although evidence is beginning to accumulate in favour of this suggestion (Marx, 1988), such a motif has not yet been conclusively demonstrated to play a major role in transcription activation. Nevertheless, an emerging feature of the mechanism by which transcription initiation is achieved seems to be that a limited number of DNA-binding protein structures are shared by families of proteins from widely-diverged evolutionary backgrounds.

1.3.4 An evolutionary relationship of components of the transcription pre-initiation complex.

Mention has already been made of various "families" of transcription factors: either families which share common structural motifs, such as the zinc finger proteins (Section 1.3.3.2); or families recognising a common sequence element, such as the CCAAT-binding proteins (Section 1.3.1). Each of these families has representative members found in a range of species. A particularly notable cross-species relationship concerns the GCN4 transcription factor of yeast, and the product of the avian c-jun proto-oncogene, which probably encodes a homologue of the mammalian transcription factor AP-1 (Bohmann et al, 1987). The protein product encoded by the c-jun proto-oncogene shows homology with the DNA-binding domain of GCN4 (Vogt et al, 1987), can bind to the recognition sequence of GCN4, and can activate transcription in yeast cells when its DNA-binding domain is fused to the GCN4 activator domain (Struhl, 1987b).

The organisation of cis-regulatory elements of yeast genes and of higher eukaryotes also show several similarities including the presence of TATA-like sequences. More importantly, there seem to be several functional similarities between the upstream activator sequence (UAS) elements of yeast genes, and enhancers of higher eukaryotes, both classes of element being capable of functioning in both orientations and in various positions relative to the transcription initiation site and to other cis-acting sequences (reviewed by Guarente, 1988; and by Struhl, 1987a).

There is also now mounting evidence of a functional homology between yeast and mammalian transcription factors. The yeast transcription factor GAL4 can activate transcription of several chimaeric promoters containing a GAL4 recognition sequence (including a minimal promoter containing only a TATA box in addition to the GAL4 sequence) when expressed in mammalian (human HeLa or chinese hamster ovary) cells (Webster et al, 1988; Kakidani et al, 1988). Furthermore, the activation domains of GAL4 and GCN4 can activate transcription from a promoter containing a mammalian hormone response

element when expressed as chimaeric fusion proteins containing the DNA-binding domain of the human oestrogen receptor (Webster et al, 1988), suggesting a conservation of the mechanism of transcription activation between yeast and mammals. Similarly, the human oestrogen receptor can function in a hormone-dependent manner in yeast cells (Metzger et al, 1988). Furthermore, Chodosh et al (1988b) have demonstrated functional complementarity of sub-units of yeast and human CCAAT-binding proteins, while Buratowski et al (1988) and Cavallani et al (1988) have shown that a yeast protein can substitute for a human TATA-binding protein in vivo and in vitro. Thus, it seems that not just the mechanism of activation of transcription, but more than one of the so-called "general" transcription factors involved in transcription initiation have been conserved between yeasts and mammals.

The proposition that the overall mechanism of transcription initiation may have been conserved over large periods of evolutionary time is supported by the homology between the large subunit of RNA Polymerase II molecules from yeast, mouse and Drosophila (Allison et al, 1985; Corden et al, 1985; Briggs et al, 1985). This homology includes a region of multiple heptamer repeats at the C-terminus, which contains a very large proportion of hydrophilic amino acids. These have been suggested to interact with the acidic regions of activation domains in the formation of a transcription pre-initiation complex (Sigler, 1988), though there is, as yet, no firm evidence in favour of this. The C-terminal repeat is not seen in the RNA Polymerase of E. coli. There is, however, homology of the eukaryotic RNA Polymerase II molecules with the β-subunit of the prokaryotic enzyme (Allison et al, 1985).

A common ancestry of eukaryotic RNA Polymerases may also be suggested by the homology between the largest subunits of yeast RNA Polymerases II and III (Allison et al, 1985). However, there are differences between the arrangement of cis-regulatory sequences in some of the genes transcribed by these two polymerases, suggesting that other components of the transcription complex involving these polymerases may be different. For example, 5S rRNA genes contain an

internal regulatory element (IRE) within the transcription unit, rather than in 5' flanking sequences (Sakonju et al, 1980) for binding by TFIIIA. Nevertheless, regulatory elements for PolIII transcription also reside upstream of the transcription initiation site, and there is increasing evidence of structural and functional similarities between the promoters of PolII and PolIII genes (reviewed in Sollner-Webb, 1988) including the discovery of the importance of a TATA-like sequence for in vitro transcription by RNA Polymerase III of the gene encoding human 7SK RNA, a stable nuclear RNA of unknown function (Murphy et al, 1987), and the presence of two copies of the immunoglobulin octamer in the Xenopus U6 snRNA promoter (Carbon et al, 1987). Thus, different eukaryotic RNA Polymerases may use common transcription factors. Furthermore, it appears that both RNA Polymerase II and III can initiate transcription of the first exon of c-myc in vitro (Chung et al, 1987) and that, when transfected into cells of another species, rRNA genes, normally transcribed by PolI, can be transcribed by PolII (Jordan, 1987). Not all RNA Polymerases seem to have the same origin, however. The highly homologous RNA Polymerases of yeast mitochondria and of T3 and T7 bacteriophages do not show any homology with the E. coli RNA Polymerase sub-units (Masters et al, 1987).

1.4 GENE REGULATION IN ERYTHROID CELLS.

One of the greatest challenges facing modern biology is the understanding of how gene expression is regulated during development. The unicellular zygote is totipotent, containing, either within its genome or as maternally-derived molecules, the whole range of developmental programmes required to produce the wide range of cell types of the adult organism. Development involves processes of cellular proliferation and differentiation. The emerging picture from studies of differentiation in a wide range of organisms, including nematode worms, sea urchins, the fruit-fly Drosophila and various mammalian tissues, is that differentiation involves a hierarchy of cells with increasingly restricted developmental potential, as cells proceed towards terminal differentiation. Terminally differentiated cells can be distinguished by their characteristic phenotypes, which

are the product of the combinations of genes which are, or have been, expressed in that cell type. The co-expression of the various "differentiation marker" genes characteristic of a given cell type involves several interesting regulatory features, including induction in response to particular intra- or extra-cellular stimuli, expression only at a particular differentiation stage(s), and expression only in a specific cell type, or limited sub-set of cell types. Much recent work has been directed towards understanding cell type-specific co-expression during differentiation, since an understanding of the mechanisms by which sets of genes come to be co-expressed in a particular cell type during differentiation should allow a fuller understanding of the process of differentiation itself.

1.4.1 Erythropoiesis as a model system for the study of cell type-specific co-expression during differentiation.

As will be discussed in detail below, the erythroid lineage represents a valuable model for the study of cell type-specific co-expression during differentiation. Cells from various stages of erythroid differentiation can readily be isolated (Section 1.4.1.1), and permanent cell lines are available which can be induced by chemicals to undergo differentiation *in vitro*, closely mimicking *in vivo* maturation (Section 1.4.1.2). Differentiation of erythroid cells is accompanied by alterations in gene expression which include the appearance of several well-characterised differentiation "markers", including haemoglobin. Regulation of expression of the globin gene family has been widely studied (Section 1.4.2) and shows characteristic alterations in which members of the family are expressed during development. One of the principal challenges of the erythroid system is the understanding of the mechanisms whereby the globin and other, non-globin erythroid differentiation markers come to be co-expressed during differentiation and development (Section 1.4.3).

1.4.1.1 Ontogeny of erythroid cells.

Studies employing chromosome and enzyme markers (reviewed in

Till and McCulloch, 1980) have demonstrated that the different types of blood cells, including the erythroid lineage, arise from the same haemopoietic pluripotent stem cell (Figure 1), as defined by the classic *in vivo* spleen clony-forming assay (Till and McCulloch, 1961), whereby injection of bone marrow cells into lethally irradiated mice results in the formation of macroscopic colonies in the spleen which contain mixtures of differing blood cell types. There has been controversy as to whether cells of the T and B cell lymphoid lineages arise from the same stem cells as the erythroid, granulocyte, macrophage, monocyte and megakaryocyte cells which comprise the myeloid lineages, but it is now widely accepted that they do share a common origin.

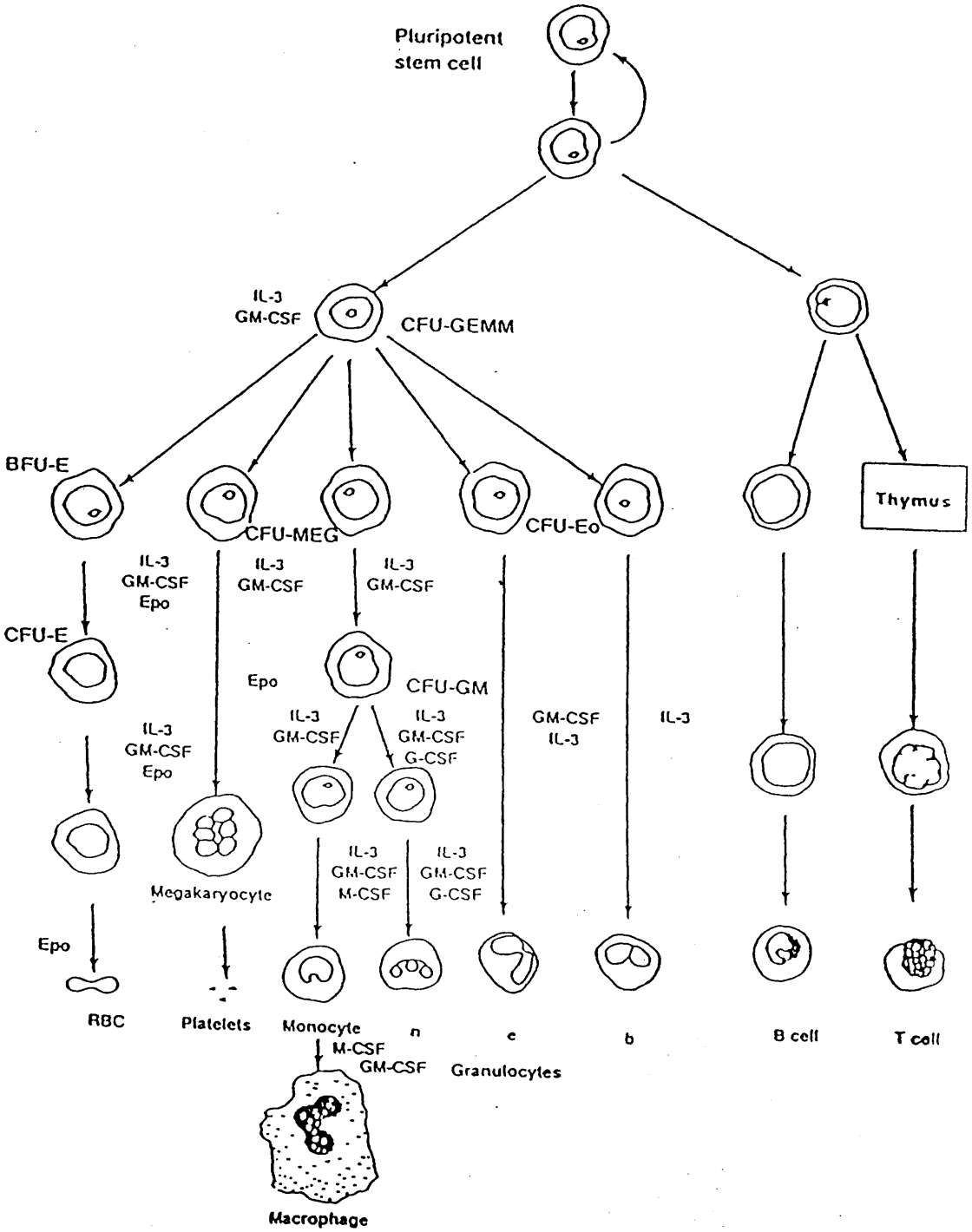
While the self-renewal capacity of pluripotent stem cells means that cell division results in the production of progeny which retain the characteristics of the pluripotent stem cell, they also produce so-called multipotent progenitor cells with limited self-renewal capacity, but with extensive proliferative capacity and the ability to differentiate down any one of several haemopoietic lineages. Further cell divisions result in still relatively undifferentiated precursor cells "committed" to particular lineages. The mechanism by which cells become committed to a particular haemopoietic lineage is uncertain, but it seems likely that commitment is determined by stochastic processes, and that once the commitment "decision" has been taken, differentiation down a particular lineage is facilitated by micro-environmental influences, probably including cell-cell interactions and the short-range actions of diffusible growth factors (reviewed in Harrison 1982a, b).

The earliest committed erythroid precursors are known as burst-forming unit-erythroid (BFU-E) and colony-forming unit-erythroid (CFU-E) cells (Figure 1), and were identified by *in vitro* cell culture of bone marrow explants (Axelrad *et al*, 1973; Stephenson *et al*, 1971) in the presence of the glycoprotein hormone, erythropoietin, the primary factor involved in erythroid differentiation (reviewed in Spivak, 1986). BFU-E represents a more primitive erythroid precursor than CFU-E, having greater proliferative capacity ("bursts" of several

Figure 1. Differentiation of multiple cell lineages during haemopoiesis.

Haemopoiesis involves a hierarchy of cells of increasingly restricted differentiation potential. A pluripotent stem cell with self-renewal capacity (top) gives rise to multipotential progenitor cells - a lymphoid progenitor of B and T cell lymphocytes and a myeloid progenitor, CFU-GEMM (colony-forming unit-granulocyte/ erythrocyte/ macrophage/ megakaryocyte). These in turn proliferate and differentiate to form committed precursor cells of the different blood cell lineages: BFU-E and CFU-E (burst-forming unit- and colony-forming unit-erythroid, respectively); CFU-MEG (colony-forming unit-megakaryocyte); CFU-GM (colony-forming unit-granulocyte/ macrophage); and CFU-Eo (colony-forming unit-eosinophil). Further proliferation and differentiation produces populations of terminally-differentiated cells: erythrocytes (RBC); megakaryocytes; monocytes; neutrophils (n); eosinophils (e); and basophils (b).

Proliferation and differentiation are regulated by a variety of haemopoietic growth factors, varying in the specificity of their activities; interleukin-3 (IL-3); Granulocyte/ macrophage-colony stimulating factor (GM-CSF); erythropoietin (Epo); granulocyte-colony stimulating factor (G-CSF); and macrophage-stimulating factor (M-CSF). (Reproduced from Clark and Kamen, 1987).


Pluripotent stem cell
IL-3
GM-CSF
CFU-GEMM
BFU-E
CFU-MEG
CFU-Eo
Thymus
IL-3
GM-CSF
Epo
IL-3
GM-CSF
IL-3
GM-CSF
CFU-E
Epo
CFU-GM
IL-3
GM-CSF
IL-3
GM-CSF
G-CSF
GM-CSF
IL-3
IL-3
GM-CSF
Epo
Megakaryocyte
IL-3
GM-CSF
M-CSF
IL-3
GM-CSF
G-CSF
Epo
RBC
Platelets
Monocyte
n
e
b
B cell
T cell
M-CSF
GM-CSF
Granulocytes
Macrophage

thousand mature erythrocytes are formed over a period of 7 to 10 days in culture), but is less sensitive to erythropoietin - CFU-Es respond to one-tenth of the amount of erythropoietin required for burst-formation, undergoing approximately 5 or 6 cell divisions to from colonies of several dozen erythrocytes after 2 to 3 days in culture. There have been doubts as to whether BFU-Es have a requirement for erythropoietin in order to undergo further differentiation (see Harrison 1982b for a discussion), but it seems likely that expression of the erythropoietin receptor occurs at an intermediate stage between BFU-E and CFU-E, which are separated by around 6 cell divisions. CFU-Es, however, have an absolute requirement for erythropoietin (Iscove et al, 1980) if they are to proliferate further and undergo terminal differentiation, cell death occurring in the absence of the hormone.

BFU-E and CFU-E cells can be separated from mouse bone marrow by so-called velocity sedimentation, under unit gravity (Heath et al, 1976). The same technique had been used by Denton and Arnstein (1973) to separate more mature erythroid cells from the bone marrow of anaemic rabbits. In normal bone marrow, erythroid precursors are only a minority of cells present, but anaemia results in homoeostatic increased production of erythroid cells. In mice, this anaemic stress results in recruitment of the spleen as a major erythropoietic organ: in rabbits, however, it is the bone marrow itself which becomes enriched in erythroid precursors, to as much as 85% (Thiele et al, 1979). Anaemic stress, induced either by bleeding or phenylhydrazine treatment, is a particularly valuable method for obtaining, from peripheral rabbit blood, cells at the penultimate, enucleate, reticulocyte stage of maturation, since the release of reticulocytes from bone marrow into the peripheral circulation is facilitated by the increase in erythropoietin levels which results from anaemic stress.

Separation of erythroid precursors by the velocity sedimentation technique depends upon a decrease in buoyant density which accompanies maturation of the erythroid precursors. This decrease in buoyant density results from morphological changes which accompany mammalian

erythroid differentiation, including condensation and, finally, extrusion of the nucleus, and subsequent clearance of most of the intracellular architecture. The ability to separate different rabbit erythroid precursors on the basis of their buoyant density allowed Denton and Arnstein (1973) to correlate metabolic changes prior to loss of the nucleus, including the cessation of mRNA and of DNA synthesis, and the commencement of haemoglobin production, with the different erythroblast stages (pro-, basophilic, polychromatic and orthochromatic erythroblasts) between CFU-Es and reticulocytes.

1.4.1.2 Murine erythroleukaemia (MEL) cells.

One of the most useful tools in studying erythroid differentiation has been the isolation of permanent cell lines including the murine erythroleukaemia (MEL), or Friend, cell lines (Friend *et al*, 1971), whose differentiation is arrested at a stage roughly corresponding to the first stage of development following the CFU-E, the pro-erythroblast stage. These cells can be induced to undergo terminal differentiation (including, under appropriate conditions, the extrusion of the nucleus) by a variety of chemicals including dimethyl sulphoxide (DMSO). Whilst this induced differentation is not perfectly analogous to *in vivo* maturation, being independent of erythropoietin, it mimics many of the morphological and biochemical changes which accompany normal differentiation, including decrease in cell size, condensation of nuclear chromatin, and the expression of various characteristic erythroid differentiation markers including haemoglobin (reviewed in Tsiftsogolou and Robinson, 1985). The molecular mechanism of MEL cell induction, however, is still unclear, though it appears to involve a lag period during which commitment to differentate is reversible if the inducer is removed within a short time after initial treatment (reviewed in Harrison, 1982b). Analogous human leukaemic cell lines have been isolated, such as HEL, KMOE and K562 (which shows some of the characteristics of a foetal erythroid cell, including expression of embryonic and foetal globins - Charnay and Maniatis, 1983). Such established erythroid cell lines have been extremely useful in the study of mechanisms of erythroid gene expression, particularly of the globin genes (Spandidos

and Paul, 1982; Allan et al, 1983 and 1985; Wright et al, 1984; Charnay et al, 1984 and 1985).

1.4.2 Regulation of globin gene expression.

The mammalian and avian globin genes show complex patterns of expression, including absolute cell type-specificity and tightly-regulated developmental switching. The unlinked clusters of α-like and β-like genes which give rise to the two types of polypeptide present in each haemoglobin tetramer are not co-ordinately regulated in the same sense as the genes encoding the enzymes of galactose or general amino acid metabolism in yeast which respond to the same transcription activation signal (Giniger et al, 1985; Hill et al, 1986 and references therein). Rather, it is now widely accepted that there must be major differences in the mechanism by which α-like and β-like globins come to be expressed (Charnay et al, 1984). For example, the human adult α-globin gene is expressed at a much earlier stage of development than the adult β-globin gene. Consequently, until recently, it has been unclear whether the α-like and β-like globin genes share any common mechanisms in their erythroid-specific co-expression.

1.4.2.1 Promoter sequences of globin genes.

Globin genes from a variety of mammalian species were amongst the first to be cloned and have subsequently been the focus for much investigation of their mechanisms of regulation. Sequence comparisons of the 5′ flanking regions of various β-like globin genes (van Ooyen et al, 1979; Lacy and Maniatis, 1980; Efstatiadis et al, 1980) revealed the presence of three conserved regions, containing the ATAAAA, CCAAT and CACCC motifs.

All three of these regions seem to be important for efficient, high-level transcription of β-globin genes, but not for cell-type-specificity of transcription, as revealed by deletions and point mutations of the promoter region. The importance of regions containing ATAAAA and CCAAT sequences for high levels of transcription

of rabbit β-globin gene have been demonstrated by the expression of various promoter deletion mutants in mouse L cells and in human HeLa non-erythroid cells (Dierks et al, 1981; Grosveld et al, 1982a, b). Dierks et al (1983) exended this analysis by demonstrating that point mutations in the ATAAAA, CCAAT and CACCC regions all reduce the level of transcription in mouse 3T6 cells. Subsequently, similar results have been obtained from saturation mutagenesis of the whole of the first 100nt of mouse β^{major}-globin 5′ flanking sequences (Charnay et al, 1985; Myers et al, 1986). It should be noted, however, that the CCAAT box does not seem to be essential for high-level expression of the rabbit β-globin gene in vitro, using a HeLa cell extract (Grosveld et al, 1981b).

1.4.2.2 Upstream sequences involved in the modulation of globin gene expression.

While the elements present within the first 100nt of 5′ flanking sequence appear to allow constitutive expression of globin gene promoter deletions in both erythroid and non-erythroid cells, there is evidence that sequences further upstream of globin genes may be involved in cell type-specific expression from the globin gene promoter. By linking human β- and γ-globin sequences within 400nt of the transcription initiation site to the neo (G418 resistance) gene and assessing transcription by scoring for neo expression, Rutherford and Nienhuis (1987) demonstrated that the human β-like globin promoters are active in human and murine erythroid (K562 and MEL) cells, but not in non-erythroid (HeLa and 3T3) cells. Recently, studies in which fragments of 5′ flanking sequence of the mouse α1-globin gene linked to a bacterial chloramphenicol acetyltransferase (CAT) gene have revealed several elements within 4kb of upstream sequence which contribute to cell type-specific expression from this promoter in both transient and stable expression experiments in murine erythroid (MEL) and non-erythroid (STO fibroblast) cells (Frampton et al, manuscript in preparation). Other 5′ flanking sequences which appear to be involved in cell type-specific expression of globin genes have been analysed by Grosveld et al (1987), who have demonstrated that the domain boundaries of the human β-globin gene cluster can

confer cell type-specific, position-independent, high-level expression of the human β-globin gene in transgenic mice. Further, they have suggested that the 3′ boundary may not be required for this regulation.

1.4.2.3 Sequences downstream of the transcription initiation site involved in regulation of globin gene expression.

Experiments in which various promoter deletions and hybrid genes containing fragments of human and rabbit adult β-globin genes were stably introduced into MEL cells, which were then induced to undergo differentiation by DMSO treatment, suggest that not all cis-regulatory sequences controlling globin gene expression are situated in the 5′ flanking region (Charnay et al, 1984; Wright et al, 1984). Wright et al (1984) demonstrated that rabbit β-globin promoter deletions containing no more than 58bp of 5′ flanking sequence can be correctly regulated upon induction of MEL cell differentiation, suggesting that 5′ flanking sequences are not necessary for regulation of the rabbit β-globin gene during erythroid maturation. Charnay et al (1985) reached a similar conclusion for the mouse β^{major}-globin gene from experiments in which none of the linker scanning mutations introduced into the promoter region affected transcription of the gene following induction of MEL cell differentiation. Further support for this conclusion comes from experiments in transgenic mice in which human β-globin promoter deletions containing as little as 48bp of 5′ flanking sequence were sufficient for cell type-specific, developmentally-regulated expression, in a few cases approaching levels of endogenous β-globin mRNA expression (Townes et al, 1985).

Experiments involving transfection of α/β hybrid globin genes (Charnay et al, 1984) suggest that sequences sufficient for increased β-globin expression during MEL cell differentiation reside downstream of the transcription initiation sites of both genes. In the case of the β-globin genes, however, it may be that both 5′ and 3′ sequences are involved in developmental regulation and that either alone is sufficient. Using hybrid genes containing the 5′ flanking sequences from non-inducible γ-globin and murine H-2K major histocompatability

(MHC) gene promoters fused to the 3′ sequences of human β-globin and vice versa, Wright et al (1984) concluded that either 5′ or 3′ sequences of the human β-globin gene were sufficient for regulated expression from a linked promoter, but that the presence of both is not necessary. The existence of an induction-specific regulatory element in the 5′ end of human β-globin gene has recently been confirmed by Antoniou et al (1988) who have demonstrated that a promoter element between 138 and 164nt upstream of the transcription initiation site increases transcription from a linked H2-K gene following induction of MEL cell differentiation.

The realisation that globin genes contain regulatory sequences downstream of the transcription initiation site has culminated in the discovery of downstream enhancers in a variety of globin genes. The 3′ enhancer of the chicken adult β^{A} globin gene, which lies approximately 400nt downstream of the polyadenylation signal, confers erythroid cell-specific and correct temporal regulation when assayed by the so-called "enhancer trap" experiment, in which potential enhancer sequences are linked in cis to the SV40 T antigen gene transcribed from the SV40 early promoter region (Choi and Engel, 1986). Cell type-specific and developmental expression is also conferred upon the human adult β-globin gene by two downstream enhancers (Antoniou et al, 1988; Kollias et al, 1987): one is located between 0.5 and 1.2kb downstream of the poly(A) addition site, and the other within the structural gene. Another enhancer also seems to be present within the human β-globin gene cluster, between 400bp and 1.2kb downstream of the polyadenylation site of the $^{A}\gamma$ gene, as determined by its effect upon the $^{G}\gamma$ promoter used to drive the CAT reporter gene, but, in this case, the 3′ enhancer does not appear to confer cell type-specificity (Bodine and Ley, 1987).

1.4.3 Co-expression of genes in erythroid cells.

Thus far, studies of the globin genes have provided little information concerning mechanisms by which co-expression of genes expressed in erythroid cells is achieved. The mechanisms by which human α- and β-globin genes are activated during differentiation would

seem to be fundamentally different (Charnay et al, 1984). Not only is the α-globin gene switched on at an earlier stage of ontogeny in the normal physiological situation, but cloned human β-globin sequences in MEL cells are correctly regulated during chemically-induced differentiation (Chao et al, 1983; Wright et al, 1983), whereas cloned human α-globin sequences are constitutively expressed in both uninduced and induced MEL cells, unlike the endogenous mouse α-globin gene, which is activated only following induction (Charnay et al, 1984). In contrast, human α- and β-globin genes introduced into MEL cells on intact chromosomes by either chromosome-mediated gene transfer or cell fusion hybrids are both appropriately regulated (Willing et al, 1979; Pyati et al, 1980; Deisseroth et al, 1980a, b; Deisseroth and Hendrick, 1978).

Recent experiments concerning the regulation of globin and non-globin genes, however, may yield important clues as to the mechanisms of erythroid co-expression. A 3′ enhancer has recently been discovered downstream of the transcription initiation site of the chicken histone H5 gene (Trainor et al, 1987) involved in the erythroid-specific expression of histone H5, which replaces histone H1 during erythroid maturation. Interestingly, this cell type-specific enhancer shows homology with a 34bp segment of the erythroid-specific chicken adult β-globin 3′ enhancer, though this region alone does not seem to be sufficient on its own to confer cell-type specific expression in the "enhancer trap" experiment (Trainor et al, 1987). Nevertheless, the finding of sequence similarities between 3′ enhancers of globin and non-globin genes which are expressed in an erythroid-specific fashion may imply an evolutionary conservation of elements involved in erythroid-specific co-expression. This is supported by the observation that the chicken adult β-globin enhancer also shows similarity, in a different region, with the enhancer of the human $^{A}\gamma$-globin gene. Although is does not seem to confer cell type-specific expression upon a linked CAT gene, the $^{A}\gamma$-globin enhancer is, however, associated with two cell type-specific DNaseI HSS, indicating that it is available for the binding of trans-acting factors in erythroid cells (Bodine and Ley, 1987).

In addition, a cell type-specific DNaseI HSS upstream of the chicken α^D-globin has recently been footprinted (Kemper et al, 1987), and may be the binding site for a factor involved in the co-expression of genes in erythroid cells (see below). Nuclear extracts from erythroid nuclei of both embryonic and adult chicken protect a region within the DNaseI HSS which contains the sequence element GATAAG. Similar sequences corresponding to the consensus sequence (T/A)GAPyAPuPu are seen in the promoters and enhancers of a variety of globin genes, including the mouse α1- , human γ-, chicken α^D-, and chicken β^H-globin promoters, and chicken β^A- and human β-globin enhancers (Mantovani et al, 1988b). GATAAG-like sequences have been footprinted in a variety of globin and non-globin genes from different species, including human γ-globin, murine α1- and β^{major}-globin genes, and the erythroid-specific promoter of the human porphobilinogen deaminase gene, and appear to be the binding sites for an erythroid-specific DNA-binding activity (Mantovani et al, 1988b; Plumb et al, submitted for publication). The GATAAG motif may thus represent an element which is involved in co-expression of genes in erythroid cells.

The finding of GATAAG-like sequences in both promoter and enhancer regions of globin genes provides an interesting analogy with the immunoglobulin octamer sequence present in both promoter and enhancer sequences of immunoglobulin genes, and which is involved in the cell type-specific expression of immunoglobulin genes (see Section 1.2.4.3 above). However, unlike the immunoglobulin octamer, which can confer lymphoid-specificity upon expression from a minimal mouse renin-1 promoter containing only a TATA box (Dreyfus et al, 1987), the GATAAG motif in isolation does not seem to be sufficient to confer cell type-specificity: in transient expression experiments, mouse α1-globin promoter deletions which remove the GATAAG element also reduce the erythroid-specific expression of a linked CAT reporter gene; but the GATAAG element alone does not result in erythroid-specific expression in the same system when linked to a minimal murine α1-globin promoter (Plumb et al, submitted for publication).

Another gene expressed in erythroid cells, the erythroid-specific lipoxygenase (RBC 15-LOX) of rabbit reticulocytes, is the subject of this thesis, which is partly concerned with investigating whether the erythroid-specific expression of this gene shows any common features with the transcriptional regulation of other genes expressed in erythroid cells.

1.5 THE LIPOXYGENASE FAMILY OF ENZYMES.

The lipoxygenases (EC 1.13.11.12 linoleate: oxygen oxidoreductases) (LOXs) are a family of non-haem iron-containing enzymes which share the common function of peroxidation of polyunsaturated fatty acids by the addition of molecular oxygen to substrates containing a cis,cis-1,4-pentadiene arrangement (Figure 2).

Thus, the simplest naturally-occurring substrate is linoleic acid, an 18-carbon fatty acid containing two C=C double bonds. Arachidonic acid, a 20-carbon fatty acid containing four double bonds is an important substrate for many functions in mammalian cells, including the LOX-catalysed generation of biological mediators known as leukotrienes (see Section 1.5.2). In principle, peroxidation of arachidonic acid can occur at either end of any of the four double bonds, at 5, 6, 8, 9, 11, 12, 14, and 15 positions relative to the carboxylated end of the chain. Catalysis results in the abstraction of hydrogen from a carbon atom two positions from the site of peroxidation and a consequent shift of a double bond, forming a conjugated 1,3-cis,trans double bond arrangement (Figure 2).

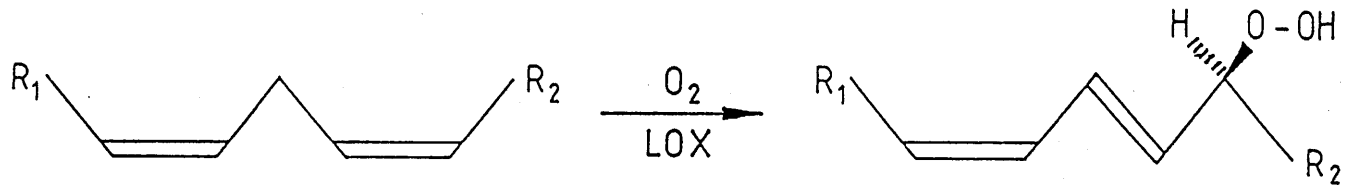
Originally isolated from plants (Theorell et al, 1947), LOXs have since been found to be widely distributed in both animal and plant kingdoms, and may be ubiquitous in mammalian tissues. As detailed in the sections which follow, they show many common features, but exhibit some diversity, most notably in their specificity for the position to which they add molecular oxygen, and in the different roles which they perform in different tissues (see Schewe et al, 1986 for a recent review).

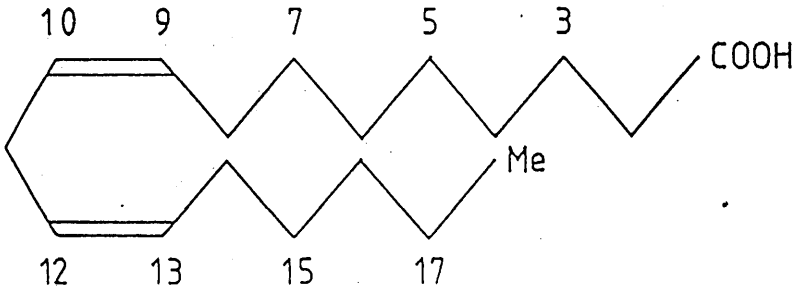
Figure 2. Catalytic activity and substrate specificity of the lipoxygenase enzyme family.

Top: Generalised catalytic function of lipoxygenases (LOXs). LOXs catalyse the addition of molecular oxygen to fatty acid substrates containing a 1,4-cis,cis-pentadiene arrangement of C=C bonds, forming lipid hydroperoxides in which there has been a shift of one of the double bonds, forming a 1,3-cis,trans-pentadiene conjugated double bond system. (R_1 and R_2 are aliphatic chains, one of which terminates in a methyl group (Me); the other terminates in a carboxyl group (COOH).

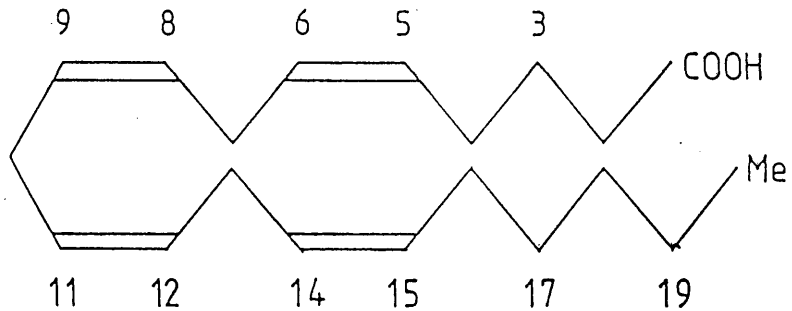
Centre: Structure of linoleic acid, the simplest naturally-occurring substrate for the LOX enzymes, which contains a single 1,4-cis,cis-pentadiene arrangement. LOXs showing different specificities for the position to which they add molecular oxygen can produce a range of hydroperoxide products from this one substrate. In principle, molecular oxygen can be added to any one of the four double-bond carbon atoms (9, 10, 12 and 13). Numbers designate the positions of carbon atoms relative to the terminal carboxyl group (C=1).

Bottom: Structure of arachidonic acid, the most common LOX substrate in mammalian cells, which contains four 1,4-cis,cis-pentadiene arrangements, permitting 8 possible locations for single peroxidation reactions (5, 6, 8, 9, 11, 12, 14, and 15). Different members of the LOX enzyme family catalyse peroxidation at these different positions.





Linoleic acid
(9,12-di-cis-octadecadienoic acid)



Arachidonic acid
(5,8,11,14-all-cis-eicosatetraenoic acid)

1.5.1 Diversity in the positional specificity of various lipoxygenases.

While the various different members of the lipoxygenase family all add molecular oxygen to a given fatty acid substrate, forming fatty acid hydroperoxides, they catalyse peroxidation at different points on the fatty acid chain, relative to the carboxyl terminus. Experiments with the erythroid-specific RBC 15-LOX of rabbit reticulocytes, using substrates of different fatty acid chain length indicate that the position for addition of molecular oxygen is determined by the distance from the terminal methyl group of the chain (Bryant et al, 1982; Schewe et al, 1986). Thus 18-carbon linoleic acid, and 20-carbon arachidonic acid are both peroxidised at the (n-6) position relative to the terminal methyl group. The main product with arachidonic acid is hence a 15-hydroperoxyeicosatetraenoic acid (the number 15 is, by convention, measured relative to the opposite, carboxylated end of the chain).

Different tissues frequently contain LOXs with differing positional specificities. For example, human tissues contain at least three major LOX activities. A 5-LOX is the predominant form in leukocytes; a 12-LOX in platelets; and a 15-LOX in skin and epithelial tissues. However, not all LOXs have absolute positional specificity, the rabbit RBC 15-LOX forming 15-hydroperoxy and 12-hydroperoxy products of arachidonic acid in a ratio of 15:1 (Bryant et al, 1983; Schewe et al, 1986). Nor does any one tissue necessarily contain only one lipoxygenase - for example, in human neutrophils, the major LOX activity shows 5-specificity, but minor 8-, 9-, 11- and 15-LOX activities are also present (Parker, 1984). Indeed, in mammalian leukocytes, the presence of both 5- and 15-LOXs by mammalian leukocytes is vital for the formation of lipoxins, a recently-discovered species of biological mediators, which, together with the leukotrienes, are involved in inflammation and immunity (reviewed in Samuelsson et al, 1987). Similarly, in plants, soybeans contain four LOX isozymes, differing in various characteristics, including positional specificity. Although linoleic acid is the preferred substrate, soybean LOX-1 isozyme peroxidises arachidonic

acid at the 15-position, while LOX-2 and LOX-3 isozymes peroxidise both 11- and 15-positions (Axelrod et al, 1981).

1.5.2 Diversity in the function of various lipoxygenases.

The physiological functions of LOXs in plants are not well understood. Possible roles in the production of biological regulators (e.g. jasmonic acid) involved in germination, fruit ripening, and senescence have been suggested, as well as an involvement in resistance to pathogens (see Yenofsky et al, 1988; Start et al, 1986 and references therein). Both commercially desirable and undesirable effects of LOX action in crop plants are observed, most notably the bleaching of carotenoid pigments during bread-making and the spoilage of soy oil.

The biological roles of LOXs in mammals have been much more widely studied, due to the importance of certain LOX products as biological mediators. The 5-LOX of human leukocytes catalyses the first two steps in the production of so-called leukotrienes (LTs) from arachidonic acid. (Arachidonic acid is also the substrate for the cyclo-oxygenase enzyme, generating a variety of other biological mediators, including prostaglandins, thromboxanes and prostacyclins). Leukotrienes are a family of substances containing a conjugated triene arrangement of double bonds as the result of peroxidation of archidonic acid by LOXs, followed by formation of an unstable epoxide intermediate known as leukotriene A_4 (LTA_4) which is the precursor of the other lekotrienes. All of the leukotrienes have pro-inflammatory effects: LTB_4 primarily affects leukocytes, while LTC_4, LTD_4 and LTE_4 mainly affect cells having contractile capacity (reviewed in Samuelsson et al, 1987). LTC_4, LTD_4 and LTE_4 are collectively referred to as "slow-reacting substance of anaphylaxis" (SRS-A) and are important as the major mediators of the immediate hypersensitivity imflammation of the bronchial tract in response to allergens which we know as asthma. Interactions between 5- and 15-LOXs are involved in the production of other mediators of inflammation and immunity, the lipoxins, which have similar, but distinct, biological activities to those of the leukotrienes (Samuelsson et al, 1987).

As will be discussed in more detail below (Section 1.6.1), a completely different function is served by the RBC 15-LOX enzyme of rabbit reticulocytes. Here, the LOX enzyme is not involved with the production of biological mediators, but with the inactivation and degradation of mitochondria in red blood cells during their maturation from reticulocyte to mature erythrocyte, following extrusion of the erythroid nucleus. Several aspects of the regulation of its expression uniquely suit this enzyme to its function, resulting in degradation of mitochondria only in erythroid cells, and only in the final phase of their maturation (see Section 1.6.3).

1.5.3 Common features of the lipoxygenase enzyme family.

Despite these differences in positional specificity and function, the various members of the LOX enzyme family show many common features (reviewed in Schewe et al, 1986). All are capable of catalysing a variety of reactions, resulting in the formation of a wide variety of products. As well as the common feature of catalysing single dioxygenation reactions of substrates containing the cis,cis--1,4-pentadiene arrangement, all are capable, under appropriate, often non-physiological conditions, of catalysing multiple dioxygenations; anaerobic "lipohydroperoxidase" reductions of fatty acid hydroperoxides to their reduced hydroxyderivatives; co-oxidations involving the free-radicals generated as a normal part of catalysis; and anaerobic formation of ketoacids and fatty acid dimers. Most notably, all LOXs, even those whose physiological role does not involve leukotriene synthesis, are capable of catalysing the formation of LTA_4 (Schewe et al, 1986).

These similar catalytic activities are reflected in similar properties of the enzymes. All seem to require "product-activation" by hydroperoxy fatty acids, to show similar behaviour with various inhibitors, including "suicidal substrates", and to show similar kinetics. Most undergo "self-inactivation", involving loss of enzyme activity as a result of attack by free radicals formed during catalysis, and a similar catalytic mechanism appears to apply to all

members of the family studied (reviewed in Schewe et al, 1986).

1.5.4 Structure-function relationships of the lipoxygenases.

All the similarities and differences between various members of the LOX enzyme family are presumably the result of their primary amino acid sequences. Concerted efforts are being made to understand features such as differing positional specificities in terms of the protein sequence of the LOX enzymes, and, at the nucleic acid level, to investigate whether the family of enzymes are encoded by an evolutionarily-related family of genes.

At the commencement of this project, in addition to the four purified soybean LOX isozymes (Axelrod et al, 1981), the only mammalian LOX enzyme to have been purified to homogeneity and characterized to any significant extent was an erythroid-specific, RBC 15-LOX from rabbit reticulocytes (Rapoport et al, 1979; Thiele et al, 1982; Bryant et al, 1983; Schewe et al, 1986). Subsequently, characterisation of several other mammalian LOX proteins has been reported, including 5-LOXs from rat basophilic leukaemia cells, human and porcine leukocytes, and murine mast cells (Hogaboom et al, 1986; Rouzer and Samuelsson, 1985; Ueda et al, 1986; Shimizu et al, 1986). Further, the genes for the rabbit RBC 15-LOX (Thiele et al, 1987 and Section 1.6.5 below) and the soybean LOX-3 isozyme (Yenofsky et al, 1988) have been obtained, and, very recently, cDNA clones for the human leukocyte 5-LOX (Dixon et al, 1988; Matsumoto et al, 1988) and the soybean LOX-1 isozyme (Shibata et al, 1987). The availability of protein sequences, predicted from nucleic acid sequences, will allow investigation of the structure-function relationships of the various LOXs (see Discussion), while the cloning of LOX genes will allow investigation of the regulation of their expression.

1.6. THE ERYTHROID-SPECIFIC RBC 15-LOX OF RABBIT RETICULOCYTES.

As will be discussed in more detail below, the expression of a 15-LOX from rabbit reticulocytes shows several interesting features, including erythroid-specific expression of both mRNA and protein

(hence the "RBC" 15-LOX) and restriction to the final stages of erythroid maturation.

1.6.1 Biological function of the rabbit RBC 15-LOX.

Unlike the 5-LOX of mammalian leukocytes, the major role for rabbit RBC 15-LOX is not in the synthesis of leukotrienes. Rather, it serves a vital function in the final stages of maturation of erythroid cells, participating in the inactivation and degradation of mitochondria in reticulocytes (Schewe et al, 1975; Rapoport et al, 1979). The primary function of the erythrocyte is the carriage of oxygen in the form of oxygenated haemoglobin. In order to achieve this function with maximum efficiency, the maturing erythroid cell produces haemoglobin in huge quantities and, following extrusion of the nucleus, is cleared of most of its intracellular structures, including the mitochondria. Inactivation and degradation of the mitochondria is important in two respects. Firstly, if oxygen is to be transported with maximum efficiency, then its utilisation by the process of aerobic respiration in the mitochondria must be inhibited; having achieved synthesis of large quantities of haemoglobin and of other proteins important in erythroid function, the mature erythrocyte no longer has need of large quantities of energy for synthetic processes. Secondly, removal of the bulk of the mitochondria and other intracellular organelles allows accumulation of the maximum quantity of haemoglobin within the cell and offers minimal resistance to deformation of the erythrocyte in passing through capillaries of the vascular system.

The RBC 15-LOX appears to be involved at several stages in the inactivation and degradation of mitochondria (reviewed in Schewe et al, 1986). Unlike the situation for the human leukocyte 5-LOX in leukotriene synthesis, all of these activities appear to be direct effects of peroxidation of fatty acids by the RBC 15-LOX, rather than indirect effects mediated by biologically active products of LOX catalysis. Firstly, the RBC 15-LOX is involved in the inhibition of the respiratory chain, possibly at several points (Rapoport et al, 1979; Schewe et al, 1981; Wiesner et al, 1981; Schewe et al, 1986).

This action may involve changes in the electical properties and fluidity of the inner mitochondrial membrane as a result of lipid peroxidation (Schewe et al, 1986). Secondly, it is involved in the lysis of the mitochondria (Schewe et al, 1975; Rapoport et al, 1979). This activity is dependent upon a feature of the RBC 15-LOX which is unique among the LOXs - the ability to directly attack intact mitochondrial membranes, without prior release of the fatty acid chains from the membranes by phospholipase action, as is necessary for the activity of other LOXs. Nevertheless, the RBC 15-LOX can peroxidise free fatty acids with even higher efficiency (Rapoport et al, 1979). The enzyme preferentially attacks mitochondrial membranes rather than the plasma membrane of the reticulocyte, probably due to differences in membrane composition, including the higher proportion of cholesterol in the plasmalemma (Rapoport et al, 1979). Finally, RBC 15-LOX appears to be involved in triggering degradation of proteins from the lysed mitochondria by ubiquitin-ATP-dependent proteolysis (reviewed in Schewe et al, 1986).

1.6.2 Characterisation of the RBC 15-LOX protein and its mRNA.

RBC 15-LOX enzyme activity, as detected by the spectrophotometric change in absorbance at 234nm which accompanies the formation of a conjugated diene during fatty acid peroxidation, is virtually undetectable in peripheral blood or in bone marrow of normal rabbits. However, during anaemia, induced either by bleeding or by phenylhydrazine treatment, there is a large increase in RBC 15-LOX protein synthesis, with the result that the enzyme is one of the most abundant non-globin proteins in reticulocytes of the peripheral blood during anaemia. Thiele et al (1979) have shown, by incorporation of ^{35}S-methionine into cellular proteins in density gradient-separated populations of reticulocytes from peripheral blood, that, between the third and fifth days of bleeding anaemia, synthesis of RBC 15-LOX protein increases from 14% to 30% of new non-globin protein synthesis, accounting for approximately 4% of total protein synthesis and 17% of total non-globin protein at this stage. The concentration of the enzyme at this point is higher even than that of catalase (Rapoport et

al, 1979).

The protein has been purified to homogeneity from reticulocytes of anaemic peripheral blood by ammonium sulphate precipitation followed by anion-exchange chromatography and isoelectric focussing (Rapoport et al, 1979), and polyclonal antibodies have been raised (Thiele et al, 1979; Wiesner et al, 1977). Its molecular weight by denaturing polyacrylamide gel electrophoresis is 68kD (Thiele et al, 1987), though earlier reports quote a value of 78kD (Rapoport et al, 1979), of which about 5% represent neutral sugars (Rapoport et al, 1979). Like other LOXs, it contains 1 atom of iron per molecule, but this is not present as haem or iron-sulphur (Wiesner et al, 1983; Rapoport et al, 1979). Denaturing polyacrylamide gel electrophoresis and N-terminal protein sequencing (Thiele et al, 1987) show that it contains a single polypeptide chain. However, it can form enzymatically-inactive oligomers (Hohne and Andree, 1980). This means that amounts of protein, as detected by immunological methods, and amounts of enzyme activity are not always in proportion. The N-terminal amino acid is glycine (Rapoport et al, 1979) and it shows C-terminal micro-heterogeneity, possibly due to carboxypeptidase activity present in reticulocytes (Wiesner et al, 1983; Rapoport et al, 1979). Its amino acid composition has been determined (Wiesner et al, 1983), and, recently, the sequence of 30 N-terminal amino acids has been obtained (Thiele et al, 1987).

The mRNA encoding the RBC 15-LOX has been isolated and characterised (Thiele et al, 1982). mRNA was isolated from polysomes by protease digestion followed by poly(U)-Sepharose affinity chromatography. It was enriched by three cycles of density-gradient centrifugation, each followed by determination of fractions containing the RBC 15-LOX mRNA by in vitro translation and size comparison with the purified 15-LOX protein. The calculated molecular weight of the mRNA, averaged from determinations by sucrose density gradient centrifugation, gel electrophoresis and electron microscopy corresponds to $1x10^6$D, or approximately 2.9kb. Estimates of the proportion of RBC 15-LOX mRNA by cell-free translation suggest that it represents about 2% of polyadenylated RNAs in reticulocytes from

anaemic peripheral blood (Thiele et al, 1982).

1.6.3 Erythroid-specific expression of the RBC 15-LOX.

Evidence discussed below shows that expression of immunologically-detectable RBC 15-LOX in rabbits is restricted to the erythroid cell type, and to late stages in the erythroid maturation, following extrusion of the nucleus. Accumulation of the RBC 15-LOX mRNA is also erythroid-specific (Section 1.6.4 below), and the differentiation stage-specific appearance of the protein seems to involve translational inactivation. The regulation of RBC 15-LOX expression thus presents several interesting features.

As outlined earlier, LOXs are found in a wide variety of mammalian tissues, but the RBC 15-LOX protein is detected only in reticulocytes, using antibodies raised against the purified RBC 15-LOX protein or by assaying for the inhibition of mitochondrial respiration characteristic of the RBC 15-LOX. The RBC 15-LOX is not enzymically active, nor immunologically-detectable in peripheral blood from mature erythrocytes; it is also very low in anaemic bone marrow, but high in peripheral blood reticulocytes of anaemic animals (Rapoport et al, 1979). Thiele et al (1979 and 1982) have demonstrated that the appearance of LOX activity in reticulocytes is due to new protein synthesis, not to a post-translational modification. Incorporation of ^{3}H-leucine into a newly-synthesized non-globin protein of the same size as the purified RBC 15-LOX protein was observed in peripheral blood reticulocytes after 3 - 5 days of anaemia induced by bleeding, but not in erythroblasts and reticulocytes from anaemic bone marrow. Furthermore, immuno-precipitation of non-globin proteins with anti-RBC 15-LOX anti-serum resulted in precipitation of a labelled protein of the same size as the pure RBC 15-LOX protein in extracts of peripheral reticulocytes, but no immunologically-reactive material was obtained from bone marrow, suggesting that there is not even a precursor protein synthesized in bone marrow (Thiele et al, 1979). Thiele et al (1982) further demonstrated that, within the reticulocyte population of anaemic animals, RBC 15-LOX synthesis occurs in older peripheral reticulocytes, but not in younger peripheral reticulocytes or

reticulocytes from anaemic bone marrow. Reticulocytes were separated from other bone marrow cell populations by sucrose gradient sedimentation, and peripheral reticulocytes of different ages were separated according to size by serum albumin density gradient centrifugation. Cellular incorporation of radio-label into newly-synthesised protein of the same size as the purified RBC 15-LOX was seen only in older reticulocytes.

The appearance of RBC 15-LOX protein synthesis only in the enucleate reticulocyte phase of erythroid maturation suggests that there must be some translational regulation of RBC 15-LOX expression, since synthesis of the RBC 15-LOX mRNA must have ceased prior to loss of the nucleus. Thiele et al (1982) have shown that RBC 15-LOX mRNA is present, but in a translationally-inactive form, in younger reticulocytes of anaemic peripheral blood. Older reticulocytes, but not younger ones, were capable of cellular incorporation of ^{35}S-methionine into an RBC 15-LOX-sized band; whereas in vitro translation of mRNA isolated from reticulocytes of both age ranges resulted in incorporation of radio-label into a protein of the appropriate size.

The mechanism by which this translational inactivation might occur is unclear. However, inactivation as an mRNP complex is a possibility (Thiele et al, 1982), since, when mRNPs, ribosomes and ribosomal sub-units from anaemic peripheral reticulocytes were separated from polysomes by sucrose gradient centrifugation, RBC 15-LOX mRNA was found to be present predominantly, but not exclusively, in the mRNP fraction. In vitro translation produced an RBC 15-LOX-sized protein only after purification of mRNA from the mRNP fraction (Thiele et al, 1982). If inactivation of RBC 15-LOX mRNA does occur via formation of an mRNP particle, it is unknown whether masking from the translational apparatus is achieved by a protein, removed by protease K treatment during mRNA isolation, or by some other RNA present in the complex (Thiele et al, 1982).

1.6.4 Cloning of the RBC 15-LOX mRNA.

Using a cDNA library prepared from RBC 15-LOX-enriched mRNA (Thiele *et al*, 1982 and Section 1.6.2 above) a series of overlapping RBC 15-LOX cDNAs have been obtained (Thiele *et al*, 1987; J. Fleming, unpublished results) by screening with two complementary mixed-sequence 17-mer oligonucleotide probes whose sequences were based upon the N-terminal sequence of the purified RBC 15-LOX protein. One cDNA recombinant, pGF6, which contains a 356bp cDNA insert, has been shown to represent a genuine RBC 15-LOX cDNA clone by a variety of methods (Thiele *et al*, 1987). Most importantly, its nucleic acid sequence accurately predicts the first 30 amino acids of N-terminal RBC 15-LOX protein sequence. Furthermore, this cDNA clone hybrid-selects an mRNA from anaemic peripheral blood reticulocytes which can be translated *in vitro* in the rabbit reticulocyte lysate system to produce a radio-labelled protein of the same size as the purified RBC 15-LOX. Furthermore, immuno-precipitation of the products of this hybrid-selection/*in vitro* translation, using affinity-purified polyclonal anti-RBC 15-LOX antibodies, also precipitates a protein of the correct size.

