

**REGULATION OF EXPRESSION OF THE HUMAN
CARBONIC ANHYDRASE I GENE**

A Thesis Submitted in Partial Fulfilment of the
Requirements for Admission to the Degree of
Doctor of Philosophy of the University of London

by

Hugh Joseph Martin Brady
Department of Biochemistry
University College London

ProQuest Number: 10609052

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 10609052

Published by ProQuest LLC (2017). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

ABSTRACT

The HCAI gene is expressed in a developmental stage-specific manner and also exhibits tissue-specific expression patterns. This thesis describes the cloning and characterisation of the 5'-end of the HCAI gene which contains a large intron within its 5' untranslated region. S1 nuclease and primer extension analysis were used to define the transcription start site and the site of 3'-end maturation.

Bandshift assays have been used to show that there are at least six DNA sequences, based on the consensus [5'-T^T/_AATC^A/_T-3'] and flanking the HCAI gene, which bind the erythroid-specific factor, GF-1. The presence of GF-1 binding sites is shown to increase expression from a eukaryotic promoter in erythroid cells and not in non-erythroid cells.

A transient heterokaryon system was set up by fusing the erythroleukaemic cell lines MEL C88 (expressing MCAI) and K562 SAI (a human cell line with an embryonic / foetal phenotype, not expressing CAI). RNAase mapping of RNA from the fused cells showed activation of the human CAI gene. This indicates the developmental stage-specific expression of HCAI to be regulated by trans-acting factors.

Expression of HCAI mRNA in colon tissue was confirmed. Furthermore as in the case of mouse CAI, the HCAI colon mRNA is transcribed from a different promoter to erythroid HCAI mRNA.

ACKNOWLEDGEMENTS

I must begin by thanking Dr Mina Edwards for her great help in all aspects of the cell culture work described in this thesis. Without her expertise, persistence and encouragement much of the work would not have reached fruition.

I am grateful to the other members of our laboratory for their help. I especially thank Nick Lowe for his patient help and technical guidance throughout the course of my project. I also thank Jane Sowden, with whom I collaborated on the bandshift assays, Dr Vernon Spencer and Jon Barlow for their help and advice.

I am also indebted to other members of the Biochemistry Department for their contributions particularly, Dr Irving Johnston, Dr Michael Rosemeyer, Dr Chris Taylorson, Rita Barallon, Anna Straker, Eric Castle and Peter Laverack. I also gratefully acknowledge the advice of Drs Yvonne Edwards, Jon Frampton and A.M. Ali Imam.

I thank Professor David Linch for his generous help and advice during my time at U.C.L.

Finally, I owe a deep debt of gratitude to my supervisor Professor Peter Butterworth. Without his advice, encouragement and financial support it would have been impossible for me to complete my work on this thesis. Thanks for putting up with me.

I dedicate this thesis to my parents, Ellen and Diarmuid.

ABBREVIATIONS

A	Adenine
amp	ampicillin
AMV-RT	Avian myeloblastosis virus reverse transcriptase
APRT	adenophosphoribosyltransferase
ATP	adenine triphosphate
bp	DNA base pairs
BSA	bovine serum albumin
C	cytosine
CAI	carbonic anhydrase I
CAT	chloramphenicol acetyltransferase
cDNA	complementary DNA
CIP	calf intestinal phosphatase
CM	chloramphenicol
ddNTP	dideoxynucleotide triphosphate
dNTP	deoxynucleotide triphosphate
DDW	double distilled water
DEPC	diethylpyrocarbonate
DHFR	dihydrofolate reductase
DMEM	Dulbecco's minimal essential medium
DNA	deoxyribonucleic acid
DNAase	deoxyribonuclease
EBV	Epstein Barr virus
EDTA	ethylenediamine tetra-acetic acid
EGTA	ethylenebis(oxyethylenenitrilo)tetra-acetic acid
FCS	foetal calf serum
G	Guanine
HbA	Adult haemoglobin
HbF	Foetal haemoglobin
HBS	HEPES buffered saline
HCAI	Human carbonic anhydrase I
HLA	Human leukocyte antigen
HPRT	hypoxanthine phosphoribosyltransferase
HSV	Herpes Simplex Virus
IDMEM	Iscove's Modified Dulbecco's Medium
Ig	immunoglobulin
IPTG	isopropyl β -D-thiogalactopyranoside
kd	kilodaltons
Klenow	Klenow fragment of DNA polymerase I
MCAI	Mouse carbonic anhydrase I
μ Ci	microcuries
min	minutes
mRNA	messenger RNA
M.W.	molecular weight
NP-40	Nonidet-P 40
nt	nucleotide
OD _x	optical density at x nm
PAGE	polyacrylamide gel electrophoresis
PBGD	porphobilinogen deaminase
PBSa	phosphate buffered saline
PEG	polyethylene glycol
pfu	plaque forming units
phage	bacteriophage

PIPES piperazine-N,N'-bis[2-ethanesulphonic acid]
PMSF phenyl methyl sulphonyl fluoride
PNK polynucleotide kinase
poly(A+) polyadenylated
Pu purine
Py pyrimidine
RCAI Rabbit carbonic anhydrase I
RNA ribonucleic acid
RNAase ribonuclease
rpm revolutions per minute
RT room temperature
SDS sodium dodecyl sulphate
SnRNP small nuclear ribonucleoprotein
SSC saline sodium citrate
T Thymine
TBS Tween blocking solution
TCA trichloroacetic acid
T_d dissociation temperature
TEMED N,N,N',N'-tetramethyl ethylenediamine
tk thymidine kinase
tlc thin layer chromatography
TO tryptophan oxygenase
TPA 12-O-tetradecanoyl phorbol 13-acetate
tRNA transfer RNA
TTBS Tris-containing Tween blocking solution
u unit(s)
uv ultraviolet
wk week(s)
Wu Weiss units
X-gal 5-bromo-4-chloro-3-indoyl β-D-galactopyranoside

CONTENTS

<u>Section</u>	<u>page</u>
Abstract	2
Acknowledgements	3
Abbreviations	4
Contents	6
Figures	10
CHAPTER ONE - INTRODUCTION	14
1.1 General structure of eukaryotic promoters	18
1.2 Specific transcription factors	26
1.3 Mode of action of transcription factors	34
1.4 Erythropoiesis and globin genes	40
1.5 Heterokaryons	51
1.6 Carbonic anhydrases	53
CHAPTER TWO - MATERIALS AND METHODS	
2.1 Materials	59
2.2 Methods	60
2.3 Restriction enzyme digests	61
2.4 DNA modification reactions	61
2.5 Agarose gel electrophoresis and recovery of DNA fragments	62
2.6 Preparation and transformation of competent <u>E.Coli</u>	62
2.7 Large scale and rapid plasmid preparations	63
2.8 Purification of oligonucleotides	64
2.9 ³² P labelling of DNA	65

2.10 Library plating and screening	67
2.11 Large scale preparation of λ bacteriophage DNA	69
2.12 Hybridisation to DNA in gel membranes	71
2.13 Rapid restriction mapping of DNA cloned in λ phage vectors	71
2.14 DNA sequencing in M13 phage by the dideoxy chain-termination method	72
2.15 Isolation of total and mRNA	74
2.16 Northern analysis	76
2.17 Primer extension analysis	77
2.18 S1 nuclease mapping	78
2.19 RNA sequencing	80
2.20 RNAase protection analysis	81
2.21 Analysis of CAI protein by SDS PAGE, isoelectric focusing and Western blotting	83
2.22 Tissue culture and cell lines	85
2.23 Cell fusion	86
2.24 Transfection of tissue culture cells	87
2.25 CAT and β -galactosidase assays	88
2.26 Protein preparation for bandshift assays	90
2.27 Bandshift assays	92

CHAPTER THREE - ISOLATION AND CHARACTERISATION OF THE 5' END OF
THE HUMAN CARBONIC ANHYDRASE I GENE

3.1 Isolation and mapping of λ 104 and H24 clones	93
3.2 Nucleotide sequence analysis of λ 104 and H24 DNA	99
3.3 Isolation, mapping and nucleotide sequencing of λ 202-204 clones	109
3.4 Mapping of the transcription start site of HCAI	117

3.5 Nucleotide sequence analysis of the HCAI gene 5'- flanking region	123
3.6 Mapping of the 3'-end of the HCAI transcription unit and nucleotide sequence analysis	126

CHAPTER FOUR - INVESTIGATION OF THE BINDING OF AN ERYTHROID-
SPECIFIC FACTOR TO REGIONS FLANKING THE HCAI
GENE AND ITS ROLE AS A POSSIBLE TRANSCRIPTION
FACTOR

4.1 Origin and preparation of the DNA probes used in bandshift assays	133
4.2 Bandshift assays using nuclear protein extracts	137
4.3 Bandshift assays using whole cell extracts	144
4.4 Summary of the results of the bandshift assays	150
4.5 The effect of TPA on HCAI expression in HEL cells	151
4.6 A transfection system to assay transcriptional activity	156
4.7 HCAI constructs to study transcriptional activity	158
4.8 Transient expression studies of pHCAICAT T/R & R/T in MEL and HeLa cells	163