Thiele *et al* (1987) have demonstrated by hybridisation of the pGF6 cDNA insert to Northern blots of RNAs from various rabbit tissues that the RBC 15-LOX mRNA accumulates in an erythroid-specific fashion, being expressed in reticulocytes, but not in bone marrow, heart, spleen or brain.

Hybridisation of the cDNA inserts of various RBC 15-LOX cDNAs with restriction digestions of rabbit liver genomic DNA produces a simple hybridisation pattern indicative of hybridisation to a single gene (Thiele *et al*, 1987; J. Fleming, unpublished results). This might suggest either that mRNAs encoding the various LOXs found in different rabbit tissues are transcribed from the same gene, or that they are transcribed from other genes, with which the RBC 15-LOX cDNA inserts do not hybridise. The observation that none of the RBC 15-LOX cDNA inserts derived from different regions of the RBC 15-LOX mRNA hybridise with RNAs from various rabbit tissues (Thiele *et al*, 1987;

J. Fleming, unpublished results) would seem to favour the idea that the LOX activities found in various other rabbit tissues are encoded by different genes.

1.6.5 Cloning of the rabbit RBC 15-LOX gene.

The RBC 15-LOX cDNA insert from the pGF6 recombinant described above has been used to screen a rabbit genomic library prepared from rabbit liver genomic DNA fragments cloned into a bacteriophage lambda vector. Two virtually identical clones, the longer of which contains a rabbit genomic insert of 17kb, have been isolated (Thiele *et al*, 1987). The 5′ end of the pGF6 insert, including sequences encoding the N-terminal region of the RBC 15-LOX protein, hybridises to a 1.5kb EcoRI fragment extending between 2.5 and 4kb from one end of this genomic fragment (Thiele *et al*, 1987 and Figure 3).

The cloning of the rabbit RBC 15-LOX gene will permit the investigation of structure-function relationships of the enzyme, and comparisons to be made with other members of the LOX enzyme family expressed in other tissues and in other species. Furthermore, it will allow the first investigation of the organisation of a mammalian LOX gene and of LOX gene expression.

1.6.6 Aims of this project.

The aims of this thesis have been to investigate transcription of the RBC 15-LOX gene and whether expression of the mRNA shares any common features with the expression of other genes in erythroid cells, particularly of globin genes. As a first step towards identifying possible cis-regulatory sequences, the transcription initiation site of the RBC 15-LOX mRNA has been mapped, allowing identification and characterisation of the 5′ flanking region. Various properties of the 5′ flanking sequences have been studied, including the binding of nuclear proteins and the ability of proximal upstream sequences to function as a promoter *in vivo*. Functional experiments designed to locate regions of the gene which

may be involved in erythroid-specific regulation of transcription from the RBC 15-LOX promoter have also been performed.

Chapter 2.

MATERIALS AND METHODS.

2.1 Materials.

All radio-isotopes, and the M13 sequencing kit were obtained from Amersham International PLC, Amersham, Buckinghamshire.

All restriction enzymes, polymerases (Avian Myeloblastosis Virus reverse transcriptase, and Klenow fragment of Eschericia coli DNA polymerase) and DNA modifying enzymes (S1 nuclease, calf intestinal alkaline phosphatase, polynucleotide kinase, T4 DNA ligase, DNaseI), oligo(dT)-cellulose, and proteinase K were obtained from Boehringer Mannheim, Lewes, East Sussex.

DNA size markers (λ/HindIII and ϕX174/HaeIII), RNA size markers, M13 Universal primer, agarose, low melting-point agarose and urea were obtained from Gibco/Bethesda Research Laboratories, Paisley.

Deoxy- and dideoxy-ribonucleotides, and mixed hexanucleotides for random-primed radio-labelling were from Pharmacia Ltd., Milton Keynes, Buckinghamshire.

Synthetic oligodeoxyribonucleotides were synthesised on an Applied Biosystems Model 381A automated DNA Synthesiser.

Biogel A-50m agarose beads for gel filtration column chromatography, and Bio-Rad Protein Assay Kit were from Bio-Rad Laboratories Ltd., Watford, Hertfordshire.

GeneClean kit for DNA purification was supplied by Stratech Scientific Ltd., London.

Elutip-d minicolums were from Schleicher and Schull, Dassel, FRG

RNAzol kit for RNA preparations was obtained from Biogenesis Ltd., Bournemouth.

Serum, media and supplements for cell culture were supplied by Gibco/Bethesda Research Laboratories, Paisley.

Bacto-tryptone, Bacto-agar and yeast extract were from DIFCO Laboratories, Detroit, Michigan, USA.

Phenol was obtained as a water-saturated liquid from Rathburn Chemicals Ltd, Walkerburn, Peebleshire

Absolute ethanol was supplied by James Burroughs (F.A.D) Ltd, Witham, Essex.

All other chemicals and solvents were obtained from BDH LTD., Poole, Dorset or from Sigma Chemical Co. Ltd., Poole, Dorset.

Plastic-ware for cell culture was supplied by Nunc InterMed, Roskilde, Denmark and by Sterilin Ltd., Feltham, Middlesex.

Thin layer chromatography plates were supplied by Camlab, Cambridge.

Chromatography columns were from Bio-Rad, Watford, Hertfordshire.

Biodyne membranes for nucleic acid blotting were from Pall, Glen Cove, New York, USA.

2.2 General procedures involving nucleic acids.

2.2.1 Handling of nucleic acids.

All solutions used in the preparation and manipulation of nucleic acids were made up using water purified using the Millipore Milli-Ro and Milli-Q reagent grade water purification system. All solutions were autoclaved before use, or, in the case of thermolabile substances, filter-sterilised using 0.22μm filters then stored in sterile plastic containers. In addition, all solutions for RNA work were treated overnight with diethylpyrocarbonate (DEPC) as a 0.1% v/v solution to minimise degradation of RNA due to endogenous or contaminating RNases, and then autoclaved to remove traces of DEPC. All glassware for work with RNAs was soaked in DEPC-treated water then

baked. Microcentrifuge tubes and disposable pipette tips for handling nucleic acids were autoclaved before use, and disposable rubber gloves worn at all times.

2.2.2 Organic extraction of nucleic acid solutions.

Phenol, obtained as a water-saturated liquid, was equilibrated with TE buffer (10mM Tris-HCl pH8.0; 1mM EDTA) following sequential extractions with 1M Tris-HCl pH8.0 then twice with 0.1M Tris-HCl pH8.0. On each occasion, phenol was mixed thoroughly with an equal volume of buffer, which was removed after separation of the phases. TE-saturated phenol was made 0.1% w/v with respect to 8-hydroxyquinoline and stored in the dark at room temperature.

Chloroform was prepared as a 25:1 v/v mixture with isoamyl alcohol.

Nucleic acid solutions were deproteinised by vortexing with an equal volume of a 1:1 mix of the phenol and chloroform preparations described above. Phase separation was achieved by centrifuagation (12000g, 5min, room temperature).

2.2.3 Concentration of nucleic acids.

Solutions containing nucleic acids were adjusted to 0.3M with respect to monovalent cations using a 3M sodium acetate pH5.0 stock. Precipitation was achieved by addition of 2.5 volumes ethanol followed by chilling at -20°C for >2hours or at -70°C for 10min. Precipitates were collected by centrifugation at 12000g for 10min at room temperature. The nucleic acid pellet was then washed with 70% v/v ethanol and dried under vacuum before resuspension in water or an appropriate buffer.

Volumes of aqueous solutions were reduced by repeated cycles of addition of equal volumes of butanol, and removal of the upper, organic phase.

2.2.4 Spectrophotometric determination of nucleic acid concentrations.

Nucleic acid concentrations were measured by determining the absorbance of an aqueous solution at 260nm using the convention that an absorbance of 1 unit is equivalent to a double stranded DNA

concentration of 50μg/ml and an RNA concentration of 40μg/ml (Maniatis et al, 1982).

2.3 Transformation of bacterial cells with DNA.

E. coli K12-derived bacterial strains used (Yanisch-Perron et al, 1985) were:

JM83: ara, Δ(lac-proAB), rpsL (=straA), φ80, lacZΔM15

JM101: thi, SupE, Δ(lac-proAB), [F' traD36, proAB, $lacI^{q}ZΔM15$]

Preparation of transformation-competent cells was carried out essentially as described by Mandel and Higa (1970).

Fresh overnight cultures were diluted 1:100 in 50ml medium (L-broth (1% w/v bacto-tryptone; 0.5% w/v yeast extract; 1% w/v NaCl) for JM83 cells; 2x TY (1.6% w/v Bacto-tryptone; 1% w/v yeast extract; 0.5% w/v NaCl) for JM101 cells) and grown to an OD_{600} of 0.4-0.6. Cells were pelleted by centrifugation (1000g, 5 min, 4°C), resuspended in 1/2 volume of pre-cooled 50mM $CaCl_2$, incubated at 0°C for 30 min, then pelleted and resuspended again in 1/10th original volume of ice-cold 50mM $CaCl_2$. Transformation-competent cells were stored for up to 2days at 4°C.

Half of each ligation mixture, containing up to 1μg DNA, was added to 200μl competent cells, incubated at 0°C for 30 min and then heat shocked at 42°C for 3min.

For plasmid transformation, heat-shocked JM83 cells were added to 1ml of L-broth and the culture incubated at 37°C for 1 hour to allow expression of plasmid-encoded functions before spreading onto L-broth supplemented with 1.5% w/v agar and 100μg/ml ampicillin contained in 9cm Petri dishes.

For M13 transfection, heat-shocked JM101 cells were added to a mixture containing 0.2ml exponentially growing cells; 40μl 2% X-gal (5-bromo-4-chloro-3-indolyl-β-galactoside) in dimethylformamide; 40μl 100mM IPTG (isopropyl-β-D-thiogalactopyranoside). The mixture was immediately added to 3ml molten H top agar (1% w/v Bacto-tryptone; 0.8% w/v NaCl; 0.8% w/v agar), mixed quickly and poured into a 9cm Petri dish containing solidified H agar (1% w/v Bacto-tryptone; 0.8% w/v NaCl; 1.2% w/v agar).

Plates were incubated, inverted at 37°C overnight.

2.4 Preparation of nucleic acids.

2.4.1 Eukaryotic RNA

2.4.1.1 From rabbit reticulocytes.

Peripheral blood was enriched for reticulocytes by rendering adult rabbits anaemic by five daily injections of a 2.5% solution in PBS (Phosphate Buffered Saline - see Section 2.1.3.1) of 1-acetyl-2-phenylhydrazine to 7mg/kg body weight. Animals were sacrificed on the seventh day after commencement of anaemia induction and blood collected in BSS (balanced salt solution) containing 10units/ml heparin. Total cytoplasmic RNA was prepared from peripheral blood essentially as described by Minty and Gros (1980). Cells were pelleted by centrifugation (2000g, 20min., 4°C) and the supernatant and buffy coat layer removed from above the red cell pellet. Red cells were washed in BSS/heparin as above, followed by a second centrifugation and removal of the remaining buffy coat cells. The red cell pellet was resuspended in an equal volume of 140mM NaCl; 1.5mM $MgCl_2$; 10mM Tris-HCl pH7.5, and 1/20th volume vanadyl-ribonucleoside complexes added to inhibit RNAses. Lysis was then achieved by addition of NP-40 to 0.5% v/v followed by incubation at 0°C for 5 min. Nuclei were pelleted by centrifugation (4000g, 5min, 4°C) and RNA precipitated from the supernatant by the addition of an equal volume of 8M urea; 4M LiCl and incubation at 4°C for 24hours. RNA was then pelleted by centrifugation (16000g, 15min, 4°C), dissolved in 25ml 8M urea and reprecipitated by the addition of 25ml 4M LiCl and incubation at 4°C for 24 h. RNA was again pelleted by centrifugation (16000g, 15min, 4°C), resuspended in 25ml 10mM Tris-HCl pH7.5; 0.2% w/v SDS followed by addition of sodium acetate to 0.3M and additional SDS (sodium dodecyl sulphate) to a final concentration of 0.5% v/v and the resulting solution extracted thoroughly with 1:1 phenol: chloroform. Phase separation was achieved by centrifugation (16000g, 15min, 20°C), after which the RNA was ethanol precipitated and stored under 70% ethanol at -20°C.

2.4.1.2 From other rabbit tissues.

Tissues were dissected and immediately plunged into liquid

nitrogen. During dissection, gall bladders were removed from livers.

RNA was obtained essentially according to Krieg et al (1983). Tissues were ground under liquid nitrogen using a mortar and pestle. The ground powder was then added to a lysis/deproteinising mixture of 10ml 66% phenol; 0.3M sodium acetate pH7.5; 0.5% w/v SDS; 5mM EDTA. The mixture was shaken gently but thoroughly for 5min before extraction with an equal volume of 25:1 chloroform: isoamyl alcohol for 5min with shaking. Phase separation was achieved by centrifugation (12000g, 5min, 4°C) and the upper, aqueous phase re-extracted with chloroform: isoamyl alcohol prior to a further extraction with an equal volume of chloroform (5min with shaking). Following separation of phases by centrifugation as above, nucleic acids were precipitated from the upper, aqueous phase with two volumes of absolute ethanol overnight at -20°C and resuspended in 1ml of DEPC-treated water. RNA was selectively precipitated from this solution with an equal volume of 4M LiCl overnight at 4°C and pelleted by centrifugation (10000g, 20min, 4°C). Pellets were rinsed with 2ml 70% ethanol and resuspended in 0.5ml DEPC-treated water. Magnesium chloride was added to a final concentration of 5mM and the solution treated with RNase-free DNaseI at 25µg/ml for 20min at room temperature to remove any contaminating DNA. The solution was then extracted once with an equal volume of 1:1 phenol chloroform and RNA precipitated by the addition of two volumes of absolute ethanol.

2.4.1.3 From cultured cell lines.

RNA for CAT mRNA analysis was obtained from adherent cell lines by a modification of the method of Chomczynski and Sacchi (1987) using the RNAzol kit.

Adherent cultured cells were collected in 1ml PBS from Petri dishes by scraping (Section 2.14.2). 0.2ml of each suspension was removed for assay of CAT enzyme activity (Section 2.14.4). The remaining 0.8ml was pelleted by centrifugation (500g, 30sec, room temperature) and the supernatant discarded. Pelleted cells were lysed in 0.8ml RNAzol solution (containing guanidinium isothiocyanate and phenol) by vortexing. The lysate was mixed with 1/10th volume of chloroform by vortexing and incubated on ice for 15min. Phases were separated by centrifugation (12000g, 15min, 4°C) and RNA precipitated

from the upper, aqueous phase by addition of an equal volume of isopropanol at -20°C for 3hours and pelleted by centrifugation (12000g, 15min., 4°C). The pellet was washed once in 75% ethanol, dried under vacuum, and resuspended in 0.1ml 0.5% w/v SDS, followed by precipitation with two volumes of absolute ethanol overnight at -20°C. The final RNA pellet was resuspended in DEPC-treated water at 1mg/ml.

2.4.1.4 Selection of poly A^+ RNA.

This was performed according to Maniatis et al (1982). Oligo(dT)-cellulose in loading buffer (0.1% w/v SDS; 1mM EDTA; 0.5M NaCl; 20mM Tris-HCl pH 7.6) was poured to form a 1ml column in a sterile 10ml BioRad plastic disposable column. The column was then washed thoroughly in loading buffer. RNA was heated in loading buffer at 65°C for 5min, cooled and applied to the column. The flow-through was collected, re-heated and re-applied to the column twice. The column was then washed once with 5ml loading buffer and then three times with 2ml (loading buffer containing 0.1M NaCl). The poly A^+ RNA was then eluted from the column by the application of three successive 1ml aliquots of elution buffer (0.05% w/v SDS; 1mM EDTA; 10mM Tris-HCl pH7.5), ethanol precipitated and stored under ethanol at -20°C.

2.4.2 Prokaryotic plasmid DNA.

2.4.2.1 Small scale preparation.

For preliminary analysis, plasmid DNA was prepared essentially according to the method of Birnboim and Doly (1979).

A single transformed bacterial colony (Section 2.3) containing the plasmid of interest was inoculated into 5ml SuperBroth pH7.2 (Gorman, 1985) [SuperBroth is a mixture of 9 volumes of Solution A (1.3% w/v Bacto-tryptone; 2.7% w/v yeast extract; 0.5% v/v glycerol) and 1 volume of Solution B (0.72M K_2HPO_4; 0.28M KH_2PO_4)]. Ampicillin was added to 100µg/ml and the culture incubated at 37°C overnight in an orbital shaker. Bacterial cells were pelleted from 1ml of the overnight culture by microcentrifugation (12000g, 15sec, room temperature), resuspended in 0.1ml 50mM glucose; 25mM Tris-HCl pH8.0; 10mM EDTA; 2mg/ml lysozyme by vortexing and incubated at room temperature for 5min to allow bacterial lysis. 0.2ml freshly-prepared

0.2M NaOH; 1% w/v SDS was added and mixed gently without vortexing, followed by incubation on ice for 5min. The solution was neutralised by addition of 0.15ml ice-cold 5M potassium acetate pH4.8 followed by thorough mixing. After a further 5-minute incubation on ice, precipitated genomic DNA and proteins were removed by microcentrifugation for 1 minute. Plasmid DNA was precipitated from the supernatant with 2 volumes of ethanol at room termperature for 2min, pelleted by microcentrifugation for 1 minute at room temperature, and, after rinsing with 70% ethanol resuspended in 20µl water by vortexing and stored at -20°C . 2µl aliquots were used for analysis by digestion with restriction enzymes (Section 2.7.1).

If required for later use, 0.5ml of the remaining 4ml bacterial culture was diluted with 0.5ml SuperBroth; 30% glycerol and stored at -20°C.

2.4.2.2 Bulk preparation

This was prepared essentially as a scaled-up version of the method of Birnboim and Doly (1979) described for small-scale preparations above (Section 2.4.2.1).

Bacteria containing the plasmid of interest were taken from storage (Section 2.4.2.1) and 100µl inoculated into 10ml of SuperBroth (Section 2.4.2.1) supplemented with 100µg/ml ampicillin and grown overnight at 37°C in an orbital shaker. 5ml of overnight culture was then transferred to an additional 500ml of SuperBroth containing 100µg/ml ampicillin and incubated for 36 hours at 37°C in an orbital shaker. Bacteria were then pelleted by centrifugation (4000g, 10min, 4°C), resuspended in 10ml of lysis solution (50mM glucose; 25mM tris-HCl, pH8.0; 10mM EDTA; 5mg/ml lysozyme) and left at ambient temperature for 5min to allow bacterial lysis. 20ml of freshly prepared 0.2M NaOH; 1% w/v SDS was then added, mixed gently and the whole left on ice for 10 min. The solution was then neutralised by addition of 15ml 5M potassium acetate pH4.8, mixed thoroughly and incubated at 0°C for 10 min. Precipitated genomic DNA was then removed by centrifugation (16000g, 10min, 4°C) and plasmid DNA precipitated from the supernatant by addition of 0.6 volumes of isopropanol and incubation at ambient temperature for 15min. After centrifugation (16000g, 10min, room temperature) the pellet was

resuspended in 2ml TE buffer (10mM Tris-HCl pH8.0; 1mM EDTA). RNA was degraded by the addition of 50µl of a boiled solution containing 10mg/ml RNaseA and incubation for 30min at 37°C. Following extraction with an equal volume of 1:1 phenol: chloroform, and ethanol precipitation, the pellet was resuspended in 1ml TE buffer.

DNA was separated from degraded RNA by gel filtration columnn chromatography. After addition of 0.1ml 10x gel loading buffer (Section 2.5.1) as a marker for the rate of migration, the DNA-containing solution was loaded on to a 30cm, 1.5cm diameter Biogel A-50 column equilibrated in TE buffer. The column was run in TE buffer, and 2ml fractions collected (approximately 6-10min per fraction). Fractions containing the RNA-free peak were determined by spectrophotometry at 260 and 280nm (Section 2.2.4) and pooled. The pooled fractions were extracted once with an equal volume of 1:1 phenol: chloroform, and ethanol precipitated twice. The purified DNA was resuspended in water to 1mg/ml and stored at -20°C

2.4.3 Single-stranded M13 viral DNA.

E. coli JM101 cells were transfected as described in Section 2.3 and single, colourless plaques transferred to 1.5ml of a 1:100 dilution in 2xTY (Section 2.3) of JM101 overnight culture, using a disposable plastic pipette tip. Growth was at 37°C with vigorous shaking for 5hours. Cultures wer transferred to 1.5ml microcentrifuge tubes and cells removed by microcentrifugation (12000g, 5min, room temperature). After a second centrifugation to ensure complete removal of cells, the supernatant was decanted into a second tube and 0.2ml 2.5M NaCl; 20% PEG (polyethylene glycol) 6000 added and mixed. Precipitated viral particles were pelleted as above and the supernatant removed. The tube was re-spun for 2min and all remaining traces of PEG supernatant removed. The pellet was redissolved in 0.1ml TE buffer (10mM Tris-HCl pH8.0; 1mM EDTA) and 50µl of TE-saturated phenol added to remove viral proteins. After vortexing for 15sec, the tube was allowed to stand for 15min at room temperature, vortexed for a further 15sec, and centrifuged for 3min. The aqueous phase was removed, and M13 single-stranded DNA recovered using the GeneClean kit as described in Section 2.6.1.1.

2.5 Gel electrophoresis of nucleic acids.

2.5.1 Agarose gel electrophoresis of DNA.

Agarose gel electrophoresis of DNA for either analytical or preparative purposes was performed using a flat bed apparatus. Gels were made from 0.8-2.0% w/v agarose dissolved and cast in buffer: 1x TAE (40mM Tris-HCl pH7.8; 20mM sodium acetate; 1mM EDTA) when preparative isolation of fragments from low melting-point agarose was required; 1x TBE (90mM Tris-HCl; 90mM boric acid; 2.5mM EDTA, pH8.3) in all other instances. Gels were cast by pouring agarose directly into the electrophoresis apparatus with a well-forming comb in place and the electrodes separated from the gel-forming chamber by removable plastic barriers. Once solidified the gel was submerged in the appropriate buffer and plasmid DNA solution (0.1-10µg depending upon well size) containing 1/10th volume of 10x gel loading buffer (50% glycerol; 0.4% w/v bromophenol blue; 0.4% xylene cyanol; 1mM EDTA) loaded into the wells. The current applied was dependant upon the dimensions of the gel and the urgency of the immediate task but was generally in the range of 10-100mA for 0.5-18 h for gels cast and run in 1x TBE buffer; 25 - 200mA for 2-18 h for gels cast and run in 1x TAE buffer. Following electrophoresis, gels were soaked in ethidium bromide solution (3µg/ml) for 10min and excess ethidium bromide removed by soaking the gel in water for 10min. DNA was then visualised by illumination with short wave ultra-violet light and photographed through a red filter using Polaroid type 57 high-speed film.

2.5.2 Agarose gel electrophoresis of RNA.

Separation of RNA for "Northern" transfer was in formaldehyde denaturing gels prepared according to Maniatis et al (1982). Agarose was dissolved in DEPC-treated water by boiling at 100°C. Following cooling to 60°C, 5x running buffer (200mM MOPS (morpholinopropane-sulphonic acid) pH7.0; 50mM sodium acetate; 1mM EDTA) and formaldehyde were added to give a final gel composition of 1.4% w/v agarose; 2.2M formaldehyde; 1x running buffer. Gels were immediately cast as described in Section 2.5.1 and submerged under 1x running buffer. RNA samples were prepared for electrophoresis by lyophilisation and resuspension in 1x running buffer followed by addition of formaldehyde

and formamide to 2.2M and 50% v/v respectively. These mixtures were then heated at 70°C for 15 min, cooled immediately on ice and 1/10th volume of gel loading buffer (Section 2.5.1) added before loading. Electrophoresis was at 150mA with recirculation of buffer between electrode chambers until the dye had travelled 3/4 of the length of the gel. RNA size markers (9.5, 7.5, 4.4, 2.4, 1.4, and 0.3kb) were then stained in 3µg/ml ethidium bromide for 15 min, following separation of the lane containing the RNA markers from the rest of the gel, destained in 1x running buffer for several hours and visualised as in Section 2.5.1 above.

2.5.3 Native polyacrylamide gel electrophoresis.

5% polyacrylamide gels for separation of fragments <500bp in length were prepared by polymerization of a de-gassed solution containing 4.75% w/v acrylamide; 0.25% w/v N,N'-methylene bis-acrylamide (1:19 bis:acryl); 1x TBE pH8.3; 0.2% v/v N,N,N',N'-tetra-methylethylenediamine catalysed by the addition of ammonium persulphate to 0.1% w/v. Immediately after addition of ammonium persulphate, gels were cast by pouring between two glass plates separated by 1-2mm, and a well-former inserted. Once solidifed, the gel was placed on a vertical apparatus with each end submerged in a reservoir of 1x TBE buffer. DNA solution containing 1/10th volume of 10x gel loading buffer (Section 2.5.1) was loaded and gels were run at 25-100mA for 1-4 h. DNA was then visualised as in Section 2.5.1.

Non-denaturing strand-separating gels were prepared and run in a similar manner, except that the polymerised gels contained 5% (1:50 bis:acryl) polyacrylamide; 0.5x TBE, and were run at 100mA for 24-48 hours.

2.5.4 Denaturing polyacrylamide gel electrophoresis.

6% polyacrylmide denaturing gels (1:19 bis:acryl) were prepared as above (Section 2.5.3), except that the gel mix contained 8M urea and the separation of glass plates was only 0.6mm. Samples were denatured by heating at 100°C for 5min in 80% v/v formamide; 0.08% w/v xylene cyanol; 0.08% w/v bromophenol blue then cooled immediately on ice. Gels were run at 1500V for 1.5-5 hours. After running, gels were transferred onto Whatman 3MM paper for drying.

Gels containing ^{35}S labelled DNA were exposed to Kodak X-AR5 film at room temperature; those containing ^{32}P labelled DNA were exposed to Kodak XAR-5 film at -70°C with intensifying screens.

2.6 Purification of DNA restriction fragments

Depending upon the size of the DNA fragment of interest and the size of contaminating DNA fragments, electrophoresis was carried out either through low melting point agarose or polyacrylamide gels (Sections 2.5.1 and 2.5.3 above). Following ethidium bromide staining and visualisation of stained bands, the appropriate region of the gel containing the DNA of interest was excised using a scalpel and the DNA purified by one of the following methods.

2.6.1 From low melting point agarose.

2.6.1.1 Using silica beads.

In most cases, DNA fragments were recovered using the silica matrix suspension provided in the GeneClean kit using a modification of the method of Vogelstein and Gillespie (1979).

The excised agarose slice was dissolved in 2.5 volumes of the sodium iodide solution provided (to an approximate final concentration of 4M NaI) at 55°C for 3min. 5µl of the Glassmilk suspension supplied was added to the resulting solution, followed by incubation on ice for 5min. DNA bound to the silica matrix was pelleted by centrifugation (12000g, 5sec, at room temperature) and the pellet washed by resuspension in 0.4ml of NEW (sodium chloride; ethanol; water) solution, prepared according to the manufacturer's instructions. After 2 more cycles of centrifugation and washing, the final pellet was resuspended in 10µl water and DNA eluted from the silica matrix by incubating at 55°C for 3min. The silica matrix was pelleted by centrifugation as above for 30sec and the DNA solution removed to a fresh tube. A second cycle of elution was performed on the silica matrix pellet, and the two eluates pooled. The DNA solution was stored at -20°C until required.

2.6.1.2 Using Elutip-d mini-columns.

End-labelled fragments for DNaseI footprinting (Section

2.1.2) were recovered using Schleicher and Schull Elutip-d mini-columns.

The gel slice containing the radio-labelled fragment was melted in an equal volume of solution I (0.2M NaCl; 20mM Tris-HCl pH7.5; 1mM EDTA) by heating at 70°C, and diluted to 10ml in Solution I at 42°C. The Elutip-d mini-column was hydrated by washing through with 3ml of solution II (1M NaCl; 20mM Tris pH7.5; 1mM EDTA) using a disposable plastic syringe, then equilibrated with 5ml of solution I at 42°C. The molten agarose solution was then passed slowly through the column and agarose removed by washing with a further 5ml of solution I at 42°C. Finally, radio-labelled DNA was eluted from the column with 0.6ml of solution II and ethanol precipitated with 30µg yeast tRNA as carrier.

2.6.2 From polyacrylamide.

DNA fragments smaller than about 400bp were eluted from the gel slice in 0.5ml 0.5M ammonium acetate; 1mM EDTA at 37°C overnight.

DNA fragments larger than about 400 b.p. were recovered by electro-elution. The gel slice was placed in a piece of dialysis tubing containing 3ml 0.1x TAE buffer and the sealed dialysis tube placed in an electrophoresis tank containing 0.1x TAE buffer. Elution of DNA was for 3hours at 100mA (600V). During this period the DNA moves to the side of the dialysis tube, so a reversal of the current was applied for 3 min at the end of the electro-elution period. DNA was ethanol precipitated with 10µg yeast tRNA as carrier, and resuspended in water.

2.7 Manipulation of DNA

2.7.1 Digestion of DNA with restriction endonucleases.

Plasmid DNA was digested with 5-10 units enzyme/µg DNA for 1-3 hours under conditions specified by the supplier. In most cases, digests were phenol:chloroform extracted and ethanol precipitated (Sections 2.2.2 and 2.2.3) before proceeding to the next step.

2.7.2 Conversion of 5′ single-stranded regions of duplex DNA to double-stranded DNA.

Blunt-ended DNA for cloning was obtained by end-filling in the presence of deoxyribonucleotide triphosphates (dNTPs) using Klenow fragment of E. coli DNA polymerase I. DNA was incubated at 14°C for 2hours in 20µl of 10mM Tris-HCl pH7.9; 60mM NaCl; 7mM magnesium sulphate; 0.5mM β-mercaptoethanol; 0.5µM each dNTP and 2 units Klenow fragment.

2.7.3 Ligation of DNA fragments into vectors.

Ligations were performed in 66mM Tris-HCl pH7.5; 1mM spermidine; 5mM $MgCl_2$; 5mM DTT; 1mM ATP overnight at 4°C using 1 unit of T4 DNA ligase. Reaction volumes (usually 10µl) were such that the concentration of 5′ ends was 0.1-1.0µM. The molar ratio of vector to insert was generally in the range of 1:5 to 1:20 (in blunt-end ligations a greater proportion of insert was used to reduce religation of vector).

2.7.4 Removal of 5′ terminal phosphate groups.

In order to reduce the number of bacterial colonies containing non-recombinant plasmid molecules during sub-cloning, and as a prelude to 5′ kinase end-labelling, the 5′ phosphate groups of linear double-stranded DNA were removed by treatment with calf intestinal alkaline phosphatase (CIP) according to Maniatis et al (1982).

At the end of a restriction enzyme digestion, the volume of the DNA solution was adjusted to 45µl, and 5µl 10x CIP buffer (0.5M Tris-HCl pH9.0; 10mM $MgCl_2$; 1mM $ZnCl_2$; 10mM spermidine) added. A ten-fold excess (0.1 units per pmole 5′ ends) CIP (1 unit/µl) was added to ensure complete dephosphorylation. For dephosphorylation of 5′ overhangs, incubation was for 30min at 37°C, followed by addition of a second aliquot of enzyme and a second 30-minute incubation. For blunt ends and 3′ overhangs, incubation was at 37°c for 15min, then 56°C for 15min before addition of the second aliquot of enzyme and a second cycle of 37 and 56°C incubations. The dephosphorylation reaction was terminated by addition of 50µl water and SDS to 0.5% w/v, followed by heating at 68°C for 15min. CIP was removed by two rounds

of 1:1 phenol: chloroform extraction, and dephosphorylated DNA was recovered by ethanol precipitation.

2.7.5 Labelling of DNA with radio-isotopes.

2.7.5.1 Random priming using Klenow polymerase.

Double-stranded fragments for probing DNA and RNA blots (Section 2.11) were labelled using mixed hexadeoxyribonucleotide primers of random sequence essentially as described by Feinberg and Vogelstein (1983 and 1984).

An aqueous solution containing 50ng of DNA was boiled at 100°C for 7min then labelled in 50µl of a solution containing 50mM Tris-HCl pH8.0; 5mM $MgCl_2$; 10mM β-mercaptoethanol; 4mM each of dATP, dGTP and dTTP; 0.2M HEPES [N-2-hydroxyethyl-piperazine-N'-2'ethane sulphonic acid] pH6.6; 110µg/ml mixed hexadeoxynucleotides; 0.4mg/ml bovine serum albumin. Labelling was with 1.85×10^6Bq (1.1×10^{14}Bq/mmol) of [α-^{32}P]dCTP using 10 units Klenow enzyme (labelling grade). Incubation was for 2.5hours at room temperature. The reaction was stopped by the addition of 34µl of 0.25M NaCl; 0.25M Tris-HCl pH7.5; 25mM EDTA; 3% w/v SDS. Labelled products were ethanol precipitated with 10µg yeast tRNA as carrier for 30min at -20°C. The labelled pellet was resuspended in water, heated at 100°C in boiling water for 5min and rapidly cooled on ice before addition to the hybridisation solution (Section 2.1.1.4).

2.7.5.2 Uniform labelling of S1 probes.

Single-stranded, uniformly-labelled probes for S1 analysis of CAT and lacZ mRNAs were obtained essentially according to the method of Bentley (1984).

Fragments were ligated (Section 2.7.3) into M13 mp19 vector. The orientation of the fragments was such that primer extension on the M13 + strand template would generate probes complementary to the mRNA being analysed. Single-stranded M13 templates were generated as in Section 2.4.3. One quarter of each template preparation (5µl) was added to 1µl (2µg) of M13 Universal primer and 1µl of 10x Klenow buffer (100mM Tris-HCl pH8.0; 50mM $MgCl_2$) and annealed by heating to 75°C in a beaker of water and then allowing to cool slowly to room

temperature. Nucleotides were added to the annealed mixture to final concentrations of 100μM unlabelled dATP, dGTP and dTTP, 50μM unlabelled dCTP and 3μM [α-^{32}P]dCTP (i.e $4.6x10^6$Bq of $1.1x10^{14}$Bq/mmol; $3.7x10^8$Bq/ml, dried down and resuspended in 2.9μl of water) in a total volume of 14μl. Primer extension from the M13 Universal primer was catalysed by 2units E. coli Klenow fragment poymerase (sequencing grade) at room temperature for 15min. The reaction was terminated by heating to 60°C for 10min then probe released from the duplex by restriction enzyme cleavage at 37°C for 1hour, using 2μl of the appropriate 10x buffer and 20units of enzyme in a final volume of 20μl. Products were ethanol precipitated with 10μg yeast tRNA as carrier, resuspended in 4μl of water and an equal volume of formamide sequenceing dye mix (Section 2.10.2.2) added. The probe was denatured by heating at 100°C in a boiling water-bath for 5min, followed by rapid cooling on ice, then isolated on a 6% denaturing polyacrylamide (1:19 bis:acryl); 8M urea; 1x TBE pH8.3 gel (Section 2.5.4). Recovery of labelled probe from the polyacrylamide gel slice was by elution in 0.2ml of 0.5M ammonium acetate; 10mM magnesium acetate; 0.1mM EDTA; 0.1% w/v SDS at 37°C overnight, followed by two rounds of ethanol precipitation of the eluate using 10μg tRNA as carrier.

2.7.5.3 Kinase 5′ end-labelling DNA.

Labelling of probes for primer extension and S1 analyses, and for DNaseI footprinting was performed as described by Maniatis et al (1982).

Purified CIP-treated (Section 2.7.4) DNA (<20 pmoles 5′ ends) was heated in 40μl of 1mM spermidine; 0.1mM EDTA; 20mM Tris-HCl pH9.5 at 70°C for 5min, cooled immediately on ice and left for 5min. before addition of 5μl of 0.5M Tris pH9.5; 0.1M $MgCl_2$; 50mM dithiothreitol; 50% glycerol and $3.7x10^6$Bq (20pmoles) of $1.85x10^{14}$Bq/mmol [γ-^{32}P]ATP. 20 units (2μl) of T4 polynucleotide kinase were then added and the mixture incubated at 37°C for 30min. The reaction was stopped by addition of 2μl 0.5M EDTA and DNA precipitated with 10μg yeast tRNA as carrier.

Kinase 5′ end-labelled strands for use in S1 and primer extension analyses were separated either on denaturing or non-denaturing

polyacrylamide gels (Sections 2.5.3 and 2.5.4), depending upon whether the two strands were of different lengths.

2.8 S1 nuclease protection analysis.

This was performed essentially according to the method of Berk and Sharp (1977), as modified by Weaver and Weissmann (1979).

Kinase 5' end-labelled single-stranded probe (Section 2.7.5.3) was heated at 100°C in 10µl 80% v/v formamide for 2 min and cooled rapidly on ice. RNA was precipitated, dried and the total amount of RNA made up to 50µg with yeast tRNA where appropriate. 1µl of 4M NaCl; 0.4M PIPES [Piperazine-N,N'-bis (2-ethanesulphonic acid)] pH6.4; 10mM EDTA was added to the RNA, dried and the whole resuspended in the 10µl of probe/formamide mix and hybridisation performed at an appropriate temperature (usually 52°C) overnight in sealed glass micro-capillary tubes. Hybridised RNA:DNA hybrids were expelled into 250µl 250mM NaCl; 1mM Zn SO_4; 30mM sodium acetate, pH4.6. S1 nuclease (usually 60 units) was then added and the mixture incubated at 37°C for 1hour. Products were then ethanol precipitated, and resuspended in water. Following addition of 4 volumes of formamide sequencing dye mix (Section 2.10.2.2), samples were denatured by heating at 100°C in boiling water for 5min, and electrophoresed through a 6% denaturing polyacrylamide (1:19 bis:acryl); 8M urea; 1x TBE pH8.3 gel (Section 2.5.4).

2.9 Primer extension analysis.

RNA:DNA hybrids were set up as for S1 nuclease analysis above (Section 2.8). After hybridisation overnight, hybrids were expelled into 100µl of ice-cold 03M sodium acetate and ethanol precipitated. Hybrids were then resuspended in 50µl 100mM Tris-HCl pH8.3; 10mM $MgCl_2$; 140mM KCl; 20mM β-mercaptoethanol; 1mM each dNTP and primer extension performed using 10units AMV (Avian Myeloblastosis Virus) reverse transcriptase at 42°C. 10µl 1M NaOH was then added to degrade RNA and incubation at 42°C continued for a further 1hour. Neutralisation was achieved by addition of 10µl of 1M HCl and products ethanol precipitated. Primer extension products were pelleted by microcentrifugation (12000g, 10min, room temperature), resuspended in water and 4 volumes of formamide dye mix (Section 2.10.2.2) added.

Products were denatured by heating at 100°C in boiling waterfor 5min, cooled rapidly on ice, then resolved on a 6% denaturing polyacrylamide (1:19 bis:acryl); 8M urea; 1xTBE pH8.3 gel.

2.10 Sequencing of nucleic acids.

2.10.1 mRNA sequencing.

mRNA sequencing was performed exactly as for primer extension (Section 2.9), but reverse transcription was performed in the presence of one of the four dideoxynucleotide triphosphates (ddNTPs) in each of four reactions. ddNTP concentrations were varied from experiment to experiment in the following ranges: ddATP 0.1-0.8mM; ddCTP 0.05-0.8mM; ddGTP 0.05-0.4mM; ddTTP 0.2-1.5mM

2.10.2 DNA sequencing.

Sub-cloning of fragments for sequencing, preparation of single-stranded templates and sequencing by the dideoxy chain-termination method of Sanger *et al* (1977) were performed essentially as described in the M13 cloning and sequencing manual provided by Amersham.

2.10.2.1 Sub-cloning of fragments for sequencing.

Fragments to be sequenced were ligated into linearised double-stranded, replicative form M13 mp10, mp18 and mp19 viral vectors (Messing and Vieira, 1982; Norrander *et al*, 1983; Yanish-Perron *et al*, 1985) as described in Section 2.7.3).

9pmoles of fragment were circularised for 90min at 4°C in 50µl of 66mM Tris-HCl; 5mM $MgCl_2$; 5mM dithiothreitol; 1mM ATP using 2 units of T4 DNA ligase to try to ensure random generation of fragments during sonication. Fragments for shot-gun sequencing were then generated by sonication (six 10sec bursts at low power - 10µ amplitude - using an MSE Soniprep 150 sonicator). Sonicated fragments were size-fractionated by agarose gel electrophoresis, and fragments within the range 200-400bp recovered as in Section 2.6.1.1, and end-repaired as in Section 2.7.2 before cloning into SmaI-cut vector.

2.10.2.2 Dideoxy chain-termination sequencing.

Sequencing of single-stranded M13 viral DNA templates, prepared as in Section 2.4.3, was performed according to the method of Sanger et al (1977) in the presence of [α-^{35}S]dATPαS (Biggin et al, 1983).

7.5µl (approx. 0.4pmoles; 1/4 of the total) single-stranded template was annealed with 1µl of M13 Universal primer (2µg/ml; 5'-GT-AAAACGACGGCCAGT-3') in a total volume of 10µl of a 15mM Tris-HCl pH8.0; 7.5mM $MgCl_2$ solution by incubating at 60°C for 2 hours. 1.5µl of 2.2×10^{13}Bq/mmol; 3.7×10^{8}Bq/ml [α^{35}-S]dATPαS (25pmoles) and 1µl (1 unit) E. coli Klenow polymerase were then added and incorporation of radio-label into a complementary strand to the M13 template was initiated by adding 2.5µl of the resulting mixture to each of four tubes containing 2µl of appropriate mixtures of deoxyribonucleotides (dNTPs) and dideoxyribonucleoties (ddNTPs) as follows:

Tube A (dATP°/ddATP) 2.5mM Tris-HCl pH8.0; 0.25mM EDTA; 0.12mM dCTP; 0.12mM dGTP; 0.12mM dTTP; 0.05mM ddATP.

Tube C (dCTP°/ddCTP) 3.3mM Tris-HCl pH8.0; 0.33mM EDTA; 8.3µM dCTP; 0.17mM dGTP; 0.17mM dTTP; 0.05mM ddCTP.

Tube G (cGTP°/ddGTP); 3.3mM Tris-HCl pH8.0; 0.33mM EDTA; 0.17mM dCTP; 8.3µM dGTP; 0.17mM dTTP; 0.02mM ddGTP.

Tube T (dTTP°/ddTTP) 3.3mM Tris-HCl pH8.0; 0.33mM EDTA; 0.17mM dCTP; 0.17mM dGTP; 8.3µM dTTP; 0.25mM ddTTP.

(Where appropriate, sequencing reactions were fine-tuned to achieve incorporation into larger or smaller products by adjustment of the quantities of ddNTPs)

After incubation at room temperature for 20min, 2µl of chase solution (0.5mM each dNTP) was added and incubation continued for a further 15min at room temperature. The sequencing reaction was terminated by addition of 4µl of formamide dye mix (95% formamide; 10mM EDTA; 0.1% w/v bromophenol blue; 0.1% xylene cyanol).

Products of the four sequencing reactions were denatured by heating at 100°C in boiling water for 5min, cooled rapidly on ice and resolved on a 6% denaturing polyacrylamide (1:19 bis:acryl); 8M urea; 1x TBE pH8.3 gel.

2.10.2.3 Computer analysis of sequences.

Sequence analysis was performed on an IBM PC-AT microcomputer using "MicroGenie" software (Beckmann).

2.11 DNA:DNA and DNA:RNA hybridisation techniques.

2.11.1 Capillary transfer from agarose gels by Southern blotting.

DNA fragments separated by agarose gel electrophoresis (Section 2.5.1) were transferred to Biodyne Membranes by the method of Southern (1975) prior to hybridisation (Section 2.11.4) with radio-labelled DNA probes prepared as in Section 2.7.5.1. Following ethidium bromide staining and photography, DNA was denatured by rinsing the gel twice in 0.5M NaOH; 1.5M NaCl for 30min each, followed by neutralisation in 3M sodium acetate pH5.5 for 1 hour. The gel was then transferred to a raised platform covered with two sheets of Whatman 3MM paper soaked in 20xSSC (3M NaCl; 0.3M sodium citrate pH7.0) such that the ends of the filter paper extended below the platform into a reservoir of 20xSSC. Care was taken to ensure that there were no visible bubbles between the filter paper and the gel. The complete assembly around the gel was then sealed off from the space above using plastic sheets to ensure that all movement of liquid occurred through the gel. A sheet of dry Biodyne membrane was then placed in direct contact with the gel, without any intervening bubbles. Two sheets of Whatman 3MM filter paper (pre-soaked in 2xSSC) were then placed directly on top of the membrane. Finally, the contents of a box of tissues were placed on top of the filter paper and compressed with a 1kg weight. The complete assembly was left overnight to allow transfer of DNA. The position of the gel wells was marked on the blotted membrane. Completion of transfer of DNA from the gel was checked by re-staining the gel in ethidium bromide. The filter was air dried, baked at 80^{o}C for 2-3 hours and then stored dry in a sealed plastic bag at room temperature until required.

2.11.2 Electro-blotting.

Short DNA restriction fragments (<500bp) separated in polyacrylamide gels were transferred to Biodyne membranes by electro-blotting.

Fragments were resolved on a 5% non-denaturing polyacrylamide gel (1:19 bis:acryl), stained in ethidium bromide solution and bands visualised as in Section 2.5.1. DNA was then denatured by soaking in 0.2M NaOH; 0.5M NaCl for 30min, followed by neutralisation by washing twice in 5x TBE pH8.3 buffer (Section 2.5.1) for 10min each. The gel was then placed in contact with a sheet of pre-cut, dry Biodyne membrane between two sheets of Whatman 3MM filter paper soaked in 0.5x TBE. This assembly was inserted upright in the electro-blot apparatus, between two sponge pads held in a plastic holder, such that the membrane was on the anode side of the gel. Electrophoretic transfer in 0.5x TBE was for 14hours at 250mA (30V), after which the filter was baked at 80^{o}C for 2hours. Successful transfer was confirmed by re-staining the electro-blotted gel in ethidium bromide solution.

2.11.3 Northern blotting.

RNA separated in denaturing agarose gels (Section 2.5.2) was transferred to nitrocellulose as described in Section 2.11.1, except that no pretreatment of the gels was performed before assembly into the transfer apparatus.

2.11.4 Hybridisation of nucleic acids immobilised on filters.

Blotted filters were wetted in 2xSSC and transferred to plastic bags. Prehybridisation buffer (50% v/v formamide; 5xSSC; 5x Denhardt's solution [100x Denhardt's is 2% bovine serum albumin; 2% polyvinylpyrollidone; 2% Ficoll pH8.0]; 10µg/ml polyA; 10µg/ml polyC; 0.25 mg/ml denatured salmon sperm DNA; 0.1% w/v SDS) was then added, the bag heat-sealed and incubated overnight at 42^{o}C with shaking. Prehybridisation buffer was then removed and replaced with hybridisation buffer (50% formamide; 10% w/v dextran sulphate; 5x SSC; 1x Denhardt's solution; 10µg/ml polyA; 10µg/ml polyC; 100µg/ml salmon sperm DNA; 0.1% w/v SDS) containing denatured radio-labelled probe of specific activity approx. 2x10^{9}cpm/µg (prepared as in Section 2.7.5.1) such that the probe concentration was 10-20ng/ml. Bags were then sealed and incubated at 42^{o}C overnight. Filters were then washed three times in 2xSSC; 0.1% w/v SDS at room temperature for 5min. and then twice in 0.1xSSC; 0.1% w/v SDS at 65^{o}C for 30 min. Filters were

air dried, wrapped in thin plastic and exposed to Kodak XAR-5 film at -70°C with intensifying screens.

2.12 In vitro DNaseI digestion.

DNaseI "footprinting" was performed as described by Plumb and Goodwin (1987). Plasmid DNA was linearised by restriction endonuclease cleavage, phosphatased (Section 2.7.4) and kinase 5' end-labelled (Section 2.7.5.3). An appropriate double-stranded fragment, end-labelled on only one strand was obtained by a secondary restriction digestion at an appropriate site, isolated by agarose gel electrophoresis (Section 2.5.1), localised by autoradiography and recovered as in Section 2.6.1.2.

Various amounts (0-30µg protein) of nuclear protein extract, prepared as in Section 2.15, were incubated at 4°C for 90min. with approximately 2ng of end-labelled restriction fragment in 100µl of 50mM NaCl; 200mM HEPES (N-2-hydroxyethyl-piperazine-N'-2-ethane sulphonic acid) pH7.9; 5mM $MgCl_2$; 0.1mM EDTA; 20% glycerol; 1mM EDTA; 1µg/ml poly(dI:dC):(dI:dC). DNase I digestion (0.5-4µl of 0.2mg/ml enzyme in 50mM NaCl; 200mM HEPES pH7.9; 5mM $MgCl_2$; 0.1mM EDTA; 0.1mM $CaCl_2$ for DNA-protein complexes; 2-4µl of 0.02mg/ml enzyme for zero-protein controls) was for 30sec at room temperature. The reaction was terminated by addition of 100µl of 80mM Tris-HCl pH8.0; 8mM EDTA; 1% w/v SDS; 250mM NaCl; 0.4mg/ml proteinase K; 70µg/ml yeast tRNA and incubation at 37°C for 30min, followed by heating to 90°C for 2min. Products were extracted once with 1:1 phenol: chloroform and once with chloroform, then precipitated at -70°C for 10min following addition of 1/16th volume 4M LiCl and 3 volumes absolute ethanol. The pellet was washed with absolute ethanol, dried under vacuum, and resuspended in formamide sequencing dye mix (Section 2.10.2.2). Products were heated to 90°C for 2min., cooled rapidly on solid CO_2, vortexed, heated and cooled again, then resolved on a 6% denaturing polyacrylamide (1:19 bis:acryl); 8M urea; 1x TBE pH8.3 gel and visualised by autoradiography.

2.13 Eukaryotic cell culture.

2.13.1 Culture of cells.

All cells were obtained from Beatson Institute stocks.

The adherent murine erythroleukaemia (MEL) Friend cell line F4-12B2 and murine STO fibroblast line were grown in Special Liquid Medium (SLM) supplemented with 4mM glutamine; 10% v/v bovine foetal serum; 37.5µg/ml penicillin; 100µg/ml streptomycin in 5% CO_2/95% air at 37°C. Cells were grown as adherent monolayers in sterile plastic flasks until subconfluent, then harvested and subcultured at 1.7-3x10^4 cells/cm^2.

To remove adherent cells from culture flasks, medium was first removed by aspiration and the cells washed once in PBS (phosphate buffered saline: 0.14M NaCl; 27mM KCl; 10mM Na_2HPO_4; 15mM KH_2PO_4. Trypsin solution (0.025% w/v in citrate buffer, pH7.8) was then added to the monolayer and cells dislodged by briefly rinsing cells. Cells were then mixed with an equal volume of medium to inactivate trypsin and an aliquot removed for determination of cell numbers using a Coulter counter. Appropriate dilutions of the cells were made in SLM plus supplements (as above) following pelleting of cells by centrifugation at 1500rpm for 5min at room temperature in sterile plastic Universal containers.

All cell lines were examined for the presence of mycoplasma using the Hoechst 33258 staining method of Chen (1977). Only cells in which mycoplasma was undetectable were used for experimentation. Any cell stocks in which mycoplasma was detected were removed from the Institute and destroyed.

2.13.2 Storage of cells in liquid nitrogen.

Cells were trypsinised, pelleted and resuspended at 0.5-1x10^7 cells/ml in medium containing 10% v/v dimethylsulphoxide. 1ml aliquots were placed in plastic 1.5ml Nunc tubes and these were wrapped in cotton wool inside a polystyrene box and placed at -70°C overnight. (Cooling of cells is at approximately 1°C per minute under these conditions). Frozen cells were then stored in liquid nitrogen.

Recovery from storage was effected by rapid thawing at 37°C. Thawed cells were then washed in 10ml of DMSO-free medium, harvested

as above, and transferred to a small flask of DMSO-free medium.

2.14 Assay of transient expression of exogenous DNA in cultured cells.

2.14.1 DNA mediated gene transfer into eukaryotic cells.

Adherent cells were transfected essentially according to the calcium phosphate co-precipitation method of Graham and van der Eb (1973) as modified according to Gorman (1985).

24 hours before transfection, cells were harvested and counted as in Section 2.13.1, and plated out at 10^6 cells per 9cm Petri dish ($1.6x10^4$cells/cm^2 in 10ml medium and incubated in an atmosphere of 5% CO_2/95% air at 37°C.

For each precipitate, 2.4pmoles of the CAT recombinant plasmid under test (11-20µg, depending on size) and 4pmoles (20µg) of internal control plasmid, pHSV-βgal (Fig. 17) were added to 1ml of 1mM Tris-HCl pH8.0; 0.01mM EDTA; 0.25M $CaCl_2$, and pUC18 plasmid DNA added to a total of 40µg DNA. This mixture was then added slowly to 1ml of 2xHBS (280mM NaCl; 50mM HEPES [N-2-hydroxyethylpiperazine-N'-2'-ethane sulphonic acid]; 1.3mM $Na_2HPO_4.2H_2O$ pH7.1) using a sterile plastic pipette tip, and aerated by pipetting up-and-down vigorously. Precipitates were allowed to stand for 30min at room temperature then half of each precipitate (1ml; 20µg DNA) was added to a dish of each of the two cell types being transfected.

Cells were returned to the incubator for 18-24hours before removing the precipitate and medium, replacing with 10ml fresh medium, and returning to the incubator for a further 18-24hours.

2.14.2 Harvesting transfected cells and preparation of soluble protein extracts

Cells were harvested approximately 48hours after addition of calcium phosphate:DNA co-precipitates. Medium was removed by aspiration and cells were rinsed briefly with 10ml PBS. The cells were then scraped off the dish surface in 1ml PBS using a disposable spatula, transferred to a microcentrifuge tube and pelleted by centrifugation (500g, 30sec, room temperature). The cell pellet was then resuspended in 200µl of 250mM Tris-HCl pH7.8 by vortexing.

Cells were disrupted by three freeze-thaw cycles, alternating between -70°C for 5 min and 37°C for 5 min. After pelleting debris by microcentrifugation at 12000g for 2min at room temperature, the supernatant was removed to a fresh tube and stored at -20°C.

2.14.3 Determination of protein concentration of extracts.

This was determined by the method of Bradford (1976) using prepared reagents purchased from Bio-Rad.

The absorbance caused by complex formation between 0.9ml of Bio-Rad reagent and protein from an aliquot of extract (the equivalent of 1µl of extract made up to 100µl in water) was measured spectrophotometrically at 595nm and compared with that due to a range of concentrations of bovine γ-globulin prepared at the same time. Absorbances were measured within 15min of mixing of the reagents to avoid distortion of results due to formation of a turbid precipitate.

2.14.4 β-galactosidase assay.

Transfected cell lysates, prepared as in Section 2.14.2 were assayed for β-galactosidase activity as internal control for transfection efficiency essentially as described by Gorman (1985).

1ml of 60mM Na_2HPO_4; 40mM NaH_2PO_4; 10mM KCl; 1mM $MgCl_2$; 50mM β-mercaptoethanol and 0.2ml of 60mM Na_2HPO_4; 40mM NaH_2PO_4; 2mg/ml ONPG (o-nitrophenyl-β-galactopyranoside) were mixed in plastic spectrophotometer cuvettes and the reaction commenced by addition of 60µl (from a total of 200µl) of transfected cell lysate. After incubation at 37°C for 1hour, the reaction was terminated by addition of 0.25ml 1M Na_2CO_3, and ONP (o-nitrophenol) production measured spectrophotometrically at 420nm (E_{420} = 2.13x10^4 ltr.$mole^{-1}$.cm^{-1}).

2.14.5 Measurement of chloramphenicol acetyltransferase (CAT) activity

Aliquots of cell extracts prepared as detailed in Section 2.14.2 were diluted to 89µl with 250mM Tris-HCl pH7.8 such that the same amount of β-galactosidase activity (as determined in Section 2.14.4) was taken from each of the extracts prepared from a given cell line in a given experiment. These were then heated at 60°C for 7mins. to denature proteins other than CAT which utilise acetyl CoenzymeA as

a substrate. Assays were then performed by the addition of 11µl of a mixture consisting of 1 volume of D-threo-(dichloro-acetyl-1-^{14}C)-chloramphenicol ($2x10^9$Bq/mmol; $7.4x10^6$Bq/ml) and 10 volumes 10mM acetyl CoA to each sample and incubation at 37°C for 30 min. Reactions were terminated by the addition of 300µl ethyl acetate and vortexing. Phase separation was by microcentrifugation (12000g, 2min, room temperature) after which the organic phase was removed to a fresh tube and dried down on a Savant speedivac concentrator (20 min).

The mixture of unacetylated and acetylated chloramphenicol was then dissolved in 20µl ethyl acetate, transferred to plastic-backed thin layer chromatography (TLC) plates and chromatography performed using 95% v/v chloroform; 5% v/v methanol as the mobile phase. TLC plates were then exposed to Kodak XAR-5 at room temperature. Quantitation of CAT activity in each extract was achieved by separately counting the radioactivity in the unacetylated and acetylated spots, cut from the TLC plate, in an LKB 1217 Rackbeta liquid scintillation counter.

2.14.6 Correction of CAT activities for β-galactosidase activity.

As described above, CAT assays in a given experiment were performed such that amounts of extract from a given cell line used in the CAT assay contained equivalent amounts of β-galactosidase activity. However, different β-galacosidase activities were used for different cell lines and for different experiments. So, to correct all results to the same level of β-galactosidase activity, results obtained by scintillation counting were adjusted so that they represented the CAT activity present in that amount of extract which produced $1x10^{-8}$moles ONP under the standard conditions of the β-galactosidase assay (Section 2.14.4). The number of moles produced can be calculated from the concentration of ONP in 0.85ml of the stopped reaction, as determined from the OD_{420} and extinction co-efficient of ONP ($A = Ecl$, where: A is the OD_{420}; E is the extinction co-efficient of ONP ($2.13x10^4$ltr.mole^{-1}.cm^{-1}; c is the concentration of ONP; and l is the path length of the spectrophotometric measurement (1cm)).

2.15 Preparation of nuclear protein extracts from eukaryotic cells.

Nuclear extracts for DNaseI protection studies, kindly provided by M. Plumb, H. Wainwright and K. Macleod, were prepared essentially according to Plumb et al (1985).

Cells were lysed in Solution A (10mM Tris-HCl pH7.5; 5mM $MgCl_2$; 0.25M sucrose; 0.2% w/v Triton X-100; 0.5mM PMSF (phenylmethyl sulphonyl fluoride) and nuclei obtained by centrifugation (1600g, 30min, 4°C) after filtration through three layers of gauze. The nuclei were washed twice with solution A, then resuspended at 5-10mg/ml in 10mM Tris-HCl pH7.5; 5mM $MgCl_2$; 0.25M sucrose; 0.5mM PMSF at 4°C. 4M NaCl at 4°C was added drop-wise to a final concentration of 0.3M, followed by incubation at 37°C for 10min. After centrifugation (4000g, 10min, 4°C), the supernatant was kept at 4°C while the extraction procedure was repeated on the pellet. The two final supernatants were then pooled and clarified by centrifugation at 100000g for 60min at 4°C. Partial purification and concentration of nuclear proteins was achieved by addition of solid ammonium sulphate to a concentration of 0.35mg/ml, followed by centrifugation (10000g, 30min, 4°C) and resuspension of pelleted proteins in E_{50} buffer (50mM ammonium sulphate; 20mM HEPES (N-2-hydroxy-ethyl-1-piperazine-N'-2-ethane sulphonic acid) pH7.9; 5mM $MgCl_2$; 0.1mM EDTA; 0.1% Brij35; 1mM dithiothreitol; 20% v/v glycerol. The suspension was then dialysed overnight against E_{50} buffer at 4°C and the dialysate clarified by centrifugation (100000g, 60min, 4°C). The "crude" nuclear protein extracts were stored at -20°C, until required.

Chapter 3.

RESULTS.

3.1 SEQUENCE SURROUNDING THE ATG TRANSLATION INITIATION CODON OF THE RBC 15-LOX GENE.

The approximate location of the first protein-coding exon within the rabbit RBC 15-LOX genomic clone, λ8, has been determined by hybridisation of labelled pGF6 cDNA insert, which accurately predicts the first 30 N-terminal residues of the rabbit protein, to Southern blots of various restriction digests of the genomic clone (Thiele et al, 1987). The 5′ end of the cDNA insert of pGF6 hybridises to a 276bp SstI fragment contained within a 1.5kb EcoRI fragment of the genomic clone (Fig. 3).

The polarity of the RBC 15-LOX genomic clone has been determined by comparing a detailed restriction map of this region with the restriction map of pGF6 cDNA insert, determined from its nucleotide sequence. The same order of restriction sites (5′-TaqI-BalI-SstI-3′) seen in both clones indicates that the 276bp SstI fragment lies within the 5′ half of the 1.5kb EcoRI genomic fragment (Fig. 3).

A 633bp SmaI fragment containing the 276bp SstI fragment was isolated from the 1.5kb EcoRI genomic fragment and sequenced according to the strategy shown in Figure 3. Appropriate double-stranded fragments were sub-cloned into double-stranded replicative form M13 virus (mp10, mp18 or mp19) (Messing and Vieira, 1982; Norrander et al, 1983; Yanisch-Perron et al, 1985), and single-stranded M13 templates were sequenced by the dideoxy. chain termination method of Sanger et al (1977). Sub-clones containing the entire 633bp SmaI fragment permit sequence to be established from either end of the fragment, while the central region of the SmaI fragment was sequenced from sub-clones containing the 276bp SstI fragment. However, there are no convenient restriction sites available for cloning into M13 the region spanning the more 5′ SstI site. To be sure that this SstI site is not, in fact, two closely-spaced sites, it was decided to perform "shot-gun" cloning of sonicated, end-filled fragments of the 633bp SmaI fragment into SmaI-cut M13mp10 (Fig. 3). Thus, all regions have been sequenced on both strands, and almost the entire sequence has been determined from a minimum of three independent M13 sub-clones.

Figure 3. Strategy for sequencing genomic sequences flanking the translation initiation site.

The genomic recombinant clone λ8 was originally obtained by screening of a rabbit genomic library with the cDNA insert of pGF6. The region of hybridisation is shown at the top of the figure. An incomplete exon map of the genomic clone has been obtained by hybridisation of other RBC 15-LOX cDNA clones with the genomic clone. A 276bp SstI fragment within a 633bp SmaI fragment contains a cluster of TaqI, BalI and SstI sites identical to that seen in the pGF6 cDNA insert. The position of the ATG codon relative to these sites, and of the transcription initiation site, as determined in Section 3.2 are shown. Sequencing was performed on fragments sub-cloned into double-stranded M13 RF viral DNA.

Restriction sites: R = EcoRI; K = KpnI; Ss = SstI; X = XbaI; H = HindIII; Sm = SmaI; Sa = Sau3A; T = TaqI; B = BalI. Shaded boxes are exons.

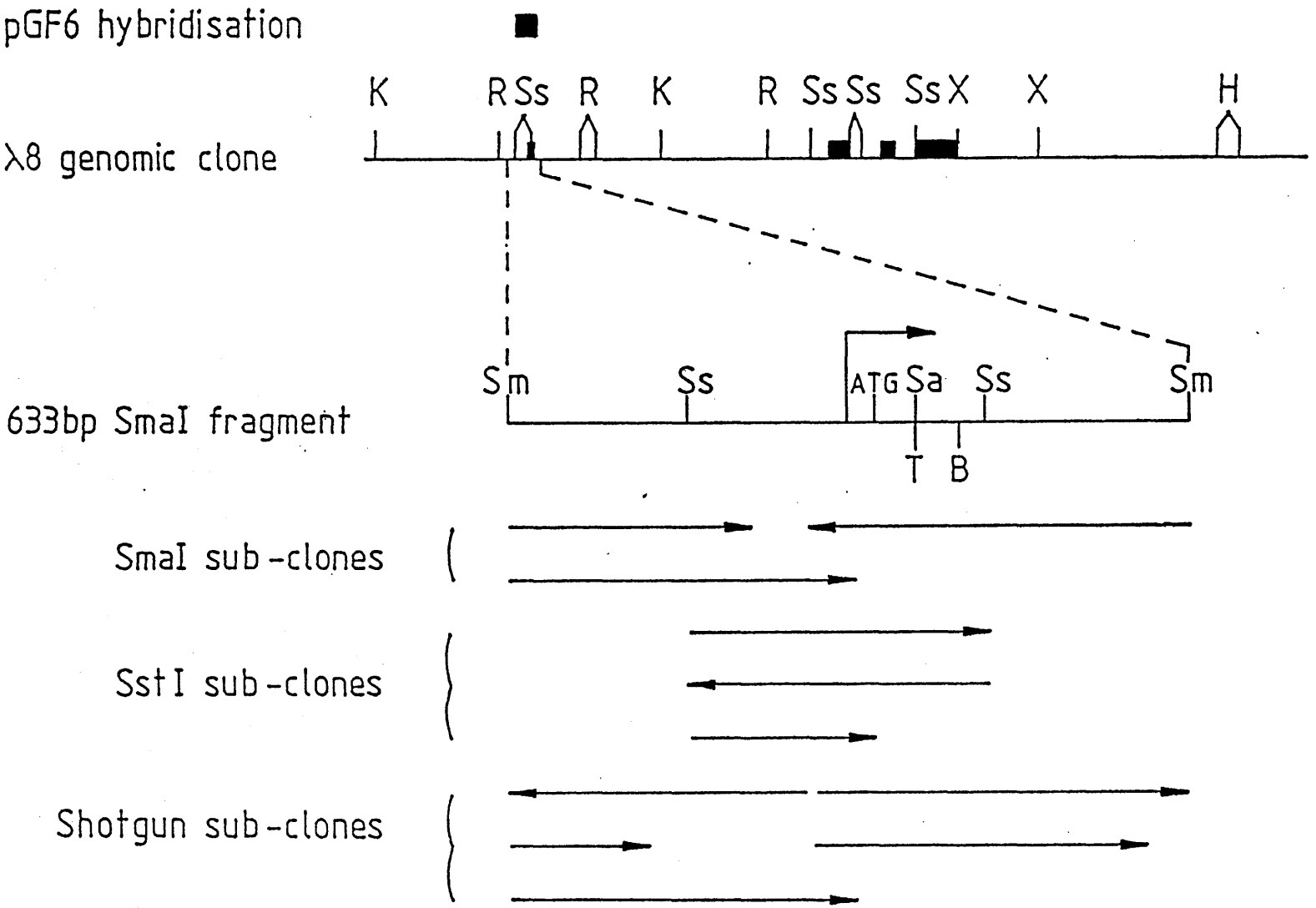
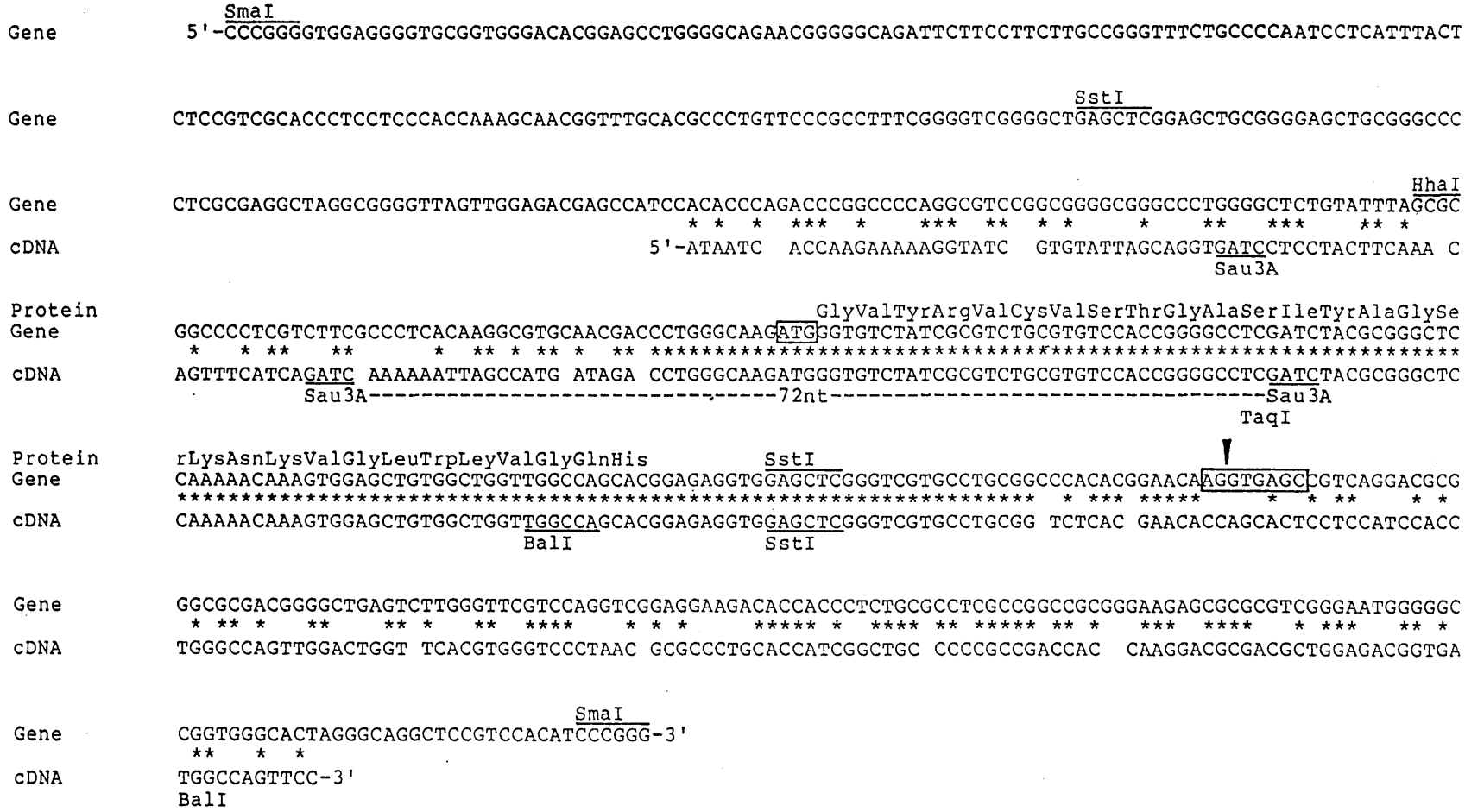

pGF6 hybridisation
λ8 genomic clone
K
R Ss
R
K
R
Ss Ss
Ss X
X
H
633bp SmaI fragment
Sm
Ss
ATG Sa
Ss
Sm
T
B
SmaI sub-clones
SstI sub-clones
Shotgun sub-clones

Figure 4. Comparison of rabbit RBC 15-LOX genomic and cDNA sequences.

The optimal nucleic acid sequence alignment of the 633bp SmaI genomic fragment sequenced here, and the 356bp cDNA insert from the cDNA insert of pGF6 (Thiele *et al*, 1987; J. Fleming, unpublished results) is shown. Identical nucleotides in the two sequences are indicated by asterisks. These include a region of 13 consecutive residues containing an ATG codon (boxed) immediately followed by a sequence which accurately predicts the peptide sequence of the purified RBC 15-LOX protein (Thiele *et al*, 1987) as indicated above the genomic sequence. The genomic sequence also includes a possible splice donor consensus sequence (Senapathy, 1988) which is boxed with a downward-pointing arrow indicating the likely splice point. The 72nt Sau3A cDNA fragment used as a probe for S1 analysis in Figure 5 is shown by a broken line under the cDNA sequence.


SmaI
Gene 5'-CCCGGGGTGGAGGGGTGCGGTGGGACACGGAGCCTGGGGCAGAACGGGGGCAGATTCTTCCTTCTTGCCGGGTTTCTGCCCCAATCCTCATTTACT
SstI
Gene CTCCGTCGCACCCTCCTCCCACCAAAGCAACGGTTTGCACGCCCTGTTCCCGCCTTTCGGGGTCGGGGCTGAGCTCGGAGCTGCGGGGAGCTGCGGGCCC
HhaI
Gene CTCGCGAGGCTAGGCGGGGTTAGTTGGAGACGAGCCATCCACACCCAGACCCGGCCCCAGGCGTCCGGCGGGGCGGGCCCTGGGGCTCTGTATTTAGCGC
cDNA 5'-ATAATC ACCAAGAAAAAGGTATC GTGTATTAGCAGGTGATCCTCCTACTTCAAA C
Sau3A
Protein GlyValTyrArgValCysValSerThrGlyAlaSerIleTyrAlaGlySe
Gene GGCCCCTCGTCTTCGCCCTCACAAGGCGTGCAACGACCCTGGGCAAGATGGGTGTCTATCGCGTCTGCGTGTCCACCGGGGCCTCGATCTACGCGGGCTC
cDNA AGTTTCATCAGATC AAAAAATTAGCCATG ATAGA CCTGGGCAAGATGGGTGTCTATCGCGTCTGCGTGTCCACCGGGGCCTCGATCTACGCGGGCTC
Sau3A----------72nt----------Sau3A
TaqI
Protein rLysAsnLysValGlyLeuTrpLeyValGlyGlnHis
SstI
Gene CAAAAACAAAGTGGAGCTGTGGCTGGTTGGCCAGCACGGAGAGGTGGAGCTCGGGTCGTGCCTGCGGCCCACACGGAACAAGGTGAGCCGTCAGGACGCG
cDNA CAAAAACAAAGTGGAGCTGTGGCTGGTTGGCCAGCACGGAGAGGTGGAGCTCGGGTCGTGCCTGCGG TCTCAC GAACACCAGCACTCCTCCATCCACC
BalI
SstI
Gene GGCGCGACGGGGCTGAGTCTTGGGTTCGTCCAGGTCGGAGGAAGACACCACCCTCTGCGCCTCGCCGGCCGCGGGAAGAGCGCGCGTCGGGAATGGGGGC
cDNA TGGGCCAGTTGGACTGGT TCACGTGGGTCCCTAAC GCGCCCTGCACCATCGGCTGC CCCCGCCGACCAC CAAGGACGCGACGCTGGAGACGGTGA
SmaI
Gene CGGTGGGCACTAGGGCAGGCTCCGTCCACATCCCGGG-3'
cDNA TGGCCAGTTCC-3'
BalI

3.1.1 The genomic clone, λ8, is a genuine RBC 15-LOX recombinant, accurately predicting the N-terminal amino acid sequence of the RBC 15-LOX protein.

The sequence of the 633bp SmaI genomic fragment (Fig. 4) conclusively proves that the λ8 genomic recombinant represents a genuine clone of the RBC 15-LOX gene. The sequence accurately predicts the first 30 amino acids of the RBC 15-LOX protein, the codon for the N-terminal glycine residue being immediately preceded by an ATG codon.

The sequence of the 633bp genomic SmaI fragment is identical to that of the pGF6 cDNA insert for 130nt, including the whole of the region which encodes the known N-terminal amino acid sequence of the RBC 15-LOX protein. However, a complete comparison of the genomic and cDNA sequences in this region, aligned according to the apparent protein-coding regions, shows two regions of non-homology, one 5′ of a point 11bp upstream of the translation initiation ATG codon, the other beginning 119bp downstream of the ATG (Fig. 4).

The sequence mismatch downstream from the ATG codon probably represents an exon-intron boundary. The genomic sequence AGGTGAGC shows a 7/8 match with the splice donor consensus sequence (C/A)GGT(G/A)(A/G)GT of Senapathy (1988), the mismatched nucleotide being the final one of the eight. However, comparison of the genomic and pGF6 cDNA sequences show 4 mismatches within the 13nt preceding this motif. Recently, cDNA clones covering the entire length of the RBC 15-LOX mRNA have been obtained from an independently-constructed cDNA library (J. Fleming and B. Thiele, unpublished results). One of these, PCR-AH25, matches the genomic sequence all the way to the putative splice donor sequence, after which point it matches neither genomic nor pGF6 cDNA sequences. This sequence is significant in 2 respects. Firstly, it suggests that the genomic sequence motif does indeed represent a splice donor sequence, and that the mismatches with pGF6 insert sequences are due to errors in the sequencing of the pGF6 cDNA. Secondly, the difference between PCR-AH25 and pGF6 cDNA

sequences beyond the putative splice point suggests that either these two cDNAs correspond to alternatively-spliced transcripts, or that one or other represents a recombinant cDNA. In fact, sequencing of other RBC 15-LOX cDNA clones has very recently revealed that the pGF6 cDNA insert contains sequences which are not contiguous in a full-length cDNA: its final 126nt (nt 231 to 356), from immediately following the putative splice site to its 3′ end, correspond to sequences 1620 to 1745nt beyond the ATG translation initiation codon in the full-length cDNA. While one cannot be sure that the pGF6 cDNA insert does not represent a shorter alternatively-spliced transcript, it seems more likely that it is a recombinant cDNA in which sequences from further 3′ in the mRNA may have become linked to sequences encoding the N-terminal region of the protein during the blunt-end ligation of double-stranded cDNA into blunt-ended vector in construction of the cDNA library. This suggestion is supported by the observation, in the same cDNA library from which pGF6 was obtained, of other recombinant clones containing RBC 15-LOX sequences from distant parts of the mRNA, joined end-to-end, which do not correspond to splice junctions in genomic sequences (J. Fleming and B. Thiele, unpublished results).

Whether the genomic sequence containing the AGGT motif does indeed represent a genuine splice donor site could easily be confirmed by performing an S1 protection experiment using a 3′ end-labelled genomic probe. This experiment has not yet been performed, as interest has been concentrated upon the 5′ end of the mRNA and the transcription initiation site of the RBC 15-LOX gene.

3.1.2 The 5′ end of the pGF6 cDNA insert is not derived from a RBC 15-LOX mRNA.

The other region of mismatch between genomic and cDNA sequences begins 11bp upstream of the ATG translation initiation codon. This corresponds to a point 90bp from the 5′ end of the cDNA. There can be no doubt, as explained above, that the central portion of the pGF6 cDNA insert arises from a genuine RBC 15-LOX mRNA, since its sequence correctly predicts the first 30 amino acids of the RBC 15-LOX protein (Thiele *et al*, 1987). So, the question arises as to what the

most 5′ 90bp of the pGF6 cDNA insert represent. They may be derived from an upstream exon, in which case the first mis-matched nucleotide of the genomic sequence represents the final nucleotide of an intron. Alternatively, the 5′ end of the pGF6 cDNA insert may be the result of an artefactual cloning event of the sort discussed above.

If the whole of the pGF6 cDNA insert 5′ of the putative splice point described in Section 3.1.2 above is derived from a single, genuine RBC 15-LOX messenger, the first 90nt being transcribed from one or more short exons upstream of the first coding exon, then the full length of any fragment from the 5′ end of pGF6 should be protected when used in an S1 protection experiment with reticulocyte RNA.

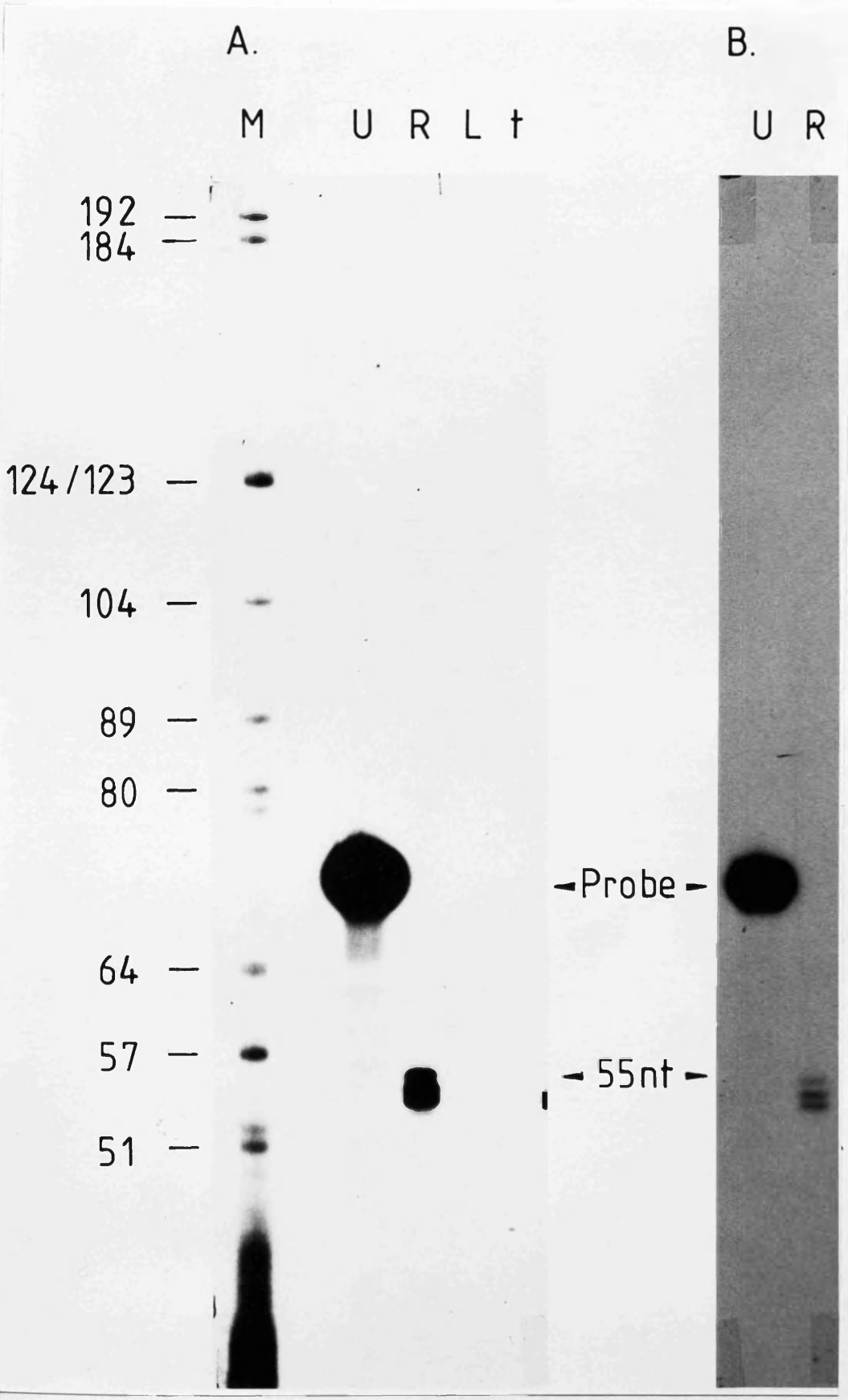
So, a 72bp Sau3A fragment corresponding to nucleotides 66-137 of the pGF6 cDNA insert (Fig. 4), was end-labelled and used as a probe for S1 protection (see Fig. 5). This fragment spans the point at which the mismatch between cDNA and genomic sequences begins. Several attempts to separate the two strands of the fragment were unsuccessful. So, a double-stranded cDNA probe was hybridised with RNA. Since complete degradation of probe was seen in controls, where double-stranded probe was hybridised with rabbit total liver RNA and yeast tRNA (Fig. 5, tracks L and t), protection seen with reticulocyte RNA is not due to renaturation of the two strands of the probe.

No protection of full-length probe by reticulocyte total RNA is detectable. Rather, the reticulocyte RNA protects a cluster of fragments, the longest of which is 55nt long. As the 5′ end of the probe's non-coding strand lies 42nt downstream of the complement of the ATG codon, the reticulocyte RNA population protects the cDNA probe to a position only 13nt upstream of the ATG codon. This corresponds to the point at which the sequence mismatch begins, and indicates that the cDNA sequences in the upstream mismatched region are not present in a genuine RBC 15-LOX messenger. It seems likely that, during construction of the cDNA library, there has been a linking of another cDNA fragment on to the 5′ end of a genuine RBC 15-LOX cDNA.

Figure 5. S1 nuclease analysis using a 72bp Sau3A probe derived from pGF6 cDNA insert.

The dephosphorylated 72bp Sau3A fragment was kinase 5′ end-labelled and ethanol precipitated then denatured and loaded on to a 5% polyacrylamide non-denaturing strand-separating gel. After unsuccessful strand-separation, double-stranded end-labelled fragment was recovered by electro-elution followed by ethanol precipitation using 100µg yeast tRNA as carrier. Double-stranded probe was then denatured and hybridised with 50µg of the appropriate RNA at 52°C overnight in 80% formamide; 0.4M NaCl; 40mM PIPES pH6.4; 1mM EDTA. Digestion of hybrids was performed in 0.25ml of 250mM NaCl; 30mM sodium acetate pH4.6; 1mM zinc sulphate at 37°C for 1 hour using 60 Units of S1 nuclease. Digestion products were ethanol precipitated and resolved on a 6% denaturing polyacrylamide gel prior to autoradiography. Panel A, overnight exposure; panel B, 1 hour exposure of tracks U and R from panel A, showing micro-heterogeneity of products.

RNAs: rabbit reticulocyte total (R); rabbit liver total RNA (L); yeast tRNA (t). U, undigested probe; M, 5′ end-labelled pBR322/HaeIII markers (Sizes as shown).


A.
B.
M
U
R
L
t
U
R
192
184
124/123
104
89
80
64
57
51
Probe
55nt

In summary, it seems certain that the pGF6 cDNA insert is a hybrid cDNA, consisting of three portions: a central portion which is undoubtedly derived from an RBC 15-LOX mRNA; a 3′ portion which represents a region from further 3′ in the full-length mRNA; and a 5′ portion of unknown origin which is not derived from a RBC 15-LOX mRNA.

3.2 MAPPING OF THE TRANSCRIPTION INITIATION SITE OF THE RABBIT RBC 15-LOX GENE.

In order to investigate the organisation and the function of the promoter region of the RBC 15-LOX gene, it is first necessary to establish the major site of transcription initiation.

Various pieces of evidence suggest that the first nucleotide of a mature mRNA corresponds to the site of transcription initiation by RNA Polymerase II. Ziff and Evans (1978) mapped the transcription initiation site of Adenovirus type 2 (Ad-2) late transcripts by hybridising the shortest fraction of *in vivo* pulse-labelled transcripts to fragments of the Ad-2 genome and demonstrated that the 5′ terminal sequence of Ad-2 late mRNAs (determined by direct sequencing using RNases) corresponds to the genomic sequence in the region of transcription initiation. Consistent with the coincidence of transcription initiation site and the 5′ end of the mRNA are observations that, for a variety of genes including mammalian globin genes, the 5′ ends of mRNA precursors and mature mRNAs map to the same position in the genomic sequence (Weaver and Weissmann, 1979; Roop *et al*, 1980; Grosveld *et al*, 1981a). Thus, mapping of the 5′ end of the mature mRNA can be used to identify the site of transcription initiation.

3.2.1 Strategy for mapping the 5′ end of the RBC 15-LOX mRNA.

To determine whether the transcription initiation site of the rabbit RBC 15-LOX gene lies within the 276bp SstI genomic fragment (Fig. 4) which contains N-terminal protein-coding sequence (see Section 3.1.1 above), or whether it lies further upstream, a combination of S1 nuclease protection and primer extension analyses,

using RNA from rabbit reticulocytes, were performed.

S1 nuclease analysis involves the formation of partially double-stranded hybrids between RNA and a labelled DNA probe under conditions which favour the formation of RNA:DNA hybrids. S1 nuclease degrades single-stranded nucleic acids, but, at low temperatures and in buffers of high ionic strength, double-stranded regions are resistant to digestion. Thus, RNA protects complementary regions of the radio-labelled DNA probe from S1 digestion, but non-homologous sequences are removed from the probe, resulting in a shortened probe.

For mapping 5′ and 3′ termini of mRNAs, or splice points, an end-labelled genomic DNA restriction fragment is denatured and hybridised with mRNA prior to S1 nuclease treatment. If the RNA terminus or a splice point does not lie within this fragment, full-length probe protection is seen. Otherwise, the reduced size of the protected fragment corresponds to the distance between the labelled end of the probe and the end of the mRNA or splice point.

Primer extension involves the generation in vitro of the cDNA complement of an RNA using the enzyme reverse transcriptase. The 5′-> 3′ RNA-dependent DNA polymerase activity of reverse transcriptase requires a primer to initiate polymerisation. The size(s) of primer extension products obtained using a primer which is complementary to transcribed sequences indicate the distance between the primer and the 5′ end of the RNA. Primer extension can also be used to sequence an RNA. The dideoxynucleotide analogue of one of the four naturally--occurring deoxynucleoside triphosphates can be included in each of four primer extension reactions, producing a sequence "ladder" in a manner analogous to that employed in dideoxy chain-termination sequencing reactions (Sanger et al, 1977). For mRNA sequencing, however, an end-labelled primer is used to copy the single-stranded (mRNA) template, rather than incorporating a labelled deoxynucleotide into the polymerase reaction.

The transcription initiation site of the RBC 15-LOX gene has been determined by several independent methods (Fig. 6). The 5′ end

Figure 6. Summary of strategy and results in mapping of the 5' end of the rabbit RBC 15-LOX mRNA.

Top: Restriction map of the λ8 genomic clone with (incomplete) exon map. The scale refers to the number of kilobases from the transcription initiation site, as mapped by the combination of methods shown here.

Centre: Restriction map of a 760bp fragment containing the 633bp SmaI fragment sequenced in Section 3.1. The transcription initiation site as mapped by the methods shown here is indicated by an arrow, 28nt upstream of the translation initiation ATG codon. The first exon is indicated by a box. Coding sequences are shaded.

Bottom: Primers and probes used in the mapping of the 5' end of the RBC 15-LOX mRNA. Sizes of primer extension/S1 protection products are indicated beneath the appropriate probes. An asterisk denotes the ^{32}P-labelled 5' end of each primer or probe. The limits of fragments used to probe Northern blots are shown by open boxes.

Restriction sites: K = KpnI; R = EcoRI; Ss = SstI; X = XbaI; H = HindIII; Sm = SmaI; Hh = HhaI; Sa = Sau3A; T = TaqI; B = BalI.

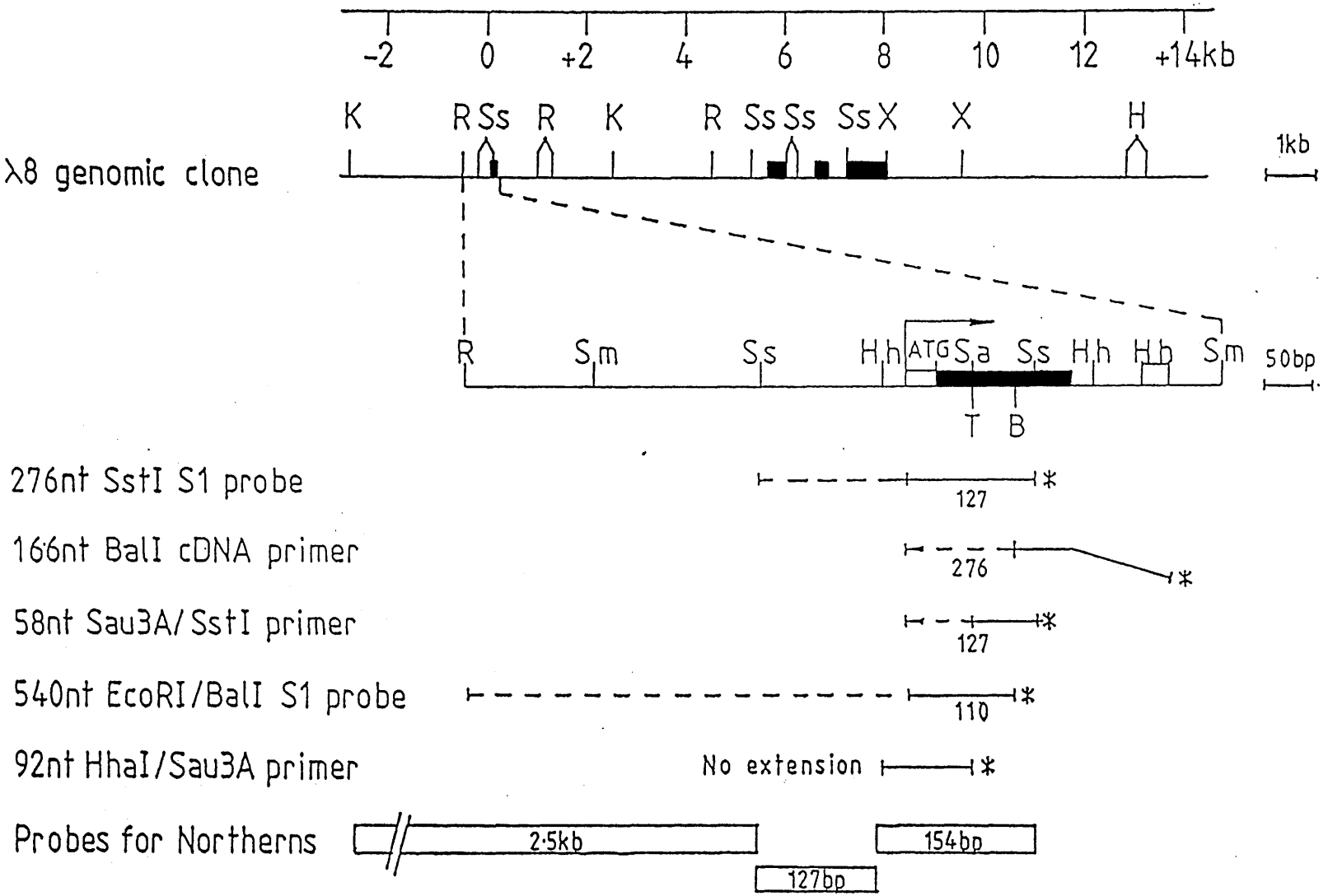

-2
0
+2
4
6
8
10
12
+14kb
λ8 genomic clone
K
R Ss
R
K
R
Ss Ss
Ss X
X
H
1kb
R
S m
S s
H h
ATG
S a
S s
H h
H h
S m
50bp
T
B
276nt SstI S1 probe
127
166nt BalI cDNA primer
276
58nt Sau3A/SstI primer
127
540nt EcoRI/BalI S1 probe
110
92nt HhaI/Sau3A primer
No extension
Probes for Northerns
2·5kb
154bp
127bp

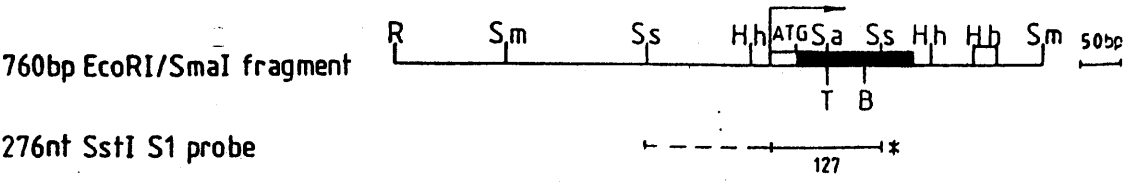
Figure 7. S1 nuclease mapping of the 5′ end of the first protein-coding exon using a single-stranded 276bp SstI genomic probe.

The 276bp SstI fragment was 5′ kinase end-labelled, denatured, and strand-separated on a 5% non-denaturing polyacrylamide gel. After electro-elution, separate strands were hybridised with 50µg of the appropriate RNA at 52°C overnight in 10µl of 80% formamide; 0.4M NaCl; 40mM PIPES pH6.4; 1mM EDTA. S1 digestion and resolution of protected products were as in Figure 5.

Panel A: S1 protection of end-labelled coding (c) and non-coding (nc) strands of the 276bp SstI fragment using various RNAs. R, rabbit reticulocyte total RNA; B, rabbit brain total RNA; t, yeast tRNA; U, equivalent amount of undigested probe to that used in each of the hybridisations. M, pBR322/HaeIII 5′ end-labelled size markers (sizes as shown).

Panel B: shorter exposure of tracks U(nc) and R(nc) from Panel A, showing micro-heterogeneity of protected fragments.

Panel A, overnight exposure at -70°C; Panel B, 1 hour exposure at -70°C.



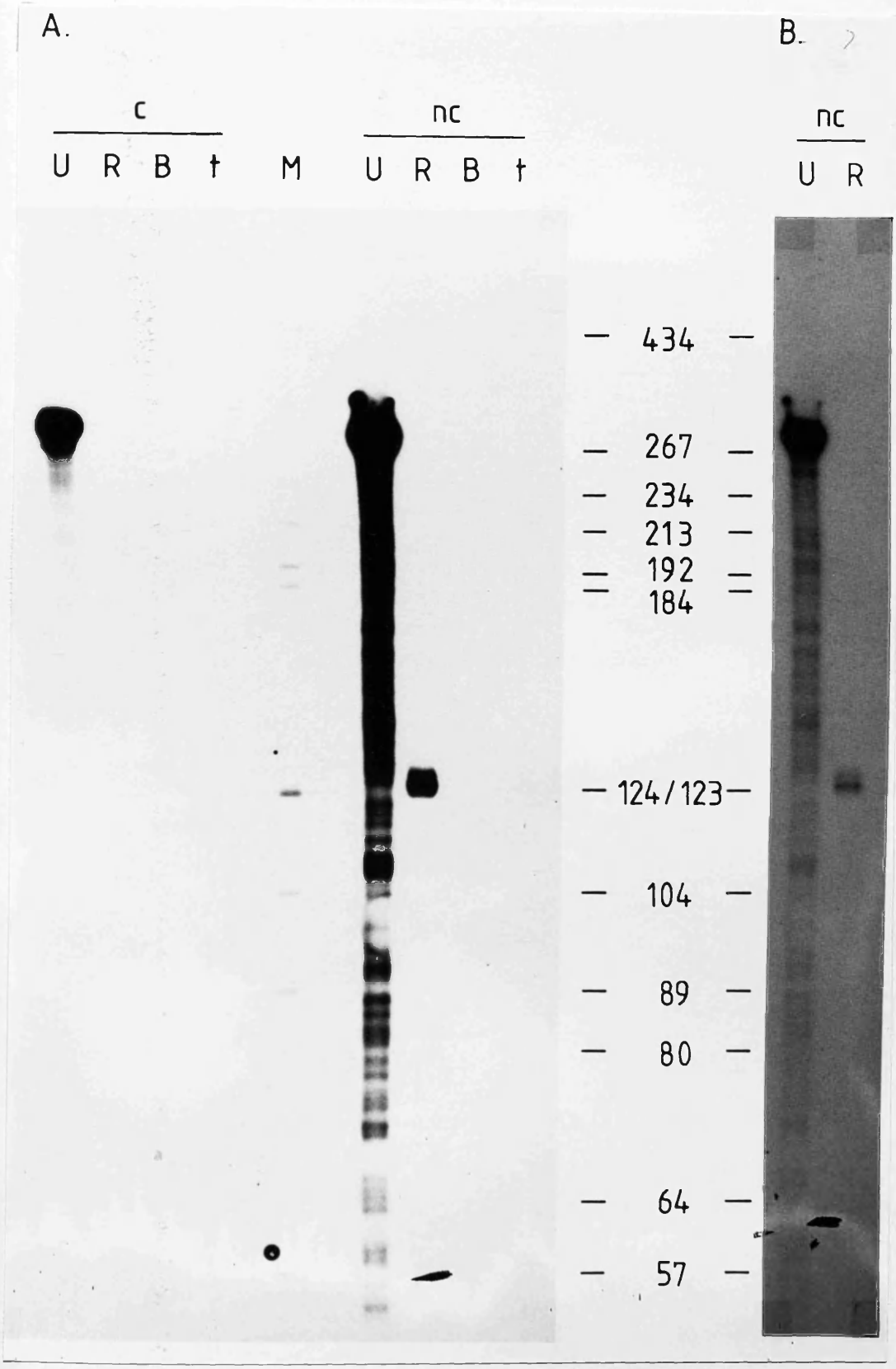

A.
B.
c
nc
nc
U R B t M U R B t
U R
434
267
234
213
192
184
124/123
104
89
80
64
57

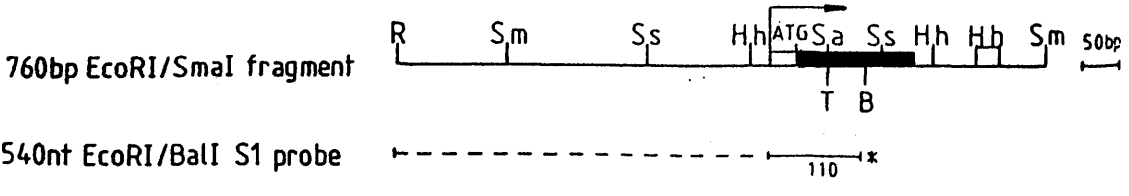
Figure 8. S1 nuclease analysis using a single-stranded 540nt EcoRI/BalI genomic probe.

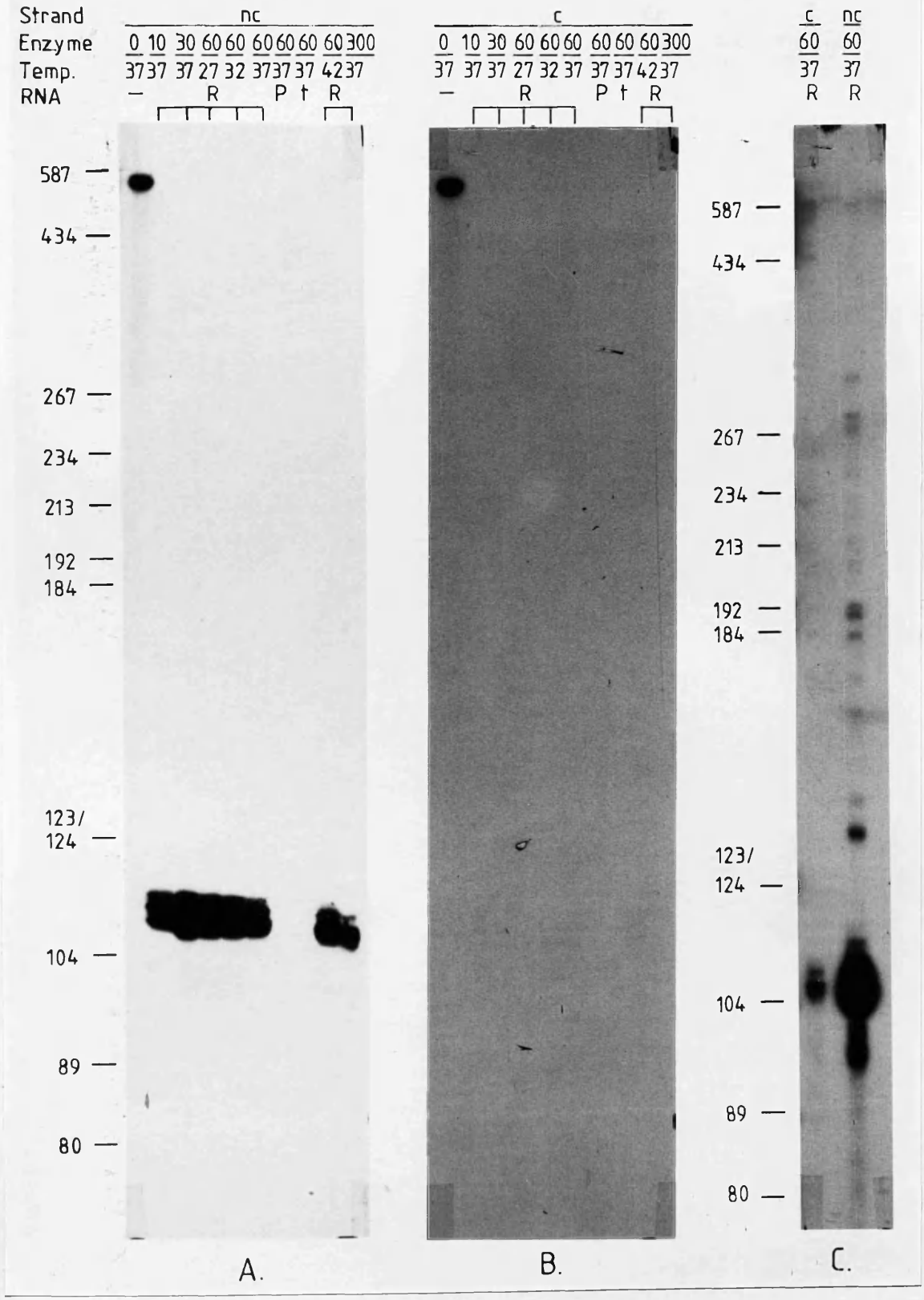
Panel A: Protection of non-coding (nc) strand at various digestion temperatures (Temp.) and using various amounts of enzyme with rabbit reticulocyte RNA (R), and with rabbit pancreas total RNA (P) or yeast tRNA (t) controls.

Panel B: Equivalent exposure of protection using the coding (c) strand as probe under identical conditions to those for the non-coding strand in panel A.

Panel C: Long exposure (13 days) of a separate experiment involving protection of non-coding and coding strands by rabbit reticulocyte total RNA. The same fragment as in A and B, but labelled on a separate occasion, was used as probe. Protected fragments up to 200nt longer than the major protected species can be seen.

The 540nt EcoRI/BalI fragment was kinase 5′ end-labelled, and strands were denatured, then separated on a 5% polyacrylamide non-denaturing gel. Isolated strands were recovered by electro-elution and ethanol precipitated with 100µg yeast tRNA as carrier. Single-stranded probes were hybridised with the appropriate RNA overnight at 52°C in 10µl 80% formamide; 0.4M NaCl; 40mM PIPES pH6.4; 1mM EDTA. Hybrids were subjected to S1 nuclease digestion for 1 hour at various temperatures in 0.25mls 250mM NaCl; 30mM sodium acetate pH4.6; 1mM zinc sulphate using various amounts of enzyme. Products were separated on a 6% denaturing polyacrylamide gel prior to autoradiography.




Strand
Enzyme
Temp.
RNA
nc
0 10 30 60 60 60 60 60 60 300
37 37 37 27 32 37 37 37 42 37
— R P t R
587
434
267
234
213
192
184
123/
124
104
89
80
A.
c
0 10 30 60 60 60 60 60 60 300
37 37 37 27 32 37 37 37 42 37
— R P t R
587
434
267
234
213
192
184
123/
124
104
89
80
B.
c nc
60 60
37 37
R R
587
434
267
234
213
192
184
123/
124
104
89
80
C.

of the mRNA has been mapped using two different S1 probes and two different primers. The localisation has been confirmed by mRNA sequencing, by the lack of primer extension seen with a primer which overlaps the transcription initiation site, and by hybridisation of probes from immediately upstream and immediately downstream of the transcription initiation site to Northern blots.

3.2.2 S1 nuclease analysis using a single-stranded 276nt SstI probe: the 5′ end of the first coding exon lies 28nt upstream of the ATG translation initiation codon.