CHAPTER FIVE - STUDIES ON THE TRANS-ACTIVATION OF ERYTHROID GENE
EXPRESSION IN TRANSIENT HETEROKARYONS

5.1 Assay of cell lines for absence or presence of CAI and confirmation of stage-specific activation of HCAI mRNA	170
5.2 Transient heterokaryon formation	177
5.3 Analysis of mouse ϵ -globin gene activation	178
5.4 Construction of a HCAI-specific probe for analysis of cell fusion RNA	183

5.5 Analysis of HCAI gene activation	186
--------------------------------------	-----

CHAPTER SIX - TRANSCRIPT ANALYSIS OF RABBIT CAI AND FURTHER
TRANSCRIPT ANALYSIS OF HCAI, REVEALING A
SECOND, NON-ERYTHROID PROMOTER

6.1 Completing the nucleotide sequence of the protein- coding region of the rabbit CAI cDNA	191
6.2 The RCAI cDNA is homologous to a novel transcript in kidney cells	197
6.3 Human colon expresses HCAI: Northern and primer extension analysis	201
6.4 HCAI and MCAI: nucleotide sequence comparison of colon promoter region	209

CHAPTER SEVEN - DISCUSSION

7.1 Structure of the HCAI gene and transcription unit(s)	213
7.2 Several GF-1 binding sites flank the HCAI gene erythroid-specific transcription unit	217
7.3 What is the role of GF-1?	227
7.4 Further remarks on transcription of the HCAI gene	234
7.5 <u>Trans</u> -acting factors confer stage-specificity on erythroid HCAI gene expression	237
7.6 Further work	242

REFERENCES	244
------------	-----

FIGURES

<u>Fig.</u>	<u>page</u>
1.1 The relative transcription levels of mutant β -globin promoters	19
1.2 A <u>Trans</u> -acting factors binding to the TATA box region B Model for the assembly of TFII factors preceding transcription	21
1.3 Comparison of the CACCC box regions of β -globin and β -like globin genes	32
1.4 Developmental changes in the site of erythropoiesis and of globin chain synthesis	39
1.5 The different progenitors present in the haemopoietic lineages	41
1.6 The appearance of precursors during red cell maturation	43
1.7 Organisation of the human globin genes and haemoglobins produced in each stage of human development	45
1.8 Representation of the reprogramming of globin gene expression in transient heterokaryons	52
1.9 Active site residues of animal carbonic anhydrases	55
3.1 Restriction enzyme digests of H24 and λ 104 DNA and autoradiographs of these digests probed with ^{32}P labelled oligo #3	94
3.2 A "cos" mapping gel of λ 104 DNA B A map of the restriction enzyme sites in the cloned inserts of λ 104, H24 and λ HGCAI 5.1 DNA	97
3.3 A Sequencing strategy for the first exon of the HCAI protein coding region in λ 104 and H24 DNA	
3.3 B The nucleotide sequence obtained 5' and 3' to the	

first exon of the protein coding region of λ 104 and H24	100
3.3 C HCAI cDNA sequence showing the origin of the oligonucleotides used as primers	101
3.4 A Northern analysis of erythroid RNAs using ^{32}P labelled oligo #5	
3.4 B Tertiary screen of HCAI genomic clones	104
3.5 Restriction enzyme digests of λ 201-204 and λ 104 DNA and autoradiographs of these digests probed with ^{32}P labelled oligo #5	106
3.6 A "cos" mapping gel of λ 200 series clones 202 & 203	
3.6 B Restriction enzyme sites in the cloned inserts of λ 201-204 DNA	110
3.7 A Probing of pBSc204SH restriction digests with ^{32}P labelled oligo #7	
3.7 B Restriction map of the <u>SstI</u> / <u>HindIII</u> insert in pBSc204SH	112
3.8 A Sequencing strategy for the region containing exon 1a of the HCAI gene	
3.8 B An example of a typical DNA sequencing gel	114
3.9 S1 mapping of the 5'-end of the HCAI mRNA	118
3.10 Primer extension analysis of the 5'-end of HCAI mRNA	121
3.11 5'-region flanking the transcription HCAI start site	124
3.12 S1 mapping of the 3'-end of the HCAI mRNA	128
3.13 A The region flanking the 3'-end of the HCAI gene	
3.13 B The human carbonic anhydrase I gene	130
4.1 DNA probes used in bandshift assays	134
4.2 Bandshift assays using nuclear protein extracts	138
4.3 Bandshift assays using whole cell extracts	145

4.4 Photomicrographs and Northern analysis of untreated and TPA-treated HEL cells	152
4.5 HCAI promoter-CAT gene constructs	159
4.6 Results from CAT assays of HCAI promoter constructs transfected into HeLa cells and F412B2 MEL cells	164
5.1 Western blot analyses of MCAI protein in MEL C88 cells	171
5.2 A Northern analysis to detect HCAI expression in various cell lines	
B Northern analysis of the developmental expression of HCAI and globin mRNA	
C Schematic representation of cell fusion experiments	174
5.3 Detection of mouse ϵ -globin transcripts in transient heterokaryons	179
5.4 Detection of HCAI transcripts using RNAase mapping	184
5.5 Detection of HCAI transcripts in transient heterokaryons	187
6.1 Primer extension analysis of RCAI mRNA	192
6.2 RNA sequencing gel	195
6.3 A Comparison of the nucleotide sequence and protein sequence over the first 25 amino acids of RCAI with MCAI and HCAI	
6.3 B Northern analysis for CAI homologous transcripts in kidney cell lines	198
6.4 Northern analysis on colon RNA for HCAI mRNA	202
6.5 Primer extension analysis of HCAI mRNA in colon	204
6.6 A Sequence of the region flanking the postulated HCAI gene transcription start site in colon tissue	207
6.6 B Comparison of the HCAI and MCAI colon promoter region	

6.6 C Structure of the 5'-region of the HCAI gene	207
7.1 Structure of the HCAI gene	212
7.2 A Chicken α -globin gene locus	
7.2 B Chicken β -globin gene locus	218
7.3 A NF-E1 binding sites in the human β -globin gene 3' enhancer	
B Comparison of Eryf1 and GF-1 amino acid sequences	221
7.4 DNAaseI hypersensitive sites in the human β -globin gene locus	231

CHAPTER ONE

INTRODUCTION

One of the fundamental objectives of molecular biology is to understand the control of eukaryotic gene expression. Why can two cells with identical genetic information produce different protein profiles? How can this profile be changed during cell development.

The mechanisms of the flow of genetic information between DNA and mRNA and from mRNA to protein are largely understood. Research is now concentrated on elucidating the nature of the regulation of these processes.

Gene expression requires transcription, RNA processing and translation. There are many levels at which these events can be controlled. These involve the different orders of DNA, RNA and protein organisation and their interaction with different cellular factors.

Chromatin contains both open and condensed regions whose organisation is thought to be controlled by differential binding of histones, or protamines in the case of sperm chromatin. An open chromatin structure usually denotes a transcriptionally active region. This has been observed directly at "puffs" in the polytene chromosomes of Drosophila melanogaster (discussed in MacLean et al., 1983). Another indicator of an open chromatin structure is the presence of regions that are hypersensitive to DNAaseI digestion. These regions are typically orders of magnitude more nuclease sensitive than other regions in bulk chromatin and represent a

minor (<1%) but highly selective fraction of the genome (reviewed in Gross & Garrard, 1988).

A correlation between histone acetylation and transcriptional activity of chromatin has been shown (Allfrey, 1977). Such acetylation is a reversible modification to the histones, especially of the core histones, and converts lysine residues to acetyl-lysine. It remains to be seen whether the acetylation is causative or, found as a result of transcriptional activity.

Methylation patterns can also be indicative of areas of chromatin in which gene expression is taking place. Methylation of the DNA itself has been linked with potentially active regions of the genome. In particular, the presence of so called "HTF islands", which are stretches of DNA where CpG dinucleotides are abundant and not methylated (Bird, 1986). Evidence suggests that regions associated with HTF-islands become transcriptionally inactive if the islands are methylated.

Recombinant DNA techniques have allowed the regulation of expression to be examined at the level of individual genes. Although transcriptional termination and nuclear-cytoplasmic transport of RNA are not well defined, a detailed picture has emerged of the DNA or RNA sequences directly involved in transcription, polyadenylation and splicing.

Three types of RNA polymerase transcribe eukaryotic genes. RNA polymerase II is responsible for the transcription of all protein coding genes. Cis-acting promoter and enhancer sequences direct the efficient and accurate transcription of these genes by the RNA polymerase. They do so via interactions

with trans-acting factors (reviewed by Maniatis et al., 1987).