Figure 7 shows the mapping of the 5′ end of the first coding exon of the RBC 15-LOX gene by an S1 nuclease protection experiment using as S1 probes separated strands of the 276bp SstI genomic fragment sequenced in Section 3.1, which includes sequences predicting the first 30 amino acids of N-terminal protein sequence. Protection is seen with only one strand, indicating that this region of the gene does not show bi-directional transcription. With the non-coding strand, rabbit reticulocyte total RNA, but not rabbit brain total RNA or yeast tRNA, protects a cluster of fragments, two stronger bands being flanked by two weaker ones. Micro-heterogeneity in the size of protected fragments is frequently seen in S1 nuclease analyses and may be due to exonucleolytic "nibbling" of the ends of double-stranded hybrids (Williams and Mason, 1985). As exonuclease "nibbling" will result in slightly shorter products, the longest protected fragment represents the most 5′ limit possible for the start of the first coding exon. Here, the longest protected fragment is at 127nt. As the labelled 5′ end of the 276nt non-coding strand of the SstI genomic fragment lies 100nt downstream of the ATG translation initiation codon (Fig. 4), the RBC 15-LOX mRNA thus protects the probe to 27nt beyond the ATG codon. This protection indicates that the first coding exon begins 28nt upstream of the A nucleotide of the ATG translation initiation codon.

3.2.3 S1 nuclease analysis using a single-stranded 540nt EcoRI/BalI probe confirms the localisation of the 5′ end of the first coding exon, and reveals that microheterogeneity of protected fragments is not a function of digestion temperature or enzyme concentration.

The localisation of the 5′ end of the first coding exon has been confirmed using a longer probe (Fig. 8). A 540nt single-stranded EcoRI/BalI genomic probe produces a cluster of fragments, the longest of which is 110nt. This experiment also demonstrates that the same micro-heterogeneity of products is seen when S1 digestion is performed under a variety of conditions. There is no alteration in the pattern of products when the digestion temperature is varied over a range of 15^{o}C, and the amount of enzyme varied by 5-fold either side of the 60 Units used for the 276nt SstI probe in Figure 7.

It is not possible to establish by S1 nuclease analysis using these probes whether the first coding exon is the most 5′ exon of the RBC 15-LOX gene, or whether there might be one or more upstream exons, as, for example, in the case of the 5′ non-coding exon of c-myc (Stanton et al, 1984). However, the complementary technique of primer extension can be used to determine whether the 5′ end of the first coding exon mapped by S1 nuclease analysis represents the 5′ end of the mRNA, or only the first nucleotide of a downstream exon.

3.2.4 Primer extension using a single-stranded 58nt genomic DNA primer: transcription initiation occurs 28nt upstream of the ATG translation initiation codon.

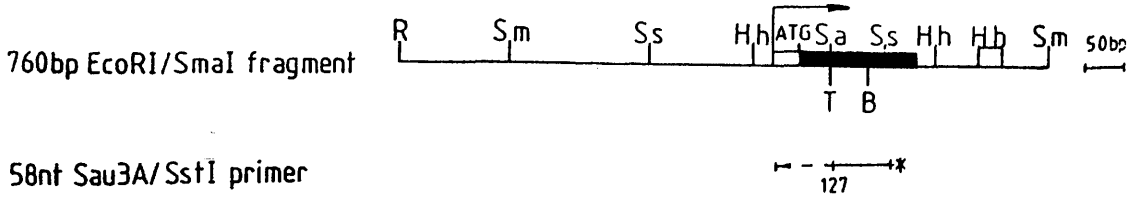
Results with a 58nt Sau3A/SstI genomic DNA primer firmly establish the 5′ end of the RBC 15-LOX mRNA as being 28nt upstream of the ATG translation initiation codon. A preliminary experiment (Fig. 9) demonstrates that the 58nt primer has an optimum hybridisation temperature of approximately 52^{o}C. Primer extension is considerably more efficient following hybridisation with reticulocyte RNA at 52^{o}C than at 42^{o}C or 60^{o}C i.e. the same pattern of products is seen in all cases, but, at 52^{o}C, the largest proportion of the primer is converted

Figure 9. Optimum hybridisation temperature for primer extensions using a single-stranded 58nt Sau3A/SstI genomic primer.

Overnight autoradiograph of primer extension reactions with rabbit reticulocyte total RNA (R) and yeast tRNA (t), using the non-coding strand strand (58nt) of the Sau3A/SstI fragment. (The 66nt coding strand does not extend - see Fig. 10).

The double-stranded Sau3A/SstI fragment was kinase 5′ end-labelled, denatured, and strands separated by running on a 6% denaturing polyacrylamide gel. Separated strands were eluted, ethanol precipitated and hybridised at a variety of temperatures (37, 42, 52 and 60°C) with 50µg of the appropriate RNA at 52°C overnight in 10µl 80% formamide; 0.4M NaCl; 40mM PIPES pH6.4; 1mM EDTA. Hybrids were ethanol precipitated and primer extension performed using AMV reverse transcriptase. Primer extension products were resolved on a 6% denaturing polyacrylamide gel, followed by autoradiography.

Sizes indicated are of 5′ end-labelled pBR322/HaeIII size markers.



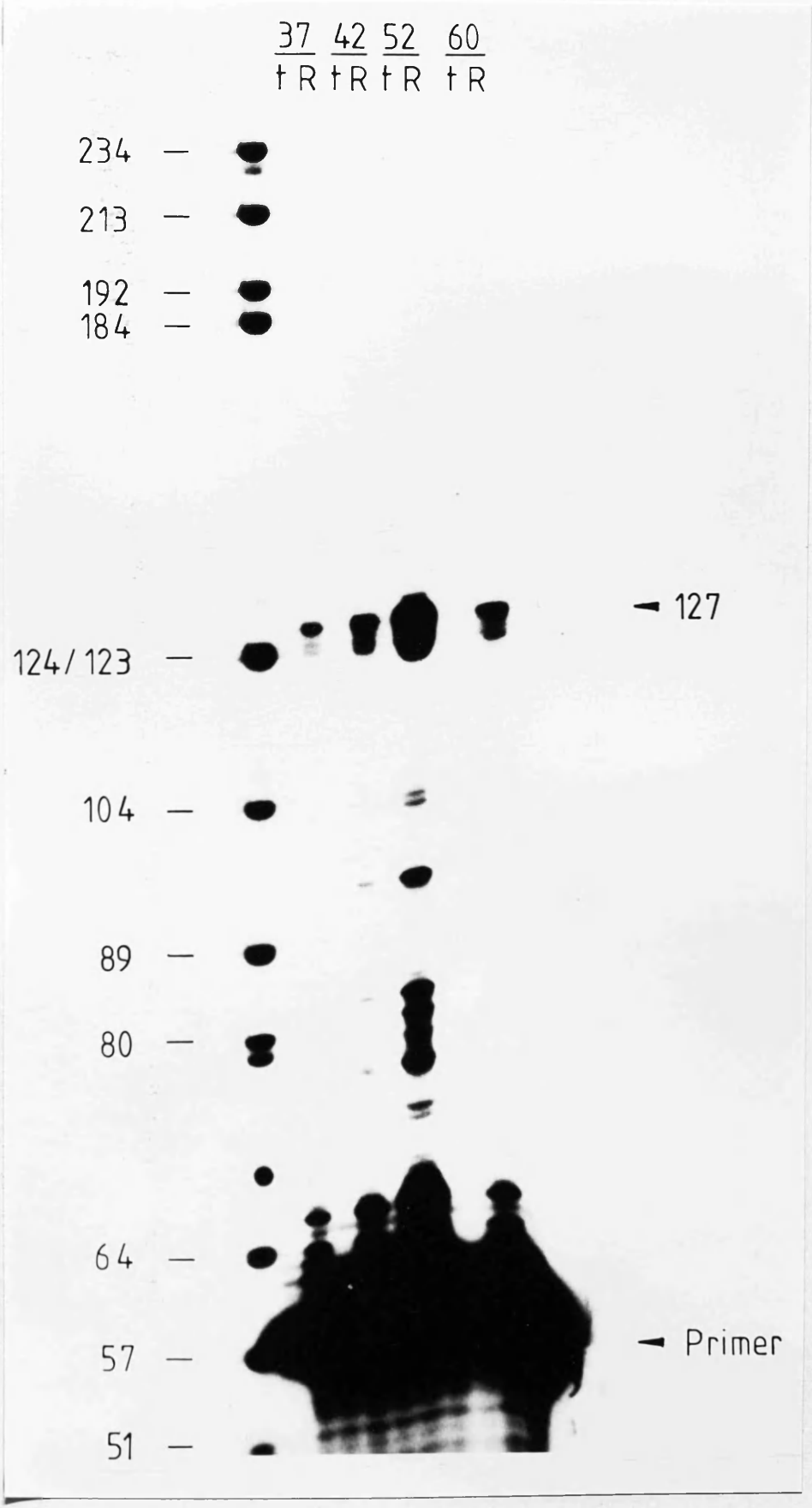

37 42 52 60
t R t R t R t R
234
213
192
184
124/123
104
89
80
64
57
51
127
Primer

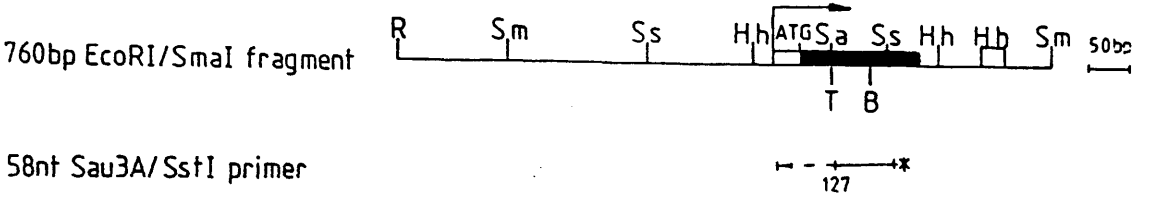
Figure 10. Primer extension and mRNA sequencing using a single-stranded 58nt Sau3A/SstI genomic primer.

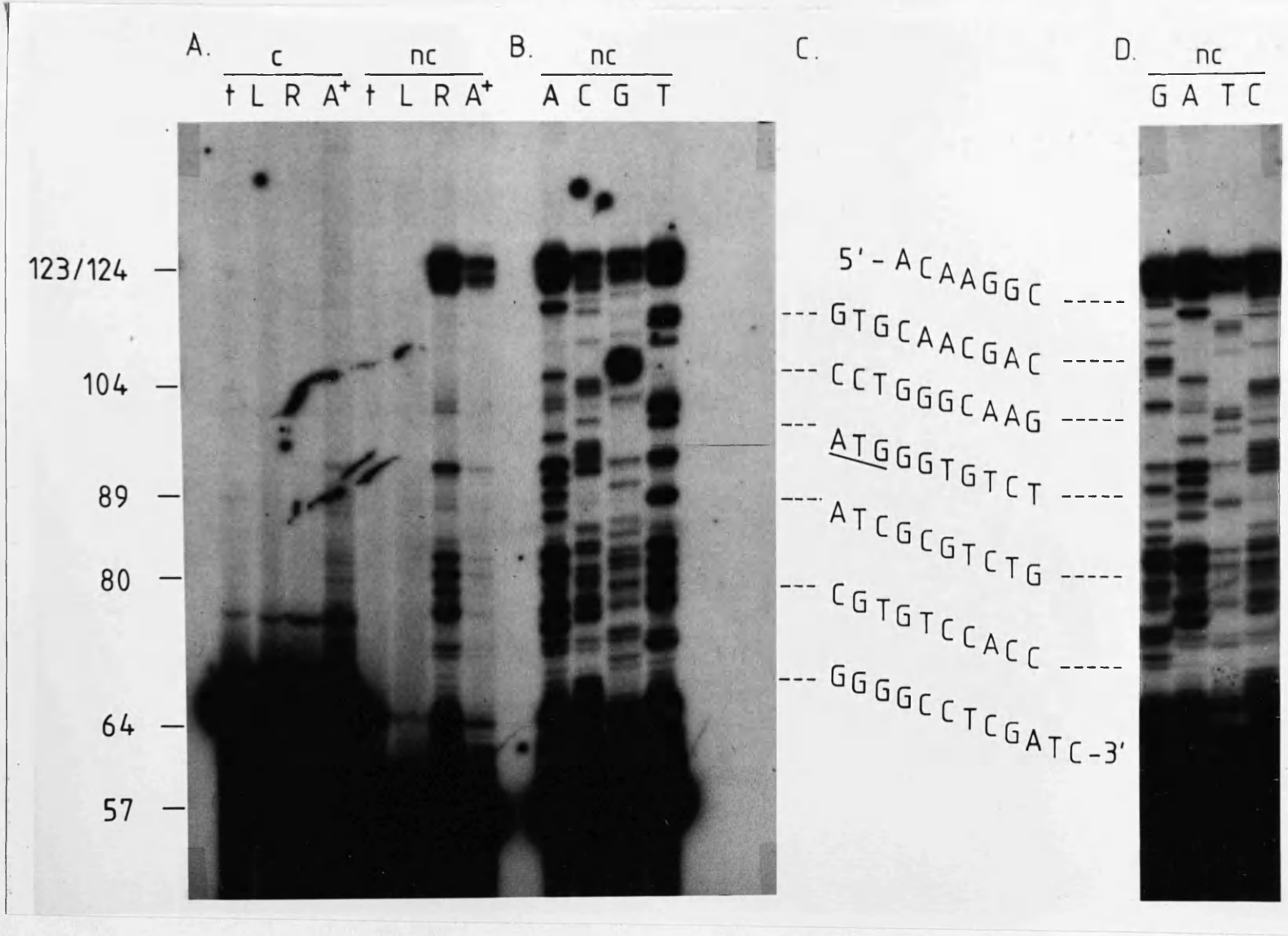
Preparation of probes, hybridisation with RNAs and primer extension were performed as in Figure 9. mRNA sequencing was performed in exactly the same way, except for the inclusion of appropriate concentrations of one of the four dideoxynucleoside triphosphates (different dideoxynucleotide concentrations were used in Panels B and D). Primer extension products were run out on a 6% denaturing polyacrylamide gel prior to autoradiography. Panels A and B, 29 day exposure; Panel D, 14 day exposure.

Panel A: Primer extension on the 58nt non-coding (nc) strand, and on the 66nt coding (c) strand with both total (R) and polyadenylated (A^+) rabbit reticulocyte RNAs, and with rabbit liver total RNA (L) or with yeast tRNA (t).

Panels B and D: mRNA sequencing by primer extension in the presence of differing quantities of dideoxynucleoside triphosphates. A, C, G and T: primer extension in the presence of ddATP, ddCTP, ddGTP and ddTTP respectively.

Panel C: Genomic sequence around the translation initiation codon (Section 3.1). The sequence is the complement of the sequence which can be read from panel D.




A.
c
nc
t L R A+ t L R A+
B.
nc
A C G T
C.
D.
nc
G A T C
123/124
104
89
80
64
57
5'-ACAAGGC -----
--- GTGCAACGAC -----
--- CCTGGGCAAG -----
--- ATGGGTGTCT -----
--- ATCGCGTCTG -----
--- CGTGTCCACC -----
--- GGGGCCTCGATC-3'

to extension products. Consequently, all subsequenct primer extensions using this primer were performed after hybridisation of primer with RNA at 52°C.

As seen in Figure 10, extension occurs from one strand only. Primer extension with the non-coding strand produces the same pattern of products with both total and polyadenylated reticulocyte RNAs, indicating that extension is occurring on an mRNA. No extension is seen with control RNA from a rabbit non-erythroid tissue (adult liver) or with yeast tRNA, consistent with the observation that accumulation of the RBC 15-LOX mRNA is erythroid-specific (Thiele et al, 1987). Micro-heterogeneity is seen in the longest primer extension products. Such micro-heterogeneity is frequently seen in primer extension experiments, and is thought to be due to premature termination of reverse transcription due to the methylated residue situated next to the mRNA cap (Williams and Mason, 1985). Consequently, the largest product is most likely to be that which represents the 5′ end of the mRNA. There are also multiple products which are considerably shorter than the longest. Comparison with products of the S1 nuclease analyses indicates that these are not genuine full-length primer extension products (such as those which might be reverse transcribed from any RBC 15-LOX mRNAs which might be truncated at their 5′ ends), since there are no corresponding S1-protected fragments. Rather they are likely to represent products of premature termination of reverse transcription.

The longest extension product is 127nt long. This is exactly the same length as the longest protected fragment in the S1 nuclease analysis using the 276nt SstI probe. Since both the 58nt primer and the 276nt probe have the same labelled 5′ end, the exact agreement of the lengths of products indicates that the 5′ end of the first coding exon in fact represents the 5′ end of the RBC 15-LOX mRNA. The major transcription initiation site of the RBC 15-LOX gene is thus located 28nt upstream of the ATG translation initiation codon.

3.2.5 mRNA sequencing using the 58nt genomic primer: priming is from a unique mRNA sequence which matches that of the RBC 15-LOX gene.

One possible objection to the use of primer extension to confirm the identity of the 5′ end of the first exon as the 5′ end of the mRNA is that the primer might conceivably be hybridising to an RNA other than the RBC 15-LOX mRNA, and that the products of primer extension and S1 analysis are the same size by coincidence. Although confirmatory results obtained with a different primer (see Section 3.2.6 below) might be considered to effectively exclude this possibility, a more direct method to establish that primer extension is occurring on the correct mRNA is to sequence the mRNA by primer extension.

mRNA sequencing using the 58nt genomic primer (Fig. 10B, C and D) conclusively proves that priming is occurring from hybridisation to the RBC 15-LOX mRNA and that mRNA alone. The mRNA sequence is that of an mRNA capable of encoding a protein whose first 12 amino acids are identical to those of the RBC 15-LOX protein. The sequence which can be read directly from the autoradiograph is unambiguous through most of its length (Fig. 10D) and is the exact complement (Fig. 10C) of 69nt of RBC 15-LOX genomic sequence (see Section 3.1.1), including the 27nt upstream of the ATG codon. Comparison of the primer extension mRNA sequence with genomic sequence identifies the precise nucleotide at which the mRNA corresponding to the longest primer extension product is initiated (an A residue 28 nt upstream of the ATG codon).

Confirmation of the exact nucleotide at which transcription initiation occurs could be obtained, if required, by using the same primer to sequence an appropriate genomic DNA fragment and to run the products of the two sequencing reactions next to one another on the same sequencing gel.

3.2.6 Primer extension using a single-stranded 166nt cDNA-derived BalI primer: confirmation that the 5′ end of the RBC 15-LOX mRNA lies 28nt upstream of the ATG translation initiation codon.

A 166nt primer corresponding to nucleotides 183-348 of the 356bp pGF6 cDNA insert (ATG = 100-102) overlaps the non-priming end of the 58nt genomic primer by 17nt (see Fig. 11) and confirms the localisation of the transcription initiation site made using the 58nt genomic primer. Since the first use of this primer in establishing the 5′ end of the RBC 15-lOX mRNA, nucleotides 231-356 of the pGF6 cDNA insert have been discovered to be derived from a region of the RBC 15-LOX mRNA far downstream from that which corresponds to the more 5′ pGF6 sequences. Hybridisation of the 166nt to the RBC 15-LOX mRNA may therefore involve "looping out" of a region of the mRNA. However, mRNA sequencing using the 166nt primer (Fig. 11) shows that this does not have any effect upon hybridisation of the 48nt (nt 183-230) at the priming end. Although not completely unambiguous, the sequence is readable in several regions right up to the longest primer extension products, and matches the genomic sequence, demonstrating that primer extension is occurring from the correct site on the mRNA, as predicted by comparison of the genomic and cDNA sequences. Confirmation of this mRNA sequence has been obtained in subsequent experiments (see Section 3.2.7 below). Figure 11 also shows that the 166nt primer has a similar optimum hybridisation temperature with the RBC 15-LOX mRNA to that for the 58nt primer. Subsequent experiments with the 166nt primer were, therefore, performed with hybridisation of the primer to RNA samples at 52^{o}C.

Figure 12A shows primer extension using both coding and non-coding strands of the 166nt BalI fragment after hybridisation with RNAs at 52^{o}C. As with the 58nt genomic primer above (Section 3.2.4), primer extension on reticulocyte total and polyadenylated RNAs produces the same pattern, showing micro-heterogeneity in the longest extension products. The longest of these is 276nt long i.e. primer extension has added 110 nt to the length of the primer. Inspection of the pGF6 cDNA sequence (Fig. 4) shows that the BalI site at the

Figure 11. Optimum hybridisation temperature for primer extension using the 166nt BalI cDNA-derived primer; preliminary mRNA sequencing; and detection of minor transcripts.

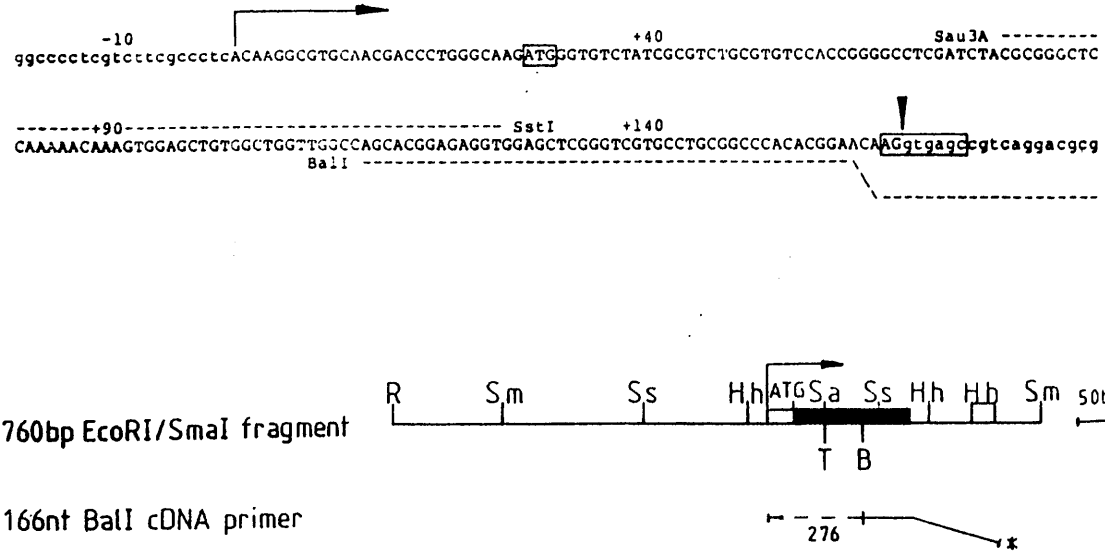
The double-stranded 166bp BalI fragment from pGF6 was kinase 5' end-labelled, denatured, and strands were separated by running on a 5% non-denaturing polyacrylamide gel. Separated strands were eluted, ethanol precipitated and hybridised with 50µg of the appropriate RNA at the appropriate temperature (37, 42, 52, or 60°C) overnight in 10µl 80% formamide; 0.4M NaCl; 40mM PIPES pH6.4; 1mM EDTA. Hybrids were ethanol precipitated and primer extension was performed using AMV reverse transcriptase. Products were then separated on a 6% denaturing polyacrylamide gel prior to autoradiography for 23 days (Panel A) or for 1 day (Panel B).

mRNA sequencing was exactly as for primer extension except for the incorporation of one of the four dideoxynucleoside triphosphates (A, ddATP; C, ddCTP; G, ddGTP; T, ddTTP).

Panel A: Primer extension on rabbit reticulocyte total RNA (R) and yeast tRNA (t) templates, using the non-coding strand of the 166nt BalI cDNA fragment from pGF6 (the coding strand does not extend - see Figure 12).

Panel B: Shorter exposure of tracks 52t and 52R from panel A, showing microheterogeneity within the major extension product.

The relationship between the 166nt cDNA-derived primer and the 58nt genomic primer used in Figures 9 and 10 is shown below. The mismatch between cDNA and genomic sequences beyond the splice donor site (Section 3.1.1) is indicated by a bent line.



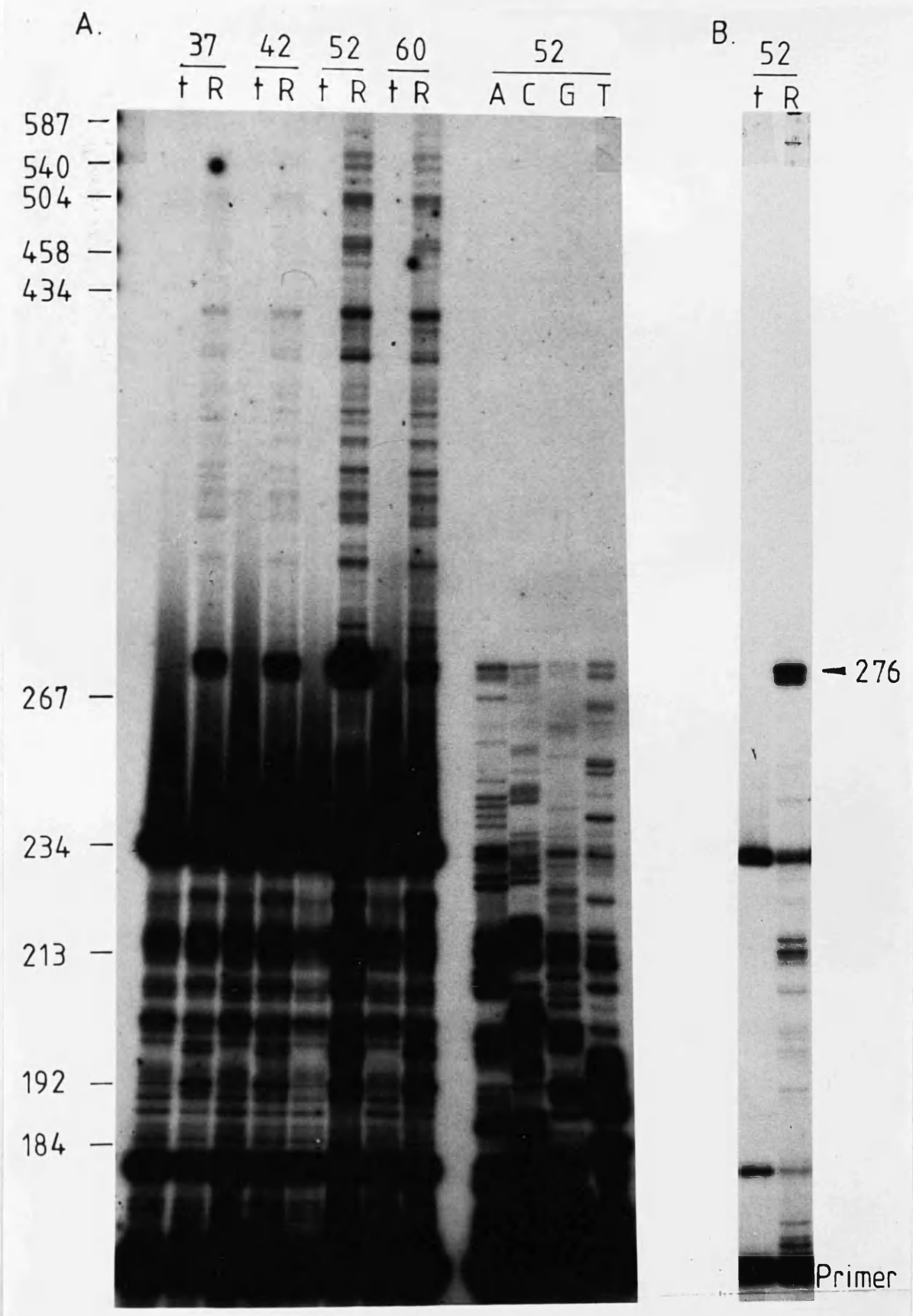

A.
37
42
52
60
52
t R t R t R t R A C G T
587
540
504
458
434
267
234
213
192
184
B.
52
t R
276
Primer

Figure 12. Primer extension on a 166nt BalI cDNA-derived primer: mRNA sequence and minor transcripts.

Preparation of single-stranded primers, hybridisation with RNAs, primer extension and mRNA sequencing were performed as in Figure 11. Primer extension products were run out on a 6% denaturing polyacrylamide gel, followed by autoradiography. Panels A and B, overnight exposure; Panels C and D, 21 days exposure.

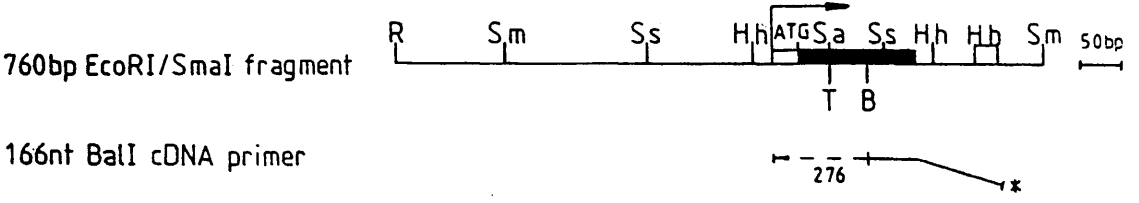
Panel A: Primer extension on non-coding (nc) and coding (c) strands of the 166nt primer with total (R) and polyadenylated (A^+) rabbit reticulocyte RNAs, and with rabbit pancreas (P) RNA and yeast tRNA (t) controls.

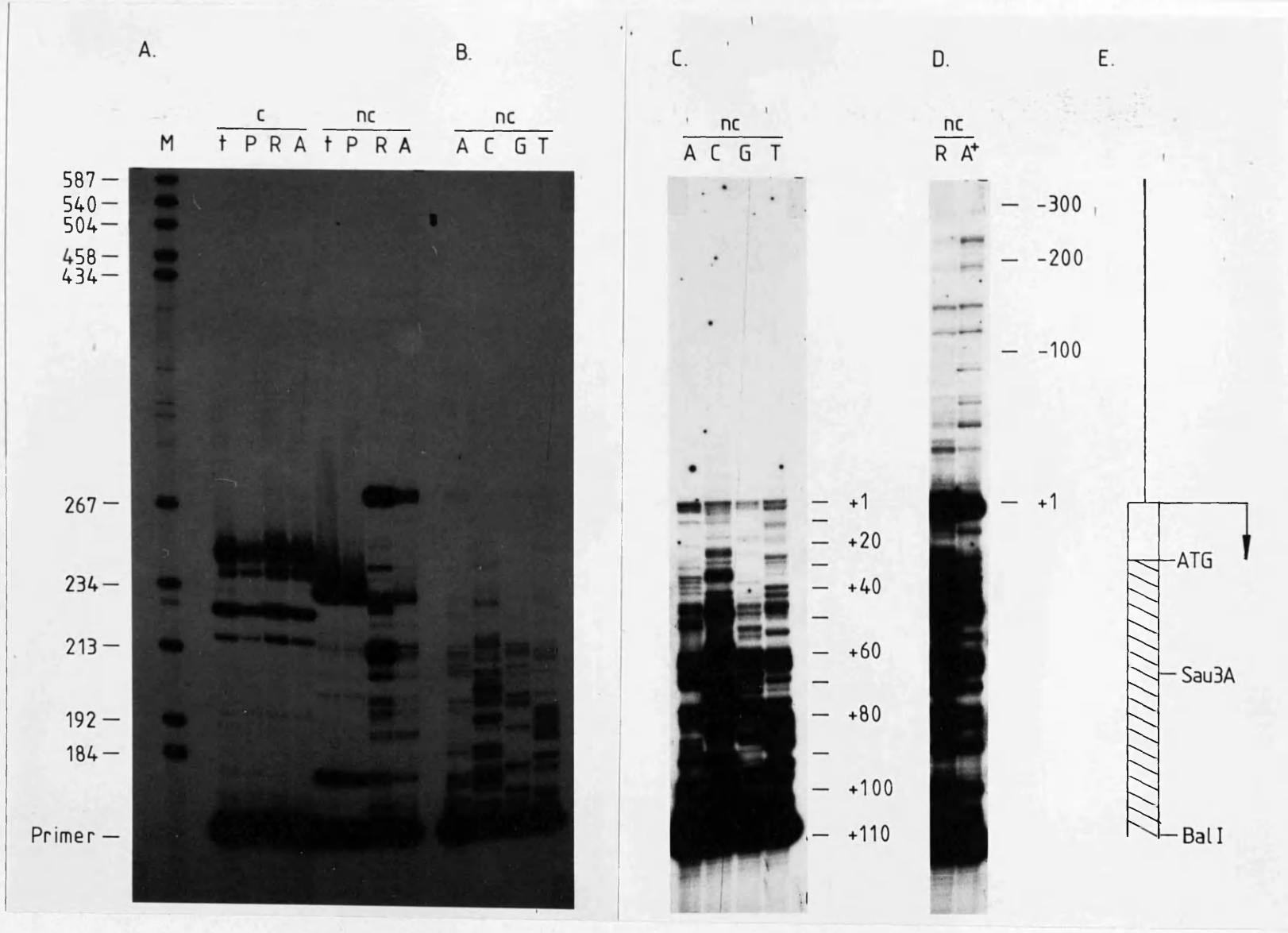
Panel B: Sequencing of the RBC LOX mRNA by primer extension in the presence of dideoxynucleoside triphosphates, as in Figures 10 and 11.

Panel C: Longer exposure of mRNA sequence shown in Panel B, showing that sequencing terminates at the apparent 5' end of the mRNA, as identified in Panel A.

Panel D: Detection of minor transcripts using the 166nt primer. A longer exposure of tracks R(nc) and A^+(nc) from panel A shows multiple minor extension products beyond the transcription initiation site identified in Panel A.

Panel E: Schematic diagram, showing the position of the transcription initiation site relative to the ATG translation initiation codon.




A.
B.
C.
D.
E.
M
c
nc
t P R A t P R A
nc
A C G T
nc
A C G T
nc
R A+
587
540
504
458
434
267
234
213
192
184
Primer
+1
+20
+40
+60
+80
+100
+110
-300
-200
-100
+1
ATG
Sau3A
BalI

priming end of the non-coding strand primer lies 82nt downstream of the first nucleotide of the ATG translation initiation codon. These results therefore confirm that transcription initiation occurs 28nt upstream of the first nucleotide of the ATG codon.

In the left-hand half of Figure 12A, products seen with the coding strand as primer are presumably not the result of primer extension on the RBC 15-LOX mRNA, since they are also seen in controls using yeast tRNA and RNA from a non-erythroid rabbit tissue (pancreas). These products may represent "self-priming" of the primer i.e. DNA-dependent DNA polymerase activity of the reverse transcriptase enzyme. Although secondary structure within the primer should be minimised under the conditions of hybridisation used, such a long primer might be able to form such secondary structures. Some primer extension protocols (e.g. Jones et al, 1985) include actinomycin D to reduce the degree of self-priming (Williams and Mason, 1985), but actinomycin D at 50µg/ml had little effect upon the pattern of primer extension products obtained using this primer (data not shown).

3.2.7 mRNA sequencing using the 166nt BalI cDNA-derived primer.

Figures 12B and 12C confirm the mRNA sequencing with the 166nt primer seen in Figure 11A and, thus, the results obtained with the 58nt genomic primer. Readable, though not completely unambiguous, mRNA sequence matches that of the RBC 15-LOX gene from 83nt downstream of the ATG translation initiation codon, right the way up to the transcription initiation site.

As described in more detail in Section 3.2.10, primer extension with the 166nt cDNA-derived primer also suggests the existence of multiple minor transcipts containing RBC 15-LOX mRNA sequences (Figs. 11A and 12D).

In summary, the transcription initiation site of the RBC 15-LOX gene has been mapped by a combination of complementary techniques. Primer extension on a 58nt genomic DNA primer indicates that the

products of S1 nuclease analysis represent the 5′ end of the transcription unit and not an intron/exon boundary. Comparison with the S1 analyses provides evidence that only the longest primer extension products represent genuine RBC15-LOX mRNA 5′ ends. Finally these primer extension results have been confirmed using a second (cDNA-derived) primer, and mRNA sequencing has confirmed that priming is occurring on the RBC 15-LOX mRNA alone.

3.2.8 A genomic primer which extends beyond the predicted 5′ end of the RBC 15-LOX mRNA does not primer extend on rabbit reticulocyte RNAs.

The localisation of the transcription initiation site of the RBC 15-LOX gene can be confirmed in a number of ways (see also Section 3.2.9 below). For example, a primer which extends beyond the 5′ end of the messenger would not be expected to be capable of primer extension - even though its 5′ end is annealed to the mRNA, its 3′ end, from which primer extension occurs, overhangs the end of the mRNA. So, no primer extension would be expected using a primer consisting of genomic sequences from both sides of the predicted transcription initation site.

In designing an experiment to investigate whether this is the case for the RBC 15-LOX mRNA, it should be borne in mind that an absence of primer extension products might be due to a variety of factors other than an overhang of the 3′ end of the primer, including failure of the primer to anneal with the RNA during hybridisation, to inappropriate buffer conditions, or the inability of reverse transcriptase enzyme to catalyse polymerisation for some reason. So, a positive control experiment was performed using the 58nt Sau3A/SstI primer described above. The two primers were annealed with the same RNA preparation under identical hybridisation conditions, at a range of hybridisation temperatures (37, 42, 52 and 60^{o}C). All hybrids were then resuspended in aliquots from the same (enzyme + buffer) solution and primer extension using the two primers was performed simultaneously under identical conditions. Neither strand (86nt and 92nt respectively) of a HhaI/Sau3A fragment extending 23nt 5′ of

the transcription initiation site (see Fig. 6) is capable of primer extension with rabbit reticulocyte total RNA at any of these temperatures (data not shown). The 58nt positive control primer, however, extended to give the same qualitative result after hybridisation at all 4 temperatures.

The differing lengths of the test and positive control primers is unlikely to have been a factor in the efficiency of hybridisation of primers, since 67 nt of the 86nt non-coding strand primer under test are complementary to the mRNA, compared with 58nt of the Sau3A/SstI positive control primer. Furthermore, in a separate experiment (Section 3.2.6 above), the much longer 166nt BalI cDNA primer extended at all four of the temperatures used here, with the same qualitative results at each temperature, suggesting that primers of very different lengths can hybridise over the same range of temperatures.

Thus, results using the 86/92nt HhaI/Sau3A primer are consistent with the determination of the 5′ end of the RBC 15-LOX mRNA made above.

3.2.9 None of the available upstream sequences show detectable levels of hybridisation with Northern blots of rabbit reticulocyte RNAs.

To provide further confirmation of the location of the 5′ end of the mRNA, it has been possible to check that genomic sequences immediately 5′ of the predicted transcription initiation site are not present in the cellular population of stable, mature mRNAs, while sequences immediately downstream of this site are expressed in the final messenger. This has been done by probing duplicate Northern blots of total cellular RNAs with genomic DNA fragments from immediately upstream and immediately downstream of the predicted transcription initiation site (Fig. 13).

Hybridisation of sequences between 150 and 24bp upstream of the transcription initiation site, contained within a 127bp SstI/HhaI

Figure 13. Northern blot analysis of RBC 15-LOX transcripts.

Panel A: Probing of duplicate Northern blots of RNAs from various rabbit tissues using genomic probes derived from immediately upstream and immediately downstream of the rabbit RBC 15-LOX transcription initiation site.

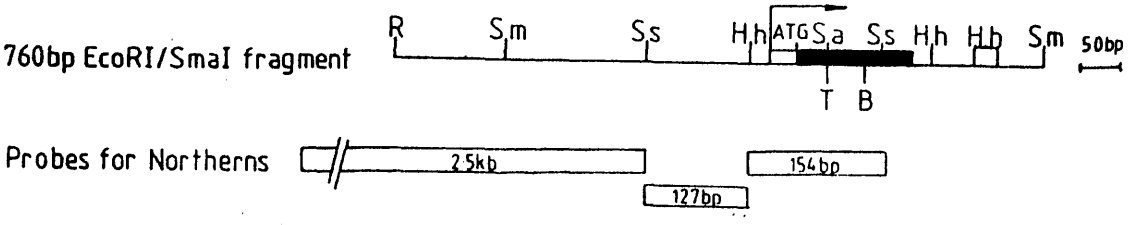
Probes: The 127bp SstI/HhaI fragment, extending from 150 to 24bp upstream of the transcription initiation site, and the 154bp HhaI/SstI fragment extending from 23bp upstream to 131bp downstream of this site, were both obtained from pLOX6/1.5R (Fig. 17), a sub-clone of the 1.5kb EcoRI genomic fragment (Fig. 3) ligated into EcoRI-cut pUC18.

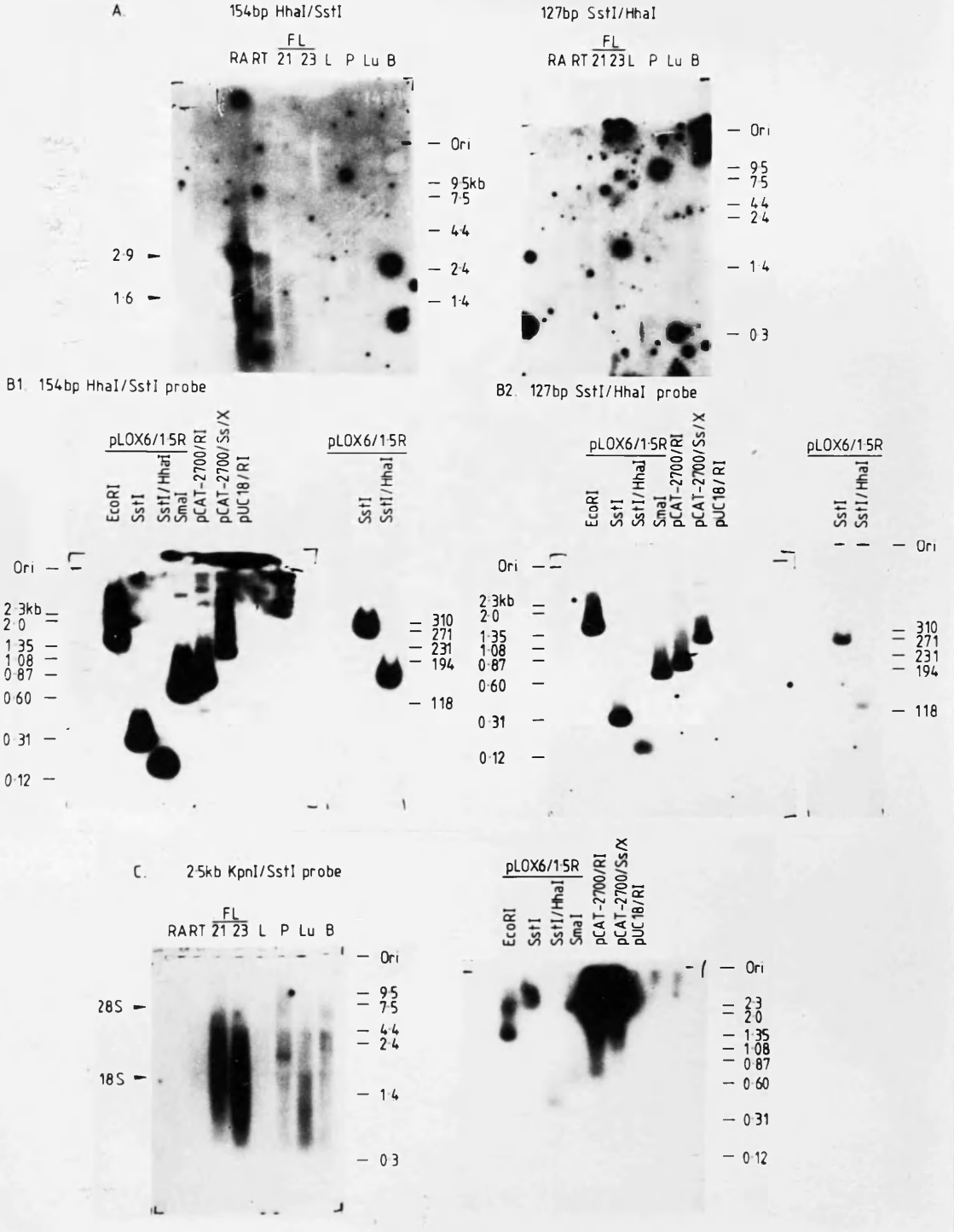
RNAs: R, rabbit reticulocyte total RNA; A^+, rabbit reticulocyte polyadenylated RNA; FL21 and FL23; total RNA from liver of 21- and 23-day gestation rabbit foetuses; L, total RNA from rabbit adult liver; P, total RNA from rabbit pancreas; Lu, total RNA from rabbit lung; B, total RNA from rabbit brain.

Panel B: Probing of duplicate Southern blots of various digests of pLOX6/1.5R (see Panel A above) and pCAT-2700 (Panel C below), using the same probes as in A. EcoRI-cut pUC18 (pUC18/RI) was used as a negative control for contamination of probes with vector sequences. pCAT-2700 plasmid digests: Ss/X = SstI and XbaI double-digest; R1 = EcoRI digest.

Panel C: Probing of Northern and Southern blots (duplicates of those used in Panels A and B above) with a 2.5kb KpnI/SstI fragment derived from pCAT-2700 (Fig. 18), an RBC 15-LOX/CAT recombinant which contains the entire 2.7kb of cloned RBC 15-LOX 5' flanking sequences, linked upstream of the bacterial CAT gene in pUC19 vector sequences.

154 and 127bp fragments were isolated using an 8% non-denaturing polyacrylamide gel. The 2.5kb KpnI/SstI fragment was isolated by preparative agarose gel electrophoresis (0.7% low melting-point agarose). Fragments were labelled by the random priming method and hybridised with the appropriate filters overnight at 42°C in 50% formamide; 10% dextran sulphate; 5xSSC; 1x Denhardt's solution; 100µg/ml polyA; 10µg/ml polyC; 0.1% SDS. Southern blots were hybridised in the same bag as Northern blots and filters were rinsed in 2x SSC; 0.1% SDS at room temperature, then washed in 0.1x SSC; 0.1% SDS at 65°C prior to autoradiography at -70°C (6-18 hours for Northerns; 1-3 hours for Southerns).




A.
154bp HhaI/SstI
127bp SstI/HhaI
FL
RA RT 21 23 L P Lu B
Ori
9·5kb
7·5
4·4
2·4
1·4
0·3
2·9
1·6
B1. 154bp HhaI/SstI probe
B2. 127bp SstI/HhaI probe
pLOX6/1·5R
EcoRI
SstI
SstI/HhaI
SmaI
pCAT-2700/RI
pCAT-2700/Ss/X
pUC18/RI
2·3kb
2·0
1·35
1·08
0·87
0·60
0·31
0·12
310
271
231
194
118
C.
2·5kb KpnI/SstI probe
28S
18S

fragment was compared with hybridisation of a 154bp HhaI/SstI genomic fragment containing sequences extending from 23bp upstream of the transcription initiation site to 131bp downstream. Results are shown in Figure 13A. Southern blots of various restriction digests of genomic sequences were hybridised with the same probes in the same bag as the Northern filters to control for the possibility that a lack of hybridisation might be due to poorly-labelled probe, or to inappropriate hybridisation conditions (Fig. 13B).

The 154bp HhaI/SstI genomic fragment hybridises strongly to two bands in the polyadenylated reticulocyte RNA track: one at 2.9kb, and one at 1.6kb. The 2.9kb band is of a similar size to previous estimates for the rabbit RBC 15-LOX mRNA employing three independent methods (agarose gel electrophoresis, sucrose gradient centrifugation and electron microscopy of an enriched mRNA population) (Thiele et al, 1982). A 2.9kb mRNA would be sufficient to encode the rabbit RBC 15-LOX protein, estimated by denaturing polyacrylamide gel electrophoresis to have a molecular weight of 68kD (Thiele et al, 1987). A 1.6kb messenger, on the other hand, is unlikely to contain sufficient coding sequence to be translated into the 68kD RBC 15-LOX glycoprotein, unless glycosylation is responsible for a large proportion of the molecular weight. Rapoport et al (1979), however, have calculated that the protein contains only 5% neutral sugars. Thus, it seems unlikely that the 1.6kb mRNA is responsible for producing the mature protein. The strength of the hybridisation, however, does suggest that this 1.6kb mRNA contains sequences very similar to those in this region of the RBC 15-LOX gene. Unlike the 2.9kb band, hybridisation to the 1.6kb band is not seen in the total reticulocyte RNA track. Therefore, the most likely explanation for the 1.6kb band is that it represents a degradation product present in the poly(A)$^+$ preparation, but not in the total RNA sample.

No clear hybridisation of the 154bp probe is seen with any of the total RNAs from other rabbit tissues investigated here, though there is perhaps the suggestion of weak hybridisation to the RNA of liver from 21-day gestation foetus, the major erythropoietic tissue of the rabbit foetus. The presence of RBC 15-LOX mRNA in foetal liver

has since been confirmed (J. Fleming, unpublished results). Thus, within the limitations of sensitivity of Northern blots, this result confirms the observations from primer extension and S1 nuclease protection experiments above, and the results of Thiele et al (1987), that the accumulation of RBC 15-LOX mRNA is erythroid-specific.

The 127bp SstI/HhaI probe, containing sequences between 150 and 23bp upstream, does not show hybridisation to poly(A)$^+$ or to total RNA from reticulocytes, nor to any of the RNAs from other rabbit tissues tested (liver, pancreas, lung and brain). The lack of hybridisation is not due to any defect in the labelling of the probe, as can be seen by the hybridisation of the probe to the expected bands in the two Southern blots shown in Fig. 13B. This failure of the 127bp HhaI/SstI fragment to hybridise with reticulocyte RNA is in agreement with the designation of the transcription initiation site by primer extension and S1 nuclease analyses above.

Sequences upstream of the transcription initiation site identified above would not be expected to be present in the mRNA population as RBC 15-LOX mRNAs. An exception to this would be if there existed a minor upstream RBC 15-LOX promoter whose transcript is spliced to an acceptor site downstream of the restriction sites used above to generate primers and S1 probes, and which would not, therefore,be detected using these probes. Alternative promoter usage combined with alternatively-spliced mRNA products is known in the expression of other genes, including a gene expressed in erythroid cells, the human porphobilinogen deaminase (PBG-D) gene (Grandchamp et al, 1987; Chretien et al, 1988). In erythroid cells, the 5′ end of the PBG-D mRNA is transcribed from an exon which is expressed exclusively in erythroid cells. In other cell types, however, transcription of the PBG-D gene is initiated from an alternative upstream promoter, and the 5′ end of the mRNA is derived from an alternative exon to that used in erythroid cells. All sequences other than those from the alternative 5′ exons are common to both mRNA species.

To investigate whether such an alternative transcription

initiation site might be present in the RBC 15-LOX gene, it was decided to test whether any of the available 2.7kb of genomic sequences upstream of the transcription initiation site identified above are present in the mRNA populations of various rabbit tissues. This was attempted by hybridising a 2.5kb KpnI/SstI genomic fragment, extending from the 5' limit of cloned RBC 15-LOX genomic sequences to 150bp upstream of the transcription initiation site, with Northern blots of RNAs from various rabbit tissues (Fig. 13C)

None of the sequences in this 2.5kb fragment hybridise to either total or poly(A)$^+$ RNAs from reticulocytes, suggesting that there is no upstream promoter in this region which is utilised in erythroid cells. Total RNAs from foetal liver, adult lung, pancreas and brain, however, show a smear of hybridisation with the 2.5kb fragment. The smearing does not seem to be due to degradation of the RNAs. The filter is a duplicate of those used in Fig. 13A, where discrete bands result, and all the total RNAs were judged to be intact by the appearance of 28S and 18S ribosomal RNAs after ethidium bromide staining of a fourth sample of each of the RNAs, run on the same gel. The integrity of the RNAs on this filter has not been checked by re-hybridisation with a cDNA probe from any ubiquitous rabbit mRNA, however. Smearing is not the result of some defect in preparation of the probe, since hybridisation to the Southern blot control is as expected. However, studies by other members of the group have shown that the 2.5kb region being used as a probe contains a repetitive sequence (B. Thiele, unpublished results). It is conceivable that the sequence of such a repetitive region might hybridise to sequences present in a variety of messengers, ranging across the entire 0.5kb to 7.0kb size range of the smear seen here. However, why none of these messengers should be present in reticulocytes or in adult liver is unclear.

The finding that none of the 2.7kb of available genomic sequences upstream of the putative transcription initiation site hybridise to rabbit reticulocyte RNA confirms that the 1.6kb mRNA detected using the 154bp probe (Fig. 13A), and suggested earlier to represent a RBC 15-LOX mRNA degradation product, does not arise from an alternative promoter within this 2.7kb region.

Overall, the data obtained from the probing of Northern blots with sequences from the 5′ end of the RBC 15-LOX genomic clone suggests that, within the limits of sensitivity of this technique, there is no transcription occurring in reticulocytes of sequences within the most 5′ 2.7kb of genomic sequences, other than from the transcription initiation site mapped above. This data does not exclude the possibility of alternative transcription sites lying outwith the available cloned sequences, nor the possibility that blotting has failed to detect low-level transcription of very short regions of 5′ flanking sequences.

3.2.10 Possible multiple minor sites of upstream transcription initiation detected by primer extension and S1 protection analyses.

A small proportion of the total population of transcripts from a variety of genes, including human and murine globins, have been shown by a combination of primer extension and S1 analyses to arise from sequences upstream of the transcription initiation site (Allan et al, 1983; Frampton et al, 1987). These minor upstream transcription initiation sites have been suggested to reflect sites which are important for the regulation of transcription, possibly by providing entry sites for the transcriptional machinery, including RNA Polymerase (Allan et al, 1983). Transcription from upstream sites of globin genes have been found to be associated with cell type-specific DNaseI hypersensitive sites (Zhu et al, 1984) and to alter with changes in development (Frampton et al, 1987). Whether this reflects initiation by mechanisms which differ from that at the major transcription initiation site, and which are of regulatory importance, or whether they reflect low-level, non-specific initiation of transcription by RNA Polymerase at sites of open chromatin structure is unclear.