The development of in vitro systems has given great impetus to the understanding of RNA processing. The 3' terminus of mature mRNA is the result of cleaving a larger precursor back to a site which is then polyadenylated (reviewed by Birnstiel et al., 1985). This polyadenylation site is located 10-30 nt downstream of a highly conserved AAUAAA sequence (Proudfoot & Brownlee, 1976). Other more variable signal sequences have also been proposed (McLauchlan et al., 1985; Hart et al., 1985). Antibodies against snRNPs (small nuclear ribonucleoproteins) have been shown to inhibit polyadenylation in vitro (Moore & Sharp, 1985) and the RNA sequences precipitated by snRNP antibodies can be demonstrated to include the polyadenylation sites of mRNA (Hashimoto & Steitz, 1986). Therefore, one possibility is that snRNP-cleavage/polyadenylation signal interactions may regulate the expression of genes with multiple polyadenylation sites. For example, developmental control of immunoglobulin μ or δ chain expression (Mather et al., 1984) or the tissue-specific expression of calcitonin/calcitonin gene related peptide (Amara et al., 1984). Recent data suggests that in the latter case the alternative splice site selection is primarily regulated by cis-acting elements near the calcitonin-specific 3'-splice junction (Emeson et al., 1989).

Studies on the regulation of gene expression by splicing have highlighted the role of trans-acting factors. Interactions have been demonstrated between snRNPs and conserved sequences at the intron/exon junctions and lariat branchpoint. These interactions begin the formation of the

spliceosome, the multicomponent complex in which splicing occurs. Different splicing profiles for identical transcripts (e.g. the human fibronectin gene, Kornblihtt et al., 1984 and rat α -tropomyosin gene, Ruiz-Opaza et al., 1985) produced in different cell types, or in the same cell type at different stages of development, must reflect the effects of trans-acting factors. snRNPs are the most likely candidates to act as these factors though other possibilities are RNA binding proteins and anti-sense RNAs which could alter the conformation of the pre-mRNA and/or block splice site recognition.

As well as differential polyadenylation of splicing, a further event controlling gene expression may be a "process or discard" choice for nuclear RNA (Darnell, 1982). This remains to be demonstrated.

Differential mRNA stability or differential translation efficiency can also regulate gene expression. An example of the former is seen in the class of mRNAs that are associated with only very transient expression, such as interferon, c-myc, c-fos and various growth factor mRNAs. All of these mRNAs have AU-rich sequences in their 3' non-coding regions. For example, in the case of the 3' non-coding region of the granulocyte macrophage colony stimulating factor gene, the AU-rich sequence has been shown to confer instability on the mRNA (Shaw & Kamen, 1986). The best known example of translational control is that of mRNA "stored" in inactive cytoplasmic ribonucleoprotein particles during embryogenesis. For example, the mRNA of sea urchin eggs become at least 50-fold more active in protein synthesis after fertilisation than before

(Hofer & Darnell, 1981).

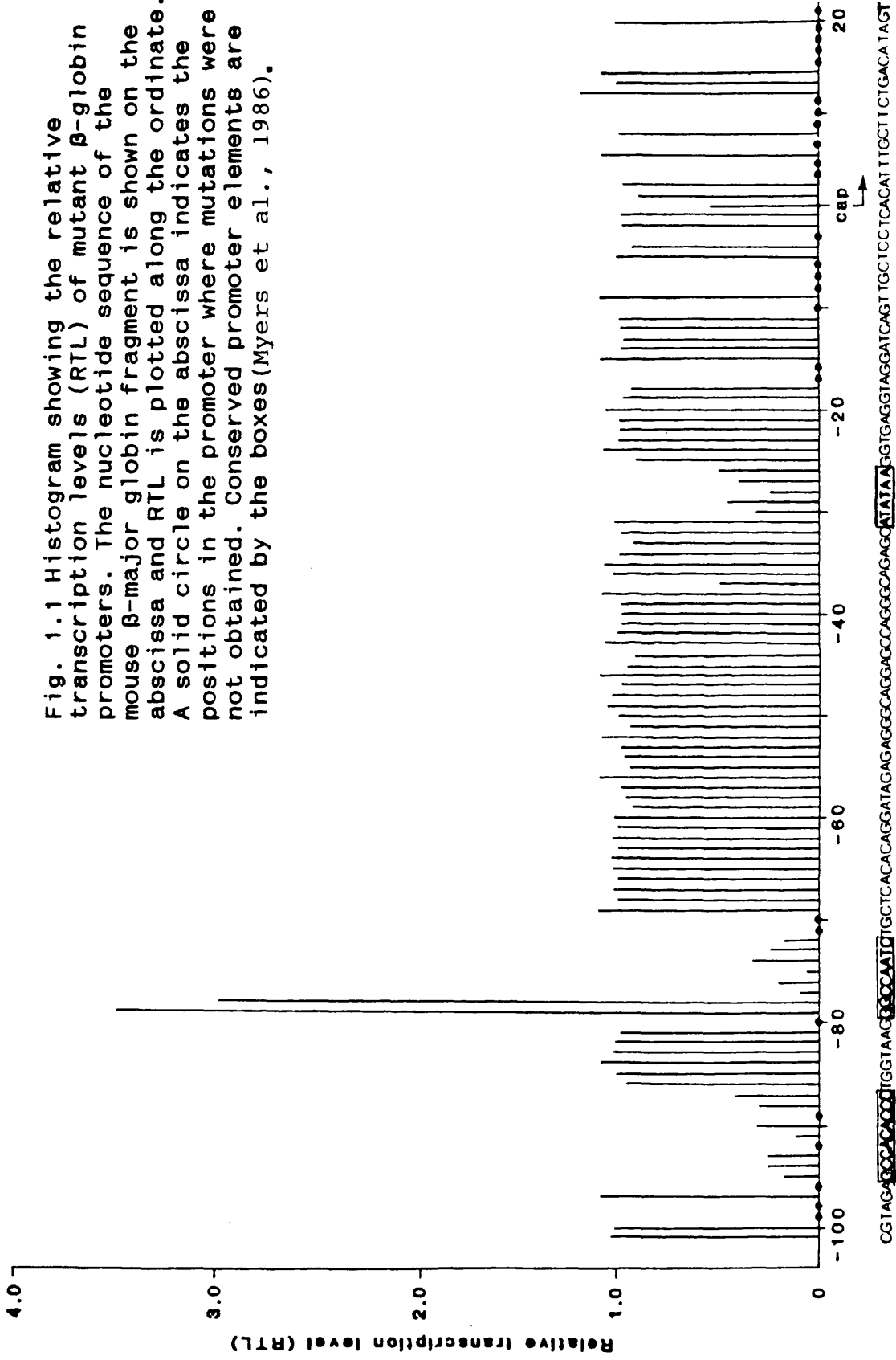
A further mechanism of control of gene expression is where RNA degradation is coupled to ribosome attachment and translation. The best characterised example of this in eukaryotes is the β -tubulin gene. The principal subunits of microtubules are α - and β -tubulin. An increase in the concentration of the tubulin subunits results in rapid and specific degradation of β -tubulin mRNA. This has been shown to be due to increased instability of β -tubulin mRNA in the cytoplasm (Yen et al., 1988). It appears that the regulatory event for the instability is a protein-protein interaction of tubulin subunits binding to the nascent, ribosome-bound tubulin polypeptide which stimulates degradation.

In conclusion, the complexity of the pathway from eukaryotic gene to protein provides numerous possibilities for control mechanisms. It is not unreasonable to expect that the first steps in the pathway are very important. This has been borne out by the advances made in understanding the control of transcription. It is also towards this end that the majority of the work in this thesis is directed. Therefore, the rest of the introduction will focus on the regulation of eukaryotic transcription, in particular of protein-coding cellular genes. This will be surveyed to include general, tissue-specific and developmental stage-specific transcription. Finally, the carbonic anhydrase gene family will be introduced and the aims of the thesis outlined.

1.1 General structure of eukaryotic promoters

Eukaryotic genes can be classified according to the type

Fig. 1.1 Histogram showing the relative transcription levels (RTL) of mutant β -globin promoters. The nucleotide sequence of the mouse β -major globin fragment is shown on the abscissa and RTL is plotted along the ordinate. A solid circle on the abscissa indicates the positions in the promoter where mutations were not obtained. Conserved promoter elements are indicated by the boxes (Myers et al., 1986).



of RNA polymerase by which they are transcribed. The three known types of RNA polymerase, I, II and III, transcribe ribosomal RNA genes, protein coding and small nuclear U RNA genes, and 5S and tRNA genes, respectively. This introduction only concerns the protein-coding genes of eukaryotes transcribed by RNA polymerase II.

Two DNA sequence elements are found in the regulation of transcription of the protein-coding genes: promoters and enhancers. Promoters are located immediately upstream from the start site of transcription and are typically about 100 bp in length (Dyana & Tjian, 1985). The promoter is required for accurate and efficient initiation of transcription, whereas enhancers increase the rate of transcription from promoters. Enhancers are distinct from promoters in that they can act on cis-linked promoters at great distances in an orientation-independent manner and can also function downstream from the transcription unit. However, the basic components of enhancers and promoters share many properties, and the mechanisms by which they regulate transcription are probably identical.

A common pattern of organisation has been revealed from the results of detailed molecular genetics analyses of different promoters. A typical promoter includes an AT-rich region designated the TATA box and one or more upstream DNA sequence elements. The TATA box functions primarily to ensure that transcripts are accurately initiated, whereas the upstream elements affect the rate of transcription. An example of how these sequences have been delineated is shown in Fig. 1.1. Single base substitutions within the promoter of the mouse β^{maj} globin gene only change the level of transcription

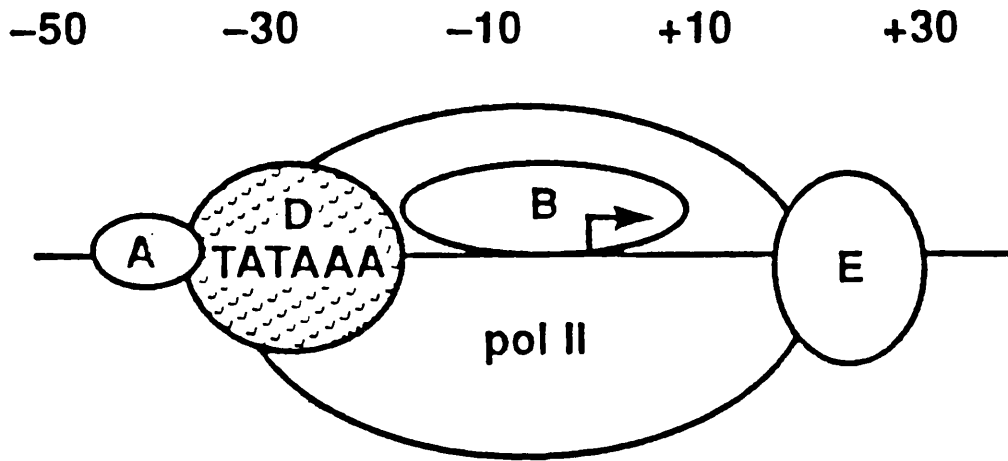


Fig. 1.2A Diagram of the trans-acting factors binding to the TATA box region.

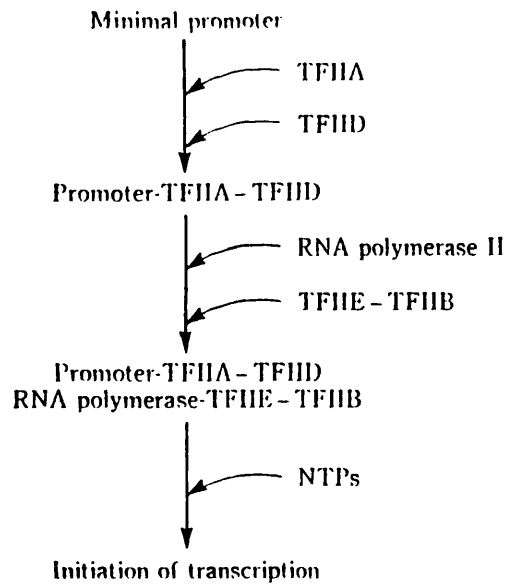


Fig. 1.2B A proposed model to describe the assembly of the TFII factors preceding transcription (from LaThangue & Rigby, 1988).

in HeLa cells when they are within the TATA box or in two upstream promoter elements (Myers et al., 1986).

The TATA box is a general feature of many promoters. It has a [5'-TATAA-3'] consensus sequence and is located 25-30 bp upstream from the transcription start site (Breathnach & Chambon, 1981). It has become clear that the TATA box, like all cis-acting DNA sequences, exerts its effect on transcription via interactions with trans-acting, or transcription factors.

Fig. 1.2A illustrates the trans-acting factors binding to the TATA box region. These factors are necessary for RNA polymerase II to accurately initiate transcription in vitro from a purified template (Matsui et al., 1980) and are referred to as TFIIA, B, D and E. The most well characterised of these factors is TFIID, which is the only one of the four to bind directly to DNA. TFIID recognises the TATA box and surrounding sequences. Part of its function may well be to interact with other transcription factors binding to upstream promoter elements. This has been shown to be the case for the major late transcription factor/upstream stimulating factor (MLTF/USF) of the human adenovirus major late promoter. The binding of the MLTF/USF is enhanced by the presence of TFIID, possibly because of protein-protein interactions (Garcia et al., 1987). A model has been proposed to describe the assembly of the TFII factors preceding transcription (Fig. 1.2B, LaThangue & Rigby, 1988). Initially, TFIIA influences the DNA in some unknown way to facilitate the binding of TFIID into a complex resistant to low levels of the detergent Sarkosyl, allowing promoter recognition by RNA polymerase II. This is

followed by TFIIE and TFIIB binding to the polymerase, generating a pre-initiation transcription complex capable of responding rapidly to the addition of nucleoside triphosphates.

However, it should be noted that higher eukaryotic promoters have been found which lack a TATA box but instead have a GC-rich region. Examples of this are the human hypoxanthine phosphoribosyltransferase (HPRT) gene (Melton et al., 1986) and the mouse dihydrofolate reductase (DHFR) gene (Sazer & Schmike, 1986). The HPRT gene has the sequence [5'-GGGGCGGAGC-3'] centred around position -33 and the DHFR gene 45-48 bp GC-rich sequence repeats upstream of the start site. Both of these types of GC-rich sequence contain binding sites for another promoter-specific transcription factor, Sp1. In the case of the DHFR gene, Sp1 has been directly shown to activate transcription by directly binding to the GC-rich repeats (Dyanan et al., 1986). The lack of a TATA box seems specifically to be associated with "housekeeping" genes that are expressed in all cells. A feature of these genes is heterogeneity of the 5'-end of the mRNA.

If not exactly a substitute for the TATA box, Sp1 binding sites are involved in promoter-specific transcription. Sp1 binding sites are an example of upstream promoter elements. These can be divided into a more general type like Sp1 sites and CCAAT boxes, which are found in many different promoters, and others, discussed later, which confer developmental or tissue specificity on transcription. Both types of element act regardless of their orientation with respect to the TATA box. It should now be emphasised that apart from the TATA box there

is no clear division between enhancer and promoter motifs because certain motifs with promoter activity in one situation can be important functional enhancer elements in another. Additionally, tandem copies of some promoter motifs can acquire properties of enhancers. This allows the hypothesis that it is the context of a given sequence motif that determines its functional characteristics rather than them being an intrinsic property of the binding protein.

The Sp1 binding site (Briggs et al., 1986) occurs in many viral and cellular promoters. It is based on [5'-GGGCGG-3'], though functional variants are found. The Sp1 site may act together with other promoter motifs such as the CCAAT box. For example, the Herpes Simplex Virus-1 thymidine kinase (HSV-1 tk) promoter has two Sp1 binding sites surrounding a CCAAT box and all three elements are required for maximal promoter activity (Jones et al., 1985). The mouse DHFR promoter (Dyanan et al., 1986) and human metallothionein II_A (Dyanan et al., 1985) promoter have three and one binding sites respectively. Sp1 binding sites occur in either orientation. Promoters that contain appropriately positioned sites can be stimulated 10-50 fold in in vitro transcription assays (Kadonaga et al., 1988). When more than one site is present, it is that nearest to the transcription start site that mediates the strongest transcriptional stimulation in vitro.

The CCAAT box as a promoter element was identified by deletion / transfection studies like that shown in Fig. 1.1 (Myers et al., 1986). This motif is located in a similar position to Sp1 sites, at -40 to -100, and is required for efficient transcription. The CCAAT box is functional in both

orientations, and can occur either alone or in the presence of other motifs, for example, Sp1 sites in the tk promoter (Jones et al., 1985) and heat shock regulatory elements in the human heat shock protein 70 gene promoter (Morgan et al., 1987).

It is clear that multiple factors recognise the CCAAT motif. One such factor, CTF, was originally defined by studying the transcription of the HSV-1 tk promoter in infected and uninfected cell extracts (Jones et al., 1985). Analysis of CTF binding sites in a variety of promoters indicated that the sequence matched a portion of the recognition sequence for NFI, a protein required for the initiation of adenovirus DNA replication in vitro (Jones et al., 1987). It would appear that NFI and CTF are related factors, able to participate in both DNA replication and transcription, as in the case of the SV40 large T antigen.

Another factor, CBP, binds to the CCAAT box in a variety of promoters. Rat liver CBP, for instance, generates a different DNAaseI-protected footprint to CTF (Johnson et al., 1986). For example, the single nucleotide change CCAAT to GCAAT enhances CBP binding but diminishes CTF binding (Graves et al., 1987). Another factor that binds to the CCAAT motif is NF-Y, which recognises [5'-ATTGG-3'], i.e. CCAAT in the opposite orientation, in the Y box of major histocompatibility complex class II genes (Dorn et al., 1987).

Along with the TATA box, the Sp1 binding site and CCAAT box motif are the major promoter specific motifs. At least one of them is found in the promoter region of every eukaryotic protein-coding gene. The transcription factors which bind to them are found in virtually all cell types. However, other

cis-acting motifs, found only in specific types of genes, bind factors that are not ubiquitous, and play a role in limiting the transcription of these genes to certain cell types.