3.2.10.1 Primer extension using the 166nt cDNA primer identifies multiple minor transcripts.

Although no transcripts containing sequences upstream of the RBC 15-LOX transcription initiation site were detected by Northern blotting (Section 3.2.9 above), apparent minor transcripts were seen in primer extension experiments using the 166nt cDNA-derived primer described in Section 3.2.6. Multiple minor extension products, above those corresponding to the RBC 15-LOX mRNA transcribed from the transcription initiation site identified above, can be seen in Figure 11A and in Figure12D, which is a longer exposure of the primer extensions on reticulocyte total and polyadenylated reticulocyte RNAs seen in Figure 12A, tracks ncR and ncA^{+}. Some of these minor products are 250-300nt longer than the major products.

Minor extension products were not seen with the 58nt genomic primer. This may be due to the fact that the specific activity of this primer was approximately 5-fold lower than that of the 166nt primer, as estimated by the duration of autoradiography required to obtain bands of roughly-equal intensity. Results obtained in mRNA sequencing using the 166nt primer suggest that the longer extension products seen in that case were all due to genuine RBC 15-LOX transcripts. In Figures 11A, 12B and 12C, a unique sequence can be read in several regions, all the way up to the transcription initiation site, suggesing that primer extension was occurring on only a single mRNA species. Nevertheless, one cannot be sure that the minor primer extension products are not artefacts resulting from low efficiency hybridisation of the primer to sequences in other RNAs which are similar, but not identical, to those of the RBC 15-LOX mRNA. If this were the case, it is possible that the signal resulting from sequencing of these other transcripts is simply not being detected.

In order to investigate whether any or all of these longer extension products do indeed represent genuine minor transcripts from the RBC LOX 15-gene, rather than coincidental hybridisation of the 166nt primer to other mRNAs, it is necessary to establish, as for the major transcription initation site, whether the extension products

correspond to protected fragments in an S1 nuclease analysis.

3.2.10.2 S1 nuclease analysis using the 540nt EcoRI/BalI genomic S1 probe also produces multiple minor products.

One problem with using S1 nuclease protection as a method to confirm the existence of minor transcripts is that any minor "protected" fragments seen by this method might represent incomplete digestion products. However, Figure 8A demonstrated that, using the longer 540nt EcoRI/BalI S1 probe, there were no detectable minor products using a variety of digestion conditions. There appeared to be very little effect on the nature of protected products of varying the amount of S1 nuclease used by 30-fold. Nor did a range of 15°C in the digestion temperature (27 to 42°C) appear to affect the distribution of products. It therefore seems unlikely that any S1 protected fragments larger than the main products are due to incomplete digestion.

Figure 8C shows a different protection experiment using the same 540nt EcoRI/BalI fragment as probe, but end-labelled and strand-separated on a separate occasion. The specific activities of the two preparations of probe were similar on each occasion. A small amount of protection is seen with the coding strand, but this is very small in comparison with the protection seen using the non-coding strand, and is presumably due to imperfect separation of the strands. With the non-coding strand, minor protected fragments are seen at least 150-200nt longer than the major protected species. However, the minor protected products seen in Figure 8C are from a three-times longer exposure than that in Figure 8A, as well as having been obtained on a separate occasion with a different preparation of end-labelled probe. In the absence of a longer exposure of the calibration experiment in Figure 8A, it is possible that minor products were not originally detected due to an insufficiently long exposure, and that they represent incompletely-digested products.

Furthermore, none of the protected fragment sizes seen in the S1 protection experiment in Figure 8C correlate with any of the more

intense minor products of the 166nt primer extension reaction (Figs. 11 and 12). Consequently, there is no firm evidence for genuine RBC 15-LOX minor upstream transcription initiation sites, since the possibility that the minor products of S1 protection and primer extension analyses are merely the result of two different sets of artefacts cannot be completely excluded, despite the attempts to do so in the experiments shown here.

As the question of what the minor products represent does not affect the conclusion as to the major site of transcription initiation, based on two sets of primer extension and two sets of S1 protection experiments, it was decided not to pursue the phenomenon of the minor products any further.

3.3 ORGANISATION OF RABBIT RBC 15-LOX GENE 5′ FLANKING SEQUENCES AND BINDING OF NUCLEAR PROTEINS.

Mapping of the transcription initiation site of the rabbit RBC 15-LOX gene to a position 28nt upstream of the ATG translation initiation codon (Section 3.2 and Fig. 14) permits analysis of the 316bp of currently-available upstream sequence for the presence of sequence elements found in the 5′ flanking regions of other genes. Of particular interest are the promoter regions of genes which are co-expressed with the RBC 15-LOX in erythroid cells and especially globin gene promoters.

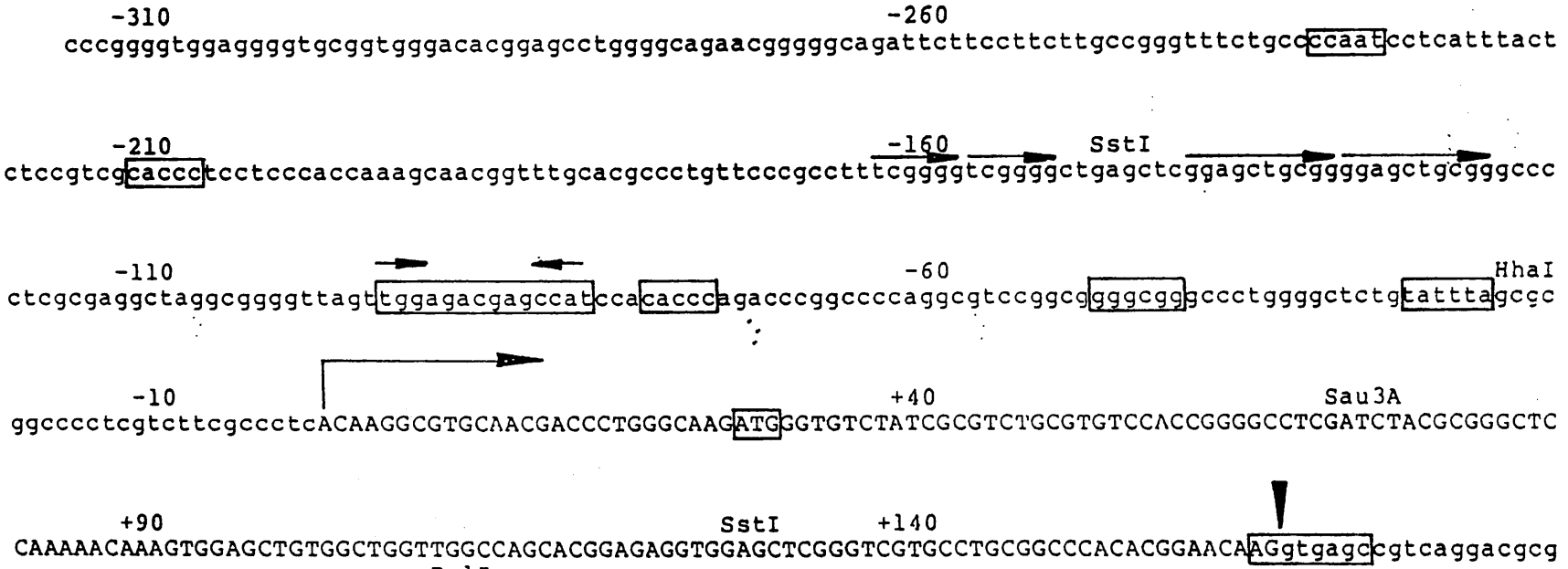
3.3.1 Sequence elements upstream of the RBC 15-LOX transcription initiation site.

As described in greater detail below, the 316bp of available 5′ flanking sequence are G/C-rich (69%) and include (Fig. 14) an A/T-rich TATA-like sequence, TATTTA, between nucleotides 30 and 25 upstream of the transcription initiation site, as well as a CCAAT, two CACCC, and GGGCGG sequences, and a sequence similar to the CTF/NF-1 recognition sequence, all of which have been found in a variety of other genes.

Figure 14. Position of the transcription initiation site and 5' flanking sequences of the RBC 15-LOX gene.

The position of the transcription initiation site (Section 3.2) within the 633bp SmaI genomic fragment sequenced in Section 3.1 is indicated by large angled arrow. The translation initiation codon is boxed. The putative splice donor sequence shown in Figure 4 is indicated by a downward-pointing arrow above the boxed splice consensus sequence. The 633bp SmaI fragment thus contains 316bp of 5' flanking sequences and 161bp of first exon sequence before the putative splice donor site. Exon sequences are shown in upper case letters; 5' flanking sequences and the first 18nt of likely first intron sequence are in lower case.

Sequences of interest within the 5' flanking sequences (Section 3.3.1) are boxed. These include CCAAT, CACCC and TATA-like sequences, and motifs similar to the recognition sequences for CTF/NF-1 and Sp1 transcription factors. Two pairs of direct tandem repeats are over-lined by horizontal arrows.


-310
-260
cccggggtggaggggtgcggtgggacacggagcctggggcagaacgggggcagattcttccttcttgccgggtttctgccccaatcctcatttact
-210
-160
SstI
ctccgtcgcacccctcctcccaccaaagcaacggtttgcacgccctgttcccgcctttcggggtcggggctgagctcggagctgcggggagctgcgggccc
-110
-60
HhaI
ctcgcgaggctaggcggggttagttggagacgagccatccacacccagacccggccccaggcgtccggcgggcgggccctggggctctgtatttagcgcc
-10
+40
Sau3A
ggcccctcgtcttcgccctcACAAGGCGTGCAACGACCCTGGGCAAGATGGGTGTCTATCGCGTCTGCGTGTCCACCGGGGCCTCGATCTACGCGGGCTC
+90
SstI
+140
CAAAAACAAAGTGGAGCTGTGGCTGGTTGGCCAGCACGGAGAGGTGGAGCTCGGGTCGTGCCTGCGGCCCACACGGAACAAGgtgagccgtcaggacgcg
BalI

The A/T-rich element, TATTTA, differs from the canonical TATA sequence originally identified in Drosophila histone genes (Goldberg, 1979; Breathnach and Chambon, 1981), as does the conserved ATA(A/T)AA element of mammalian globin genes (Efstratiadis et al, 1980) and A/T-rich regions in a similar position relative to the transcription initiation site in a variety of other genes. The similarity of position relative to the transcription initiation site with other A/T-rich, TATA-like elements in other genes may indicate that the TATTTA element of the RBC 15-LOX gene serves a similar function in determining the position of transcription initiation (Grosschedl and Birnstiel, 1980; Benoist and Chambon, 1981). It is unclear at present whether the various A/T-rich, TATA-like sequences are recognised by a family of transcription factors which recognise slightly different sequences, or whether it is the A/T-richness of this region which is important for interaction with a factor(s) involved in positioning RNA Polymerase II for transcription initiation.

CCAAT sequences are found in a wide variety of genes (Benoist et al, 1980; Efstratiadis et al, 1980; McKnight and Tjian, 1986) and there appear to be a family of transcription factors which bind to sequences which include a CCAAT motif (Santoro et al, 1988; Chodosh et al, 1988a, b; Dorn et al, 1987; Barberis et al, 1987). The CCAAT motif at -236 to -232nt in the RBC 15-LOX gene is further upstream of the TATA-like sequence than, for example, the CCAAT sequence of mammalian globin genes (Efstratiadis et al, 1980), but differing separations of CCAAT and TATA boxes are also seen in comparisons of other genes. For example, the separation of CCAAT and TATA sequences in mouse α1 globin is greater than that seen in the mouse β^{major}-globin gene (Nishioka and Leder, 1979; Konkel et al, 1979).

The more proximal of the two CACCC motifs (nucleotides -79 to -75 relative to the transcription initiation site) present in RBC 15-LOX 5' flanking sequences is contained within an octanucleotide sequence CCACACCC identical to that conserved between rabbit, mouse and human adult β-globin genes (Efstratiadis et al, 1980 and references therein) and similar sequences are also seen in other genes (e.g. an inverted copy in the rat tryptophan oxidase promoter; Schule

et al, 1988). In both rabbit β-globin and in mouse β^{major}-globin genes, the CACCC motif has been shown to be important for constitutive transcription initiation (Dierks et al, 1983; Charnay et al, 1985; Myers et al, 1986) and Mantovani et al (1988a) have recently reported the binding of both ubiquitous and cell type-specific proteins to the CACCC motif of the human β-globin gene.

The proximal CACCC motif of the RBC 15-LOX gene lies just downstream of a sequence element TGGAGACGAGCCAT very similar to the TGG(A/C)$(N)_5$GCCAA consensus binding sequence for the CTF/NF-1 family of transcription factors (Dynan, 1985; Jones et al, 1987; Santoro et al, 1988).

Between the proximal CACCC motif and the TATTTA, TATA-like element, there is a single copy of the sequence GGGCGG, located between -50 and -45nt. This is contained within a sequence GGGGCGGGCC which shows only one difference from the extended consensus sequence, (G/T)GGGCGGPuPuPy, for the binding of the "general" transcription factor, Sp1 (Kadonaga et al, 1986).

In addition to these sequences, present in a variety of genes, and known to be involved in the binding of transcription factors, the RBC 15-LOX 5′ flanking sequences also contain direct repeats of two short sequence elements. A perfect direct repeat of the hexanucleotide TCGGGG lies between 164 and 153nt upstream of the transcription initiation site. Immediately downstream is another direct repeat, of the decanucleotide GGAGCTGCGG. This repeat contains one perfect and one imperfect copy of the sequence element GGGAG, which is also found in the promoter regions of several other genes (See Discussion).

Sequence comparisons, in themselves, of course, tell us nothing as to whether a sequence element is important in the regulation of transcription. Ideally, the importance of particular regions should be tested in vivo (see Section 3.4 below). In addition, clues as to whether particular upstream sequences may be involved in the formation of a transcription pre-initiation complex can be obtained by analysing

the ability of nuclear proteins to bind to such sequences.

3.3.2 DNaseI in vitro footprinting of 5′ flanking sequences of the RBC 15-LOX gene.

In vitro DNaseI footprinting involves the introduction of "random" single-stranded nicks into "naked" end-labelled DNA, to produce, at appropriate low concentrations, a series of labelled fragments differing from each other by one nucleotide. In the presence of DNA-binding proteins, DNaseI cleavage is prevented in regions of the DNA which have been bound by protein, resulting in the absence of labelled fragments in a particular size range. The protected fragments can be observed by running the products of a limited DNaseI digestion on a polyacrylamide sequencing gel and comparing the products with a digestion of DNA digested in the absence of DNA-binding protein. Protected regions of the DNA appear as "footprints" in the ladder of labelled fragments on an autoradiograph (Galas and Schmitz, 1978).

In vitro DNaseI protection was performed using crude nuclear protein extracts from murine erythroleukaemia (MEL) and mouse brain cells, and from human HeLa cells. Nuclei were isolated from cells by lysis in isotonic sucrose (0.25M) containing 0.25% Triton detergent, and proteins extracted by addition of sodium chloride to 0.35M. A partial purification and concentration of nuclear proteins was achieved by precipitation using 0.35g/ml ammonium sulphate. Nuclear proteins from murine cells, rather than from rabbit tissues, were used to allow a comparison of *in vitro* protein binding data, and *in vivo* functional analysis (Section 3.4).

The crude nuclear protein extracts were used to footprint a 450bp EcoRI/BamHI fragment, 5′ end-labelled on only the non-coding strand, at the BamHI site. This labelled fragment consists entirely of RBC 15-LOX 5′ flanking sequences, except for 27nt downstream of the transcription initiation site: 20nt complementary to RBC 15-LOX mRNA 5′ untranslated sequences, and 7nt of pUC polylinker. Thus, distances of footprinted regions from the transcription initiation site were

calculated by subtracting 27nt from the size of protected regions, as measured by comparison with end-labelled pBR322/HaeIII markers.

3.3.2.1 Footprinting using murine nuclear protein extracts.

Figure 15 shows a preliminary experiment using crude nuclear protein extracts from MEL and normal mouse brain cells. Six possible regions of protection (designated FP1-6) by nuclear proteins from mouse brain and/or MEL cells are indicated, and the sequences lying within these regions are shown schematically in Figure 16. Of these, two (FP2 and FP4) show clear protection compared with the zero protein control tracks and, in the case of FP4, the acquisition of so-called "hypersensitive sites" (not to be confused with DNaseI HSS in chromatin structure) at the boundary of the protected region. In addition, FP1 may represent weak protein binding, but the possibility cannot be excluded that this merely represents a region of inefficient DNaseI digestion. Confirmation of the other three possible protected regions must await further footprinting studies.

FP2 lies between 75 and 106nt upstream of the transcription initiation site, and is protected by nuclear proteins from both the erythroid and non-erythroid cell-types. It includes the more proximal of the two CACCC motifs, as well as the TGGA$(N)_5$GCCAT sequence similar to the CTF/NF-1 recognition sequence. A combination of CACCC and CTF/NF-1-like recognition sequences are also seen, and have been footprinted, in chicken β^A- and β^H-globin genes (Plumb *et al*, 1986). The weak protection seen in the region of FP1 between 38 and 58nt upstream of the transcription initiation site, includes the GGGCGG motif, and also appears to be present with both erythroid and non-erythroid cell-types. No apparent protection was seen in either cell-type of two other sequences noted in sequence analysis of the upstream sequences - the TATTTA, TATA-like sequence and the decanucleotide repeat containing the GGGAG motif.

3.3.2.2 Footprinting using HeLa nuclear protein extract.

HeLa nuclear protein extracts have been widely used in DNaseI

Figure 15. In vitro DNaseI footprinting of the non-coding strand of the rabbit RBC 15-LOX promoter using murine and human nuclear proteins.

DNaseI protection of sequences adjacent to the transcription initiation site was performed using a 450bp EcoRI/BamHI fragment extending from 420nt upstream of the transcription initiation site. pCAT-2700 (Fig. 18) was kinase 5' end-labelled, then subjected to secondary digestion with EcoRI and the resulting 450bp EcoRI/BamHI fragment, labelled only at the BamHI site, isolated by agarose gel electophoresis. An aliquot of the recovered end-labelled fragment was incubated with varying amounts of nuclear protein extract for 90 minutes at 4°C. DNA/protein complexes were then subjected to limited DNaseI digestion for 30 seconds at room temperature. After termination of the digestion, products were extracted once with phenol/chloroform, once with chloroform, then ethanol precipitated and resolved on a 6% denaturing polyacrylamide gel. Autoradiography was for 9 days at -70°C.

Nuclear proteins used were from murine erythroleukaemia cells (MELC) and mouse brain cells and from human epithelial (HeLa) cells. Possible protected regions are designated FP1 to FP6 with increasing distance from the RBC 15-LOX transcription initiation site. Numbers 0, 20, 40, 60 and 80 indicate numbers of microlitres of nuclear protein extract used in the protein binding step prior to DNaseI digestion. M = end-labelled pBR322/HaeIII size markers. Sizes shown on the right indicate position relative to the transcription initiation site, obtained by subtracting 27nt (the distance of the transcription initiation site from the labelled BamHI end) from the size of the markers.

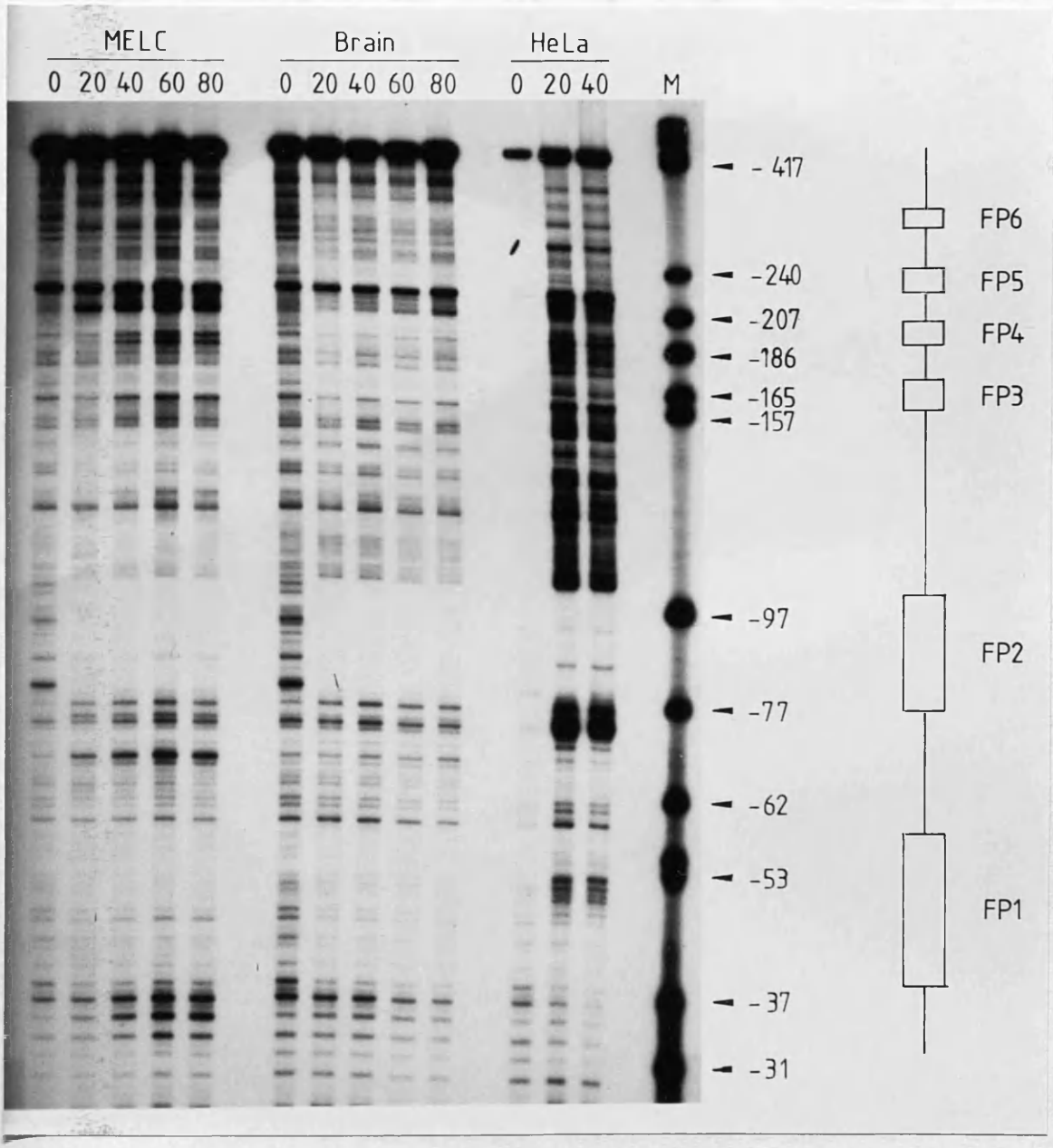
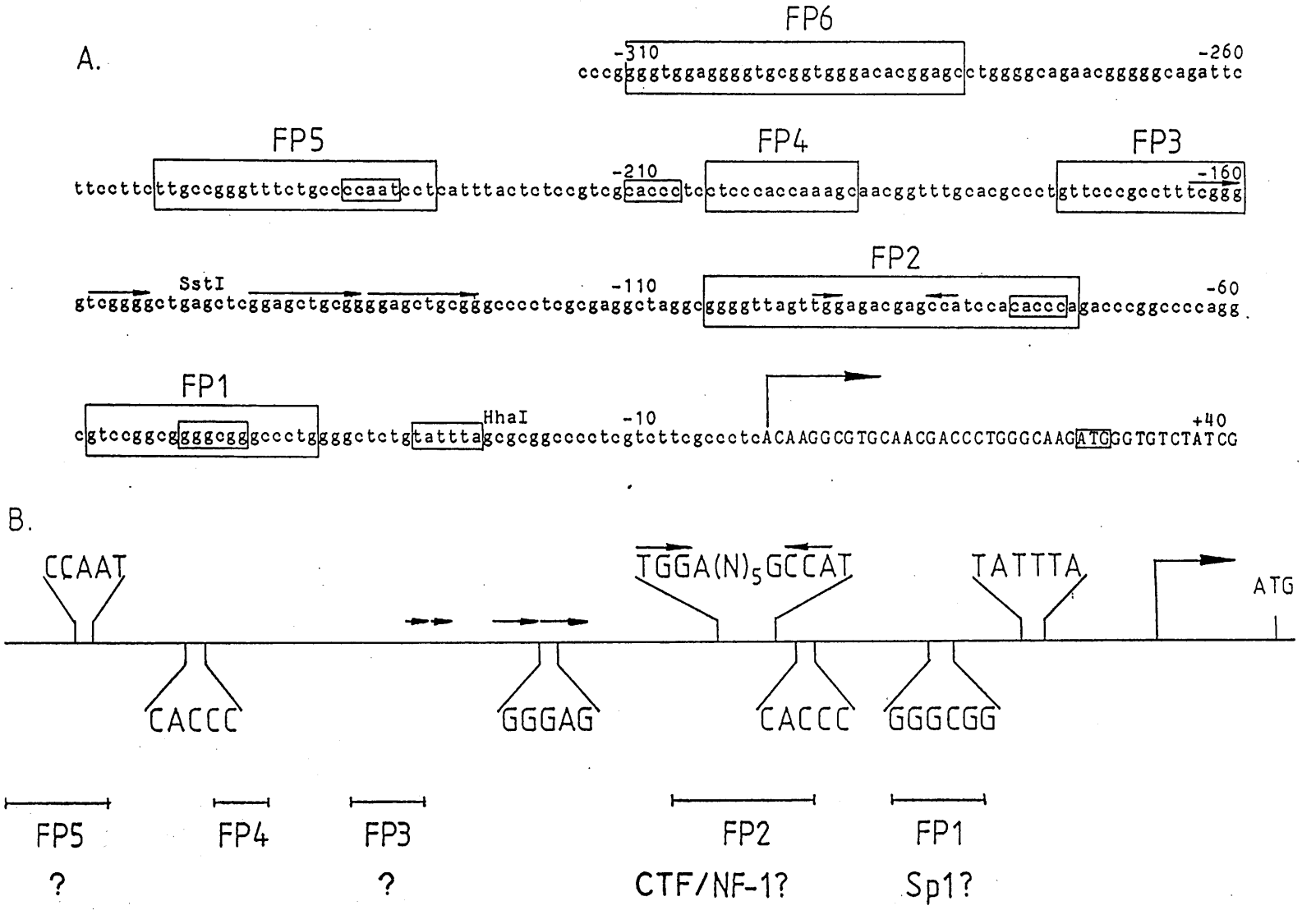

MELC
0 20 40 60 80
Brain
0 20 40 60 80
HeLa
0 20 40
M
- 417
-240
-207
-186
-165
-157
-97
-77
-62
-53
-37
-31
FP6
FP5
FP4
FP3
FP2
FP1

Figure 16. Schematic diagram of murine nuclear protein binding to the non-coding strand of the rabbit RBC 15-LOX promoter.

Panel A: Sequence of the rabbit RBC 15-LOX promoter and 5' end of exon 1. Possible protein-binding sites (FP1-6 as in Fig. 15) are enclosed in large boxes. Repeated sequences are shown overlined by arrows; other notable sequences (as in Fig. 14) are enclosed in small boxes. The transcription initiation site is indicated by a large arrow. Upper case characters, transcribed sequences; lower case characters, 5' flanking sequences.

Panel B: Schematic diagram of protected sequences, and other sequences of interest in the RBC 15-LOX promoter, as in panel A.


A.
FP6
-310
-260
cccggggtggaggggtgcggtgggacacggagcctggggcagaacgggggcagattc
FP5
FP4
FP3
-210
-160
ttccttcttgccgggtttctgccccaatcctcatttactctccgtcgcaccctcctcccaccaaagcaacggtttgcacgccctgttcccgcctttcgggg
FP2
SstI
-110
-60
gtcggggctgagctcggagctgcggggagctgcgggccccctcgcgaggctaggcggggttagttggagacgagccatccacacccagacccggccccagg
FP1
HhaI
-10
+40
cgtccggcggggcggccctgggctctgtatttagcgcggccccctcgtcttcgccctcACAAGGCGTGCAACGACCCTGGGCAAGATGGGTGTCTATCG
B.
CCAAT
TGGA(N)5GCCAT
TATTTA
ATG
CACCC
GGGAG
CACCC
GGGCGG
FP5
FP4
FP3
FP2
FP1
?
?
CTF/NF-1?
Sp1?

footprinting analyses, and for the isolation of a variety of nuclear proteins including the transcription factors CTF/NF-1 (Rosenfeld and Kelly, 1986; Jones et al, 1987) and Sp1 (Briggs et al, 1986; Kadonaga and Tjian, 1986).

Figure 15 also shows footprinting of RBC 15-LOX 5' flanking sequences using HeLa nuclear protein extract. The zero-protein control is unfortunately over-digested, but comparison of footprinted lanes can be made with the zero-protein controls next to the footprints obtained with murine nuclear proteins.

Very little can be deduced from this experiment regarding FPs 3-6. However, protection is seen in both FP1 and FP2 regions. Protection of FP2 is very similar to that seen with MEL and murine brain nuclear protein extracts; the FP1 pattern, however, is completely different. Here, there appear to be two closely-separated regions of protection. The region containing the GGGCGG motif (-50 to -45nt) lies between these two regions, and does not appear to be protected. As Sp1 was originally isolated from HeLa cells (Briggs et al, 1986; Kadonaga and Tjian, 1986) it would seem either that protection in the FP2 region with murine nuclear proteins is not due to Sp1 itself, or that binding of another protein(s) to adjacent sequences prevents binding of Sp1 to the GGGCGG motif. As noted earlier (Section 3.3.1) the decanucleotide containing the GGGCGG sequence in the RBC 15-LOX promoter differs from the extended consensus Sp1 binding sequence at one position. The affinity of Sp1 for this particular variant has not been compared with that for the consensus G/C box sequence (Kadonaga et al, 1986).

3.3.3 Proximal RBC 15-LOX upstream sequences may be involved in constitutive promoter function.

Footprinting experiments are still at a preliminary stage and much work will be needed to clarify the situation, particularly regarding sequences upstream of -150nt (FPs 3-6) where resolution is poor (a longer sequencing gel run was performed, but did not provide any additional information). Results presented here are from only a

single experiment and with only a limited number of nuclear protein extracts. Furthermore, the location of footprints has only been determined relative to end-labelled size markers; a more accurate positioning will be achieved by comparing DNaseI digestion products with a sequencing ladder obtained by chemical degradation of the same fragment (Maxam and Gilbert, 1977).

Nevertheless, the presence of "general" cis-regulatory sequences and the lack of erythroid-specific nuclear protein-binding in the proximal upstream sequences of the RBC 15-LOX gene may indicate that immediate upstream sequences constitute a basic promoter, concerned with ensuring accurate and efficient transcription initiation, while sequences located elsewhere may be involved in regulating transcription in a cell type-specific fashion. It is thus of interest to investigate whether proximal RBC 15-LOX 5′ flanking sequences do indeed function as a constitutive promoter, and whether sequences further upstream are involved in modulating promoter function, by analysing the in vivo function of the RBC 15-LOX flanking sequences in both erythroid and non-erythroid cell lines.

3.4 FUNCTIONAL ANALYSIS OF THE RABBIT RBC 15-LOX 5′ FLANKING SEQUENCES IN MURINE CELLS.

3.4.1 Strategy for investigating the function of the RBC 15-LOX promoter.

The strategy for investigating RBC 15-LOX promoter function in vivo can be summarised as follows: Various progressive deletions of the 5′ flanking sequences of the rabbit RBC 15-LOX gene were linked to the bacterial chloramphenicol acetyltransferase (CAT) gene. CAT enzyme activity was then assayed following transient expression of these RBC 15-LOX/CAT recombinants in erythroid and non-erythroid mouse cell lines, transfected by the calcium phosphate co-precipitation method. Expression of β-galactosidase from a co-transfected plasmid containing the lacZ gene was used as an internal control for variability in the efficiency of transfection.

In vivo functional analyses employing "chimaeric" genes, to investigate the ability of sequences of interest to direct the expression of a cis-linked heterologous gene, such as the bacterial CAT gene, are widely used as an assay for sequences involved in the regulation of transcription, including cell type-specific regulation (reviewed in Kelly and Darlington, 1985). Assays of transient expression following transfection of established cell lines by the calcium phosphate co-precipitation method are frequently used to rapidly assay such sequences (reviewed in Spandidos and Wilkie, 1984a). Although established cell lines in culture are very widely used in experiments designed to assess the function of nucleic acid sequences in vivo, the difficulties of extrapolating results obtained from transformed or partly-transformed cell lines, often at high passage number, to the varied cell types of normal living tissues in an intact organism, should be clearly borne in mind.

In the absence of any suitable rabbit erythroid cell lines, it was decided to express the RBC 15-LOX/CAT recombinants in murine cells. Functional analysis of heterologous gene expression has frequently been performed in cells from other species. Such studies have included the expression of various rabbit β-globin promoter deletions in both mouse 3T6 and L cell non-erythroid cell lines and in MEL erythroid cells (Dierks et al, 1981 and 1983; Wright et al, 1984). To investigate whether RBC 15-LOX 5′ flanking sequences contain any erythroid-specific regulatory elements, both erythroid and non-erythroid cell lines were used - the Friend leukaemia virus--transformed murine erythroleukaemia (MEL) line, F4.12B2, and the murine fibroblast-derived STO cell line.

The use of CAT assays in the functional analysis of transcriptional regulation is a highly-sensitive, but indirect, method of assaying levels of transcription. The effects of sequences of interest, introduced into a CAT-containing plasmid (either immediately upstream of the CAT gene in the case of promoter sequences, or at a distance in the case of enhancers) on transcription are investigated by measuring CAT enzyme activity. The CAT enzyme catalyses the acetylation, using acetyl Coenzyme A, of ^{14}C-labelled

chloramphenicol. The mobilities of acetylated chloramphenicol derivatives differ from that of the unacetylated molecule in thin layer chromatography, allowing separation of substrates and products of CAT enzyme activity. The proportion of chloramphenicol which has been converted to its acetylated derivatives provides a measure of this CAT activity. The use of radio-labelled substrate makes this an extremely sensitive assay, being capable of detecting as little as 0.02pmoles of product (0.5% conversion), but since enzyme activity, rather than the level of correctly-initiated transcripts, is measured, results must be confirmed by CAT mRNA analysis (see Section 3.4.8).

RBC 15-LOX/CAT recombinant plasmids containing progressive deletions of RBC 15-LOX 5' flanking sequences were derived from a pUC19-based vector, pCAT19 (Fig. 18) in which the CAT and lacZ transcription units are in the opposite orientation. The CAT gene fragment and transcription termination signals from the Herpes Simplex Virus (Type 2) Immediate Early-5 (HSV-2 IE-5) gene were obtained from the plasmid pLW2 (Fig. 17)

3.4.2 Variability in assays of transient expression of CAT enzyme activity.

The variability of results obtained in assays of transient CAT gene expression is well known. Most of the variability is due to variations in the efficiency with which plasmid DNA is introduced into cells. Thus, differences can be observed between the CAT activities from two transfected populations of the same cell type, not only because the plasmids may be expressed with different efficiencies, but also because there are different amounts of plasmid present in the nucleus, and, hence, available for expression. Only a very small proportion of the cell population take up and express plasmid DNA in transfections. Experiments in which thymidine kinase negative (TK^-) MEL cells were stably transformed to the TK-positive phenotype by calcium phosphate transfection of HSV TK-containing plasmids produced transformation frequencies of approximately 1 transformant per microgramme of donor DNA per 10^6 cells transfected (Spandidos and Paul, 1982). Such experiments, of course, involve an additional,

Figure 17. Plasmids used in the construction of RBC 15-LOX/CAT recombinants or as transfection controls.

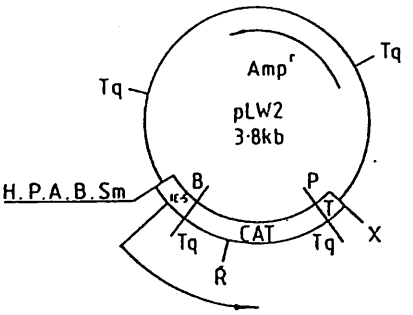
pLW2 (Gaffney et al, 1985) is a pUC9-based plasmid containing a 773bp TaqI fragment from the CAT gene. Transcription is driven by a 210bp fragment from the HSV-2 IE-5 gene promoter region (IE-5), including 90bp of 5′ flanking sequence. Downstream of the CAT gene, is a 100bp fragment (T) containing the transcription termination and 3′ end processing signals from the HSV-2 IE-5 gene. Promoter and terminator fragments are in the same orientation as in the IE-5 gene.

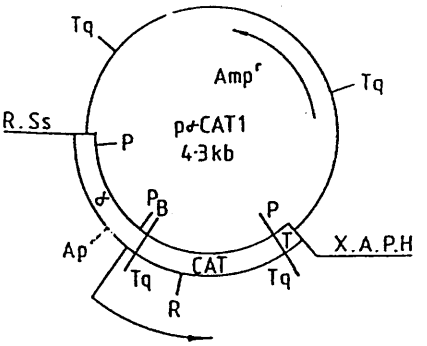
pHSV-βgal, used as internal control to standardise for transfection efficiency, is derived from the lacZ-containing plasmid pCH110 (Hall et al, 1983). Transcription of the lacZ gene is driven by the HSV-2 IE-5 promoter (IE-5), obtained from pLW2 as a 210bp BamHI fragment, via cloning into the BamHI site of the shuttle vector pIC20H (March et al, 1984), and inserted as a 310bp HindIII fragment into the HindIII site created by removal of the mouse mammary tumour virus promoter used to drive transcription in pCH110.

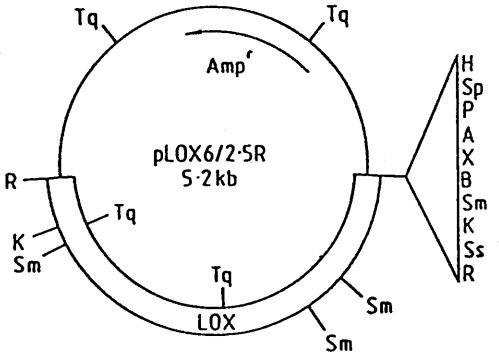
pLOX6/1.5R and pLOX2.5R contain the 1.5kb and 2.5kb EcoRI fragments respectively from the rabbit RBC 15-LOX gene (Figure 6) derived from the genomic recombinant λ8 (Thiele et al, 1987), and sub-cloned into EcoRI-cut pUC18 (Yanisch-Perron et al, 1985; Norrander et al, 1983)

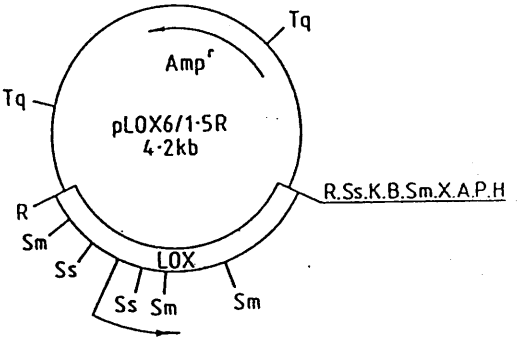
pαCAT1 contains a 750bp NcoI fragment from the mouse α1-globin gene. Since the 5′ overhang at the 3′ end of this fragment contains the ATG translation initiation codon, this sequence was removed by S1 nuclease treatment. Following end-filling, the blunt-ended fragment was inserted into SmaI-cut p22, a plasmid which was constructed (J. Lang, unpublished results) in an identical fashion to pCAT19 (Fig. 18), except that it is derived from pUC12 (Norrander et al, 1983) rather than from pUC19 (Yanisch-Perron et al, 1985) and which consequently has pUC sequences in the opposite orientation relative to the CAT gene and terminator sequences.

Restriction sites: B = BamHI; Sm = SmaI; K = KpnI; Ss = SstI; R = EcoRI; C = ClaI; E = EcoRV; X = XbaI; Bg = BglII; Xh = XhoI; N = NruI; H = HindIII; Pv = PvuII; Sp = SphI; P = PstI; A = AccI; Tq = TaqI; Ap = ApaI


Amp^r
pLW2
3·8kb
Tq
H.P.A.B.Sm
B
P
T
CAT
X
R


Amp^r
pα-CAT1
4·3kb
Tq
R.Ss
P
Ap
B
CAT
X.A.P.H
R


Amp^r
pLOX6/2·5R
5·2kb
Tq
R
K
Sm
LOX
Sm
H
Sp
P
A
X
B
Sm
K
Ss
R


Amp^r
pLOX6/1·5R
4·2kb
Tq
R
Sm
Ss
LOX
R.Ss.K.B.Sm.X.A.P.H

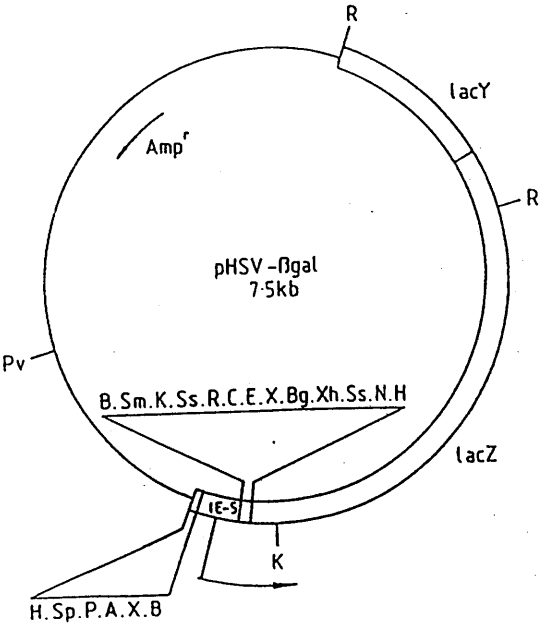

Amp^r
pHSV-ßgal
7·5kb
R
lacY
R
Pv
B.Sm.K.Ss.R.C.E.X.Bg.Xh.Ss.N.H
lacZ
IE-5
K
H.Sp.P.A.X.B

Figure 18. RBC 15-LOX/CAT recombinants used in the investigation of RBC 15-LOX promoter function.

pCAT19 is a "promoterless" CAT recombinant used as the basis for construction of all the RBC 15-LOX recombinants used in this study. It was obtained by the insertion of a 900bp XbaI/BamHI fragment from pLW2 (Fig. 17), containing the CAT gene and HSV-2 IE-5 transcription termination signals, into XbaI/BamHI-cut pUC19 (Yanisch-Perron et al, 1985), such that the lacZ and Amp^r transcription units of pUC sequences are in the opposite orientation to the CAT gene.

pCAT-40 contains RBC 15-LOX sequences extending from 40nt upstream of the transcription initiation site to 20nt downstream, obtained as a 60bp BstNI fragment from pLOX6/1.5R (Fig. 17), end-filled and blunt-end ligated into SmaI-cut pCAT19. Correct insertion of RBC 15-LOX sequences has been checked by sequencing.

pCAT-150 contains sequences from 150nt upstream of the transcription initiation site to 20nt downstream, obtained as a 170bp SstI/BstNI partial digest from pLOX6/1.5R (Fig. 17) and ligated into SstI/SmaI-cut pCAT19. Correct insertion of RBC 15-LOX sequneces has been checked by sequencing.

pCAT-2700 was obtained by co-ligation of a 1kb SstI/XbaI fragment from pCAT-150 (containing RBC 15-LOX sequences from -150 to +20, the CAT gene and HSV-2 IE-5 transcription termination signals) with a 2.5kb KpnI/SstI fragment from pLOX6/2.5R (Fig. 17) (containing sequences between 2.7kb and 150nt upstream) into KpnI/XbaI-cut pCAT19. Successful construction has been tested by extensive restriction enzyme mapping.

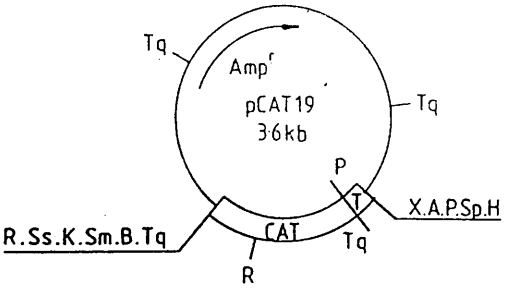
pCAT-316 contains a 330bp SmaI/BamHI fragment from pCAT-2700, including RBC 15-LOX sequences between -316 and +20, inserted into pCAT19. Reconstruction of the SmaI site has been checked by restriction digestion.

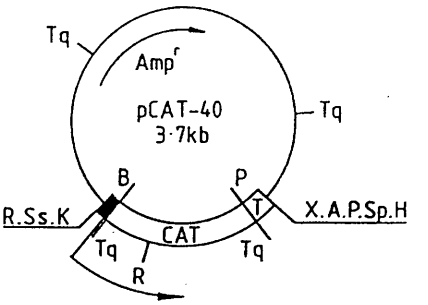
pCAT-450 contains 450nt of RBC 15-LOX 5' flanking sequences, obtained by insertion of a 700bp EcoRI fragment from pCAT-2700 into EcoRI-cut pCAT19.

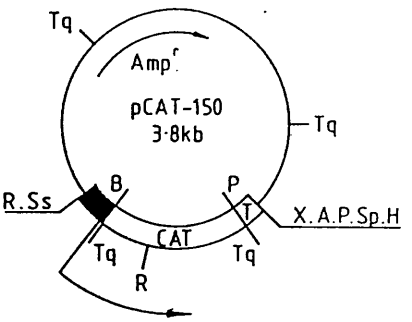
pCAT-1000 was obtained by cloning of a 700bp SmaI fragment from pCAT-2700, containing sequences between 1kb and 316bp upstream of the RBC 15-LOX gene into SmaI-cut pCAT-316. Successful insertion and reconstitution of SmaI sites was checked by restriction digestion.

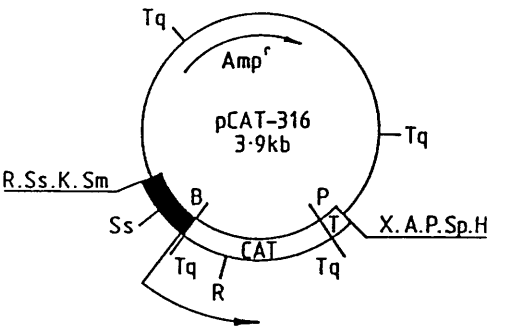
As in Figure 17, CAT is a 773bp TaqI fragment of the chloramphenicol acetyltransferase gene, and T is a 100bp fragment containing HSV-2 IE-5 gene transcription termination signals.

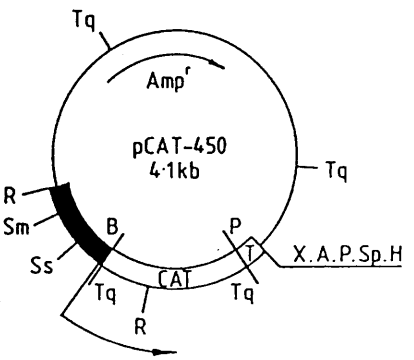
Restriction sites: R = EcoRI; Ss = SstI; K = KpnI; Sm = SmaI; B = BamHI; Tq = TaqI; X = XbaI; A = AccI; P = PstI; Sp = SphI; H = HindIII.

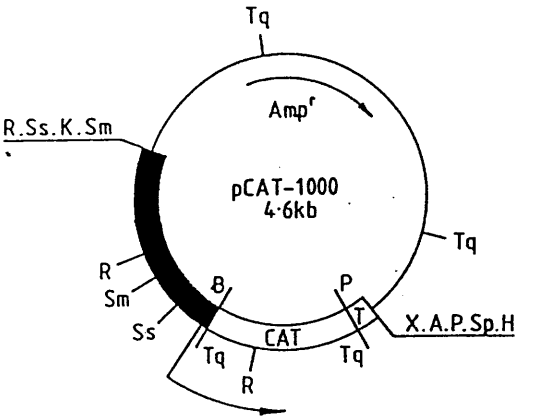

Tq
Amp^r
pCAT19
3·6kb
Tq
P
T
X.A.P.Sp.H
R.Ss.K.Sm.B.Tq
CAT
Tq
R


Tq
Amp^r
pCAT-40
3·7kb
Tq
B
P
T
R.Ss.K
X.A.P.Sp.H
CAT
Tq
Tq
R


Tq
Amp^r
pCAT-150
3·8kb
Tq
R.Ss
B
P
T
X.A.P.Sp.H
CAT
Tq
Tq
R


Tq
Amp^r
pCAT-316
3·9kb
Tq
R.Ss.K.Sm
B
P
Ss
T
X.A.P.Sp.H
CAT
Tq
Tq
R


Tq
Amp^r
pCAT-450
4·1kb
Tq
R
Sm
B
P
Ss
T
X.A.P.Sp.H
CAT
Tq
Tq
R


Tq
R.Ss.K.Sm
Amp^r
pCAT-1000
4·6kb
Tq
R
Sm
B
P
Ss
T
X.A.P.Sp.H
CAT
Tq
Tq
R

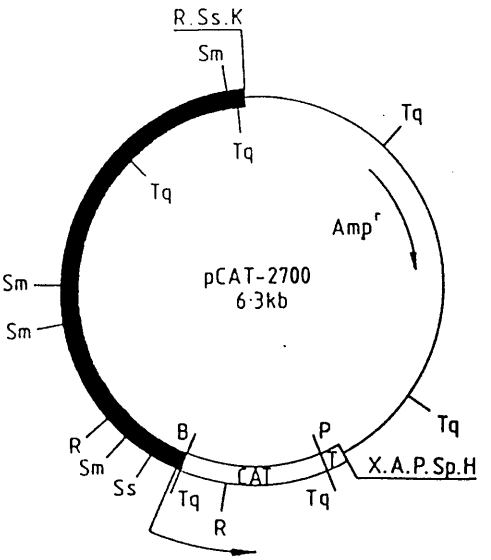
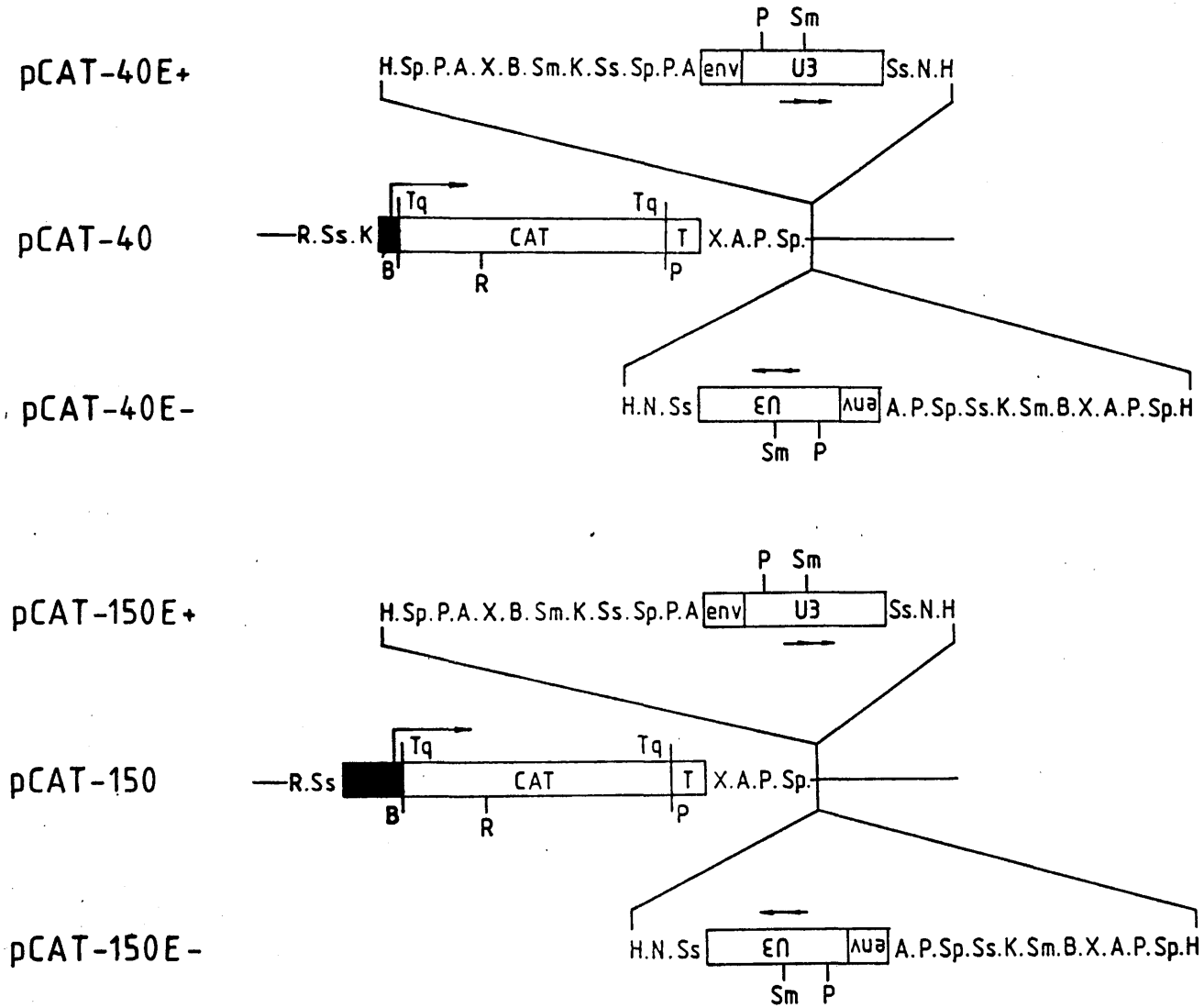

R.Ss.K
Sm
Tq
Tq
Tq
Amp^r
pCAT-2700
6·3kb
Sm
Sm
R
Sm
B
P
Tq
Ss
T
X.A.P.Sp.H
CAT
Tq
Tq
R

Figure 19. **RBC 15-LOX/CAT recombinants used in investigation of the response of the RBC 15-LOX promoter to a heterologous enhancer.**

A 550bp ClaI/SstI fragment from the 3' long terminal repeat (LTR) region of the Friend Murine Leukaemia provirus was modified by cloning into shuttle vectors (see below) such that each end of the fragment contained a HindIII site. The modified fragment was inserted, in either orientation (+ or -) relative to the direction of CAT gene transcription, into the unique HindIII site of the RBC 15-LOX/CAT recombinants pCAT-40 and pCAT-150 (Fig. 18), producing pCAT-40E (+ and -) and pCAT-150E (+ and -) plasmids. The proviral fragment includes the 3' end of the envelope gene (env) and a portion of the U3 region which includes the CCAAT motif, but not the TATA box or transcription initiation site. The imperfect 65/74bp repeats of the LTR enhancer (Bosze et al, 1986) are underlined by short arrows.