1.2 Specific transcription factors

OCT-1 and -2

The octamer motif [5'-ATGCAAAT-3'] is a cis-acting sequence that occurs in a variety of promoters. Although originally defined in the chicken histone H2B promoter (Harvey et al., 1982), it is now clear that this element can be an important motif in enhancers, for example that of SV40 (Nomiyama et al., 1987). Multimers of this sequence possess enhancer activity and cell specificity (Ondek et al., 1987), indicating that the octamer motif can be either an enhancer or upstream promoter motif.

In the human histone H2B gene, the octamer motif is located in reverse orientation 40-50 bp upstream from the TATA box. The cell cycle-specific transcription of H2B genes can be reproduced in vitro in extracts prepared from synchronised HeLa cells (Sive et al., 1986) in which they are transcribed efficiently in S phase extracts but poorly in G1 extracts (Heintz & Roeder, 1984). The octamer motif is required for the transcriptional induction in S phase extracts, whereas disruption of this motif has no effect on basal level transcription in G1 extracts (Fletcher et al., 1987). A ubiquitous octamer binding factor, OTF-1, purified from HeLa cell extracts by oligonucleotide affinity chromatography, binds to this motif and stimulates transcription from the H2B promoter when added to G2 HeLa cell extracts (ibid.). It is

unclear how OTF-1 mediates cell cycle-specific induction, but this may involve regulation of OTF-1 synthesis or post-transcriptional modification.

The octamer sequence occurs in the opposite orientation in both heavy and light chain immunoglobulin promoters and has been shown convincingly to confer lymphoid-specific expression in vivo (Wirth et al., 1987). This suggests the existence of a lymphoid-specific octamer binding protein. Extracts derived from B cells contain cell-specific octamer binding proteins and transcribe the light chain promoter in vitro in a similar fashion to that in vivo (Staudt et al., 1986; Mizushima-Sugano & Roeder, 1986). This assay enabled purification of a lymphoid-specific octamer binding factor (OTF-2), shown to consist of three polypeptides (Scheidreit et al., 1987). OTF-2 activates transcription of the light chain promoter in non-lymphoid cell extracts by binding to the octamer motif where its footprint is indistinguishable from that caused by OTF-1 binding.

The octamer motif is also important for transcriptional activation by viral trans-activators; for example, the induction of HSV gene transcription by the virion trans-activator Vmw 65 involves the formation of a complex between the octamer binding factor and Vmw 65 (O'Hare & Goding, 1988).

One of the most important questions regarding the family of octamer binding proteins is what features of a promoter determine whether ubiquitous or cell-specific factors bind. It may be that the position of the octamer motif is the critical feature since in the H2B promoter it is located between -50 and -40, in contrast to -66 and -59 in the immunoglobulin

light chain promoter. This location may determine how other proteins, for example TFIID, interact with octamer-binding proteins. If only the ubiquitous factor can cooperate with surrounding proteins, and if cooperation is required for efficient DNA binding, then the distance between motifs would be paramount in determining which octamer factor bound.

AP-1

Another specific transcription factor is designated AP-1. It was originally defined through its selective binding to cis-acting sequences in the SV40 and human metallothionein II_A enhancers (Lee et al., 1987). The AP-1 consensus recognition site is [5'-^C/_GTGACT^C/_AA-3']. A group of polypeptides with a 47 kd major species, purified from HeLa cells, have the DNA binding properties of AP-1 and, in addition, stimulate transcription of SV40 and metallothionein II_A in vitro (ibid; Angel et al., 1987). Most viral and cellular genes that respond to AP-1, such as collagenase, stromelysin, metallothionein II_A and SV40, are also induced by treating cells with the phorbol ester tumour promoter, TPA. In addition, multiple synthetic copies of the consensus AP-1 binding site act as TPA-inducible enhancers in plasmid constructs following transfection into HeLa cells. Some activation of transcription by TPA occurs by post-translational mechanisms (Imbra & Karin, 1986) and, as TPA activates protein kinase C, it seems likely that AP-1 needs to be phosphorylated in order to increase its transcription activating capability.

The avian sarcoma virus, ASV 17, induces fibrosarcomas in

chickens. Its oncogene, as v-jun, shows a structural and functional homology with the yeast transcription factor GCN4 (Struhl, 1987). The conserved region is restricted to the C-terminus of v-jun, which has 44% homology with the C-terminus of GCN4, known to be the DNA binding domain. This suggested that the v-jun oncogene was derived from a cellular gene involved in transcriptional control. A possible candidate for this arose when the AP-1 consensus binding site was found to be very similar to that of GCN4, [5'-TGACTCA-3'], (Arndt & Fink, 1986). This relationship was confirmed by isolating and sequencing the c-jun proto-oncogene, which has greater than 80% homology with v-jun sequences. Expression of the c-jun cDNA in bacteria produced a protein with sequence-specific DNA binding properties identical to AP-1 (Bohmann et al., 1987), strongly supporting the idea that c-jun encodes AP-1. It is therefore a possibility that v-jun transforms cells by transcriptionally activating cellular genes that have AP-1 binding sites and that a subset, or all, of these may also be induced by TPA through the action of AP-1.

The protein encoded by v-jun binds to DNA with the same specificity as AP-1, the c-jun gene product, indicating that the mechanism by which the proto-oncogene is activated does not involve any major alteration in DNA binding specificity but more likely a change in the way in which the transcription factor interacts with other proteins (Bos et al., 1988).

PEA-1

That transcription factors can be regulated during differentiation is illustrated by studies on murine AP-1

related proteins. The polyoma virus enhancer contains an AP-1 motif, and mutational analysis has shown this domain to be required for transcription and DNA replication (Veldman et al., 1985). The murine equivalent of AP-1, known as PEA-1 (Piette & Yaniv, 1987), is regulated during the differentiation of embryonal carcinoma stem cells such that the binding activity is increased when F9 stem cells differentiate to parietal endoderm (Kryské et al., 1987). The absence of PEA-1 in the stem cells may account for some of the host range properties of polyoma virus, which is poorly expressed in embryonal carcinoma stem cells relative to differentiated murine cells (Katinka et al., 1981). The post-translational activation of AP-1 by tumour promoters in HeLa cells is consistent with a mechanism for activating PEA-1 during stem cell differentiation.

NF κ B

An example of a transcription factor that can cause a gene to be transcribed in a tissue-specific manner is NF κ B. The κ light chain immunoglobulin gene has an enhancer located in the J-C intron (Picard & Schaffner, 1984). This enhancer contains at least 3 sites that interact with specific DNA binding proteins, one of which, the κ B site, binds a B-cell specific factor called nuclear factor κ B (NF- κ B) (Sen & Baltimore, 1986). The DNA binding activity of this factor is not detectable in nuclear extracts of the pre-B cell line 70Z/3 but, can be activated by a post-translational mechanism after treatment of cells with bacterial lipopolysaccharide or phorbol esters (ibid.). The induction of NF κ B activity by these

treatments strongly correlates with the transcriptional activity of the κ gene (ibid.). Mutational analysis of the κ enhancer and the construction of functional enhancers using oligonucleotides representing NF κ B binding sites (Pierce et al., 1988) have provided further evidence that the binding of NF κ B to the κ B motif in the κ enhancer is essential for the transcriptional activity, inducibility, and even the developmental stage specificity of the κ enhancer. Although NF κ B activity is limited in cell type it is apparently present in various other cell types in an inactive form. Phorbol ester treatment of cells, where it is inactive, produces actively binding NF κ B which is translocated to the nucleus.

Myogenic factors

An even more dramatic example of a DNA binding factor, and possible transcription factor involved in conferring tissue specificity, is that of the myogenic determinants: MyoD (Davis et al., 1987); myd (Pinney et al., 1988); myogenin (Wright et al., 1989) and myf-5 (Braun et al., 1989). Studies with MyoD have shown that expression of this gene product is able to activate myogenesis in cell lines derived from mesodermal, ectodermal and endodermal tissue (Weintraub et al., 1989). This suggests that expression of a single determination factor can specify the muscle cell lineage alone or in combination with other commonly expressed factors.

MyoD, myogenin and myf-5 are expressed exclusively in skeletal muscle and share extensive sequence homology, including both a basic region and a region of sequence similarity with the myc family of proteins. The basic and myc

ADULT β -GLOBIN GENES

RABBIT β	(-111)	C A G A -	C C T	C A C C C T	G	22 NT
	(-97)	C A G A G	C C A	C A C C C T	G	7 NT
HUMAN β	(-112)	T A G A -	C C T	C A C C C T	G	22 NT
	(-98)	T G G A G	C C A	C A C C C T	A G	6 NT
MOUSE β^{MINOR}	(-114)	T G A A G	C C T	C A C C C T	G	23 NT
MOUSE β^{MAJOR}	(-99)	T A G A G	C C A	C A C C C T	G	7 NT
GOAT β^{A} AND β^{C}	(-112)	C A A G -	C C T	C A C C C T	G	22 NT
CHICKEN β	(-141)	T G A G -	C C C	C A C C C T	G	55 NT

FETAL AND EMBRYONIC β -LIKE GLOBIN GENES

HUMAN γ^{A} AND γ^{G}	(-151)	T A A A -	C T C	C A C C C	A T G	21/48 NT
GOAT γ	(-112)	C A A G -	C C T	C A C C C	T G	22 NT
HUMAN ϵ	(-118)	T G A - -	C T C	C A C C C	C T G	20 NT

β -GLOBIN CONSENSUS PyPuPuPu- C C $\begin{matrix} \text{A} \\ \text{T} \\ \text{C} \end{matrix}$ C A C C C T G

Fig. 1.3 Comparison of the CACCC box regions of β -globin and β -like globin genes (from Dierks et al., 1983).

homology regions of MyoD have been shown to be both necessary and sufficient to convert 10T½ fibroblasts into muscle cells (Tapscott et al., 1988). The nuclear localisation of MyoD suggested it might also be a DNA binding protein.

Muscle-specific regulatory regions have been defined for several different muscle differentiation marker genes. Some of the DNA binding factors that interact with these regulatory regions have been characterised by in vitro binding assays. A differentiated myocyte nuclear factor, MEF1, was found to bind to the muscle-specific enhancers of the muscle creatine kinase (MCK) gene as well as to the enhancer region of the myosin light chain 1/3 (MLC^{1/3}) gene (Buskin & Hauschka, 1989). MEF1 binding activity was not found in myoblasts nor in non-muscle cells.

Recent data has demonstrated that MEF1 is antigenically related to MyoD, and that a bacterially produced MyoD fusion protein is capable of sequence-specific interaction with two regions of the MCK enhancer that are required for enhancer activity. These results suggest that MyoD activates the muscle differentiation programme by directly binding to muscle-specific regulatory regions (Lassar et al., 1989).

CACCC box factor

Fig. 1.1 shows the cis-acting sequences in the mouse β^{maj} globin gene promoter (Myers et al., 1986). One of these elements is the CACCC box at about -90, containing the sequence [5'-CCACACCCT-3']. This is not a sequence which occurs in the promoter region of many genes but it is found very often in globin gene promoters, particularly β - and β -

like globin genes, as illustrated in Fig. 1.3 (taken from Dierks et al., 1983). It has been suggested that a trans-acting factor which binds to this sequence is erythroid-specific and therefore, might be thought to confer erythroid-specificity on globin gene transcription (Mantovani et al., 1988). However, this would now appear not to be so. The CACCC box binding factor is not confined to cells of the erythroid lineage and is ubiquitous (deBoer et al., 1988; Plumb et al., 1989; Schule et al., 1988a). The factor has also been shown to act as a transcription factor in a cooperative manner with a glucocorticoid receptor binding site upstream of the transcription start site of the steroid-inducible rat tryptophan oxygenase (TO) gene (Schule et al., 1988a). Mutations in the CACCC box of the rat TO gene, identical to those carried out on the mouse β^{maj} globin CACCC box (Myers et al., 1986), reduce the binding of the CACCC box factor and also reduce steroid induction. In comparing the sequences of other steroid-inducible genes (e.g. human metallothionein II_A, rat tyrosine aminotransferase, human growth hormone, chicken lysozyme and rabbit uteroglobin), homologies to the TO CACCC box [5'-GCTGGGTGTGGC-3'] have been found in the vicinity of the glucocorticoid receptor binding sites.

1.3 Mode of action of transcription factors

The significance of a cis-acting DNA sequence can vary greatly in different cell types because the DNA binding factor or factors that recognise it vary in abundance or in ability to function in different tissues (e.g. NF κ B, Sen & Baltimore, 1986). Overlapping or superimposed binding sites for multiple

factors can result in different positive and negative factors competing for sites. It has been shown that, in some cases, synergistic effects are dependent on strict spacing between adjacent cis-acting elements (Schule et al., 1988b). The interaction between the various factors that may bind to the cis-acting elements is obviously of great importance since it will hold the key to understanding how the factors affect transcription. It would seem likely that combinations of factors binding to specific cis-acting elements, arrayed in unique configurations, confer on each gene an individualised tissue-specific and/or stage-specific transcription programme.

The factors that bind to cis-acting DNA sequences can be detected in cell extracts and identified according to their sequence specificities in various in vitro DNA binding assays, such as DNAaseI protection and bandshift. These low abundance proteins can be purified from nuclear extracts by sequence specific DNA affinity chromatography in order to obtain material for biochemical experiments and for protein sequence analysis (e.g. Kadonaga & Tjian, 1986). Their binding properties have also been used as a means to screen expression libraries for cDNA clones encoding them (e.g. Singh et al., 1988). Structure-function studies of cloned mammalian DNA binding transcription factors have included the general strategy of dissecting protein structure by deletion analysis, in order to compare the biochemical activities of mutant versions of the factor. Most commonly, the DNA binding properties of wild type and mutant factors are assayed in vitro, and the transcriptional activities are assayed by expressing the factor transiently in tissue culture cells and

measuring the transcription from cotransfected reporter gene promoters containing binding sites for the factor. Such approaches have shown that different DNA binding transcription factors are composed of a surprising variety of usually separable DNA binding and transcriptional activation domains.

The DNA binding activity of several mammalian factors has been localised to relatively small subregions consisting of 60-100 amino acids. These studies have shown that a DNA binding domain is necessary but not sufficient for transcriptional activation. The first type of DNA binding structures to be identified were the zinc finger motifs in the RNA polymerase III transcription factor TFIIIA, which bind to an internal control region of the 5S RNA gene (Miller et al., 1985). Since then two types of zinc fingers have been found in DNA binding factors that participate in transcription mediated by RNA polymerase II. Sp1 has TFIIIA-like zinc fingers (Kadonaga et al., 1988). This type of zinc finger consists of about 30 amino acids with two cysteine and two histidine residues that stabilise the domain by tetrahedrally coordinating a Zn^{2+} ion. A region of about 12 amino acids between the invariant cysteine-histidine pairs is characterised by scattered basic residues and several conserved hydrophobic residues. A second class of zinc finger motifs is exemplified by the DNA binding domains of steroid hormone receptors. This motif uses two pairs of cysteines rather than the cysteine-histidine arrangement typical of Sp1 (Evans and Hollenberg, 1988).

A second type of DNA binding domain is the homeobox. This domain, which encompasses about 60 amino acids, was first

identified as a conserved protein segment in several regulators of Drosophila embryogenesis and subsequently in genes of vertebrates as well (reviewed in Wright et al., 1989). Biochemical and in vivo studies with cloned mammalian transcription factors showed that homeobox domain-containing proteins can bind and directly activate transcription of target genes. Homeobox domains are found in OTF-1 (Sturm et al., 1988), OTF-2 (Ko et al., 1988) and the pituitary-specific factor, Pit-1 (Ingraham et al., 1988). The presence of such a domain in OTF-1 shows that this structure is not restricted to regulators of development, as OTF-1 appears to be expressed in all mammalian cells.

A third type of DNA binding domain was first described for the mammalian enhancer binding protein C/EBP (Landschulz et al., 1988). A search for proteins with similarities to C/EBP identified several transcription factors that contain a two part region of primary sequence similarity. This consists of a highly conserved stretch of about 30 amino acids with a substantial net basic charge immediately followed by a region containing four leucine residues positioned at intervals of 7 amino acids. The latter part, designated the "leucine zipper", is required for dimerisation and for DNA binding. It is believed that dimerisation of proteins in this group, which includes mammalian C/EBP, Jun and Fos (an oncogene now known to be a transcription factor) is stabilised by hydrophobic interactions between closely apposed α -helical leucine repeat regions of the two subunits (Turner & Tjian, 1989). The basic region adjacent to the leucine motifs seems to be necessary for DNA binding but not for dimerisation, and both subunits

appear to contribute their basic regions to form the DNA binding domain. Other proteins, including myc and OTF-2, contain leucine repeat motifs without the conserved basic region found in the C/EBP-like factors. The leucine region in myc is required for tetramerisation and for transformation of primary cells (Dang et al., 1989). As mentioned above, OTF-2 has a homeobox domain DNA binding structure; the function of the leucine repeat region is not known (Ko et al., 1988).

The transcriptional activation functions of DNA binding factors depend on regions of 30-100 amino acids that are separate from the DNA binding domain. Factors can have more than one activation domain, and several seemingly unrelated structural motifs have been identified which confer these functions. The first activation regions to be identified in eukaryotic transcription were identified by studies of the yeast factors, GAL4 and GCN4 (reviewed in Ptashne, 1988). The activation domains of these factors consist of relatively short stretches of amino acids with apparently only two features in common: they are regions with significant negative charge and can form amphipathic α -helical structures. There also appears to be some correlation between activation and negatively charged α -helical regions in the AP-1/jun transcription factors (Bohmann et al., 1987). It has been proposed that these acidic activation domains may facilitate transcription initiation by interacting in a relatively non-specific manner with a general component of the initiation complex, such as TFIID or RNA polymerase II (Ptashne, 1988).