RBC 15-LOX sequences are shaded, and the transcription initiation site is indicated by a bent arrow. As in Figures 17 and 18, T represents the transcription termination and 3' end processing signals of the HSV-2 IE-5 gene. Multiple restriction endonuclease cleavage sites on either side of the inserted fragment were added during shuttling of the ClaI/SstI fragment into the polylinker regions of pUC12 and then pIC20H vectors (J. Frampton, unpublished results) during preparation of the fragment for insertion into the HindIII site of pCAT-40 and pCAT-150.

Restriction sites: R = EcoRI; Ss = SstI; K = KpnI; B = BamHI; Tq = TaqI; P = PstI; X = XbaI; A = AccI; Sp = SphI; H = HindIII; Sm = SmaI; N = NruI.


pCAT-40E+
P Sm
H.Sp.P.A.X.B.Sm.K.Ss.Sp.P.A
env
U3
Ss.N.H
pCAT-40
Tq
Tq
R.Ss.K
CAT
T
X.A.P.Sp.
B
R
P
pCAT-40E-
H.N.Ss
U3
env
A.P.Sp.Ss.K.Sm.B.X.A.P.Sp.H
Sm
P
pCAT-150E+
P Sm
H.Sp.P.A.X.B.Sm.K.Ss.Sp.P.A
env
U3
Ss.N.H
pCAT-150
Tq
Tq
R.Ss
CAT
T
X.A.P.Sp.
B
R
P
pCAT-150E-
H.N.Ss
U3
env
A.P.Sp.Ss.K.Sm.B.X.A.P.Sp.H
Sm
P

chromosome integration, step compared with transient expression experiments, and probably represent an underestimate of transfection efficiencies obtained in this case.

Experience has shown that variations in transfection efficiency using different calcium phosphate/DNA precipitates and/or different batches of cells are major contributors to varibility in transient expression experiments. In the experiments described here, various approaches were adopted to ty to reduce the variability in CAT assay results. The most important of these was the use of an assay for β-galactosidase activity, expressed from a co-transfected lacZ-containing plasmid, as internal standard for transfection efficiency (see Sectiaon 3.4.3 below).

To try to reduce variability between different precipitates, each transfection was performed in duplicate, and each set of transfections repeated on several separate occasions. On each occasion, two plates of each cell type were transfected with different calcium phosphate precipitates, made up using the same solutions, but mixed separately. These duplicates, though made using the same reagents are assumed to represent separate experimental determinations. Different preparations of some plasmids (pCAT19, pCAT-40 and pCAT-150; Fig. 18) were each used on several occasions. These different plasmid DNA preparations were obtained by the same method (alkaline lysis, followed by gel filtration column chromotography), but on separate occasions. Plasmid concentrations were checked by OD_{260} and OD_{280} measurements, and the proportion of supercoils in each preparation were shown to be equivalent by running out 100ng aliquots on agarose gels. In all cases, the same total amount of DNA (20μg) was added to a fixed number of cells. Different precipitates, containing recombinants of different sizes, each contained the same amount (2.4pmoles) of the plasmid DNA under test. Since the RBC 15-LOX and CAT gene fragments are contained in pUC vector sequences, the amount of DNA per transfection was made up to a constant total of 20μg per 1ml precipitate using pUC18 plasmid DNA.

In order to compare the expression of a given precipitate in

erythroid and non-erythroid cell types, half of the precipitate was added to a dish of cells of each type. These are assumed to represent, as far as possible, an identical transfection, carried out on different cell types.

To try to reduce variability between transfections performed on separate occasions, using different batches of cells, low passage number cells (usually less than 20 i.e. 8-10 weeks of continuous culture) were plated out at $1x10^6$ cells per 9cm Petri dish to try to ensure healthy growth of cells and optimum expression of introduced sequences throughout the 72 hours duration of each transient expression experiment. Transfection efficiency was found to be lower with cells plated at $3x10^6$ cells per dish (data not shown). Transfections of the semi-adherent MEL cell line used were only performed on populations containing approximately 30-40% of cells showing elongated morphology. Each dish of cells was plated from the same stock of cells, maintained in logarithmic growth at all times by seeding at low cell density in fresh medium every 2-3 days.

3.4.3 β-galactosidase enzyme assay as internal standard for transfection efficiency.

The use of β-galactosidase enzyme activity as an internal standard to correct CAT activities for transfection efficiency involves co-transfection (with equal molar quantities of each of the CAT recombinant plasmids under test) of a fixed amount (10μg; 2pmoles) of a plamid containing the bacterial lacZ gene, which encodes the β-galactosidase enzyme. In the internal control plasmid used here, pHSV-βgal (Fig. 17), transcription of the lacZ gene was under the control of the HSV-2 IE-5 gene promoter. This promoter was obtained from the plasmid pLW2 (Fig. 17) which is expressed efficiently in both cell types used here (Fig. 20).

The assay for β-galactosidase activity relies upon the enzymic cleavage of the chromogenic substrate o-nitrophenyl-β-galactopyranoside, producing o-nitrophenol, which has a yellow colour, and whose concentration can thus be measured spectrophotometrically at 420nm.

Correction of CAT enzyme activities on the basis that β-galactosidase activity reflects the transfection efficiency of CAT- and lacZ-containing plasmids makes several assumptions. Firstly, it assumes, like the CAT enzyme assay, that the activity of the β-galactosidase enzyme reflects, in direct proportion, the amount of mRNA transcribed and available for translation, and the amount of plasmid taken up and available for transcription. As for CAT assays, the amount of lacZ mRNA can be checked by RNA analysis (see Section 3.4.8). Further, the efficiency of expression of the lacZ plasmid in each dish of cell is assumed to be constant for each of the transfections in a given cell type on a given occasion. No assumption is made concerning the ability of different cell types to express β-galactosidase, though the HSV IE-5 promoter used to drive the lacZ gene in pHSV-βgal is known to work well in both cell types used here when driving expression of the CAT gene in the plasmid pLW2 (J. Frampton, unpublished results; and Fig. 20). It is also assumed that, in all cases, the molar ratio of lacZ and CAT plasmids available for expression is the same. This assumption would be unnecessary if the lacZ and CAT genes were physically linked in the same plasmid. However, in this case, interpretation of results might be further confused by the possibility of transcriptional interference between the two genes.

All β-galactosidase assays were performed in substrate excess so that the amount of product produced was never more than 10% of the substrate originally present. β-galactosidase activity was then used to correct for transfection efficiency at two stages. Firstly, for each set of transfections with each cell type, amounts of extract containing equivalent amounts of β-galactosidase activity were used in CAT assays. This entailed the use of different amounts of extract in each CAT assay. As the range of the amounts of protein from a given cell type in a given set of transfections was never more than 4- to 5-fold, it was assumed that the varying amounts of protein used in each CAT assay had no effect upon reaction rate. (The amounts of protein used were such that the CAT assay reaction was always in substrate excess, percentage conversions usually being less than 20%)

Secondly, all the CAT activity results from several experiments were corrected to correspond to a standard β-galactosidase activity (see Section 2.14.6). Percentage conversions from 5 experiments, corrected in this way, are shown in Table I. In most cases, the correction factor involved was small (typically 0.5 to 2 times) with the result that the corrected percentage conversions are similar to those obtained directly by scintillation counting of the CAT assay products. However, for some CAT recombinants which are highly expressed, such as pLW2, even correction factors as small as these result in corrected percentage conversions which are greater than 100% i.e. the amount of extract which would produce the standard amount of o-nitrophenol during the β-galactosidase assay would, if used in the CAT assay, have exhausted the chloramphenicol substrate before the end of the CAT assay.

As can be seen in Table I, the use of a β-galactosidase assay as internal standard for efficiency of transfection usually results in relatively little variability between duplicates in a given experiment. Although there are large variations in CAT expression from a given recombinant plasmid in different experiments, the relative expression from different recombinants in a given cell type is much less variable.

3.4.4 Strategy to define the minimal functional promoter.

To establish the minimal region required to function as a promoter in the cell types chosen, two initial RBC 15-LOX/CAT recombinants, pCAT-40 and pCAT-150, were constructed (Fig. 18) from the plasmid, pCAT19, which contains the CAT gene and the HSV-2 IE-5 transcription termination signal from the plasmid pLW2 (Fig. 17) but lacks any heterologous promoter sequences. In the construction of pCAT19 and its derivatives, the CAT transcription unit is inserted into pUC19 polylinker sequences in the opposite orientation to the direction of transcription of the lacZ gene to try to avoid transcription initiation due to vector sequences. In fact, levels of acetylation of chloramphenicol in CAT assays using this vector do not show any significant difference from those seen with a "promoterless"

CAT vector with CAT and lacZ transcription units in the same relative orientation (data not shown).

pCAT-40 contains sequences from 40bp upstream of the transcription initiation site to 20bp beyond the initiation site. This recombinant therefore contains most of the 5′ untranslated region of the wild-type RBC 15-LOX mRNA, but LOX sequences are truncated upstream of the ATG translation initiation codon, preventing any abnormal hybrid RBC 15-LOX/CAT translation products. If transcription initiation occurs from the normal RBC 15-LOX transcription initiation site, then the 5′ untranslated region of the CAT mRNA produced from pCAT-40 should contain only 20nt of LOX sequences, 9nt of pUC polylinker, and the final 30nt of the usual 5′ untranslated region of the CAT mRNA.

pCAT-150 is identical to pCAT-40, except that it contains an additional 110bp of RBC 15-LOX 5′ flanking sequences, extending to 150bp upstream of the RBC LOX transcription initiation site. The RBC 15-LOX/pUC polylinker/ CAT junctions of both pCAT-40 and pCAT-150 have been sequenced and are identical in the two plasmids.

Expression of CAT enzyme from these two RBC 15-LOX/CAT recombinants and from the "promoterless" parent plasmid vector, pCAT19, were determined in 5 independent experiments using two different batches of each recombinant to control for variations between plasmid DNA preparations. Representative autoradiographs are shown in Fig. 20. The percentage conversions of chloramphenicol to its acetylated derivatives, corrected for equivalent amounts of β-galactosidase activity as in Section 3.4.3, are shown in Table I. The mean values for each set of transfections, and the standard error of the mean for each set, are represented as histograms in Figure 21.

Figure 20. CAT expression driven by the rabbit RBC 15-LOX promoter in murine cell lines, and response to a heterologous enhancer.

Representative autoradiographs of the products of CAT enzyme activity in extracts from MEL and STO murine cell lines transfected with various RBC 15-LOX/CAT recombinant plasmids. pCAT19 is the "promoterless" parent vector for pCAT-40 and pCAT-150, containing 40 and 150nt of RBC 15-LOX 5' flanking sequences, respectively, linked upstream of the CAT gene in pCAT19. pCAT-40E+ and - and pCAT-150E+ and - contain the Friend Murine Leukaemia virus LTR enhancer inserted in either orientation (+ or -) downstream of the CAT gene in pCAT-40 and pCAT-150 respectively. pLW2 is a positive control containing the CAT gene driven by the HSV-2 IE-5 promoter (For further details, see Figs. 17, 18 and 19, and text)

Cells were transfected with calcium phosphate/DNA co-precipitates containing 3.5pmoles of the RBC 15-LOX/CAT recombinant under test, 2pmoles (10µg) of pHSV-ßgal (Fig. 17) as internal standard, and pUC18 to a total of 20µg DNA. Results shown were obtained in a 30-minute assay at 37^{o}C using amounts of extract showing equal amounts of ß-galactosidase activity. Percentage conversions seen here range from 0.5% (pCAT19, pCAT-40, pCAT-40E+ and - in STO cells) to approximately 90% conversion (pLW2 in MEL cells). Autoradiography was for 4 days at room temperature.

Percentage conversions from this experiment, after standardisation for ß-galactosidase activity to permit comparison with results from other experiments (Section 2.14.6), are shown as Experiment 2 in Table I.

Cm = chloramphenicol; 1-Ac-Cm, 3-Ac-Cm and 1,3-di-Ac-Cm = mono- and di-acetylated chloramphenicol derivatives.

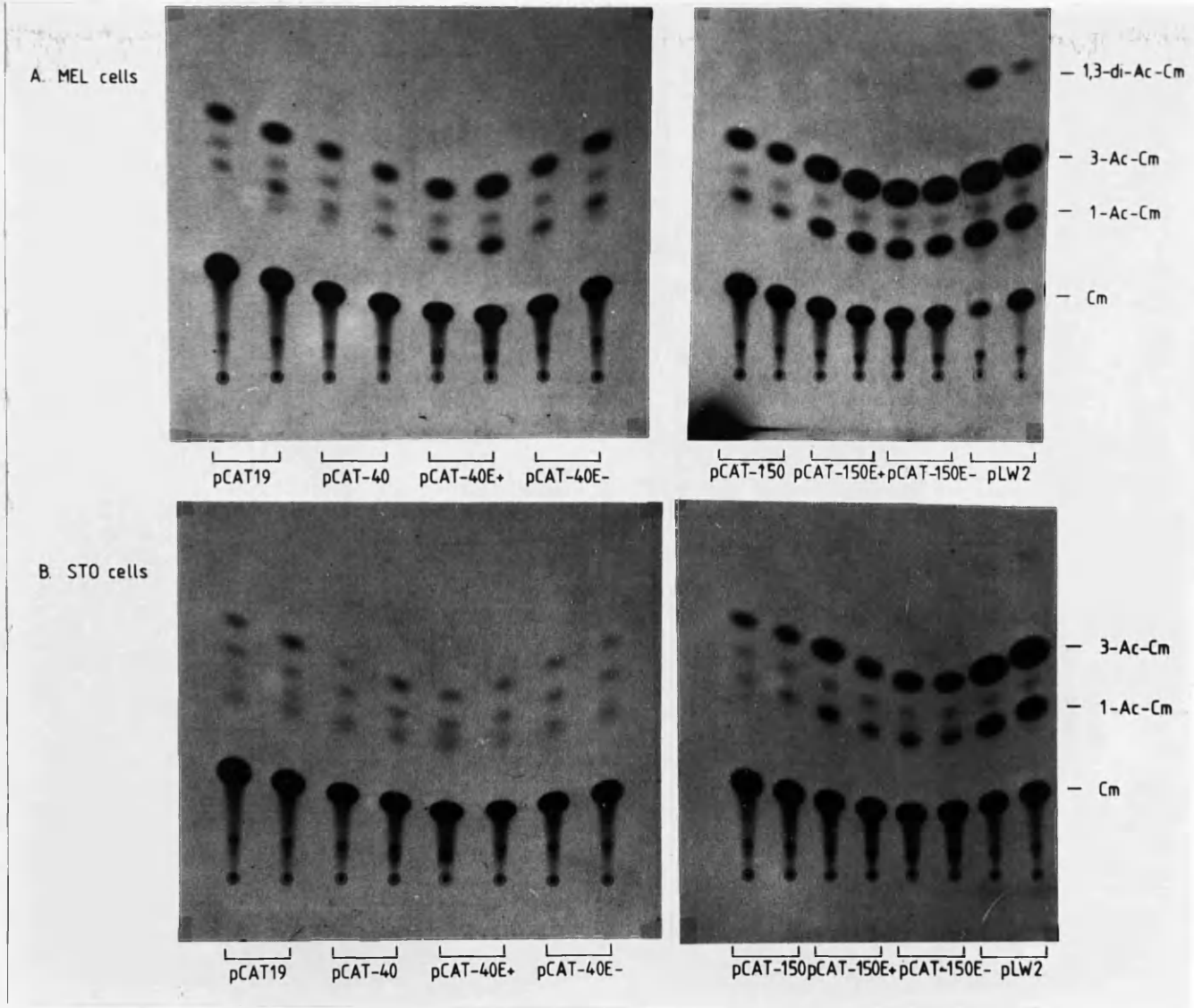

A. MEL cells
1,3-di-Ac-Cm
3-Ac-Cm
1-Ac-Cm
Cm
pCAT19
pCAT-40
pCAT-40E+
pCAT-40E-
pCAT-150
pCAT-150E+
pCAT-150E-
pLW2
B. STO cells
3-Ac-Cm
1-Ac-Cm
Cm
pCAT19
pCAT-40
pCAT-40E+
pCAT-40E-
pCAT-150
pCAT-150E+
pCAT-150E-
pLW2

Table I. CAT activities of various RBC 15-LOX/CAT recombinants, corrected for β-galactosidase activity.

Corrected percentage conversions are shown for 5 independent sets of transfections (Experiments 1-5). All except experiment 5 were performed as duplicate transfections (a and b). For each cell line, in each experiment, CAT assays were performed on amounts of extract containing equal β-galactosidase activity. After determining the proportion of the amount of radio-label converted to acetylated products by scintillation counting, this amount was expressed as a proportion of the total, and percentage conversions from all transfections were corrected so that they all represent the CAT activity which would be obtained from that quantity of extract containing a standard β-galactosidase activity under standard assay conditions (see Sections 2.14.4 and 2.14.6).

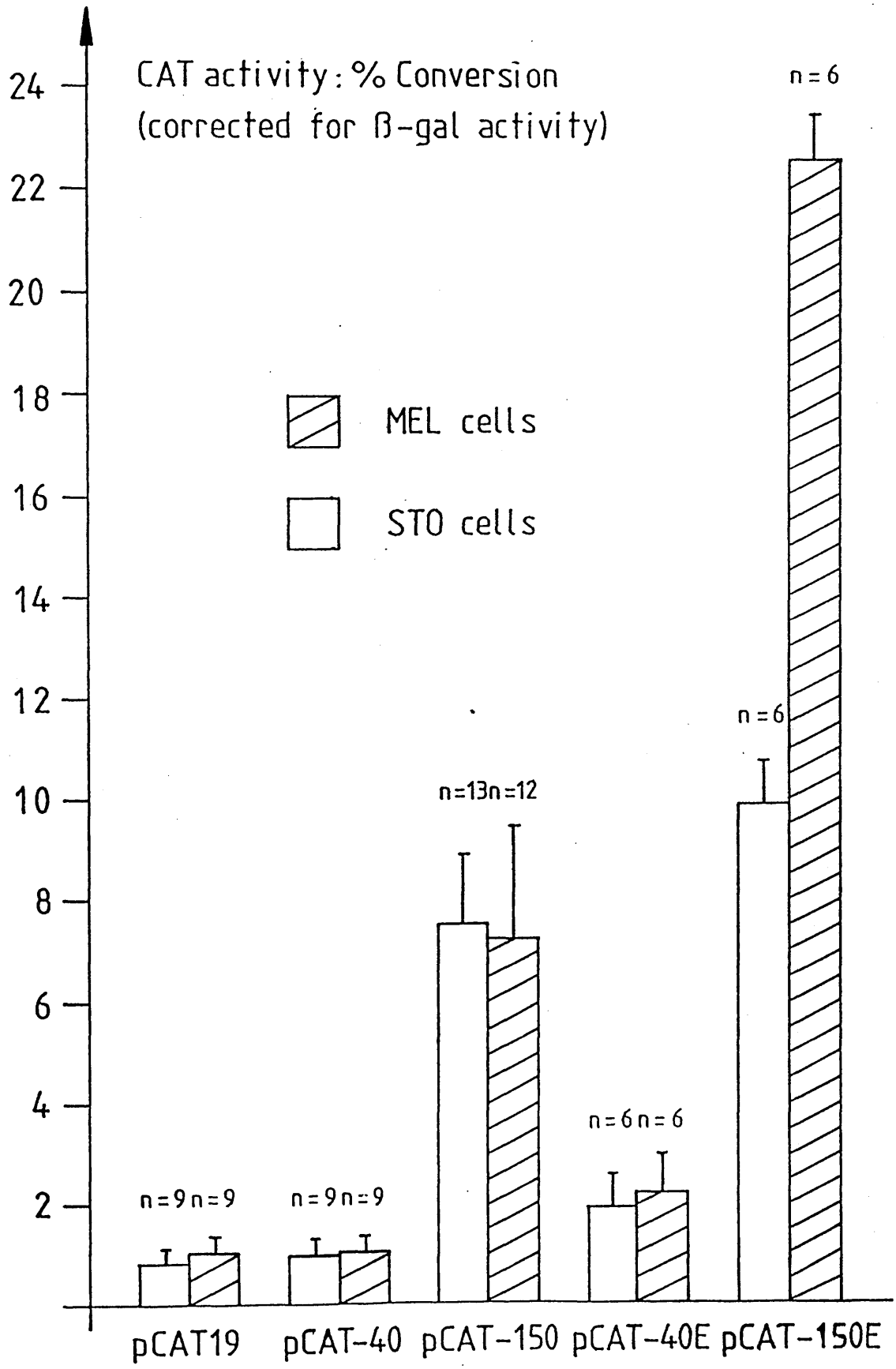
CAT ACTIVITIES

Percentage conversions corrected for β-galactosidase activity

Experiment		1a	1b	2a	2b	3a	3b	4a	4b	5
pCAT19	MEL	2.7,	1.7	0.4,	0.9	0.9,	0.1	0.4,	0.6	1.8
	STO	1.9,	2.2	0.9,	1.5	0.4,	0.5	0.1,	0.1	0.1
pCAT-40	MEL	2.9,	2.3	0.5,	0.7	0.1,	0.4	0.4,	0.3	1.7
	STO	2.0,	2.0	2.1,	0.8	0.3,	1.2	0.1,	0.2	0.3
pCAT-150	MEL	5.9,	7.5	2.3,	2.1	3.1,	2.9	7.5,	7.3	30.3
		10.8,	4.8	2.4,	—					
	STO	11.8,	9.1	3.4,	5.0	8.8,	8.9	2.2,	1.9	3.1
		17.6,	14.4	4.4,	7.7					
pCAT-330	MEL					1.0,	2.1	6.2,	5.3	22.9
	STO					5.7,	9.5	1.4,	1.4	2.0
pCAT-450	MEL					3.5,	1.7	5.4,	5.4	18.1
	STO					12.5,	7.8	1.4,	1.4	2.7
pCAT-1000	MEL							6.3,	8.7	32.5
	STO							1.3,	1.0	1.8
pCAT-2700	MEL	13.7,	9.1			1.3,	1.6	7.5,	7.4	29.0
	STO	4.3,	—			0.8,	0.8	0.3,	0.3	0.8
pαCAT1	MEL					5.9,	6.3	26.7,	24.8	11.2
	STO					3.0,	3.5	1.8,	1.2	2.1
pLW2	MEL	223,	213	112,	64					280
	STO	—	109	121,	107					91
pCAT-40E +	MEL			2.2,	1.4	5.9,	1.6			
pCAT-40E −	MEL			1.3,	1.0					
pCAT-40E +	STO			1.7,	1.4	1.4,	0.4			
pCAT-40E −	STO			1.5,	5.2					
pCAT-150E +	MEL			20.3,	24.7	21.6,	21.9			
pCAT-150E −	MEL			25.5,	21.5					
pCAT-150E +	STO			8.5,	11.9	8.6,	9.7			
pCAT-150E −	STO			12.8,	8.0					

Figure 21. Basic RBC 15-LOX promoter function and response to a heterologous enhancer.

CAT activities (after standardisation for β-galactosidase activity) from several independent transfections of pCAT-40 and pCAT-150 recombinants in MEL and STO cells, and the effect on these sequences of Friend murine leukaemia virus LTR enhancer, placed downstream of the CAT gene (Figs. 18 and 19). Sizes of bars represent the mean; error bars the Standard Error of the Mean of figures from several experiments, as presented in Table I. Figures above the bars indicate the number of determinations (n), using independent calcium phosphate/DNA co-precipitates, for each RBC 15-LOX/CAT recombinant. pCAT-40E and pCAT-150E represent the combined figures for determinations of the effect of the enhancer in both orientations (+ and -) relative to RBC 15-LOX sequences.


CAT activity: % Conversion
(corrected for ß-gal activity)
24
22
20
18
16
14
12
10
8
6
4
2
MEL cells
STO cells
n=9 n=9
n=9 n=9
n=13 n=12
n=6 n=6
n=6
pCAT19
pCAT-40
pCAT-150
pCAT-40E
pCAT-150E

3.4.5 Sequences between 40 and 150bp upstream of the rabbit RBC 15-LOX transcription initiation site are sufficient to elevate expression of a linked CAT gene in both erythroid and non-erythroid murine cells.

A low level of acetylation of chloramphenicol (0.3%) is seen, under the standard conditions used, with extracts from cells which are transfected using either pUC18 plasmid alone, or with pHSV-βgal and pUC18 sequences, but with no CAT recombinant (see Fig. 22). This background acetylation is presumably due to cellular acetylases. Such acetylase activity must be heat-stable to survive the heat treatment used prior to the assay for CAT activity. This background level was subtracted from each of the values for percentage conversion obtained with each of the transfections using CAT recombinants.

An additional low level of conversion is seen in cells of both types transfected with the "promoterless" CAT plasmid, pCAT19 (see Fig. 22). This additional level of conversion is presumed to be due to the presence of the CAT-containing plasmid and therefore due to CAT enzyme activity. As explained above, the design of the pCAT19 plasmid should prevent initiation of CAT gene transcription from either lacZ or Ampicillin-resistance promoters, but it is possible that the low level of CAT expression is due to initiation elsewhere in the pUC vector sequences.

Expression of CAT activity from pCAT-40 is not significantly higher than the pCAT19 background level in either cell type (Figs. 20 and 22 and Table I). Thus, the TATA-like element of the RBC 15-LOX promoter, on its own, is insufficient to function as an efficient promoter and requires additional sequences for high levels of transcription. This is consistent with other investigations of TATA-like regions of other genes. For example, the TATA-box of the SV40 early promoter is dispensable for high-level expression of SV40 T antigen in monkey cells (which does require additional upstream sequences), but has a key role in defining the site of transcription initiation (Benoist and Chambon, 1981)

In determining cell type-specificity in the expression of heterologous genes in different cell lines requires a comparison to be made between the recombinant under investigation, and some other reference recombinant. Choosing a reference CAT plasmid which has no preference for expression in different cell types provides a problem. pCAT19 is not suitable, since the CAT gene is "promoterless", and no specific transcription of CAT mRNA would be expected in either cell type. In experiments described here, pCAT-40 was chosen as an arbitrary reference plasmid for determining cell type-specificity.

The extra 111nt of RBC 15-LOX gene 5' flanking sequences present in pCAT-150 have a marked effect upon expression of the CAT gene in both cell types tested. In both MEL and STO cells, the CAT activity from pCAT-150-transfected cell extracts is approximately 7 or 8 times higher than that seen after transfection with pCAT-40. The observation that the difference between CAT expression from pCAT-150 and from pCAT-40 is the same in both cell types suggests that there are no sequences in the first 150bp of 5' flanking sequence of the RBC 15-LOX gene which have any cell type-specific effect in the two cell types being investigated. This is in good agreement with the footprinting data (Section 3.3) which showed that the CACCC, CTF/NF-1-like, and GGGCGG sequences within 150nt of the transcription initiation site are protected by extracts from both MEL and mouse brain cells, and lends support to the suggestion (Section 3.3.3 above) that sequences within 150nt of the transcription initiation site are concerned with basic, "constitutive" promoter function, and that any sequences involved in modulation of levels of transcription lie elsewhere.

3.4.6 Sequences present in RBC 15-LOX promoter can respond to a heterologous enhancer.

Having established that sequences between +20 and -150nt from the transcription initiation site can act as a basic functional promoter, it was decided to investigate whether expression from this basic promoter region can be modulated by other regulatory elements. In order to do so, a fragment from the U3 region of the long terminal

repeat (LTR) of the Friend murine leukaemia virus, which contains an erythroid-specific enhancer (Bosze et al, 1986) and a CCAAT box, but no TATA box or transcription initiation site, was inserted into a HindIII site downstream from the transcription termination signals of pCAT-40 and pCAT-150 (Fig. 19).

A comparison of results obtained with pCAT-40 and the two pCAT-40E (+ and -) plasmids shows that the Friend virus LTR enhancer has little effect upon expression of CAT activity in either of the cell types examined, when placed downstream of 40bp of 5′ flanking sequence. The effect of the Friend virus LTR enhancer upon CAT expression driven by RBC 15-LOX promoter sequences up to 150bp upstream of the transcription initation site, however, is markedly different from its effect upon only 40bp of upstream sequence. Comparing expression of CAT enzyme activity from the two pCAT-150E (+ and -) plasmids with that from pCAT-150 in the same cell type shows only a slight increase (of approximately 25%) in STO cells, but a larger, 2- to 3-fold, increase in either orientation in MEL cells (Figures 20 and 21, and Table I). The magnitude of this cell type-specific effect of the Friend virus LTR enhancer on the RBC 15-LOX basic promoter in murine cells is smaller than the effect on the enhancerless SV40 Early promoter in human cells observed by Bosze et al (1986), who obtained a 50-fold increase in CAT activity driven by the SV40 promoter in the presence of the Friend virus enhancer, situated downstream of the CAT gene, in K562 human erythroid cells, but only a 1.5-fold increase in non-erythroid HeLa cells. The difference between these two sets of results may be attributable either to the use of cell lines from a different species, or to different effects of the enhancer on different heterologous promoters.

These results indicate that sequences between 40 and 150bp upstream of the transcription initiation site of the rabbit RBC 15-LOX gene can respond to a heterologous enhancer in a cell type-specific fashion, and provide further evidence that sequences immediately upstream of the rabbit RBC 15-LOX transcription initiation site can act as a functional promoter, albeit for expression of a heterologous gene in heterologous cells.

3.4.7 Sequences between 1000 and 2700bp upstream of the rabbit RBC 15-LOX transcription initiation site have a cell type-specific effect upon expression of a linked CAT gene in murine cells.

Having ascertained that sequences within 150bp upstream of the RBC 15-LOX transcription initiation site can act as a basic functional promoter and can respond to heterologous regulatory elements, it was decided to investigate whether any regions in the 2.7kb of available 5′ flanking sequences of the rabbit gene could exert a regulatory influence upon expression from the +20 to -150nt region in the same experimental system. pαCAT1, a plasmid containing 700bp of 5′ flanking sequence from the mouse α1-globin gene linked upstream of the CAT gene in an analogous fashion to the RBC 15-LOX/CAT recombinants (Fig. 17), was used as a positive control for erythroid-specific expression. This region of the mouse α1-globin gene has been previously shown to have a 4- to 5-fold erythroid-specific effect in the same assay system used here (Frampton et al, manuscript in preparation).

Various progressive 5′ deletions of the RBC 15-LOX promoter were made such that, in all cases, sequences were in the same relative position and orientation as in the wild-type rabbit gene. So, in considering the results of functional expression studies, problems raised by incorrect spatial separation of regulatory elements should not arise. Figure 18 shows the strategy for construction of additional RBC 15-LOX/CAT recombinants containing 316, 450, 1000 and 2700bp of RBC 15-LOX 5′ flanking sequence, linked upstream of the CAT gene. A representative autoradiograph of transfections of these recombinants into MEL and STO cells is shown in Figure 22, and corrected CAT activities are given in Table I.

It was seen earlier (Section 3.4.5) that pCAT-150 appears to show no apparent cell type-specificity. To determine whether upstream sequences have any cell type-specific regulatory effects upon sequences within 150nt upstream of the transcription initiation site,

Figure 22. CAT expression in murine cell lines from various progressive deletions of RBC 15-LOX 5′ flanking sequences.

Representative autoradiographs of the products of CAT enzyme activity in extracts from MEL and STO murine cell lines transfected with various RBC 15-LOX/CAT recombinant and control plasmids. pCAT-40, pCAT-150, pCAT-316, pCAT-450, pCAT-1000 and pCAT-2700, containing 40, 150, 316, 450, 1000 and 2700bp of RBC 15-LOX 5′ flanking sequences respectively, are as shown in Figure 18. pUC18 and pHSV-βgal (Fig. 17) are controls for background acetylation in the absence of CAT-containing plasmid; pαCAT1 (Fig. 17), containing 700bp of 5′ flanking sequences of the mouse α1-globin gene, is a positive control for cell type-specific CAT expression.

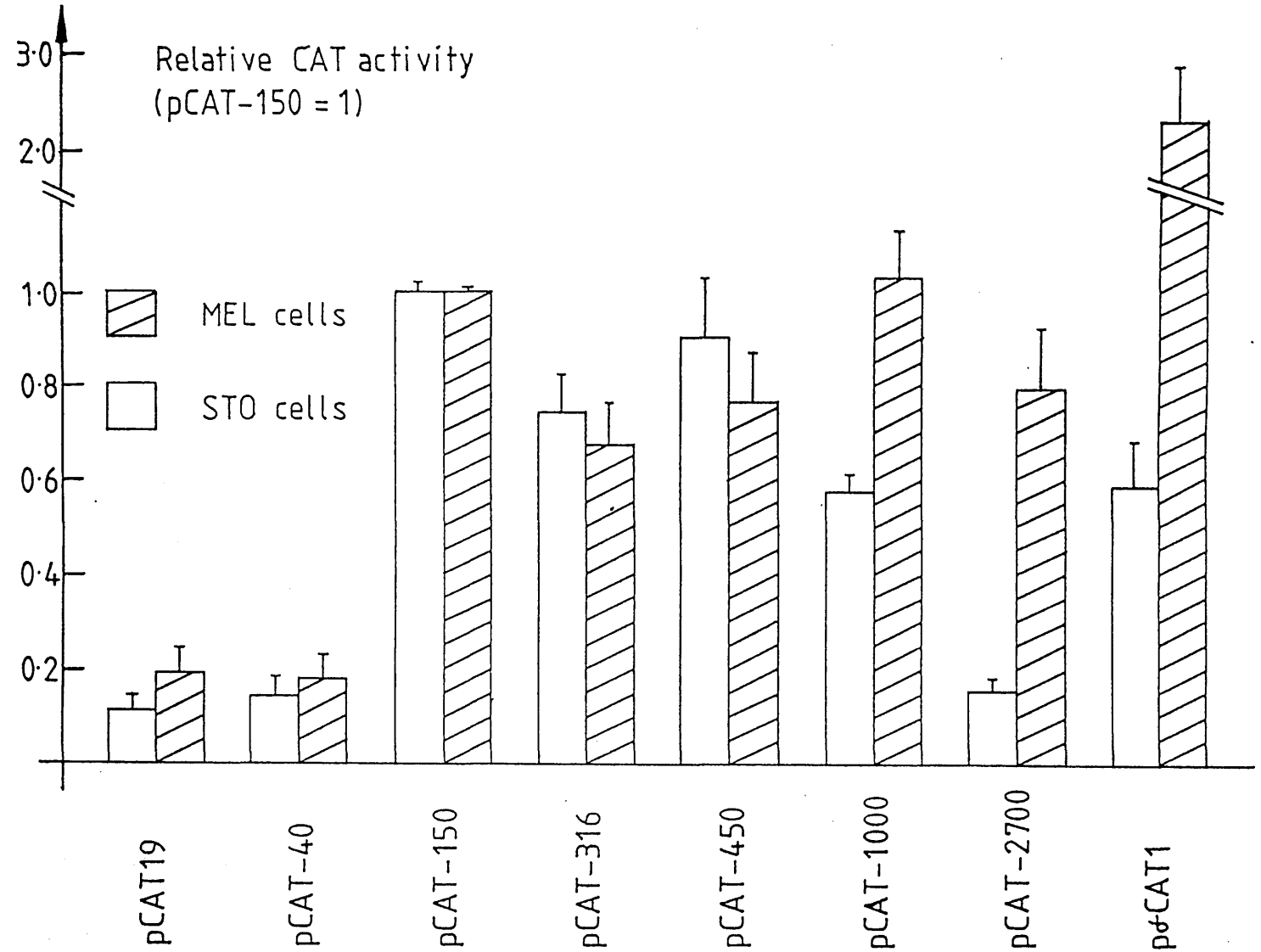
Cells were transfected by the addition of a calcium phosphate/DNA co-precipitate containing 2.4pmoles of the RBC 15-LOX/CAT plasmid under test, 2pmoles (10μg) of pHSV-βgal as internal standard, and pUC18 to a total of 20μg DNA. Results shown, obtained in a 30 minute CAT assay at 37°C using amounts of transfected cell extract showing equal amounts of β-galactosidase activity, range from 0.3% (pUC18 and pHSV-βgal in STO cells) to 32% conversion (pαCAT1 in MEL cells). Percentage conversions from this experiment, after standardisation for β-galactosidase activity to allow comparison with other experiments (Section 2.14.6), are shown as Experiment 4 in Table I. Autoradiographs shown here were exposed for 7 days at room temperature.

A. MEL cells

B. STO cells

Figure 23. Effect of sequences upstream of -150nt on CAT expression from the basic RBC 15-LOX promoter.

CAT activities of RBC 15-LOX/CAT recombinants containing progressive deletions of 5' flanking sequences within 2.7kb upstream of the transcription initiation site are expressed relative to expression from pCAT-150. For each recombinant, (except pCAT-1000; only 3 transfections) CAT activities from 5 independent transfections (Experiments 3, 4 and 5 in Table I) were corrected for β-galactosidase activity and were then each expressed relative to pCAT-150 activity in the same experiment. Bar sizes represent the mean values of these relative activities and error bars represent the Standard Error of the Mean.


Relative CAT activity
(pCAT-150 = 1)
3·0
2·0
1·0
0·8
0·6
0·4
0·2
MEL cells
STO cells
pCAT19
pCAT-40
pCAT-150
pCAT-316
pCAT-450
pCAT-1000
pCAT-2700
pσCAT1

levels of CAT expression from the various deletions were compared in each cell type with that from pCAT-150 in the same experiment, given the arbitrary value of 1. Relative CAT activities of the various recombinants are shown in Figure 23.

RBC 15-LOX/CAT recombinants containing 5' flanking sequences within 1kb upstream of the transcription initiation site do not have any significant effect upon CAT enzyme expression relative to pCAT-150 in either MEL or STO cells, with the possible exception of pCAT-1000 which shows a slight reduction in STO cells. However, expression of the CAT gene from pCAT-2700 seems to show a marked cell type-specific effect, expression of CAT activity relative to pCAT-150 in STO cells being approximately 6-fold lower than in MEL cells and barely above the expression seen from pCAT-40. Although the most obvious interpretation of this cell type-specific effect is that it is an STO-specific negative effect, this cannot yet be established with any certainty (see Discussion). The degree of this cell type-specific effect would appear to be slightly larger than the 4- to 5-fold effect seen with pαCAT1, consistent with earlier observations that the region of the mouse α1-globin gene contained within this fragment is responsible for some, but not all of the erythroid-specific regulation of transcription of this gene by sequences within 4kb of the transcription initiation site (Frampton et al, manuscript in preparation).

In summary, CAT assays on extracts from murine cells transfected with rabbit RBC 15-LOX/CAT recombinants suggest that sequences upstream of the RBC 15-LOX transcription initiation site include at least two regions important for the expression of CAT enzyme in murine cells (see Sections 3.4.5 - 3.4.7 above). One, between 40 and 150bp upstream is capable of increasing expression of the CAT enzyme in both erythroid and non-erythroid murine cell lines (Fig. 21). The other, between 1000 and 2700bp upstream, exerts a cell type-specific effect upon expression driven by the -40 to -150 region (Fig. 23).

3.4.8 Analysis of CAT mRNAs from murine cells transfected with RBC 15-LOX/CAT recombinants.

CAT assays are widely used to investigate the effects of defined sequences upon transcription. However, the CAT assay itself is only an indirect measure of transcription as it assesses only levels of CAT enzyme activity. Thus, analysis of CAT mRNAs, in addition to that of CAT enzyme activity, is important for several reasons. Firstly, it provides confirmation that CAT mRNAs transcribed from the various RBC 15-LOX/CAT recombinants are correctly initiated at the RBC 15-LOX transcription initiation site in the heterologous cell lines used. Secondly, it establishes that differences in CAT enzyme activity from different RBC 15-LOX/CAT recombinants are not due to differences in sites of initiation, in which case CAT enzyme activities might be a result of differential post-transcriptional effects upon CAT mRNAs differing in their 5' untranslated regions. Finally, quantitation of CAT mRNA levels can provide ~~and hence~~ confirmation that CAT enzyme activity accurately reflects the levels of transcripts from each CAT recombinant. Of course, such analysis still does not provide clear evidence that CAT transcripts are regulated at the level of transcription.

In order to analyse CAT mRNAs transcribed from the RBC 15-LOX promoter in STO and MEL cells, $2x10^6$ cells of each type were transfected with pCAT19, pCAT-40, pCAT-150, pCAT-1000, pCAT-2700 and pLW2. pLW2 was used as a positive control: as CAT expression from this plasmid in both cell types is approximately 10-fold higher than from any of the RBC 15-LOX recombinants (Table I), it was hoped that detection of CAT mRNA from this recombinant would be possible even if transcripts from the RBC 15-LOX/CAT recombinants were not detectable. The lacZ-containing plasmid pHSV-βgal was co-transfected with each RBC 15-LOX/CAT recombinant to allow standardisation of CAT mRNA levels by comparison with lacZ mRNA levels, analogous to the correction of CAT enzyme activities according to β-galactosidase activity.

CAT enzyme assays were performed on extracts from one fifth of the cells from each transfection as above (results are shown as

Experiment 5 in Table I). The remaining cells were used to prepare RNA. RNAs were judged to be intact by the appearance of intact 18 and 28S ribosomal RNAs on ethidium bromide-stained agarose gels. Mapping of 5′ ends and quantitation of CAT and lacZ mRNAs was attempted using both S1 nuclease protection and primer extension analyses. Unfortunately, both were unsuccessful for various reasons discussed below.

3.4.8.1 Analysis of CAT mRNAs by primer extension.

A single primer, derived from sequences within the CAT gene, can be used to map the 5′ ends of CAT transcripts from any recombinant containing the CAT gene. The primer chosen was a synthetic single-stranded 35-mer oligodeoxyribonucleotide of sequence 5′-GTGGTA-TATCCAGTGATTTTTTTCTCCATTTTAGC-3′ (34% G/C). This sequence is the complement of CAT gene sequences extending from 6 nucleotides upstream of the translation initiation codon to 29nt beyond. Although this primer is very A/T-rich, there is no region sufficiently close to the start of the CAT gene to allow efficient primer extension which has a lower A/T content. A 35-mer was thus chosen rather than a shorter oligo. in order to optimise the hybridisation to the CAT mRNA.

For standardisation of CAT mRNA levels relative to lacZ mRNA levels, a synthetic 30-mer oligodeoxyribonucleotide of sequence 5′-CC-TTCGCGATGCCGCCGCTGCCGTCCCGGT-3′ (77% G/C) was chosen as primer. This sequence is complementary to pHSV-βgal sequences extending from 38 to 67bp beyond the translation initation site, within the region of the HSV-2 IE-5 promoter which is used to drive transcription from the lacZ gene. The extreme G/C-richness of this region does not permit the use of a less G/C-rich primer.

These two primers were 5′ end-labelled and used in primer extension reactions with approximately 10μg of total RNA from transfected STO cells and 20-50μg from MEL cells. Unfortunately, results of primer extension on these RNAs were not satisfactory using either of the primers, and it was not possible to obtain quantitation

of CAT mRNAs by this method. In all RNAs tested, including a tRNA control, primer extension using the oligo. designed to prime on lacZ mRNAs gave a ladder of primer extension products, extending well beyond the normal transcription initiation site of the HSV IE-5 gene, (data not shown). The observation of an identical pattern in a tRNA control suggests that non-specific priming is occurring, probably from a tRNA species, perhaps due to a hybridisation temperature too low for the G/C-richness of the primer.

However, with the CAT primer, there were no visible primer extension products, even after a two-week exposure. In contrast to the situation with the lacZ primer, the failure to observe extension products may have been due to a hybridisation temperature that was too high for the A/T-rich primer. Alternatively, it might be due to the fact that only a small proportion of the population of cells to which DNA was added have actually taken up, and are expressing, the CAT recombinants, in which case CAT mRNAs would be present only at very low levels, representing an extremely small proportion of the total cellular RNA extracted from each transfected population of cells. It is notable in this respect that many of the reports of successful primer extension analyses of CAT mRNAs have been obtained with RNA isolated from pools of drug resistant cells obtained by co-transfection of CAT recombinants with a selectable drug-resistance marker, such that a larger proportion of cells will be expressing the CAT plasmid (e.g. Hermonat et al, 1988). Furthermore, the CAT mRNA is thought to be a relatively unstable messenger. Consequently, primer extension with an end-labelled primer might not be sufficiently sensitive to produce detectable extension products.

There are several possible solutions to this technical problem of sensitivity. A first would be to put the RBC 15-LOX promoter in all the CAT recombinants under the influence of a non-cell type-specific enhancer in the hope of increasing the level of CAT mRNA transcription. However, this might provide additional problems of interpretation of results. An alternative solution would be to increase the amount of CAT mRNA available as template for primer extension. This could be achieved either by including larger

quantities of total RNA in each hybridisation, or by using poly(A)$^+$ enriched RNA. Another possibility would be to increase the specific activity of the products. This could be achieved either by including a radio-labelled nucleotide in the primer extension reaction, or by using a primer of higher specific activity. As it would be difficult to obtain a synthetic oligo. of higher specific activity than one which is end-labelled, other than by incorporating radio-labelled deoxynucletides in the oligo. synthesis, it was decided that an attempt should be made to obtain a probe of higher specific activity by performing S1 protection experiments with uniformly-labelled probe.

3.4.8.2 Analysis of CAT mRNAs by S1 protection.

Single-stranded uniformly-labelled probes for detecting 5' ends of both lacZ and CAT mRNAs were generated by the method of Bentley (1984). This involves sub-cloning of restriction fragments into double-stranded (replicative form) M13 virus, and the generation of single-stranded recombinant M13 DNA templates, followed by the incorporation of a radio-labelled deoxynucleotide into a complementary strand by primer extension on the M13 Universal primer using Klenow enzyme, and, finally, liberation of radio-labelled single-stranded probe by cleavage of the resulting duplex with a single restriction enzyme cut. The probe can then be isolated by running out on a sequencing gel.

The fragment chosen for generation of an S1 probe for analysis of CAT mRNAs was a 419bp SstI/EcoRI fragment from pCAT-150, cloned into SstI/EcoRI-cut M13mp19. Following generation of a complementary strand, and cleavage with SstI, a probe of 436nt, including the M13 Universal primer, was obtained. Initiation of transcription from the usual RBC LOX transcription initiation site should yield a 290nt protected fragment; any transcripts initiated more than 146nt upstream of this site would produce full-length protection.

The fragment chosen to generate a single-stranded S1 probe for quantitation of lacZ mRNA levels was a 246bp XbaI/SstI fragment from pHSV-βgal, containing approximately 100bp of sequence from either side

of the HSV IE-5 transcription initiation site. Transcripts initiated at this site would be expected to protect a fragment of 167nt.

The CAT and lacZ probes were labelled using [α-^{32}P]dCTP at $1.1x10^{14}$Bq/mmol. Radio-label was present at a final concentration of 3μM in the labelling mixture, along with 50μM unlabelled dCTP, producing a final specific activity of $6.7x10^{12}$Bq/mmol. This should produce probe labelled at 6 residues in every hundred nucleotides of probe i.e. approximately 26 labelled residues per molecule ($1.7x10^{14}$Bq/mmol) of the 436nt CAT probe (assuming even distribution of the 4 bases), and 16 labelled residues per molecule ($1.1x10^{14}$Bq/mmol) of the 269nt lacZ probe. Obviously, this labelling of an S1 probe with more than one molecule of isotope per molecule produces probes of far higher specific activity than can be achieved by end-labelling.

Unfortunately, using the same quantities of RNA as in the primer extension analysis of CAT mRNAs above, no 5' end analysis or quantitation of CAT mRNAs expressed from any of the RBC 15-LOX/CAT recombinant plasmids was possible using this method either. No full-length protected fragments were seen. Protected fragments of the expected size were seen from both cell types transfected with the highly-expressed pLW2 plasmid, but no protection could be detected, even after a 2-week exposure, with RNAs from either cell type transfected with RBC 15-LOX/CAT recombinants (data not shown). This suggests that the technical problem of detecting CAT mRNAs has been at least partly overcome - the probes used are of a suitable specific activity to detect CAT mRNAs from the pLW2 plasmid, but transcripts from the RBC 15-LOX/CAT recombinants, present at lower levels, have not been detected. A repetition of this experiment using larger amounts of the CAT mRNAs (either larger quantities of total RNA, or similar quantities of polyadenylated RNA) might be expected to achieve successful quantitation of CAT mRNAs expressed from the RBC LOX/CAT recombinants. However, a lack of sufficient time has meant that this experiment has not yet been performed.

Chapter 4.

DISCUSSION.

4.1 OVERVIEW OF THE PROJECT.

As part of a group effort directed at investigating the co-expression of genes transcribed in erythroid cells, this project has been concerned with the organisation and function of the promoter region of the gene encoding the rabbit erythroid-specific RBC 15-LOX. This enzyme, until recently the only mammalian LOX to have been purified to homogeneity (Rapoport *et al*, 1979), is one of the major non-globin proteins of anaemic peripheral blood, and shows several interesting features in its expression, including an erythroid-specific accumulation of RBC 15-LOX mRNA (Thiele *et al*, 1987; and Section 3.2 above) and translational inactivation of the messenger until the final stages of erythroid maturation (Thiele *et al*, 1982; and Introduction Section 1.6.3).

This project has mainly been concerned with transcription of the recently cloned RBC 15-LOX gene, involving: mapping of the transcription initiation site; characterisation of proximal 5′ flanking sequences, including the binding of nuclear proteins *in vitro*; functional analysis of proximal upstream sequences for promoter function *in vivo*; and investigation of the effect of sequences further upstream of the transcription initiation site on promoter function. The following discussion will concentrate on possible future areas of research into RBC 15-LOX transcription, but will also deal with elements of the work presented here which, in conjunction with the recent results of others, may have implications for post-transcriptional or translational regulation of RBC 15-LOX expression and for the evolution of the LOX enzyme family.

4.2 FUNCTIONAL ANALYSIS OF THE RBC 15-LOX PROMOTER.

In investigating the ability of 5′ flanking sequences to function as a promoter in the efficient initiation of transcription in living cells, one would ideally want to perform functional analysis in cells from the appropriate species. However, in the absence of suitable rabbit cell lines, rather than attempting to clone the mouse or human RBC 15-LOX genes, it was decided to investigate promoter

function in heterologous, murine cell lines.

The system chosen to investigate promoter function involved the use of murine erythroleukaemia (MEL) and STO fibroblast cell lines. Promoter deletions of other rabbit erythroid-specific genes have, in the past, been shown to be capable of functioning in both erythroid and non-erythroid murine cell lines. For example, Dierks et al (1981 and 1983) have demonstrated that 5′ flanking sequences of the rabbit β-globin gene are required for accurate and efficient transcription initiation both in transient expression experiments in mouse 3T6 cells and in stably transfected mouse L cells. So, transfection into murine cell lines should permit investigation at least of "constitutive" cis-acting sequences involved in transcription initiation from the rabbit RBC 15-LOX promoter. Furthermore, Wright et al (1984) have demonstrated that transcription from a minimal rabbit β-globin promoter containing only 58bp of 5′ flanking sequence could be correctly regulated upon induction of differentiation of MEL cells into which the rabbit β-globin recombinants had been stably transfected. This suggests that murine cell lines may also be legitimate cells in which to study the effects of regulatory elements upon transcription from the rabbit RBC 15-LOX promoter.

4.2.1 Two regions of rabbit RBC 15-LOX upstream sequences important for expression of a reporter gene from the RBC 15-LOX promoter in murine cell lines.

Functional data presented here indicate two regions of rabbit RBC 15-LOX gene 5′ flanking sequences which affect expression of a cis-linked CAT gene in murine MEL and STO cell lines. A proximal element, between 40 and 150nt upstream of the transcription initiation site, functions as a "constitutive" promoter, increasing expression of a linked CAT gene by 7- to 8-fold in both the erythroid and non-erythroid murine cell lines studied (Section 3.4.5), while a distal element located between 1000 and 2700nt upstream has a 6- to 8-fold cell type-specific effect upon CAT expression (Section 3.4.7).

These observations are reminiscent of the situation in the

β-like globin gene family, in which 5′ flanking sequences within 100nt of the adult β-globin genes drive transcription in both erythroid and non-erythroid cell types (Dierks et al, 1983; Grosveld et al, 1982; Myers et al, 1986; Charnay et al, 1985) while sequences downstream of the polyadenylation signal appear to be involved in erythroid-specific expression (Antoniou et al, 1988; Choi and Engel, 1986). This is in contrast to the situation for the human α1-antitrypsin gene, however, where cell type-specific regulatory sequences have been found immediately upstream of the transcription initiation site. Here, sequences between -137 and -37nt upstream of the transcription initiation site have been found to be sufficient to confer liver-specific expression upon the homologous gene or upon a heterologous SV40 promoter (De Simone et al, 1987).

A priority for the immediate future is the confirmation of CAT enzyme activity results by analysis of CAT mRNAs, in order to demonstrate firstly that the increase in expression of CAT enzyme seen from the -150 and -40 regions occurs from the correct transcription initiation site and, secondly, that differences in CAT activities reflect CAT mRNA levels rather than different translational efficiencies of mRNAs transcribed from different recombinants. These studies, commenced as part of this project (Section 3.4.8), have not yet been successfully completed, and are currently being pursued by other members of the group.