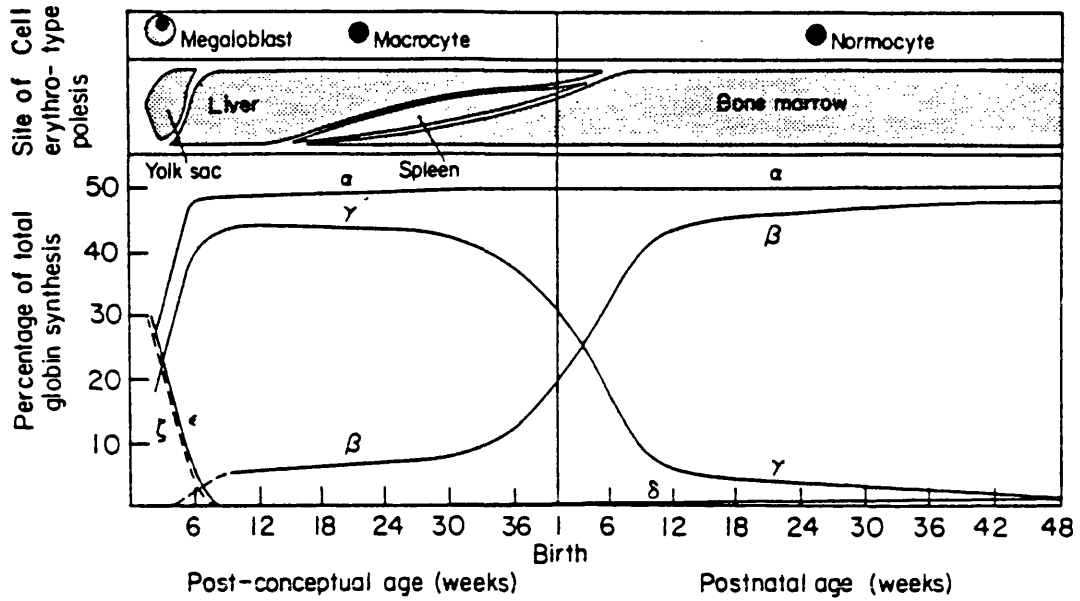


Fig. 1.4 Developmental changes in the site of erythropoiesis and of globin chain synthesis (from Weatherall & Clegg, 1981).

1.4 Erythropoiesis and globin genes

The work in this thesis deals essentially with the transcription of the human carbonic anhydrase I gene in erythroid tissue. Having detailed some of the elements of transcriptional control it is now necessary to outline the background to erythropoiesis and gene expression in erythroid tissue.

In humans, erythropoietic cells are first detected in the mesenchyme of the yolk sac. This remains the site of erythropoiesis for the first 12 weeks of embryonic life at which stage erythropoiesis is switched to the foetal liver (Ingram, 1972). At about 20 weeks of gestation, definitive haematopoiesis is observed in the bone marrow and this gradually replaces liver as the major site of blood cell production (Fig. 1.4).

All erythroid cells originate from a common haematopoietic stem cell. These stem cells proliferate and maintain a steady pool from which cells are then drawn into the differentiation pathways. The "primitive", pluripotent stem cell, termed CFU-s (colony forming units-spleen) can give rise to different haematopoietic cells (granulocytes, megakaryocytes or erythroid) depending on the microenvironment. It is now believed that even the lymphoid precursors are derived from the same original stem cell population (CFU-L-M), so that now a single type of pluripotent stem cell can give rise to all the cellular elements of the blood (Fig. 1.5).

The mechanism by which cells become committed to a particular differentiation pathway is not clear. The first committed erythroid progenitor is called the Erythropoietin Responsive Cell (ERC). This is not a uniform cell population

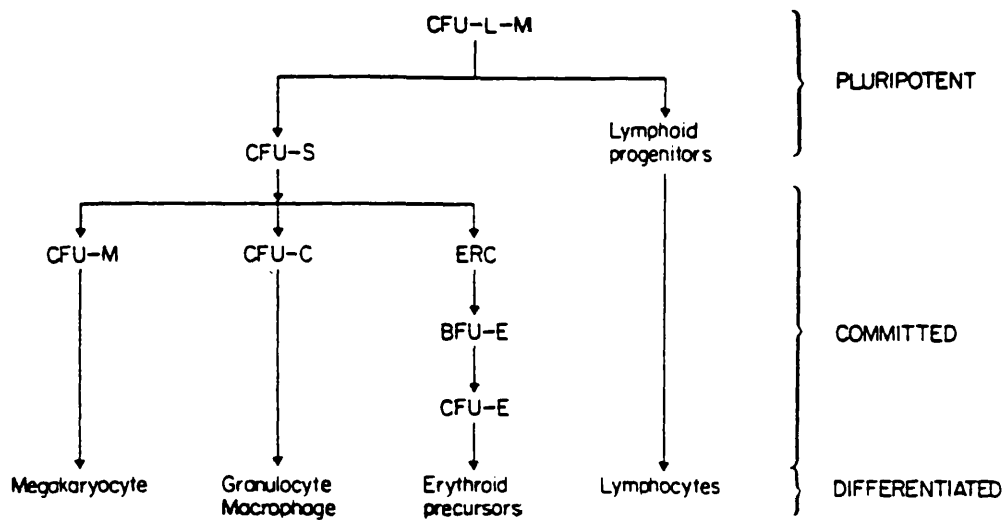


Fig. 1.5 Diagram of the different progenitors present in the haematopoietic lineages (from Weatherall & Clegg, 1981).

but probably consists of cells at different stages of differentiation which respond in varying extents to erythropoietin. Early fully committed erythroid progenitors have been identified from studies of erythroid cells in in vitro cultures (Axelrad et al., 1974). When foetal liver or bone marrow are incubated in vitro in the presence of erythropoietin, small colonies appear (CFU-E) containing 20-60 differentiated erythroid cells. Upon prolonged incubation and in the presence of higher levels of erythropoietin, much larger colonies (bursts) appear (about 10^4 cells, BFU-E). The CFU-E are thought to represent late precursors, probably just before the proerythroblast stage while the BFU-E are not as differentiated and are probably derived from more "primitive" progenitor cells with an increased proliferative capability (Dexter, 1979).

Soon after the stimulation of the committed erythroid progenitors, probably as a result of the action of erythropoietin, large proerythroblasts appear which have a limited proliferative potential and which in turn produce basophilic erythroblasts. These then give rise to polychromatic erythroblasts with a dense nucleus, at which point cell division stops. After this, there follows the maturation stage during which the precursors first undergo various morphological changes (nucleus extruded, disappearance of mitochondria, etc.) and are then released into the circulation to yield the end product, a mature erythrocyte (Fig. 1.6). The process is thought to occur in 7 days, the first four of which are mainly spent in the generation of differentiated precursors while the last three

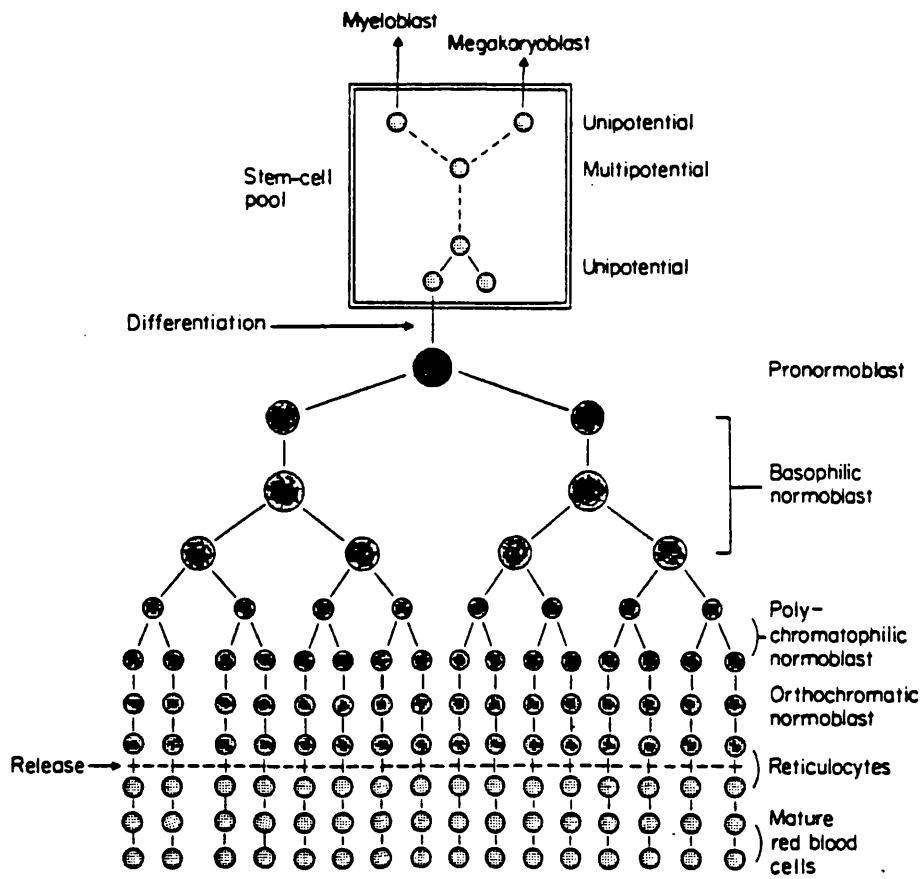


Fig. 1.6 Schematic representation of the different precursors that appear during red cell maturation (from Weatherall & Clegg, 1981).

are involved with the maturation process.