4.2.2 Potential cis-regulatory sequences in the RBC 15-LOX promoter.

One of the principal objectives in studying the structure and function of the rabbit RBC 15-LOX gene promoter has been to investigate whether there are any functionally important cis-acting sequences in common with other genes expressed in erythroid cells. The basic functional rabbit RBC 15-LOX promoter, contained within 150nt of the transcription initiation site, includes (Section 3.3) CCAAT, CACCC and TATA-like sequences seen in the promoter of the rabbit β-globin gene (Hardison et al, 1979; van Ooyen et al, 1979), and other mammalian globin genes (Efstratiadis et al, 1980), as well

as other sequences present in the promoters of a wide variety of non-globin genes, most notably a single GGGCGG, G/C-box sequence, the recognition sequence for transcription factor Sp1 (Kadonaga et al, 1986), and a potential binding site for CTF/NF-1, one of a family of transcription factors which recognise sequences which include the CCAAT motif (Santoro et al, 1988; Chodosh et al, 1988a, b; Jones et al, 1987). This assortment of sequences which are also found in the promoter regions of other genes is reminiscent of findings in a variety of other gene promoters which contain modular arrangements of cis-acting sequences recognised by "general" transcription factors involved in basic promoter function (see Introduction Section 1.2.3).

4.2.3 DNA-protein interactions in the RBC 15-LOX promoter.

Preliminary data from DNaseI footprinting of the proximal region, show that some of these "general" promoter sequence motifs are protected by nuclear protein extracts from both erythroid and non-erythroid murine cells (Section 3.3.2) and support the suggestion from functional data that this region represents a basic functional promoter, which is bound by transcription factors present in many cell types, and that any sequences which may be involved in cell type-specific regulation of the level of transcription from this region may be located elsewhere in the RBC 15-LOX gene.

A variety of experiments will be required in the near future to clarify the situation regarding the binding of transcription factors to the RBC 15-LOX promoter, particularly whether the same factors are involved in binding to the RBC 15-LOX promoter as are involved in the function of other promoters. For example, a region (FP2) which is protected by both erythroid and non-erythroid murine nuclear protein extracts contains both the more proximal of two CACCC motifs, and a sequence immediately upstream, which resembles the recognition sequence for the CTF/NF-1 family of transcription factors (Section 3.3.2.1). It will be of interest to examine whether either or both motifs are bound by factors which bind similar sequences in other genes. Mantovani et al (1988a) have reported both ubiquitous and erythroid-specific nuclear proteins which bind to the CACCC box of the

human β-globin promoter, while, as mentioned in Section 3.3.2.1, a combination of CACCC and CTF/NF-1-like sequences have also been footprinted in chicken β^A- and β^H-globin genes (Plumb et al, 1986). Protection at FP2 might be the result of the binding of factors to either or both of the CACCC and CTF/NF-1-like motifs. This could be investigated by comparing the electrophoretic mobility shift patterns obtained using erythroid nuclear proteins bound to three double-stranded synthetic oligos: one containing only the CTF/NF-1-like motif; one containing only the CACCC motif, and another containing both. If this third oligo. simultaneously binds two proteins, a mobility shift greater than that from binding to either of the other two oligos. might be expected. A comparison of the mobility shifts obtained with purified CTF/NF-1 proteins (Jones et al, 1987) or CTF/NF-1 expressed from cloned cDNAs (Santoro et al, 1988) and with murine nuclear proteins could also be used to investigate whether it is a CTF/NF-1 protein or a murine variant which is responsible for the protection observed. Further information as to the nature of the factor binding the CTF/NF-1-like sequence could be obtained by performing in vitro footprinting and mobility shift experiments using large molar excesses of competitor oligos. containing sequences known to be bound by CTF/NF-1 in other promoters.

Both MEL and mouse brain extracts also weakly protected a region (FP1) containing the GGGCGG sequence recognised by the transcription factor Sp1. However, a different pattern was obtained with HeLa cell extract (Fig. 15), the original source of purified Sp1 (Kadonaga and Tjian, 1986; Briggs et al, 1986). It therefore seems likely that the protection around GGGCGG is not due to Sp1 itself, but might be due to a murine protein with similar sequence specificity in its DNA-binding. This suggestion might also be tested using a mobility shift assay. In a similar experiment to that suggested above, the question of whether Sp1 itself is capable of binding to the RBC 15-LOX GGGCGG sequence could be investigated by observing the pattern of shifted products when a double-stranded synthetic oligonucleotide corresponding to the RBC 15-LOX G/C-box is incubated with MEL or mouse brain extracts with the pattern obtained using either purified Sp1 (Kadonaga and Tjian, 1986; Briggs et al, 1986) or with Sp1 expressed from its cloned cDNA

(Kadonaga et al, 1987) in an expression vector.

If the RBC LOX G/C box is indeed bound by some murine factor, but not by Sp1, this might be explained by the context of the G/C-box i.e. the nature of surrounding sequences. Kadonaga et al (1986) have presented an extended consensus sequence, (G/T)GGGCGG(G/A)(G/A)(C/T), for Sp1 binding, and investigated the affinity of Sp1 binding to variants of this sequence. The G/C-box contained in the RBC 15-LOX promoter differs from this consensus sequence at only one position, a C residue occupying the ninth position, rather than G or A. Unfortunately, this variant was not one of those tested by Kadonaga et al (1986). The proportion of synthetic oligonucleotide showing altered mobility with pure Sp1 could be compared with that of another oligo. corresponding to the extended Sp1 consensus sequence to determine the relative affinity of Sp1 for the G/C box of the RBC 15-LOX promoter.

To date, protein binding to the rabbit RBC 15-LOX promoter has only been investigated using murine nuclear extracts, in order to correlate results with those obtained by functional assays in murine cell lines. To examine whether such sequences are also recognised by rabbit transcription factors, it will be necessary to perform DNaseI footprinting with rabbit nuclear proteins. Foetal liver, as the principal erythropoietic organ of the developing rabbit, might prove a suitable tissue for obtaining erythroid nuclear proteins, since it represents a more homogeneous population of erythroid cells than, say, bone marrow from anaemic animals. Furthermore, Northern blot hybridisation studies presented here (Fig. 13), and subsequently confirmed in other studies (J. Fleming, unpublished results), suggest that this tissue expresses the RBC 15-LOX mRNA.

4.2.4 GGGAG: a sequence motif present in the RBC 15-LOX promoter and in upstream regions of a variety of other genes.

In addition to sequence motifs resembling TATA-like, CCAAT, CACCC, GGGCGG and CTF/NF-1-binding site which have been found in a variety of other genes, RBC 15-LOX flanking sequences also contain two

short direct repeats. One of these, a perfect decanucleotide repeat, contains a previously undescribed sequence element GGGAG (Fig. 14). This sequence element is also seen in the promoter regions of a variety of genes, including genes expressed in erythroid cells, and some which are not (Table II). As discussed below, the distribution of GGGAG sequences shows several interesting features which may suggest that it represents a general transcription regulatory element.

The RBC 15-LOX promoter decanucleotide repeat contains one perfect and one imperfect copy of the GGGAG element in the arrangement, $CGGAG(N)_5GGGAG$. Intriguingly, similar repeat arrangements are also seen in the mouse α-globin and human carbonic anhydrase III (CAIII) gene 5′ flanking sequences (see Table II for references). In mouse α-globin, the arrangement is $GGGAC(N)_5GGGAG$, while in the human CAIII gene, it is $GGGAG(N)_6GGGAG$.

Multiple copies of the sequence GGGAG are also seen in the upstream sequences of mouse band 3 anion antiport and mouse β-globin genes (Table II). In the former, two copies of GGGAG are present as an inverted repeat separated by 47nt. One of the two copies in the mouse β-globin promoter region is conserved in an identical position in rabbit and human β-globin genes.

As well as its presence in the RBC 15-LOX promoter, GGGAG is also present in four tandemly repeated copies in one of the RBC 15-LOX introns (B. Thiele, personal communication). Interestingly, a conserved GGGAG motif is also present in the first intron of human and mouse erythropoietin genes (Table II). The introns of the two erythropoietin genes show unusually highly-conserved sequences, and this sequence similarity has been suggested to imply some functional significance, perhaps the presence of an enhancer (Shoemaker and Mitstock, 1986) in a situation analogous to the enhancers present within the introns of immunoglobulin genes (Gillies et al, 1983; Banerji et al, 1983; Queen and Baltimore, 1983).

Of course, sequence comparisons are not sufficient, in themselves, to indicate that a sequence is of importance for

Table II. Occurrence of the sequence motif GGGAG in the 5' flanking sequences of various genes.

Alignment of twenty 5' flanking sequence elements which include the GGGAG motif. Numbers to the right of the nucleotide sequences indicate the position of the element relative to their respective transcription initiation sites. Three elements which contain the GGGAG motif in the opposite orientation on the non-coding strand are indicated by arrows to the left of the sequences. Three sequences which are not perfect copies of the GGGAG sequence are included in which the GGGAG-like sequence is contained in closely-spaced repeats within 5' flanking regions.

A possible "consensus" sequence, derived from these twenty sequences, is shown beneath the sequence alignment, indicating the number of occurrences of each of the four nucleotides at a particular position within a 15nt region centred on the GGGAG motif.

RBC 15-LOX	Rabbit		gagctCGGAGctgcg ctgcgGGGAGctgcg	-145 to -141 -135 to -131	Section 3.3.1
β-globin	Rabbit		ggtctGGGAGataca	-192 to -188	Hardison et al, 1979
	Human		tcagaGGGAGgggct	-185 to -181	Lawn et al, 1980
	Mouse		tcagaGGGAGtccca gtccaGGGAGaaata	-191 to -187 -138 to -134	Konkel et al, 1979
α-globin	Mouse		aaacaGGGACccaga ccagaGGGAGaggtg	-335 to -331 -325 to -321	Nishioka and Leder, 1979
Band 3	Mouse	 <-	ggtttGGGAGcctcc cagccGGGAGcccct	-187 to -183 -235 to -239	Kopito et al, 1987
Glutathione peroxidase	Mouse		tataaGGGAGcatga atgggGGGAGggtag	-355 to -351 -127 to -123	Chambers et al, 1986
Porphobilinogen deaminase	Human	<- <-	cactgGGGAAcctgt tcagtGGGAGtgctg	-144 to -148 - 35 to - 39	Chretien et al, 1988
Erythropoietin	Mouse		gtttgGGGAGccccc	First intron	McDonald et al, 1986 Shoemaker & Mitstock, 1986
	Human		gaaggGGGAGgggggg cttggGGGAGtcctt	First intron First intron	Jacobs et al, 1985
Carbonic anhydrase III	Human		ctcccGGGAGagcgc gcgcaGGGAGgggcg tgcggGGGAGctaca	-210 to -206 -199 to -195 - 30 to - 26	Lloyd et al, 1987

Consensus:

	n	n	n	c/g	n	G	G	G	A	G	c	c/g	n	c	n
A	2	6	6	1	7	-	-	-	20	1	4	1	4	1	6
C	6	5	4	8	2	1	-	-	-	1	9	7	6	10	3
G	7	3	5	8	7	19	20	20	-	18	4	7	6	5	5
T	5	6	5	3	4	-	-	-	-	-	3	4	4	4	6

regulation of transcription. Further evidence must be obtained from DNA-protein interactions and from functional studies. Intriguingly, the decanucleotide repeat of the RBC 15-LOX promoter which contains the GGGAG motif lies just within the -150 to -40nt region shown here to be important for the up-regulation of a cis-linked CAT gene in murine cells (see Section 3.4.5). However, in initial in vitro DNaseI protection studies using murine nuclear proteins (Section 3.3.2.1), the GGGAG motif does not appear to be protected by extracts from either erythroid or non-erythroid cells, as might have been expected if GGGAG were indeed a motif involved in the regulation of a wide variety of genes. Thus, at present, the possibility cannot be excluded that the presence of the GGGAG motif in the 5′ flanking sequences of the various genes shown in Table II does not merely represent random occurrences of a short, pentanucleotide sequence, an occurrence which has a probability of occurring approximately once in every kilobase in a random sequence of the four deoxyribonucleotides.

Whether the GGGAG motif is important for basic promoter function can be investigated by further deletions of the -150 to -40nt region. Further functional dissection of the -40 to -150 region using the same system is also of interest in defining the minimal promoter necessary for up-regulation of transcription in this assay system. For example, do further promoter deletions removing the footprinted CTF/NF-1-like and/or CACCC motifs abolish expression of the cis-linked reporter gene?

4.2.5 A distal element which confers cell type-specificity upon expression driven by the proximal element.

It has been demonstrated here that the RBC 15-LOX promoter sequences within 150nt of the transcription initiation site can respond to cis-regulatory sequences, either to the heterologous Friend Virus enhancer (Fig. 21), or to sequences from the -2700 to -1000nt region of RBC 15-LOX 5′ flanking sequences which confer a 6- to 8-fold cell type-specific effect (Fig. 23). It therefore seems likely that regulation of transcription may be at least one factor involved in the cell type-specific accumulation of RBC 15-LOX mRNA seen by Thiele et

al (1987), and confirmed by S1 and primer extension analyses and Northern blot hybridisation data presented here (Figs. 7, 8, 10, 12 and 13). It has thus become a matter of importance to establish whether transcriptional regulation is indeed important for control of RBC 15-LOX gene expression in rabbit erythroid cells. Nuclear run-on transcription experiments using isolated nuclei from an erythroid tissue such as rabbit foetal liver (see Section 4.2.3 above) are therefore essential in the immediate future. Further information as to the likely importance of transcription in the cell type-specific accumulation of the RBC 15-LOX mRNA will be provided by investigating whether there are cell type-specific DNaseI HSS associated with the 5′ flanking sequences of the RBC 15-LOX gene in erythroid and non-erythroid tissues. Work is currently in progress to investigate this possibility (J. Fleming, unpublished results).

Thus far, the difference in level of expression conferred by the distal, -2700 to -1000nt, element has only been observed in erythroid and fibroblast-derived cell types. It is conceivable that the lower expression seen in STO cells is characteristic only of fibroblast cells, and not of other non-erythroid cell types. Other members of the LOX enzyme family are expressed in other cell types (see Introduction Section 1.5.1), including epithelial cells. If similar mechanisms of expression apply to all LOX genes, then other cell types might contain factors similar to those which permit higher levels of expression of pCAT-2700 in MEL cells than in STO cells. Analysis of expression from the pCAT-2700 recombinant is currently being investigated in other murine cell lines (J. O'Prey, unpublished results). Data obtained thus far indicate that a similar effect to that seen with STO cells is also seen in an epithelial cell line.

4.2.6 Is the cell type-specific effect seen with the -2700 to -1000 region a positive or a negative effect?

The cell type-specific effect conferred upon CAT expression by 5′ flanking sequences between 2700 and 1000nt upstream of the RBC 15-LOX gene results in a higher level of expression in MEL cells than in STO cells. The STO cells show a lower level of expression than

that obtained with the pCAT-150 recombinant which contains the minimal functional promoter (Fig. 23). The most straight-forward explanation of these observations would be that STO cells contain a negative regulatory factor which is absent from, or prevented from binding its cognate sequence, in MEL cells. However, pCAT-2700 contains 6.3kb of DNA, while pCAT-150 is only 3.8kb in size. There is some evidence that the size of transfecting plasmid has a non-cell type-specific effect upon transfection efficiency into both MEL and STO cells (J. Frampton, unpublished results). Expression in MEL and STO cells of a recombinant CAT plasmid which contains 4kb of bacteriophage lambda sequences inserted downstream of the CAT gene in pαCAT1 (Fig. 17), shows an equivalent decrease in both cell types of CAT activity relative to that from the parent, pαCAT1, plasmid. Thus, an alternative explanation of the cell type-specific effect seen with pCAT-2700 may be that the effect of sequences between -1000 and -2700 is a positive effect seen only in MEL cells, but which is masked by a negative effect of plasmid size.

Competition experiments using sequences between -1000 and -2700 could be used to investigate whether the cell type-specific effect is a MEL-specific positive or a STO-specific negative effect. A large molar excess of these sequences, not linked to a CAT gene, would be expected to increase expression from pCAT-2700 in STO cells if it is an STO-specific negative effect, by competing for binding of a negative regulatory factor. If, however, these sequences are involved in an MEL-specific positive effect, they would be expected to compete out a positive factor, resulting in a decrease in CAT expression in pCAT-2700 transfected MEL cells. Similar competition experiments were used to define a cell type-specific negative regulatory region between -334 and -253 of the human retinol-binding protein (RBP) gene (Colantuoni et al, 1987) following a similar set of experiments to those used in this project, employing RBP promoter-deletion/CAT recombinants.

A resolution of this question may also be provided by further CAT expression experiments, using additional deletions of upstream sequences to more accurately localise sequences responsible for the

cell type-specific effect. Having done so, it should be possible to investigate the effect of such sequences on RBC 15-LOX promoter function in smaller CAT plasmids.

4.2.7 Future analysis of the cell type-specific regulatory element.

Having localised the cell type-specific effect to within two or three hundred base pairs by functional analysis of further promoter deletion/CAT recombinants, a large number of experimental approaches can be adopted. Firstly, sequences involved in the binding of transcription factor proteins could be investigated by footprinting, preferably using rabbit nuclear extracts. Secondly, the tissue-distribution of proteins binding to these sequences could be rapidly assessed by electrophoretic mobility shifts with nuclear extracts from a variety of cell types, providing information upon whether the sequences involved in cell type-specific effects are recognised by cell type-specific binding activities. Thirdly, residues which are important for the binding of the appropriate factor(s) could be assessed either by mobility shift assays using oligos. containing point mutations relative to the RBC 15-LOX sequence, or by methylation interference footprinting studies to identify which methylated residues inhibit binding of the factor(s).

It will be interesting to see whether such sequences include any sequences not previously associated with cell type-specific expression, or whether they include elements previously shown to bind cell type-specific nuclear proteins. Within the available 316bp of rabbit RBC LOX 5′ flanking sequence, there is no apparent homology, in either orientation, with the GATAAG sequence (Section 1.4.3) present in the promoter and enhancer regions of a variety of globin and non-globin gene promoters, which binds an erythroid-specific protein (Mantovani *et al*, 1988b; Plumb *et al*, submitted for publication), and which has been suggested to be involved in the co-expression of a variety of genes in erythroid cells (Plumb *et al*, submitted for publication). It will be of interest to discover whether there is a GATAAG motif within the RBC 15-LOX distal element.

The identification of sequences involved in conferring cell type-specificity will allow an investigation of whether such sequences, when linked to a reporter genes as synthetic double-stranded oligos., are sufficient to confer cell type-specificity upon expression from a minimal promoter, as is the case for the immunoglobulin octamer when linked to a region of the mouse renin-1 promoter containing only a TATA box (Dreyfus et al, 1987). Furthermore, it should be possible to investigate whether the positioning of these sequences relative to other cis-acting regulatory elements in an artificial, "designer" promoter has any effect upon expression, as has been suggested to be the case for the "enhanson" sub-units of enhancer elements, which appear to differ from intact enhancer elements in being more sensitive to alterations in spacing (Ondek et al, 1988).

4.2.8 Other cis-regulatory elements in the RBC 15-LOX gene?

In data presented here, the function of RBC 15-LOX upstream sequences in regulating expression of a cis-linked CAT gene from the RBC 15-LOX promoter has been investigated only using 5′ flanking sequences and only in an erythroid cell line which represents a relatively immature stage of differentiation. Experiments to investigate whether RBC 15-LOX 5′ flanking sequences contain regulatory elements which function only at certain stages in differentiation were commenced as part of this project, and have been continued by other members of the group. The design of these experiments is analogous to those of Wright et al (1984) which investigated the expression of the rabbit β-globin gene in MEL cells upon induction of differentiation. For the study of RBC 15-LOX expression during MEL cell differentiation, the same RBC 15-LOX/CAT recombinants used in the transient expression experiments described here were stably introduced into the genome of MEL cells and transformants obtained by selection for a co-transfected G418-resistance plasmid (Homer 6; Spandidos and Wilkie, 1984b). No results are yet available, however, for the expression of CAT activity from these RBC 15-LOX/CAT transformants prior to or following chemical induction of MEL cell maturation.

The discovery of various 3′ enhancers downstream of the transcription initiation site which play important roles in developmental and cell type-specific regulation of globin genes (Choi and Engel, 1986; Antoniou et al, 1988) raises the possibility that not all of the cis-acting sequences necessary for regulation of the RBC 15-LOX gene are contained within the 5′ flanking sequences. This possibility is supported by a set of experiments (J. O'Prey, unpublished results), in which RBC 15-LOX 3′ flanking sequences have been linked downstream of the CAT gene in pCAT-150. A region between 1.5 and 3.7kb downstream of the RBC 15-LOX polyadenylation signal confers a 2- to 3-fold cell type-specific effect upon expression of CAT in the transient assay system described here.

4.2.9 Regulation of the rabbit RBC 15-LOX gene in the intact animal.

In the long-term, one would like to study the expression of the RBC 15-LOX gene, and of various recombinants containing modifications of regulatory sequences, in intact animals. One way to do this would be to investigate the expression of the RBC 15-LOX mRNA during erythroid differentiation in lethally-irradiated animals whose bone marrow has subsequently been repopulated by bone marrow cells infected in vitro by recombinant retroviral vectors containing RBC 15-LOX sequences. Haemopoietic stem cells have been shown to be capable of repopulating bone marrow and expressing G418-resistance and human HPRT (hypoxanthine/guanine phosphoribosyl transferase) genes introduced by retroviral vectors, though only a proportion of the progeny of infected stem cells continue to express the G418-resistance gene (reviewed in Dick et al, 1986). Cone et al (1987) have further demonstrated that expression of the human β-globin gene following infection of MEL cells with a retroviral vector containing globin sequences can be correctly regulated upon chemical induction of MEL cell differentiation.

Another method for studying RBC 15-LOX expression in a

physiological situation would be the generation of transgenic animals. A variety of genes which are expressed in a cell type-specific fashion, including elastase, γ-crystallin, protamine, albumin, immunoglobulin and T cell receptor genes have been studied in transgenic animals (reviewed in Jaenisch, 1988). The erythroid-specific, developmentally-regulated expression of the human β-like globin genes has already been studied following introduction of the globin genes into the germ-line of the mouse. For example, transgenic mice have been used to demonstrate cell type-specific and developmentally-regulated expression of hybrid mouse/human β-globin genes (Chada et al, 1985; Magram et al, 1985), and in the localisation of sequences regulating transcription, where Townes et al (1985) were able to obtain appropriate expression of the human β-globin gene with as little as 48bp of 5′ flanking sequences. Following this demonstration in transgenic animals of the importance of sequences downstream of the transcription initiation site for the regulation of globin gene expression, Kollias et al (1987) have also used transgenic mice to localise at least part of the human β-globin 3′ enhancer to between 0.5 and 1.2kb downstream of the polyadenylation signal.

One of the problems with the use of retroviral infection and transgenic technologies in the study of gene expression has been that the site of integration of the exogenous gene can have a strong influence on the regulation of expression. For example, Nandi et al (1988) have observed that expression of the human β-globin gene in MEL/human cell hybrids could be induced during differentiation when targetted to the recipient cell's β-globin locus, but not when randomly inserted. Two recent advances may facilitate the study of regulated expression of genes such as the RBC 15-LOX gene in erythroid cells of intact animals. Firstly, the ability to "target" genes to the appropriate chromosomal location by the technique of "homologous recombination" (Smithies et al, 1985), as employed in the targetted correction of HPRT deficiency in mouse embryonic stem cells (Doetschman et al, 1987), may reduce the problem posed by such chromosomal "position effects". Secondly, Grosveld et al (1987) have observed that the "domain boundary" sequences lying 50kb upstream and 20kb downstream of the human β-globin gene can confer high-level,

position-independent, cell type-specific expression of the human β-globin gene in transgenic mice. It is unknown at present whether these domain boundary sequences are specific for β-globin genes, or whether they will have a similar effect when linked to other genes. If the latter is the case, such sequences might be extremely valuable in obtaining efficient erythroid-specific expression of the rabbit RBC 15-LOX gene in transgenic animals, allowing study of its regulation during erythroid differentiation by using various modified forms of the gene.

4.3 POSSIBLE ROLES FOR NON-CODING mRNA SEQUENCES IN CONTROL OF RBC 15-LOX EXPRESSION.

The cell type-specific effect conferred by sequences between -2700 and -1000nt upstream of the RBC 15-LOX transcription initiation site upon CAT expression driven from the RBC 15-LOX promoter has been suggested above (Section 4.2.5) to indicate that regulation of transcription may play a role in the observed erythroid-specific accumulation of RBC 15-LOX mRNA (Thiele *et al*, 1987; Figs. 7, 8, 10, 12 and 13 above). However, until nuclear run-on transcription assays have been performed on nuclei isolated from rabbit tissues, there is as yet no firm evidence that there is erythroid-specific regulation of transcription from the RBC 15-LOX promoter. Although there is no data concerning events immediately following transcription of the RBC 15-LOX mRNA it is conceivable that its erythroid-specific accumulation might also involve differential post-transcriptional regulation in erythroid and non-erythroid cell types, and that 5′ and 3′ untranslated sequences might play a role in such post-transcriptional regulation. The untranslated sequences may also be candidates for the translational inactivation of the RBC 15-LOX mRNA observed in younger reticulocytes (Thiele *et al*, 1982; see Introduction, Section 1.6.4).

The mapping of the RBC 15-LOX transcription initiation site reported here (Section 3.2) indicates that the RBC 15-LOX mRNA contains a short 5′ untranslated region of only 27nt. This finding may have implications for the translational inactivation and/or

possible post-transcriptional regulation of the RBC 15-LOX mRNA. Examples of an involvement of 5′ untranslated sequences in translational control are known in a variety of other genes (reviewed in Hunt, 1985). For example, in the case of the mRNA encoding the yeast GCN4 transcription factor protein, deletion from the unusually long 5′ untranslated region of yeast GCN4 mRNA of four short open-reading frames upstream of the authentic translation initiation codon almost completely abolishes the translational repression of the GCN4 messenger (Mueller and Hinnebusch, 1986). This may be due to relief from interference with authentic initiation of translation occurring in accordance with the scanning ribosome model (Kozak, 1986). Furthermore, the insertion of multiple linker sequences capable of forming strong secondary structures into the 5′ untranslated region of the HSV thymidine kinase gene can result in a decrease in translation (Pelletier and Sonenberg, 1985). The short 5′ untranslated region of the RBC 15-LOX mRNA suggests that neither of these mechanisms is a realistic possibility for translational regulation of the RBC 15-LOX mRNA, though a mechanism for translational repression which involves the inhibition of translation by annealing of a single-stranded, "anti-sense" RNA to the 5′ untranslated region, such as that seen when anti-sense globin mRNA was injected into frog oocytes (Melton, 1985), cannot be ruled out.

If the short 5′ untranslated region of the RBC 15-LOX mRNA is not responsible for the translational inactivation seen by Thiele et al (1982), then it is possible that the 3′ untranslated region may be involved. Recent cDNA sequence analysis (B. Thiele, unpublished results) of the 3′ untranslated region of the rabbit RBC 15-LOX mRNA shows that the 575nt 3′ untranslated region contains an array of ten tandem repeats of a 19nt C-rich sequence corresponding to the consensus CCCNNCCCTCTTCCCCAAG, beginning 50nt downstream of the TGA translation termination signal. It seems unlikely that such an arrangement could occur in an actively transcribed mRNA purely by chance, and no such sequence is seen in the 3′ untranslated region of any of the other cloned LOXs.

There is, as yet, no indication as to whether these sequences

are involved in the translational inactivation of the RBC 15-LOX mRNA. However, it will be interesting to see whether removal of some or all of the 10 repeats alters the translation of the RBC 15-LOX mRNA in extracts made from rabbit reticulocytes at various stages of maturation, or, conversely, whether insertion of these sequences into the 3′ untranslated region of another mRNA results in translational inhibition. It is intriguing to note that Thiele et al (1982) have observed an association in reticulocytes of some, but not all, of the RBC 15-LOX mRNA population with messenger ribonucleoprotein (mRNP) particles, separated from polysomes by sucrose density-gradient centrifugation. mRNA within these particles was translationally inactive in a cell-free reticulocyte lysate translation system, but deproteinisation of the mRNPs permitted in vitro translation of authentic RBC 15-LOX protein. It will be of interest to see whether removal of the repeat sequences from the RBC 15-LOX mRNA, or their insertion into a non-LOX mRNA results in an alteration in the distribution of the mRNA between polysomes and the mRNP population.

An alternative possibility is that the repeated sequences might be involved in the cell type-specific accumulation of the RBC 15-LOX mRNA, perhaps by influencing the stability of the messenger in erythroid and non-erythroid cells, for example. There is accumulating evidence for the importance of 3′ untranslated sequences in the determination of mRNA stability (reviewed in Raghow, 1987). For example, a 51nt A/T-rich sequence from the 3′ untranslated region of the mRNA of the human lymphokine, granulocyte/ macrophage colony-stimulating factor (GM-CSF), which includes an AUUUA motif also seen in the 3′ untranslated regions of various lymphokines, cytokines and proto-oncogenes, has been shown to have destabilising effects when introduced into the 3′ untranslated region of globin mRNAs (Shaw and Kamen, 1986). 3′ untranslated sequences have also been suggested to be involved in the rapid degradation of the c-fos mRNA, following transient transcription of the gene (Treisman, 1985).

An intriguing experimental possiblity would be to insert the repeated sequences of the RBC 15-LOX mRNA into the 3′ untranslated regions of mRNAs which are usually unstable, to examine whether the

presence of the repeats confers a decrease in mRNA stability in erythroid cells, or whether, in complementary experiments, insertion into the 3' untranslated region of an unstable messenger results in selective stabilisation in erythroid cells.

4.4 AN EVOLUTIONARY RELATIONSHIP BETWEEN THE LIPOXYGENASES.

Data presented here includes a region of the rabbit RBC 15-LOX gene which accurately predicts the first 30 amino acids of the rabbit RBC 15-LOX protein (Thiele et al, 1987). An alignment of this protein sequence with the N-terminal sequence of the human leukocyte 5-LOX as predicted from cDNA sequences (Dixon et al, 1988; Matsumoto et al, 1988) and with the peptide sequence of a 5-LOX from rat basophilic leukaemia cells (Hogaboom et al, 1986) is shown in Figure 24. As noted by Dixon et al (1988), and by Matsumoto et al (1988), the human and rat N-terminal sequences are virtually identical, showing the same amino acids at 28 of the first 30 residues. The rabbit RBC 15-LOX, however, shows identity with the human 5-LOX at only 11 of the first 30 amino acid residues (37% homology) with a further 6 residues showing "conservative" changes. The greater divergence of rabbit from human and rat sequences may represent different evolutionary pressures operating on enzymes which differ in their positional specificity, rather than a cross-species difference.

Dixon et al (1988) have noted similarities of different regions of the human 5-LOX protein with the predicted partial amino acid sequence of RBC 15-LOX (Thiele et al, 1987) and with another 15-LOX, the soybean LOX-1 isozyme (Shibata et al, 1987). Yenofsky et al (1988) have further observed 70% amino acid homology between three soybean isozymes. Although not yet completely sequenced, recently-obtained cDNA clones covering the full length of the RBC 15-LOX mRNA (B. Thiele and J. Fleming, unpublished results) allow a more detailed comparison of LOX protein sequences from these three species, and provide some clues as to the evolution and structure-function relationships of the LOXs. An alignment of the incomplete RBC 15-LOX protein sequence with the sequences predicted by the soybean LOX-3 gene and human leukocyte 5-LOX cDNA sequences is

shown in Figure 24, and schematically in Figure 25.

The predicted RBC 15-LOX amino acid sequence shows three regions of significant homology, designated C, D, and E in Figures 24 and 25, with those of the soybean and human proteins. When these three regions are aligned in each of the three proteins, the soybean protein appears to contain an N-terminal region (A) of 135 amino acids which is not present in either the human or rabbit proteins. There is limited homology between soybean and human proteins over the first 364 amino acids of the human protein (region B). Other than the N-terminal sequence, RBC 15-LOX peptide sequence is not yet available for this region.

Of the regions for which amino acid sequence is available in all three proteins, regions C and E, which are well conserved between all three species, are separated by a region, D, which does not seem to be strongly conserved between the plant and mammalian enzymes, but is strongly conserved between the human and rabbit LOXs (Figure 25).

Within regions C, D and E, the degree of conservation between soybean LOX-3 and human 5-LOX sequences, and between human 5-LOX and rabbit 15-LOX sequences are very similar (Figure 25), but the degree of homology between rabbit and soybean is less than that between human and soybean in all cases, especially in region E. This would seem to indicate that the evolutionary divergence between soybean and human LOXs is less than that between the soybean and rabbit enzymes. This might be attributed to different selection pressures operating during evolution on LOXs having different positional specificities. Intriguingly, a study of the 3' untranslated region of the soybean LOX isozymes (Yenofsky et al, 1988) has shown that LOX-1 and LOX-2 are more closely related to each other than either is to LOX-3. Yet the enzymic properties, including positional specificity, of LOX-2 are more similar to LOX-3 than to those of LOX-1. Thus, the rate of evolution of LOX genes in soybean does not seem to reflect their positional specificity. It may, therefore, reflect different biological functions. While the physiological functions of the soybean LOX isozymes is incompletely understood (see Introduction

Figure 24. Comparison of available LOX protein sequences from various species.

Top. Comparison of the N-terminal 30 amino acids of rat basophilic leukaemia 5-LOX (Hogaboom et al, 1986), human leukocyte 5-LOX (Dixon et al, 1988; Matsumoto et al, 1988), and rabbit RBC 15-LOX (Thiele et al, 1987) proteins. Identical amino acids are indicated by vertical lines between sequences; "conservative" amino acid changes are indicated by asterisks.

Bottom. Optimal alignment of available rabbit RBC 15-LOX protein sequence, as deduced from cDNA sequences (B. Thiele and J. Fleming, unpublished results), with predicted protein sequences from the soybean lipoxygenase-3 isozyme gene (Yenofsky et al, 1988) and human leukocyte 5-LOX cDNAs (Dixon et al, 1988; Matsumoto et al, 1988). Three regions of significant homology between human and rabbit sequences, corresponding to regions C, D and E in Figure 25, are indicated by over-lining. Two of these (C and E) also show significant homology with the soybean LOX-3 isozyme. Soybean residues are numbered relative to the N-terminal amino acid; in the absence of complete cDNA sequence, rabbit RBC 15-LOX residues are numbered relative to the first residue of the sequence shown in this portion of the diagram, designated as n+1; positions of homologous residues of the human 5-LOX protein are given in Figure 25.

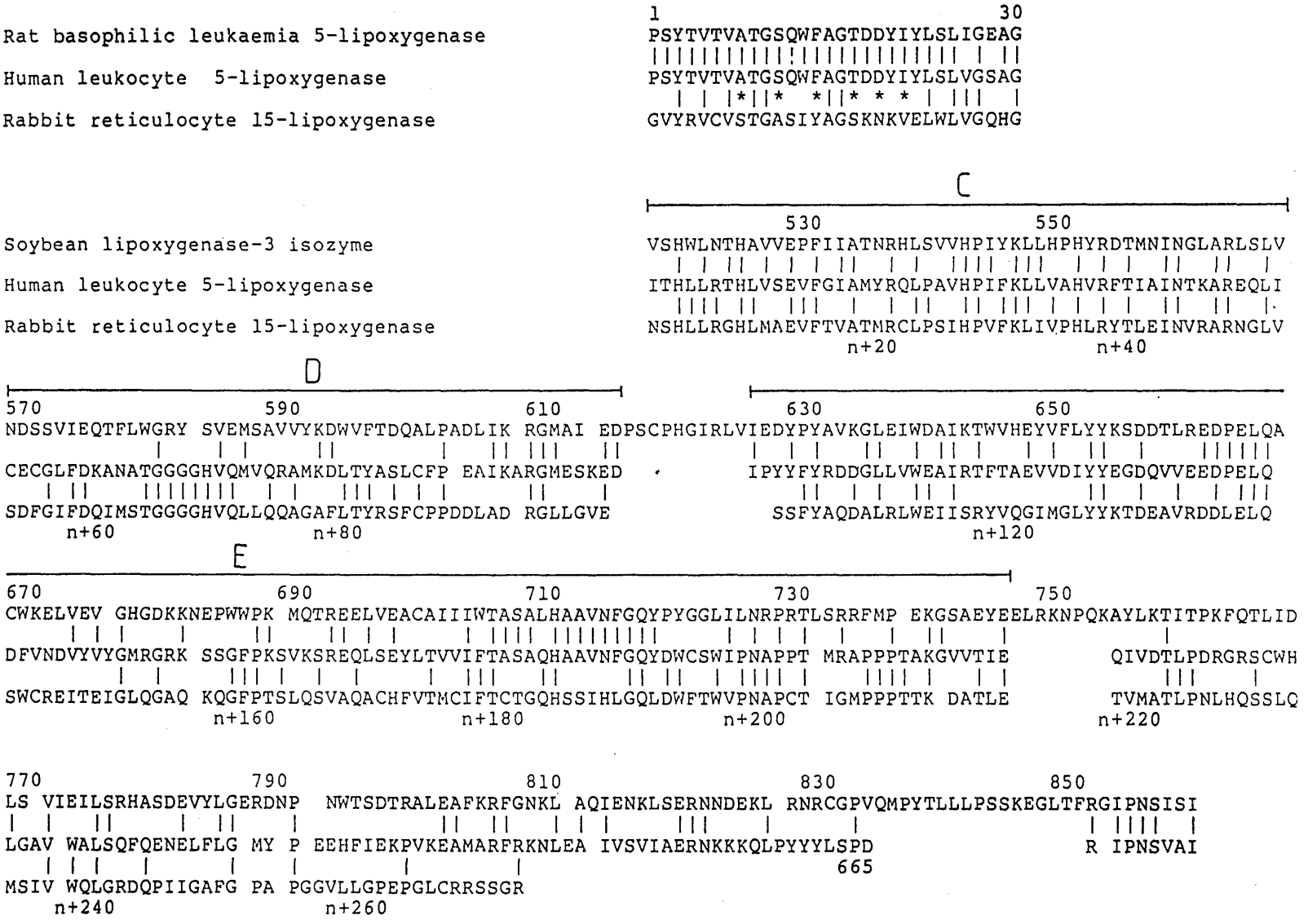
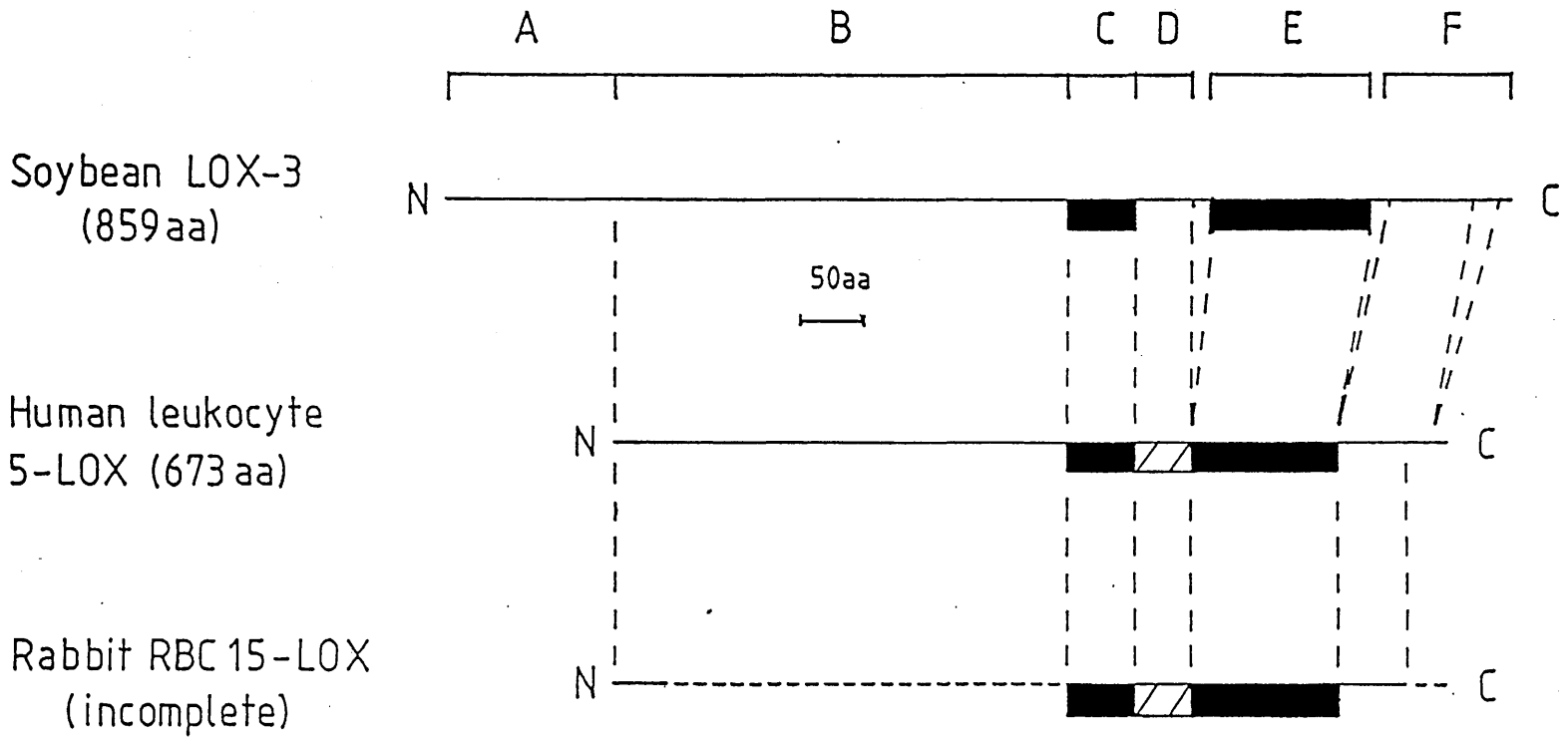

Rat basophilic leukaemia 5-lipoxygenase
Human leukocyte 5-lipoxygenase
Rabbit reticulocyte 15-lipoxygenase
1
30
PSYTVTVATGSQWFAGTDDYIYLSLIGEAG
PSYTVTVATGSQWFAGTDDYIYLSLVGSAG
GVYRVCVSTGASIYAGSKNKVELWLVGQHG
C
530
550
Soybean lipoxygenase-3 isozyme
Human leukocyte 5-lipoxygenase
Rabbit reticulocyte 15-lipoxygenase
VSHWLNTHAVVEPFIIATNRHLSVVHPIYKLLHPHYRDTMNINGLARLSLV
ITHLLRTHLVSEVFGIAMYRQLPAVHPIFKLLVAHVRFTIAINTKAREQLI
NSHLLRGHLMAEVFTVATMRCLPSIHPVFKLIVPHLRYTLEINVRARNGLV
n+20
n+40
D
570
590
610
630
650
NDSSVIEQTFLWGRY SVEMSAVVYKDWVFTDQALPADLIK RGMAI EDPSCPHGIRLVIEDYPYAVKGLEIWDAIKTWVHEYVFLYYKSDDTLREDPELQA
CECGLFDKANATGGGGHVQMVQRAMKDLTYASLCFP EAIKARGMESKED IPYYFYRDDGLLVWEAIRTFTAEVVDIYYEGDQVVEEDPELQ
SDFGIFDQIMSTGGGGHVQLLQQAGAFLTYRSFCPPDDLAD RGLLGVE SSFYAQDALRLWEIISRYVQGIMGLYYKTDEAVRDDLELQ
n+60
n+80
n+120
E
670
690
710
730
750
CWKELVEV GHGDKKNEPWWPK MQTREELVEACAIIIWTASALHAAVNFGQYPYGGLILNRPRTLSRRFMP EKGSAEYEELRKNPQKAYLKTITPKFQTLID
DFVNDVYVYGMRGRK SSGFPKSVKSREQLSEYLTVVIFTASAQHAAVNFGQYDWCSWIPNAPPT MRAPPPTAKGVVTIE QIVDTLPDRGRSCWH
SWCREITEIGLQGAQ KQGFPTSLQSVAQACHFVTMCIFTCTGQHSSIHLGQLDWFTWVPNAPCT IGMPPPTTK DATLE TVMATLPNLHQSSLQ
n+160
n+180
n+200
n+220
770
790
810
830
850
LS VIEILSRHASDEVYLGERDNP NWTSDTRALEAFKRFGNKL AQIENKLSERNNDEKL RNRCGPVQMPYTLLLPSSKEGLTFRGIPNSISI
LGAV WALSQFQENELFLG MY P EEHFIEKPVKEAMARFRKNLEA IVSVIAERNKKKQLPYYYLSPD R IPNSVAI
665
MSIV WQLGRDQPIIGAFG PA PGGVLLGPEPGLCRRSSGR
n+240
n+260

Figure 25. Conserved regions of LOX proteins from various species.

Schematic diagram of the optimal protein sequence alignment shown in Figure 24. Regions conserved between all three proteins are indicated by shaded boxes. A region of homology between rabbit and human sequences which is less strongly conserved compared with the soybean LOX-3 isozyme is indicated by cross-hatched boxes. The incomplete portion of the rabbit RBC 15-LOX sequence is shown as a broken horizontal line.

Percentage identities between the three amino acid sequences in six regions (A - F) are shown in the lower part of the diagram, beneath the limits of each region. Also shown are the degrees of conservation over larger portions of available sequence (C+D, etc.)



	A	B	C	D	E	F	C+D	C+D+E	C to F	Overall
Soybean LOX-3	1 - 135	136 - 516	517 - 566	567 - 614	625 - 745	754 - 859	517 - 614	517 - 745	517 - 859	136 - 859
Human 5-LOX		1 - 364	365 - 414	415 - 464	465 - 585	586 - 673	365 - 464	365 - 585	365 - 673	1 - 673
Rabbit 15-LOX			n+1 - n+50	51 99	100 217	218 271	1 99	1 217	1 271	
Soybean vs human	-------	26%	52%	26%	44%	27%	40%	41%	36%	31%
Human vs rabbit	-------	?	54%	44%	38%	22%	49%	43%	39%	?
Soybean vs rabbit	-------	?	46%	16%	25%	17%	31%	27%	25%	?

Section 1.5.2), the widely different functions of human 5-LOX and rabbit 15-LOXs, in leukotriene synthesis and mitochondrial degradation respectively, lend support to this idea.

The conservation of regions of LOX proteins from plants to mammals may suggest that these sequences are of functional importance, perhaps forming part of the catalytic site. In fact, Dixon et al (1988) have noted that the portion of the human 5-LOX protein sequence within region C is related to the interface-binding domain of human and rat lipases, suggesting that the mechanism of interaction of LOX and lipase enzymes with their substrates may be similar. There is no homology, however, between the available LOX protein sequences and the protein sequences predicted by the recently-published cDNA sequence of the sheep seminal vesicle cyclo-oxygenase (Merlie et al, 1988), whose main substrate, arachidonic acid, is also oxidised by all of the LOXs discussed here.

Whatever their function, the conservation of these three extended regions of protein sequence homology over such long periods of evolutionary time would seem to suggest that these three members of the LOX enzyme family, at least, are encoded by an evolutionarily-related family of genes rather than being the result of convergent evolution towards a common function.

4.5 FUTURE DIRECTIONS.

The rabbit erythroid-specific RBC 15-LOX gene thus presents several interesting avenues for further research. Firstly, the continuation, as outlined in Section 4.2 above, of the investigation begun here into the importance of transcription in the erythroid-specific accumulation of RBC 15-LOX mRNA, and further localisation of the regulatory signals involved. Secondly, to study whether the repeated region within the 3′ untranslated region of the RBC 15-LOX mRNA plays a part in post-transcriptional regulation of RBC 15-LOX expression: either by influencing the erythroid-specific accumulation of the mRNA by cell type-specific effects upon a post-transcriptional event; or by playing a role in the translational

inactivation of RBC 15-LOX mRNA translation seen in younger reticulocyte populations. Finally, comparison with the products of other LOX genes may provide important information concerning structure-function relationships in the LOX enzymes, including the nature of the catalytic site, and the basis for the different positional specificities of the various members of the LOX enzyme family.

REFERENCES.

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K. & Watson, J. D. (1983). Molecular Biology of the Cell. Garland Publishing Inc., New York.

Allan, M., Lanyon, W.G. & Paul, J. (1983). Multiple origins of transcription in the 4.5kb upstream of the ε-globin gene. Cell 35,187

Allan, M., Montague, P., Grindlay G.J., Sibbet, G., Donnovan-Peluso, M., Bank, A.W. & Paul, J. (1985). Tissue-specific transcription of the human ε-globin gene following transfection into the embryonic erythroid cell line US62. Nucleic Acids Res. 13, 6125-6136.

Allison, L. A., Moyle, M., Shales, M. & Ingles, C. J. (1985). Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA Polymerases. Cell 42, 599-610.

Alt, F. W., Blackwell, T. K. & Yancopoulos, G. D. (1985). Immunoglobulin genes in transgenic mice. Trends in Genetics 1, 231-236.

Alt, F. W., Bothwell, A. L. M., Knapp, M., Sidden, E., Mather, E., Koshland, M. & Baltimore, D. (1980). Synthesis of secreted and membrane-bound immunoglobulin μ heavy chains directed by mRNAs that differ at their 3' ends. Cell 20, 293-301.

Antoniou, M., de Boer, E., Habets, G. & Grosveld, F. (1988). The human β-globin gene contains multiple regulatory regions: identification of one promoter and two downstream enhancers. EMBO J. 7, 377-384.

Axelrad, A. A., MacLeod, D. L., Shreeve, M. M. & Heath, D. S. (1973) Properties of cells that produce erythrocytic colonies in vitro. In Proc. Second Int. Workshop on Hemopoiesis in culture (DHEW Publication No. NIH 74-205, Ed. Robinson, W. A.) pp. 226-237.

Axelrod, B., Cheesbrough, T. M. & Laasko, S. (1981). Lipoxygenase from soyabean. Methods in Enzymology 71, 441-451.

Banerji, J., Olson, L. & Schaffner, W. (1983). A lymphocyte-specific cellular enhancer is located downstream of the joining region in Ig heavy chain genes. Cell 33, 729-740.

Banerji, J., Rusconi, S. & Schaffner, W. (1981). Expression of a β-globin gene is enhanced by remote SV40 sequences. Cell 27, 299-308.

Barberis, A., Superti-Furga, G. & Busslinger, M. (1987). Mutually exclusive interaction of the CCAAT-binding factor and of a displacement protein with overlapping sequences of a histone gene promoter. Cell 50, 347-359.

Benoist, C. & Chambon, P. (1981). In vivo sequence requirements of the SV40 early promoter region. Nature 290, 304-310.

Benoist, C., O'Hare, K., Breathnach, R. & Chambon, P. (1980). The ovalbumin gene - sequence of putative control regions. Nucleic Acids Res. 8, 127-142.

Bentley, D. L. (1984). Most κimmunoglobulin mRNA in human lymphocytes is homologous to a small family of germ-line V genes. Nature 307, 77-80.

Berk, A. J. & Sharp, P. A. (1977). Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease-digested hybrids. Cell 12, 721-732.

Bienz, M. & Pelham, H. R. B. (1986). Heat shock regulatory elements function as an inducible enhancer in the Xenopus hsp70 gene and when linked to a heterologous promoter. Cell 45, 753-760.

Biggin, M.D., Gibson, T.J. & Hong, G.F. (1983). Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. Proc. Nat. Acad. Sci., USA 80, 3963-3965.

Biggs, J., Searles, L. L. & Greenleaf, A. L. (1985). Structure of the eukaryotic transcription apparatus: features of the gene for the largest subunit of Drosophila RNA Polymerase II. Cell 42, 611-621.

Bingham, P. M., Chou, T-B., Mims, I. & Zachar, Z. (1988). On/off regulation of gene expression at the level of splicing. Trens in Genetics 4, 134-138.

Bird, A. P. (1984). DNA methylation - how important in gene control? Nature 307, 503-504.

Bird, A. P. (1986). CpG-rich islands and the function of DNA methylation. Nature 321, 209-213.

Birnboim, H. C. & Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. 7, 1513-1523.

Birnstiel, M.L., Busslinger, M. & Strub, K. (1985). Transcription termination and 3' processing: The end is in sight! Cell 41, 349-359.

Blake, C. C. F. (1978). Do genes-in-pieces imply proteins-in-pieces?. Nature 273, 267.

Bodine, D. M. & Ley, T. J. (1987). An enhancer element lies 3' to the human $^{A}\gamma$globin gene. EMBO J. 6, 2997-3004.

Bohmann, D., Bos, T. J., Admon, A., Nishimura, T., Vogt, P. K. & Tjian, R. (1987). Human proto-oncogene c-jun encodes a DNA binding protein with structural and functional properties of transcription factor AP-1. Science 238, 1386-1392.

Bonner, J. J. & Pardue, M. L. (1977). Polytene chromosome puffing and in situ hybridisation measure different aspects of RNA metabolism. Cell 12, 227-234.

Bosze, Z., Thiesen, H-J. & Charnay, P. (1986). A transcriptional enhancer with specificity for erythroid cells is located in the LTR of the Friend murine leukaemia virus. EMBO J. 5, 1615-1623.

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilising the principle of protein-dye binding. Anal. Biochem. 72, 248-254.

Brand, A. H., Breeden, L., Abraham, J., Sternglanz, R. & Nasmyth, K. (1985). Characterisation of a "silencer" in yeast: a DNA sequence with properties opposite to those of a transcriptional enhancer. Cell 41, 41-48.

Breathnach, R. & Chambon, P. (1981). Organization and expression of eukaryotic split genes coding for proteins. Ann. Rev. Biochem. 50, 349-383.

Brent, R. (1985). Repression of transcription in yeast. Cell 42, 3-4.

Brent, R. & Ptashne, M. (1985). A eukaryotic transcriptional activator bearing the DNA specificity of a prokaryotic repressor. Cell 43, 729-736.

Briggs, M. R., Kadonaga, J. T., Bell, S. P. & Tjian, R. (1986). Purification and biochemical characterization of the promoter-specific transcription factor Sp1. Science 234, 47-52.

Britten, R. J. & Davidson, E. H. (1969). Gene regulation in higher cells: a theory. Science 165, 349-357.

Brock, M. L. & Shapiro, D. J. (1983). Estrogen stabilizes vitellogenin mRNA against cytoplasmic degradation. Cell 34, 207-214.

Brown, R.S., Sander C. & Argos, P. (1985). FEBS. Letts. 186, 271-274

Bryant, R. W., Bailey, J. M., Schewe, T. & Rapoport, S. M. (1982). Positional specificity of a reticulocyte lipoxygenase. J. Biol. Chem. 257, 6050-6055.

Buratowski, S., Hahn, S., Sharp, P.A. & Guarente, L. (1988). Function of a yeast TATA element-binding protein in a mammalian transcription system. Nature 334, 37-42

Busslinger, M., Portmann, R., Irminger, J. D. & Birnstiel, M. L. (1980). Ubiquitous and gene-specific regulating 5′ sequences in a sea urchin histone DNA clone coding for histone protein variants. Nucleic Acids Res. 8, 957-977.

Carbon, P., Murgo, S., Ebel, J-P., Krol, A., Tebb, G. & Mattaj, I. W. (1987). A common octamer motif binding protein is involved in the transcription of U6 snRNA by RNA Polymerase III and U2 snRNA by RNA Polymerase II. Cell 51, 71-79.

Cavallani, B., Huet, J., Plassat, J-L., Sentenac, A., Egly, J-M. & Chambon, P. (1988). A yeast activity can substitute for the HeLa cell TATA box factor. Nature 334, 77-80.

Cech, T. R. (1986). The generality of self-splicing RNA: relationship to nuclear mRNA splicing. Cell 44, 207-210.

Cedar, H. (1988). DNA methylation and gene activity. Cell 53, 3-4

Chada, K., Magram, J., Raphael, K., Radice, G., Lacy, E. & Constantini, F. (1985). Specific expression of a foreign β-globin gene in erythroid cells of transgenic mice. Nature 314, 377-380.

Chambers, I., Frampton, J., Goldfarb, P., Affara, N., McBain, W. & Harrison, P. R. (1986). The structure of the mouse glutathione peroxidase gene: the selenocysteine in the active site is encoded by the "termination" codon, TGA. EMBO J. 5, 1221-1227.

Chao, M. V., Mellon, P., Charnay, P., Maniatis, T. & Axel, R. (1983). The regulated expression of β-globin genes introduced into mouse erythroleukaemia cells. Cell 32, 483-493.

Charnay, P. & Maniatis, T. (1983). Transcriptional regulation of globin gene expression in the human erythroid cell line K562. Science 220, 1281-1283.

Charnay, P., Mellon, P. & Maniatis, T. (1985). Linker scanning mutagenesis of the 5'-flanking region of the mouse β-major-globin gene: sequence requirements for transcription in erythroid and non-erythroid cells. Mol. Cell. Biol. 5, 1498-1511.

Charnay, P., Treisman, R., Mellon, P., Chao, M., Axel, R. & Maniatis, T. (1984). Differences in human α- and β-globin gene expression in mouse erythroleukaemia cells: the role of intragenic sequences. Cell 38, 251-263.

Chen, T. R. (1977). In situ detection of mycoplasma contamination in cell cultures by fluorescent Hoechst 33258 stain. Exp. Cell. Res. 104, 255-262.

Chiu, R., Boyle, W. J., Meek, J., Smeal, T., Hunter, T. & Karin, M. (1988). The c-fos protein interacts with c-jun/AP-1 to stimulate transcription of AP-1 responsive genes. Cell 54, 541-552.

Chodosh, L. A., Baldwin, A. S., Carthew, R. W. & Sharp, P. A. (1988a). Human CCAAT-binding proteins have heterologous subunits. Cell 53, 11-24.

Chodosh, L. A., Olesen, J., Hahn, S., Baldwin, A. S., Guarente, L. & Sharp, P. A. (1988b). A yeast and a human CCAAT-binding protein have heterologous subunits that are functionally interchangeable. Cell 53, 25-35.

Chomczynski, P. & Sacchi, N. (1987). Single-step method of RNA isolation by guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem 162, 156-159.

Choi, O-R. & Engel, J. D. (1986). A 3' enhancer is required for temporal and tissue-specific transcriptional activation of the chicken adult β-globin gene. Nature 323, 731-734.

Chretien, S., Dubart, A., Beaupain, D., Raich, N., Grandchamp, B., Rosa, J., Goossens, M. & Romeo, P-H. (1988). Alternative transcription and splicing of the human porphobilinogen deaminase gene result either in tissue-specific or in housekeeping expression. Proc. Nat. Acad. Sci., USA 85, 6-10.

Chung, J., Sussman D.J., Zeller, R. & Leder, P. (1987). The c-myc gene encodes superimposed RNA polymerase II and III promoters. Cell 51, 1001-1008.

Clark, S. C. & Kamen, R. (1987). The human hematopoietic colony stimulating factors. Science 236, 1229-1237.

Cohen, P. (1982). The role of protein phosphorylation in neural and hormonal control of cellular activity. Nature 296, 613-620.

Colantuoni, V., Pirozzi, A., Blance, C. & Cortese, R. (1987). Negative control of liver-specific gene expression: cloned human retinol-binding protein is repressed in HeLa cells. EMBO J. 6, 631-636.

Cone, R. D., Weber-Benarous, A., Baorto, D. & Mulligan, R. C. (1987). Regulated expression of a complete human β-globin gene encoded by a transmissible retrovirus vector. Mol. Cell. Biol. 7, 887-897.

Corden, J., Cadera, D. L., Ahearn, J. M. & Dahmus, M. E. (1985). A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA Polymerase II. Proc. Nat. Acad. Sci., USA 82, 7934-7938.

Darnell, J. E. (1982). Variety in the level of gene control in eukaryotic cells. Nature 297, 365-371.

Deisseroth, A., Bode, U., Fontana, J. & Hendrick, D. (1980a). Expression of human α-globin genes in hybrid mouse erythroleukaemia cells depends on differentiated state of human donor cell. Nature 285, 36-38.

Deisseroth, A., Bode, U., Lebo, R., Dozy, A. & Kan, Y. W. (1980b). Isolation of hybrid cell clones that contain deletion and non-deletion defects of α-thalassaemia in man. Blood 55, 992-997.

Deisseroth, A. & Hendrick, D. (1978). Human α-globin gene expression following chromosomal dependent gene transfer into mouse erythroleukaemic cells. Cell 15, 55-63.

Denton, M. J. & Arnstein, H. R. V. (1973). Characterisation of developing adult mammalian erythroid cells separated by velocity sedimentation. Brit. J. Haematol. 24, 7-17.

Derman, E., Krauter, K., Walling, L., Weinberger, C., Ray, M. & Darnell, J. E. (1981). Transcriptional control in the production of liver-specific mRNAs. Cell 23, 231-239.

De Simone, V., Ciliberto, G., Hardon, E., Paonessa, G., Palla, F., Lundberg, L. & Cortese, R. (1987). Cis- and trans-acting elements responsible for the cell-specific expression of the human α1-antitrypsin gene. EMBO J. 6, 2759-2766.

de The, H., Marchio, A., Tiollias, P. & Dejean, A. (1987). A novel steroid thyroid hormone receptor-related gene inappropriately expressed in human hepatocellular carcinoma. Nature 330, 667-670.

Dick, J. E., Magli, M. C., Phillips, R. A. & Bernstein, A. (1986) Genetic manipulation of hematopoietic stem cells with retrovirus vectors. Trends in Genetics 2, 165-170.

Dierks, P., van Ooyen, A., Cochran, M. D., Dobkin, C., Reiser, J. & Weissmann, C. (1983). Three regions upstream from the cap site are required for efficient and accurate transcription of the rabbit β-globin gene in mouse 3T6 cells. Cell 32, 695-706.

Dierks, P., van Ooyen, A., Mantei, N. & Weissmann, C. (1981). DNA sequences preceding the rabbit β-globin gene are required for formation in mouse L cells of β-globin RNA with the correct 5′ terminus. Proc. Nat. Acad. Sci., USA 78, 1141-1145.

Dixon, R. A. F., Jones, R. E., Diehl, R. E., Bennett, C. D., Kargman, S. & Rouzer, C. A. (1988). Cloning of the cDNA for human 5-lipoxygenase. Proc. Nat. Acad. Sci., USA 85, 416-420.

Doetschman, T., Gregg, R. G., Maeda, N., Hooper, M. L., Melton, D. W., Thompson, S. & Smithies, O. (1987). Targeted correction of a mutant HPRT gene in mouse embryonic stem cells. Nature 330, 576-578.

Dorn, A., Bollekens, J., Staub, A., Benoist, C. & Mathis, D. (1987). A multiplicity of CCAAT box-binding proteins. Cell 50, 863-872.

Dreyfus, M., Doyen, N. & Rougeon, F. (1987). The conserved decanucleotide from the immunoglobulin heavy chain promoter induces a very high transcriptional activity in B-cells when introduced into an heterologous promoter. EMBO J. 6, 1685-1690.

Dynan, W. S. (1985). A single protein that binds to enhancers, promoters and replication origins? Trends in Genetics 1, 269-270.

Dynan, W. S. (1986). Promoters for house-keeping genes. Trends in Genetics 2, 196-197.

Dynan, W. S. & Tjian, R. (1985). Control of eukaryotic mRNA synthesis by sequence-specific DNA-binding proteins. Nature 316, 774-778.

Early, P., Rogers, J., Davis, M., Calame, K., Bond, M., Wall, R. & Hood, L. (1980). Two mRNAs can be produced from a single immunoglobulin μ gene by alternative processing pathways. Cell 20, 313-319.

Efstratiadis, A., Posakony, J. W., Maniatis, T., Lawn, R. M., O'Connell, C., Spritz, R. A., De Reil, J. K., Forget, B. G., Weissman, S. M., Slightom, J. C., Blechl, A. E., Smithies, O., Baralle, F. E., Shoulders, C. C. & Proudfoot, N. J., (1980). The structure and evolution of the human β-globin gene family. Cell 21, 653-668.

Emerson, B. M., Lewis, C. D. & Felsenfeld, G. (1985). Interaction of specific nuclear factors with the nuclease-hypersensitive region of the chicken adult β-globin gene: nature of the binding domain. Cell 41, 21-30.

Emerson, B. M., Nickol, J. M., Jackson, P. D. & Felsenfeld, G. (1987). Analysis of the tissue-specific enhancer at the 3' end of the chicken adult β-globin gene. Proc. Nat. Acad. Sci., USA 84, 4786-4790.

Falkner, F. G. & Zachau, H. G. (1984). Correct transcription of an immunoglobulin κ gene requires an upstream fragment containing conserved sequence elements. Nature 310, 71-74.

Feinberg, A. P. & Vogelstein, B. (1983). A technique for radiolabelling DNA restriction endonuclease fragments to high specificity. Anal. Biochem. 132, 6-13

Feinberg, A. P. & Vogelstein, B. (1984). A technique for radiolabelling DNA restriction endonuclease fragments to high specificity. Addendum Anal. Biochem. 137, 266-267

Fink, G. R. (1986). Translational control of transcription in eukaryotes. Cell 45, 155-156.

Frampton, J., Conkie, D., Chambers, I., McBain, W., Dexter, M. & Harrison, P. R. (1987). Changes in minor transcripts from the α-1 and β-major globin and glutathione peroxidase genes during erythropoiesis. Nucleic Acids Res. 15, 3671-3688.

Friend, C., Scher, W., Holland, J. G. & Sato, T. (1971). Hemoglobin synthesis in murine virus-induced leukemic cells in vitro: stimulation of erythroid differentiation by dimethylsulfoxide. Proc. Nat. Acad. Sci., USA 68, 378-382.

Fujita, T., Shibuya, H., Ohashi, T., Yamanishi, K. & Taniguchi, T. (1986). Regulation of the human interleukin-2 gene: Functional DNA sequences in the 5′ flanking region for the gene expression in activated T lymphocytes. Cell 46, 401-407.

Gaffney, D. F., McLauchlan, J., Whitton, J. L. & Clements, J. B. (1985). A modular system for the assay of transcription regulatory signals: the sequence TAATGARAT is required for herpes simplex virus immediate early gene activation. Nucleic Acids Res. 13, 7847-7863.

Galas, D. J. & Schmitz, A. (1978). DNAse footprinting: a simple method for the detection of protein: DNA binding specificity. Nucleic Acids Res. 5, 3157-3170.

Galson, D.L. Housman, D.E. (1988). Detection of two tissue-specific DNA-binding proteins with affinity for sites in the mouse β-globin intervening sequence 2. Mol. & Cell. Biol. 8, 381-392.

Garcia, J. V., Bich-Thuy, L., Stafford, J. & Queen, C. (1986). Synergism between immunoglobulin enhancers and promoters. Nature 322, 383.

Gasser, S. M. & Laemmli, U. K. (1986). The organization of chromatin loops: characterisation of scaffold attachment sites. EMBO J. 5, 511-518.

Ghosh, P. K., Lebowitz, P., Frisque, R. J. & Gluzman, Y. (1981). Identification of a promoter component involved in positioning the 5′ termini of simian virus 40 early mRNAs. Proc. Nat. Acad. Sci., USA 78, 100-104

Gill, G. & Ptashne, M. (1987). Mutants of GAL4 protein altered in an activation function. Cell 51, 121-126.

Gillies, S. D., Morrison, S. L., Oi, V. T. & Tonegawa, S. (1983). A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. Cell 33, 717-728.

Giniger, E., Varnum, S. M. & Ptashne, M. (1985). Specific DNA binding of GAL4, a positive regulatory protein of yeast. Cell 40, 767-774.

Goldberg, M. (1979). PhD Thesis. Stanford University, Stanford, California.

Goldman, M. A. (1988). The chromatin domain as a unit of gene regulation. BioEssays 9, 50-55.

Goodbourn, S., Burstein, H. & Maniatis, T. (1986). The human β-interferon gene enhancer is under negative control. Cell 45, 601-610.

Gorman, C. (1985). High efficiency gene transfer into mammalian cells. Chapter 6 in: DNA Cloning: A Practical Approach. (Ed. Glover, D. M.). IRL Press, Oxford, pp. 143-190.

Gottesfeld, J. & Bloomer, L. S. (1982). Assembly of transcriptionally active 5S RNA gene chromatin in vitro. Cell 28, 781-791.

Graham, F. L. & Van der Eb, A. J. (1973). A new technique for the assay of infectivity of human and adenovirus 5 DNA. Virology 52, 456-461.

Grandchamp, B., De Vermeuil, H., Beaumont, C., Chretien, S., Walter, O. & Nordmann, Y. (1987). Tissue-specific expression of porphobilinogen deaminase: two isoenzymes from a single gene. Eur. J. Biochem. 162, 105-110.

Grosschedl, R. & Birnstiel, M. L. (1980). Identification of regulatory sequences in the prelude sequences of an H2A histone gene by the study of specific deletion mutants in vivo. Proc. Nat. Acad. Sci., USA 77, 1432-1436.

Grosveld, F., von Assendelft, G. B., Greaves, D. R. & Kollias, G. (1987). Position-independent high-level expression of the human β-globin gene in transgenic mice. Cell 51, 975-985.

Grosveld, G. C., Koster, A. & Flavell, R. A. (1981a). A transcription map for the rabbit β-globin gene. Cell 23, 573-584.

Grosveld, G. C., Shewmaker, C. K., Jat, P. & Flavell, R. A. (1981b). Localisation of DNA sequences necessary for transcription of the rabbit β-globin gene in vitro. Cell 25, 215-226.

Grosveld, G. C., de Boer, E., Shewmaker, C. K. & Flavell, R. A. (1982a). DNA sequences necessary for transcription of the rabbit β-globin gene in vivo. Nature 295, 120-126.

Grosveld, G. C., Rosenthal, A. & Flavell, R. A. (1982b). Sequence requirements for the transcription of the rabbit β-globin gene in vivo: the -80 region. Nucleic Acids Res. 10, 4951-4971.

Groudine, M. & Weintraub, H. (1982). Propagation of globin DNase I-hypersensitive sites in absence of factors required for induction: a possible mechanism for determination. Cell 30, 131-139.

Guarente, L. (1988). UASs and enhancers: common mechanism of transcriptional activation in yeast and mammals. Cell 52, 303-305.

Guinta, D. R. & Korn, L. J. (1986). Differential order of replication of Xenopus laevis 5S RNA genes. Mol. Cell. Biol. 6, 2536-2542.

Guyette, W. A., Matusik, R. J. & Rosen, J. M. (1979). Prolactin-mediated transcriptional and post-transcriptional control of casein gene expression. Cell 17, 1013-1023.

Hagenbuchle, O., Toci, M., Schibler, U., Bovey, R., Wellaver, P. K. & Young, R. A. (1981). Mouse liver and salivary gland α-amylase mRNAs differ only in 5′ non-translated sequences. Nature 289, 643-646.

Hall, C. V., Jacob, P. E., Ringold, G. M. & Lee, F. (1983). Expression and regulation of Escherichia coli lac Z gene fusions in mammalian cells. J. Mol. Appl. Gen. 2, 101-109.

Hammer, R. E., Swift, G. H., Ornitz, D. M., Quaife, C. J., Palmiter, R. D., Brinster, R. L. & MacDonald, R. J. (1987). The rat elastase 1 regulatory element is an enhancer that directs correct cell specificity and developmental onset of expression in transgenic mice. Mol. Cell Biol. 7, 2956-2967.

Hardison, R. C., Butler, E. T., III., Lacy, E. & Maniatis, T. (1979). The structure and transcription of four linked rabbit β-like globin genes. Cell 18, 1285-1297.

Harrison, P. R. (1982a). Stem cell regulation in erythropoiesis. Nature 295, 454-455.

Harrison, P. R. (1982b). Regulation of differentiation in retrovirus-induced murine erythroleukaemias: a personal view from the fridge. Cancer Surveys 1, 231-277.

Harrison, P. R. (1988). Molecular mechanisms involved in the regulation of gene expression during cell differentiation and development. In Colony Stimulating Factors: Molecular and Cellular Biology, (Eds. Dexter, T. M., Garland, J. & Testa, N.). Marcel Dekker, New York.

Heath, D. S., Axelrad, A. A., Mcleod, D. L. & Shreeve, M. M. (1976). Separation of the erythropoietin-responsive progenitors BFU-E and CFU-E in mouse bone marrow by unit gravity sedimentation. Blood 47, 777-792.

Hermonat, P. L., Spalholz, B. A. & Howley, P. M. (1988). The bovine papillomavirus P_{2443} promoter is E2 trans-responsive: evidence for E2 autoregulation. EMBO J. 7, 2815-2822.

Herr, W. & Clarke, J. (1986). The SV40 enhancer is composed of multiple functional elements that can compensate for one another. Cell 45, 461-470.

Hesse, J. E., Nickol, J. M., Lieber, M. R. & Felsenfeld, G. (1986). Regulated gene expression in transfected primary chicken erythrocytes. Proc. Nat. Acad. Sci., USA 83, 4312-4316.

Hill, D. E., Hope, I. A., Macke, J. P. & Struhl, K. (1986). Saturation mutagenesis of the yeast his3 regulatory site: requirements for transcriptional induction and for binding by GCN4 activator protein. Science 234, 451-457.

Hogaboom, G. K., Cook, M., Newton, J. F., Vamchio, A., Shorr, R. G. L., Sarau, H. M. & Crooke, S. T. (1986). Purification, characterization, and structural properties of a single protein from Rat Basophilic Leukaemia (RBL-1) cells possessing 5-lipoxygenase and leukotriene A_4 synthetase activities. Molecular Pharmacology 30, 510-519.

Hohne, H. & Andree, H. (1980). Acta Biol. Med. Ger. 39, 1241-1242.

Hope, I. A. & Struhl, K. (1986). Functional dissection of a eukaryotic transcriptional activator protein, GCN4 of yeast. Cell 46, 885-894.

Hope, I. A. & Struhl, K. (1987). GCN4, a eukaryotic transcriptional activator protein, binds as a dimer to target DNA. EMBO J. 6, 2781-2784.

Hunt, T. (1985). False starts in translational control of gene expression. Nature 316, 580-581.

Hurst, H. C. & Jones, N. C. (1987). Identification of factors that interact with the E1a-inducible adenovirus E3 promoter. Genes & Development 1, 1132-1146.

Iscove, N. N., Guilbert, L. J. & Weyman, C. (1980). Complete replacement of serum in primary cultures of erythropoietin-dependent red cell precursors (CFU-E) by albumin, transferrin, iron, unsaturated fatty acid, lecithin and cholestrol. Exp. Cell. Res. 126, 121-126.

Jacobs, K., Shoemaker, C., Rudersdorf, R., Neill, S. D., Kaufman, R. J., Mulson, A., Seehra, J., Jones, S. S., Hewick, R., Fritsch, E. F., Kawakita, T., Shimizu, T. & Miyake, T. (1985). Isolation and characterisation of genomic and cDNA clones of human erythropoietin. Nature 313, 806-810.

Jaenisch, R. (1988). Transgenic animals. Science 240, 1468-1474.

Jones, K. A., Kadonaga, J. T., Rosenfeld, P. J., Kelly, T. J. & Tjian, R. (1987). A cellular DNA-binding protein that activates eukaryotic transcription and DNA replication. Cell 48, 79-89.

Jones, K. A., Yamamoto, K. R. & Tjian, R. (1985). Two distinct transcription factors bind to the HSV thymidine kinase promoter in vitro. Cell 42, 559-572.

Jordan, G. (1987). Transcription: at the heart of the nucleolus. Nature 329, 489.

Kadonaga, J. T., Carner, K. R., Masiaz, F. R. & Tjian, R. (1987). Isolation of a cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain. Cell 51, 1079-1090.

Kadonaga, J. T., Jones, K. A. & Tjian, R. (1986). Promoter-specific activation of RNA polymerase II transcription by Sp1. Trends in Biochemical Sciences 11, 20-23.

Kadonaga, J. T. & Tjian, R. (1986). Affinity purification of sequence-specific DNA binding proteins. Proc. Nat. Acad. Sci., USA 83, 5889-5893.

Kakidani, H. & Ptashne, M. (1988). GAL4 activates gene expression in mammalian cells. Cell 52, 161-167.

Kelly, J. H. & Darlington, G. J. (1985). Hybrid genes: molecular approaches to tissue-specific gene regulation. Ann. Rev. Genetics 19, 273-296.

Kemper, B., Jackson, P. D. & Felsenfeld, G. (1987). Protein-binding sites within the 5' DNase I-hypersensitive region of the chicken α^D-globin gene. Mol. Cell. Biol. 7, 2059-2069.

Killary, A. M. & Fournier, R. E. K. (1984). A genetic analysis of extinction: trans-dominant loci regulate expression of liver-specific traits in hepatoma hybrid cells. Cell 38, 523-534.

Klug, A. & Rhodes, D. (1987). "Zinc fingers": a novel protein motif for nucleic acid recognition. Trends in Biochemical Sciences 12, 464-469.

Kollias, G., Hurst, J., de Boer, E. & Grosveld, F. (1987). The human β-globin gene contains a downstream developmental specific enhancer. Nucleic Acids Res. 15, 5739-5747.

Konkel, D. A., Maizel, J. V. & Leder, P. (1979). The evolution and sequence comparison of two recently diverged mouse chromosomal β-globin genes. Cell 18, 865-873.

Kopito, R. R., Andersson, M. A. & Lodish, H. F. (1987). Multiple tissue-specific sites of transcriptional initiation of the mouse anion antiport gene in erythroid and renal cells. Proc. Nat. Acad. Sci., USA 84, 7149-7153.

Kozak, M. (1986). Point mutations define a sequence flanking the AUG codon that modulates translation by eukaryotic ribosomes. Cell 44, 283-292.

Krieg, P., Amtmann, E. & Sauer, G. (1983). The simultaneous extraction of high-molecular weight DNA and of RNA from solid tumors. Anal. Biochem. 134, 288-294.

Krimpenfort, P., de Jong, R., Uematsu, Y., Dembis, Z., Ryser, S., von Boemher, H., Steinmetz, M. & Berni, A. (1988). Transcription of T cell receptor β-chain genes is controlled by a downstream regulatory element. EMBO J. 7, 745-750.

Lacy, E. & Maniatis, T. (1980). The nucleotide sequence of a rabbit β-globin pseudogene. Cell 21, 545-553.

Landschulz, W. H., Johnson, P. F. & McKnight, S. L. (1988). The Leucine Zipper: A hypothetical structure common to a new class of DNA binding proteins. Science 240, 1759-1764.

Laughon, L. & Scott, M. P. (1984). Sequence of a Drosophila segmentation gene: protein structure homology with DNA-binding proteins. Nature 310, 25-31.

Lawn, R. M., Efstratiadis, A., O'Connel, C. & Maniatis, T. (1980). The nucleotide sequence of the human β-globin gene. Cell 21, 647-651.

Lee, W., Haslinger, A., Karin, M. & Tjian, R. (1987a). Activation of transcription by two factors that bind promoter and enhancer sequences of the human metallothionein gene and SV40. Nature 325, 368.

Lee, W., Mitchell, P. & Tjian, R. (1987b). Purified transcription factor AP-1 interacts with TPA-inducible enhancer elements. Cell 49, 741-752.

Leff, S. E., Evans, R. M. & Rosenfeld, M. G. (1987). Splice commitment dictates neuron-specific alternative RNA processing in calcitonin/CGRP gene expression. Cell 48, 517-524.

Lloyd, J., Brownson, C., Tweedie, S., Charlton, J. & Edwards, Y. H. (1987). Human muscle carbonic anhydrase: gene structure and DNA methylation patterns in fetal and adult tissues. Genes & Development 1, 594-602.

Ma, J. & Ptashne, M. (1987a). The carboxy-terminal 30 amino acids of GAL4 are recognised by GAL80. Cell 50, 137-142.

Ma, J. & Ptashne, M. (1987b). A new class of yeast transcriptional activators. Cell 51, 113-119.

Magram, J., Chada, K. & Constantini, F. (1985). Developmental regulation of a cloned adult β-globin gene in transgenic mice. Nature 315, 338-340.

Mandel, M. & Higa, A. J. (1970). Calcium dependant bacteriophage DNA interaction. J. Mol. Biol 53, 159-162.

Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982). Laboratory Manual of Molecular Cloning. Cold Spring Harbor Publications, New York.

Maniatis, T., Goodbourn, S. & Fischer, J. A. (1987). Regulation of inducible and tissue-specific gene expression. Science 236, 1237-1245.

Manley, J. L., Fire, A., Caro, A., Sharp, P. A. & Gefter, M. L. (1980). DNA-dependent transcription of adenovirus genes in a soluble whole-cell extract. Proc. Nat. Acad. Sci., USA 77, 3855-3859.

Mantovani, R., Malgaretti, N., Nicolis, S., Giglioni, B., Comi, P., Capellini, N., Bertero, M. T. & Caligani-Cappio, F. & Ottolenghi,S. (1988a). An erythroid-specific nuclear factor binding to the proximal CACCC box of the β-globin gene promoter. Nucleic Acids Res. 16, 4299-4313.

Mantovani, R., Malgaretti, N., Nicolis, S., Ronchi, A., Giglioni, B. & Ottolenghi, S. (1988b). The effects of HPFH mutations in the human γ-globin promoter on binding of ubiquitous and erythroid specific factors. Nucleic Acids Res. 16, 7783-7797.

Marx, J. L. (1988). Multiplying genes by leaps and bounds. Science 240, 1408-1410.

Masters, B. S., Stohl, L. L. & Clayton, D. A. (1987). Yeast mitochondrial RNA Polymerase is homologous to those encoded by bacterophages T3 and T7. Cell 51, 89-99.

Matsumoto, T., Funk, C. D., Radmar, O., Hoog, J-O., Jornvall, H. & Samuelsson, B. (1988). Molecular cloning and amino acid sequence of human 5-lipoxygenase. Proc. Nat. Acad. Sci., USA 85, 26-30. (Correction ibid 85, 3406).

Maxam, A. M. & Gilbert, W. (1977). A new method for sequencing DNA. Proc. Nat. Acad. Sci., USA 74, 560-564.

McDonald, J. D., Lin, F-K. & Goldwasser, E. (1986). Cloning, sequencing and evolutionary analysis of the mouse erythropoietin gene. Mol. Cell. Biol. 6, 842-848.

McGinnis, W., Shermoen, A. W., Heemsberk, J. & Beckendorf, S. K. (1983). DNA sequence changes in an upstream DNase I-hypersensitive region are correlated with reduced gene expression. Proc. Nat. Acad. Sci., USA 80, 1063-1067.

McKnight, G. S. & Palmiter, R. D. (1979). Transcriptional regulation of the ovalbumin and conalbumin genes by steroid hormones in chick oviduct. J. Biol. Chem. 254, 9050-9058.

McKnight, S. & Tjian, R. (1986). Transcriptional selectivity of viral genes in mammalian cells. Cell 46, 795-805.

Melton, D. A. (1985). Injected anti-sense RNAs specifically block messenger RNA translation in vivo Proc. Nat. Acad. Sci., USA 82, 144-148.

Merlie, J. P., Fogan, D., Mudd, J. & Needleman, P. (1988). Isolation and characterisation of the complementary DNA for sheep seminal vesicle prostaglandin endoperoxide synthase (cyclooxygenase). J. Biol. Chem. 263, 3550-3553.

Messing, J. & Vieira, J. (1982). A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. Gene 19, 269-276.

Metzger, D., White, J. H. & Chambon, P. (1988). The human oestrogen receptor functions in yeast. Nature 334, 31-36.

Miller, A. M., MacKay, V. L. & Nasmyth, K. A. (1985). Identification and comparison of two sequence elements that confer cell-type specific transcription in yeast. Nature 304, 598-603.

Miller, J. H. (1978). In: The Operon (Eds. Miller, J. H. & Reznikoff, W. S.). Cold Spring Harbor Labs., New York, p. 66.

Minty, A. J. & Gros, F. (1980). Coding potential of non-polyadenylated mRNA in mouse Friend cells. J. Mol. Biol. 139, 61-83.

Montminy, M. R. & Bilerzikjian (1987). Binding of a nuclear protein to the cyclic-AMP response element of the somatostatin gene. Nature 328, 175-178.

Mueller, P. P. & Hinnebusch, A. G. (1986). Multiple upstream AUG codons mediate translational control of GCN4. Cell 45, 201-207.

Murphy, S., Diliegro, C. & Melli, M. (1987). The in vitro transcription of the 7SK RNA gene by RNA Polymerase II is dependent only on the presence of an upstream promoter. Cell 51, 81-87.

Myers, R. M., Tilly, K. & Maniatis, T. (1986). Fine structure genetic analysis of a β-globin promoter. Science 232, 613.

Nandi, A. K., Roginski, R. S., Gregg, R. G., Smithies, O. & Skoultchi, A. I. (1988). Regulated expression of genes inserted at the human chromosomal β-globin locus by homologous recombination. Proc. Nat. Acad. Sci., USA 85, 3845-3849.

Nishioka, Y. & Leder, P. (1979). The complete sequence of a chromosomal mouse α-globin gene reveals elements conserved throughout vertebrate evolution. Cell 18, 875-882.

Norrander, J., Kempe, T. & Messing, J. (1983). Construction of improved M13 vectors using oligonucleotide-directed mutagenesis. Gene 26, 101-106.

Ondek, B., Gloss, L. & Herr, S. (1988). The SV40 enhancer contains two distinct levels of organization. Nature 333, 40-45.

Pabo, C. O. & Sauer, R. T. (1984). Protein-DNA recognition. Ann. Rev. Biochem. 53, 293.

Page, D. C., Mosher, R., Simpson, E. M., Fisher, E. M. C., Mardon, G., Pollack, J., McGillivray, B., de la Chapelle, A. & Brown, L., (1987). The sex-determining region of the human Y chromosome encodes a finger protein. Cell 51, 1091-1104.

Parker, C. S. & Topol, J. (1984). A Drosophila RNA Polymerase II transcription factor contains a promoter-region-specific DNA-binding activity. Cell 36, 357-369.

Parker, C. W. (1984). Lipoxygenases and leukotrienes. J. Allergy Clin. Immunol. 74, 343-348.

Parslow, T. G. & Granner, D. K. (1982). Chromatin changes accompanying immunoglobulin κ gene activation. A potential control region within the gene. Nature 299, 449-451.

Parslow, T. G., Blair, D. L., Murphy, W. J. & Granner, D. K. (1984). Structure of the 5, ends of immunoglobulin genes; a novel conserved sequence. Proc. Nat. Acad. Sci., USA 81, 2650-2654.

Pelham, H. (1985). Activation of heat shock genes in eukaryotes. Trends in Genetics 1, 31-35.

Pelletier, J. & Sonenberg, N. (1985). Insertion mutagenesis to increase secondary structure within the 5′ non-coding region of a eukaryotic mRNA reduces translational efficiency. Cell 40, 515-526.

Perlino, E., Cortese, R. & Ciliberto, G. (1987). The human α1-antitrypsin gene is transcribed from two different promoters in macrophages and hepatocytes. EMBO J 6, 2767-2771.

Petit, C., Levilliers, J., Ott, M. O. & Weiss, M. C. (1986). Tissue-specific expression of the rat albumin gene: genetic control of its extinction in microcell hybrids. Proc. Nat. Acad. Sci., USA 83, 2561-2565.

Pfeifer, K., Arcangioli, B. & Guarante, L. (1987a). Yeast HAP1 activator competes with the Factor RC2 for binding to the upstream activation site UAS1 of the CYC gene. Cell 49, 9-18.

Pfeifer, K., Presant, T. & Guarante, L. (1987b). Yeast HAP1 activator binds to two upstream activation sites of different sequence. Cell 49, 19-27.

Plumb, M. A., Nicolas, R. H., Wright, C. A. & Goodwin, G. H. (1985). Multiple sequence-specific DNA binding activities are eluted from chicken nuclei at low ionic strengths. Nucleic Acids Res. 13, 4047-4065.

Plumb, M. A., Lobanenkov, V. V., Nicolas, R. H., Wright, C. A., Zavou, S. & Goodwin, G. H. (1986). Characterisation of chicken erythroid nuclear proteins which bind to the nuclease hypersensitive regions upstream of the β^D- and β^A-globin genes. Nucleic Acids Res. 14, 7675-7693.

Plumb, M. A. & Goodwin, G. H. (1988). Detection of sequence-specific DNA-protein interactions by the DNA-footprinting technique. Chapter 12 in: Methods in Molecular Biology. Vol.4 - New Nucleic Acid Techniques (Ed. Walker, J.M.)

Preiss, A., Rosenberg, U. B., Kiemlin, A., Seifert, E. & Jackle, H. (1985). Molecular genetics of Kruppel, a gene required for segmentation of the Drosophila embryo. Nature 313, 27-32.

Ptashne, M. (1986). Gene regulation by proteins acting nearby and at a distance. Nature 322, 697-701.

Pyati, J., Kucherlapati, R. S. & Skoultchi, A. (1980). Activation of human β-globin genes from non-erythroid cells by fusion with murine erythroleukaemia cells. Proc. Nat. Acad. Sci., USA 73, 862-866.

Queen, C. & Baltimore, D. (1983). Immunoglobulin gene transcription is activated by downstream sequence elements. Cell 33, 741-748.

Raghow, R. (1987). Regulation of messenger RNA turnover in eukaryotes. Trends in Biochemical Sciences 12, 358-360.

Rapoport, S. M., Schewe, T., Wiesner, R., Halangk, W., Ludwig, P., Hohne, M., Tannert, C. & Hiebsch, C. (1979). The lipoxygenase of reticulocytes: Purification, characterisation and biological dynamics of the lipoxygenase; its identity with respiratory inhibitors of the reticulocyte. Eur. J. Biochem 96, 545-561.

Razin, A. & Riggs, A. D. (1980). DNA methylation and gene function. Science 210, 604-610.

Remmers, E. F., Yang, J-Q. & Marcu, K. B. (1986). A negative transcriptional control element located upstream of the murine c-myc gene. EMBO J. 5, 899-904.

Rogers, J., Early, P., Carter, C., Calame, K., Bond, M., Hood, L. & Wall, R. (1980). Two mRNAs with different 3' ends encode membrane-bound and secreted forms of immunoglobulin μ chain. Cell 20, 303-312.

Roop, D. R., Tsai, M-J. & O'Malley, B. W. (1980). Definition of the 5' and 3' ends of transcripts of the ovalbumin gene. Cell 19, 63-68.

Rosenberg, U. B., Preiss, A., Seifert, E., Jackle, H. & Krupple, D. C. (1985). Production of phenocopies by Kruppel antisense RNA injection into Drosophila embryos. Nature 313, 703-706.

Rosenberg, U.B., Schroder, C., Preiss, A., Kienlin, A., Cote, S., Riede, I. & Jackle, H. (1986). Structural homology of the product of the Drosophila Kruppel gene with Xenopus transcription factor IIIA. Nature 319, 336-339.

Rosenfeld, P. J. & Kelly, T. J. (1986). Purification of nuclear factor 1 by DNA recognition site affinity chromatography. J. Biol. Chem. 261, 1398-1408.

Rossi, P. & de Crombrugghe, B. (1987). Identification of a cell-specific transcriptional enhancer in the first intron of the mouse a2 (type I) collagen gene. Proc. Nat. Acad. Sci., USA 84, 5590-5594.

Rouzer, C. A. & Samuelsson, B. (1985). On the nature of the 5-lipoxygenase reaction in human leukocytes: enzyme purification and requirement for multiple stimulatory factors. Proc. Nat. Acad. Sci., USA 82, 6040-6044.

Ruiz-i-Altaba, A., Perry-O'Keefe, H. & Melton, D. A. (1987). Xfin: an embryonic gene encoding a multifingered protein in Xenopus. EMBO J. 6, 3065-3070.

Rutherford, T. & Nienhuis, A. W. (1987). Human globin gene promoter sequences are sufficient for specific expression of a hybrid gene transfected into tissue culture cells. Mol. Cell. Biol. 7, 398-402.

Sakonju, S., Bogerhagen, D. F. & Brown, D. D. (1980). A control region in the center of the 5S RNA gene directs specific initiation of transcription: 1. The 5' border of the region. Cell 19, 13-25.

Samuelsson, B., Dahler, S. E., Lindgren, J. A., Rouzer, C. A. & Serhan, C. N. (1987). Leukotrienes and lipoxins: structures, biosynthesis, and biological effects. Science 237, 1171-1176.

Sanger, F., Nicklen, S. & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Nat. Acad. Sci., USA 74, 5463-5467.

Santoro, C., Mermod, N., Andrews, P. C. & Tjian, R. (1988). A family of human CCAAT-box-binding proteins active in transcription and DNA replication: cloning and expression of multiple cDNAs. Nature 334, 218-224.

Sassone-Corsi, P., Sisson, J. C. & Verma I. M. (1988). Transcriptional autoregulation of the proto-oncogene fos. Nature 334, 314-319.

Schewe, T., Albracht, S. P. J. & Ludwig, P. (1981). On the site of action of the inhibition of the mitochondrial respitatory chain by lipoxygenase. Biochim. Biophys. Acta 636, 210-217.

Schewe, T., Halangk, W., Hiebsch, C. & Rapoport, S. M. (1975). A lipoxygenase in rabbit reticulocytes which attacks phospholipids and intact mitochondria. FEBS Letts. 60, 149-152.

Schewe, T., Rapoport, S. M. & Kuhn, H. (1986). Enzymology and physiology of reticulocyte lipoxygenase: comparison with other lipoxygenases. Adv. Enzymol. and Related Areas of Mol. Biol. 58, 191-272.

Schule, R., Muller, M., Otsuka-Murakami, H. & Renkawitz, R. (1988). Cooperativity of the glucocorticoid receptor and the CACCC-box binding factor. Nature 332, 87-90.

Sen, R. & Baltimore, D. (1986a). Inducibility of κ immunoglobulin enhancer-binding protein NF-κB by a post translational mechanism. Cell 47, 921-928

Sen, R. & Baltimore, D. (1986b). Multiple nuclear factors interact with the immunoglobulin enhancer sequences. Cell 46, 705-716.

Senepathy, P. (1988). Possible evolution of splice-junction signals in eukaryotic genes from stop codons. Proc. Nat. Acad. Sci., USA 85, 1129-1133.

Serfling, E., Jasin, M. & Schaffner, W. (1985). Enhancers and eukaryotic gene transcription. Trends in Genetics 1, 224-230.

Shaw, G. & Kamen, R. (1986). A conserved AU sequence from the 3' untranslated region of GM-CSF mRNA mediates selective mRNA degradation. Cell 46, 659-667.

Shepherd, J. C. W., McGinnis, W., Carrasco, A. E., De Robertis, E. M. & Gehring, W. J. (1984). Fly and frog homeo domains show homologies with yeast mating type regulatory proteins. Nature 310, 70-71.

Shermoen, A. W. & Beckendorf, S. K. (1982). A complex of interacting DNase I-hypersensitive sites near the Drosophila glue protein gene, sgs4. Cell 29, 601-607.

Shibata, D., Steczko, J., Dixon, J. E., Hermodson, M., Yazdanparast, R. & Axelrod, B. (1987). Primary structure of soybean lipoxygenase-1. J. Biol. Chem. 262, 10080-10085.

Shimizu, T., Izumi, T., Seyama, Y., Tadokoro, K., Radmark, O. & Samuelsson, B. (1986). Characterisation of leukotriene A_4 synthase from murine mast cells: Evidence for its identity to arachidonate 5-dipoxygenase. Proc. Nat. Acad. Sci., USA 83, 4175-4179.

Shoemaker, C. B. & Mitsock, L. D. (1986). Murine erythropoietin gene: cloning, expression and human gene homology. Mol. Cell. Biol. 6, 849-858.

Shore, D. & Nasmyth, K. (1987). Purification and cloning of a DNA binding protein from yeast that binds to both silencer and activator elements. Cell 51, 721-732.

Sigler, P. B. (1988). Transcriptional activation: acid blobs and negative noodles. Nature 333, 210-212.

Smithies, O., Gregg, R. G., Boggs, S. S., Koralewski, M. A. & Kucherlapati, R. S. (1985). Insertion of DNA sequences into human chromosomal β-globin locus by homologous recombination. Nature 317, 230.

Sollner-Webb, B. (1988). Surprises in Polymerase III transcription. Cell 52, 153-154.

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

Spandidos, D. A. (1984). Transfer of human globin genes to human erythroleukaemia cells. Mol. Biol. Med. 2, 167-175.

Spandidos, D. A. & Paul, J. (1982). Transfer of human globin genes to erythroleukaemic mouse cells. EMBO J. 1, 15-20.

Spandidos, D. A. & Wilkie, N. M. (1984a). Expression of exogenous DNA in mammalian cells. Chapter 1 in: Transcription & Translation: a practical approach. (Eds. Hames, B. D. & Higgins, S. J.). IRL Press, Oxford, pp. 3-48.

Spandidos, D. A. & Wilkie, N. M. (1984b). Malignant transformation of early passage rodent cells by a single mutated human oncogene. Nature 310, 469-475.

Spradling, A. C. & Mahowald, A. P. (1980). Amplification of genes for chorion proteins during oogenesis in drosophila melanogaster. Proc. Nat. Acad. Sci., USA 77, 1096-1100.

Spivak, J. L. (1986). The mechanism of action of erythropoietin. Int. J. Cell Cloning 4, 139-166.

Stalder, J., Larsen, A., Engel, J. D., Dolan, M., Groudine, M. & Weintraub, H. (1980). Tissue-specific DNA cleavages in the globin chromatin domain introduced by DNase I. Cell 20, 451-460.

Stanton, L. W., Farhlander, P. D. & Marcu, K. B. (1984). Nucleotide sequence comparison of normal and translocated murine c-myc gene. Nature 310, 423-425.

Start, W. G., Ma, Y., Polacco, J. D., Hildebrand, D. F., Freyer, G. A. & Altshuler, M. (1986). Two soybean seed lipoxygenase nulls accumulate reduced levels of lipoxygenase transcripts. Plant Mol. Biol. 7, 11-23.

Steiner, C., Muller, M., Baniahmad, A. & Renkawitz, R. (1987). Lysozyme gene activity in chicken macrophages is controlled by positive and negative regulatory elements. Nucleic Acids Res. 15, 4163-4178.

Stephenson, J. A., Axelrad, A. A., McLeod, D. L. & Shreeve, M. (1971). Induction of clones of hemoglobin-synthesising cells by erythropoietin in vitro. Proc. Nat. Acad. Sci., USA 68, 1542-1546.

Struhl, K. (1987a). Promoters, activator proteins and the mechanism of transcriptional initiation in yeast. Cell 49, 295-297.

Struhl, K. (1987b). The DNA-binding domains of the jun oncoprotein and the yeast GCN4 transcriptional activator protein are functionally homologous. Cell 50, 841-846.

Theisen, M., Stief, A. & Sippel, A. E. (1986). The lysozyme enhancer: cell-specific activation of the chicken lysozyme gene by a far-upstream DNA element. EMBO J. 5, 719-724.

Theorell, H., Holman, R. T. & Akeson, A. (1947). Crystalline lipoxidase. Acta. Chem. Scand. 1, 571-576.

Thiele, B. J., Andree, H., Hohne, M. & Rapoport, S. M. (1982). Lipoxygenase mRNA in rabbit reticulocytes. Eur. J. Biochem. 129, 133-141.

Thiele, B. J., Belkner, J., Andree, H., Rapoport, T. A. & Rapoport, S. M. (1979). Synthesis of non-globin proteins in rabbit erythroid cells: Synthesis of a lipoxygenase in reticulocytes. Eur. J. Biochem. 96, 563-569.

Thiele, B. J., Fleming, J., Kasturi, K., O'Prey, J., Black, E., Chester, J., Rapoport, S. M. & Harrison, P. R. (1987). Cloning of a rabbit erythroid cell-specific lipoxygenase mRNA. Gene 57, 111-119.

Till, J. E. & McCulloch, E. A. (1961). A direct measurement of the radiation sensitivity of normal mouse bone marrow cells. Radiation Research 14, 213-222.

Till, J. E. & McCulloch, E. A. (1980). Hemopoietic stem cell differentiation. Biochim. Biophys. Acta 605, 431-459.

Townes, T. M., Lingrel, J. B., Chen, H. Y., Brinster, R. L. & Palmiter, R. D. (1985). Erythroid-specific expression of human β-globin genes in transgenic mice. EMBO J. 4, 1715-1723.

Trainor, C. D., Stamler, S. J. & Engel, J. D. (1987). Erythroid-specific transcription of the chicken histone H5 gene is directed by a 3' enhancer. Nature 328, 827-830.

Treisman, R. (1985). Transient accumulation of c-fos RNA following serum stimulation requires a conserved 5' element and c-fos 3' sequences. Cell 42, 889-902.

Treisman, R. & Maniatis, T. (1985). SV40 enhancer increases the number of RNA polymerase II molecules on linked DNA. Nature 315, 72-75.

Tsiftsogolou, A. S. & Robinson, S. H. (1985). Differentiation of leukemic cell lines: a review focusing on murine erythroleukaemia and human HL-60 cells. Int. J. Cell Cloning 3, 349-366.

Ueda, N., Kaneko, S., Yoshimoto, T. & Yamamoto, S. (1986). Purification of arachidonate 5-lipoxygenase from porcine leukocytes and its reactivity with hydroperoxyeicosatetraenoic acid. J. Biol. Chem. 261, 7982-7988.

van Ooyen, A., van den Berg, J., Mantei, N. & Weissmann, C. (1979). Comparison of total sequence of a cloned rabbit β-globin gene and its flanking regions with a homologous mouse sequence. Science 206, 337-344.

Vieira, J. & Messing, J. (1982). The pUC plasmids, an M13mp7 derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene 19, 259-268.

Vogelstein, B. & Gillespie, D. (1979). Preparative and analytical purification of DNA from agarose. Proc. Nat. Acad. Sci., USA 76, 615-819.

Vogt, P. K., Bos, T. J. & Doolittle, R. F. (1987). Homology between the DNA-binding domain of the GCN4 regulatory protein of yeast and the carboxy-terminal region of a protein encoded for by the oncogene jun. Proc. Nat. Acad. Sci., USA 84, 3316-3319.

Walker, M. D., Edlund, T., Bonlet, A. M. & Rutter, L. J. (1983). Cell-specific expression controlled by the 5'-flanking region of insulin and chymotrypsin genes. Nature 306, 557-561.

Wasylyk, B., Wasylyk, C., Augereau, P. & Chambon, P. (1983). The SV40 72bp repeat preferentially potentiates starting from proximal natural or substitue promoter elements. Cell 32, 503-574.

Weaver, R. F. & Weissmann, C. (1979). Mapping of RNA by a modification of the Berk-Sharp procedure: the 5' termini of 15S β-globin mRNA precursor and mature 10S β-globin mRNA have identical map co-ordinates. Nucleic Acids Res. 7, 1175-1193.

Weber, F. & Schaffner, W. (1985). SV40 enhancer increases RNA polymerase density within the linked gene. Nature 315, 75-77.

Webster, N., Jin, J. R., Green, S., Hollis, M. & Chambon, P. (1988). The yeast UAS_G is a transcriptional enhancer in human HeLa cells in the presence of the GAL4 trans-activator. Cell 52, 169-178.

Weintraub, H., Beug, H., Groudine, M. & Graf, T. (1982). Temperature-sensitive changes in the structure of globin chromatin in lines of red cell precursors transformed by ts-AEV. Cell 28, 931-940.

Weintraub, H. & Groudine, M. (1976). Chromosomal subunits in active genes have an altered conformation. Science 193, 848-856.

Weintraub, H., Larsen, A. & Groudine, M. (1981). α-globin-gene switching during the development of chicken embryos: expression and chromosome structure. Cell 24, 333-344.

Wiesner, R., Hausdorf, G., Anton, M. & Rapoport, S. M. (1983). Lipoxygenase from rabbit reticulocytes: Iron content, amino acid composition and C-terminal heterogeneity. Biomed. Biochim. Acta 42, 431-436.

Wiesner, R., Ludwig, P., Schewe, T. & Rapoport, S. M. (1981). Reversibility of the inhibition of cytochrome c oxidase by reticulocyte lipoxygenase. FEBS Letts. 123, 123-126.

Williams, J. G. & Mason, P. J. (1985). Hybridisation in analysis of RNA. Chapter 6 in: Nucleic acid hybridisation: a practical approach (Eds. Hames, B. D. & Higgins, S. J.). IRL Press, Oxford, pp. 152-160.

Willing, M. C., Nienhuis, A. W. & Anderson, W. F. (1979). Selective activation of a human β- but not a γ-globin gene in human fibroblast and mouse erythroleukemic cell hybrids. Nature 277, 534-538.

Wilson, M. C., Fraser, N. W. & Darnell, J. E. (1979). Mapping of RNA initiation sites by high doses of UV irradiation: evidence for three independent promoters within the left 11% of the Ad-2 genome. Virology 94, 175-184.

Wingender, E. (1988). Compilation of transcription regulating proteins. Nucleic Acids Res. 16, 1879-1902.

Wright, S., de Boer, E., Grosveld, F. G. & Flavell, R. A. (1983). Regulated expression of the human β-globin gene family in murine erythroleukaemia cells. Nature 305, 333-336.

Wright, S., Rosenthal, A., Flavell, R. & Grosveld, F. (1984). DNA sequences required for regulated expression of β-globin genes in murine erythroleukaemia cells. Cell 38, 265-273.

Wu, L-C., Morley, B. J. & Campbell, R. D. (1987). Cell-specific expression of the human complement protein Factor B gene: evidence for the role of two distinct 5'-flanking elements. Cell 48, 331-342.

Wu, C., Bingham, P. M., Livak, K. J., Holmgren, R. & Elgin, S. C. R. (1979). The chromatin structure of specific genes: 1. Evidence for higher order domains of defined DNA sequence. Cell 16, 797-806.

Yamamoto, K. K., Gonzalez, G. A., Biggs, W. H. & Montminy, M. R. (1988). Phosphorylation-induced binding and transcriptional efficacy of nuclear factor CREB. Nature 334, 494-498.

Yanisch-Perron, C., Vieira, J. & Messing, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33, 103-119.

Yenofsky, R. L., Fine, M. & Liu, C. (1988). Isolation and characterization of a soybean (Glycine max) lipoxygenase-3 gene. Molecular and General Genetics 211, 215-222.

Yisraeli, J., Adelstein, R. S., Melloul, D., Nudel, U., Yaffe, D. & Cedar, H. (1986). Muscle-specific activation of a methylated chimaeric actin gene. Cell 46, 409-416.

Zenke, M., Grundstrom, T., Mathes, H., Wintzerith, M., Schatz, C., Wildeman, A. & Chambon, P. (1986). Multiple sequence motifs are involved in SV40 enhancer function. EMBO J. 5, 387-397.

Zhu, J. D., Allan, M. & Paul, J. (1984). The chromatin structure of the human ε-globin gene: nuclease hypersensitive sites correlate with multiple initiation sites of transcription. Nucleic Acids Res. 12, 9191-9204.

Ziff, E. & Evans, R. M. (1978). Coincidence of the promoter and capped 5′ terminus of RNA from the adenovirus 2 major late transcription unit. Cell 15, 1463-1475.

Zinn, K. & Maniatis, T. (1986). Detection of factors that interact with the human β-interferon regulatory region *in vivo* by DNaseI footprinting. Cell 45, 611-618.