Erythropoietin plays a central role in red cell maturation. It binds to specific membrane receptors and can only act on committed cells such as those that make up the ERC population (BFU-E, CFU-E and others). It has been implicated in several events; for instance, stimulating the proliferation of committed erythroid precursors, driving precursors towards the maturation pathway, increasing the rate of maturation and others (Quesenberry & Levitt, 1979).

As the erythroid cells progress through this differentiation pathway a number of biochemical changes also occur. The most obvious one of these is the steady accumulation of haemoglobin. Globin mRNA can first be detected in the late proerythroblasts and accumulates in basophilic erythroblasts (Clissold et al., 1977). Less abundant molecules also appear. Glycophorin A, the major sialoglycoprotein of erythrocytes, and spectrin are detected in red blood cells just before the onset of haemoglobin synthesis (Gahmberg et al., 1978; Eisen et al., 1977). Other erythroid markers, such as band 3, begin their synthesis in erythroblasts after that of haemoglobin and glycophorin A (Fukuda et al., 1980). Markers such as transferrin receptors, HLA-A, B, C and DR antigens (Robinson et al., 1981), i/I antigens, or the A, B and H blood group antigens are already present on the erythroid progenitors (Fitchen et al., 1981). These markers are not specific for the erythroid lineage and are not selectively expressed on erythroid progenitors.

Carbonic anhydrase I (CAI) is found early in normal human erythroid differentiation. Studies using a polyclonal anti-CAI

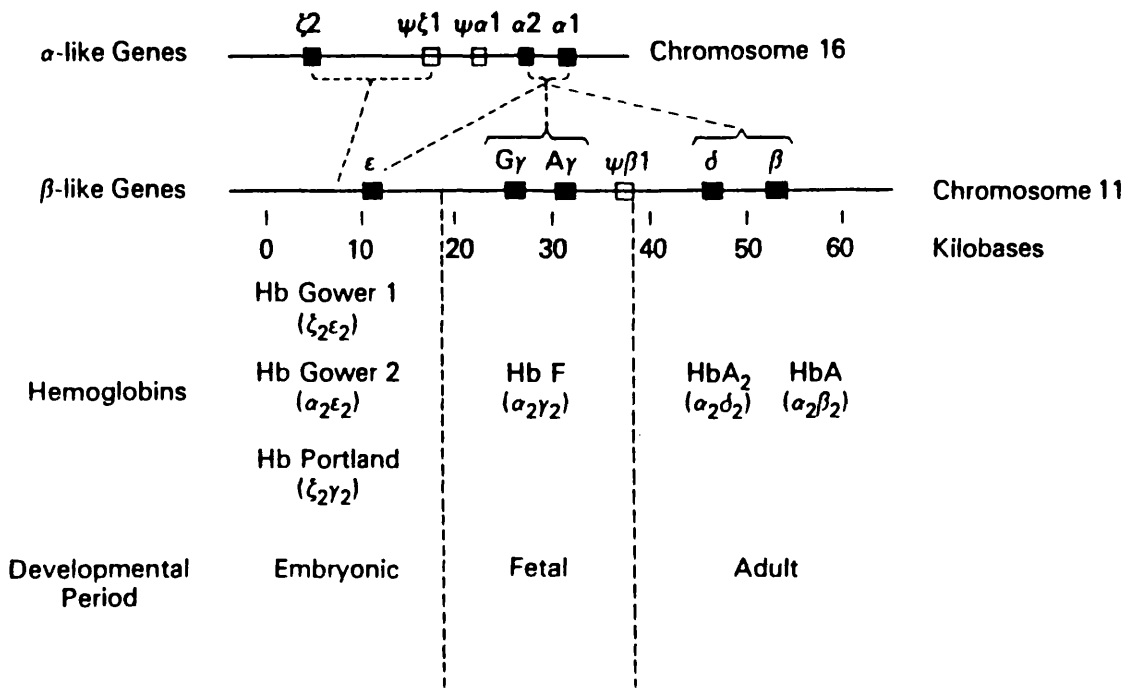


Fig. 1.7 Organisation of the human globin genes and haemoglobins produced in each stage of human development. Another α-like globin gene, θ₁ has been described. This gene is located approximately 3kb downstream from the α₁ globin gene (Leung et al., 1987).

antibody have shown the presence of CAI in cell colonies derived from both CFU-E and BFU-E erythroid progenitors (Villevall et al., 1985). CAI is therefore a specific marker of erythroid differentiation expressed at the progenitor level.

Coupled to the changes occurring during red cell differentiation and maturation, the developing organism also undergoes a series of "switches" in the expression of various genes whose functions are presumably needed at particular phases. The most intensively studied, and hence clearly defined, of these are the globin gene family.

The globin genes of man and most other species are organised into two clusters (Fig. 1.7). Three functional genes are found in the α gene cluster, those that code for the ζ and α proteins. The α_2 and α_1 genes differ slightly in sequence, but their protein coding regions are identical. In addition, two pseudogenes ($\psi\alpha$ and $\psi\zeta$) are found in the α gene cluster. Pseudogenes are defective in some essential component of gene structure and thus cannot be expressed in a globin product despite their sequence homology to functional genes. They are thought to be by-products of evolution, representing previously functional genes now inactivated by mutations in coding or regulatory regions. The β -like gene cluster includes the functional genes ϵ , $^G\gamma$, $^A\gamma$, δ , and β and one pseudogene ($\psi\beta$). The ϵ genes encode identical products with the exception of either glycine ($^G\gamma$) or alanine ($^A\gamma$) at position 136 on the polypeptide chain (Schroeder et al., 1968). The δ gene is defective in several structural elements, leading to a very low level of expression.

All normal haemoglobins are tetramers composed of four

polypeptide chains: two are encoded by a gene in the α cluster, and two are encoded by a gene in the β cluster. It is the selective expression of the various globin genes during development which forms the basis for the switches in the haemoglobin phenotype of human red cells.

In humans, the first haemoglobin tetramers produced in the embryo at about 4-5 weeks gestation were identified on starch gels by Huehns et al., 1961, and are known as Hb Gower 1 and Hb Gower 2. Hb Gower 1 consists of two α -like chains, ζ , and two β -like chains, ϵ (i.e. $\zeta_2\epsilon_2$), while Hb Gower 2 has the structure $\alpha_2\epsilon_2$ (Huehns & Farooqui, 1975). These embryonic haemoglobins gradually disappear and are replaced by foetal haemoglobin (HbF, $\alpha_2\gamma_2$) which reaches close to 100% of the total haemoglobin at about 12 weeks gestation (Fig. 1.4). It was also shown that HbF consisted of the 2 different γ -chains, $^G\gamma$ and $^A\gamma$, the ratio of which changes late in development. Adult haemoglobin (HbA) is also found at low levels in normal foetuses, as early as 6 weeks. The level of HbA remains constant until the "switch" from HbF to HbA production takes place at about 36 weeks gestation. From these observations one can identify two switching events in haemoglobin synthesis: one from embryonic to foetal haemoglobin and one from foetal to adult haemoglobin. The two switches actually require three "gene switches", one of the α -like genes, ζ to α , and two of the β -like genes, ϵ to γ and γ to β .

The amounts of HbA and HbF synthesised by erythroid cells from different erythropoietic tissue are the same at any given developmental stage (Wood et al., 1979), so the changing

pattern of globin chain synthesis is not a function of a change in the site of erythropoiesis. The exact time at which the switch from HbF to HbA occurs appears to be "programmed" or age-dependent, in that it is not related to the time of birth. Another question raised by the switch is whether a new red cell population appears, containing exclusively HbA or do HbF and HbA co-exist in all red cells? Existing experimental evidence tends to support the second alternative. It has been shown that 50% of cord blood red cells contain large amounts of both HbF and HbA (Wood, 1976). This argues against the appearance of an independent cell line synthesising exclusively HbA, and it seems that the switch occurs uniformly within the erythroid progenitor pool.

Carbonic anhydrase I expression is also "switched" on during erythroid development. The amount of CAI present in adult erythrocytes is a 100-fold higher than in the majority of foetal red cells (Boyer et al., 1983). More recent work using antibody staining of permeabilised red cells and fluorescent activated cell sorting has confirmed the switching on of CAI expression, and shown that the first detectable levels of CAI are present in red cells after 36 weeks gestation (Brady et al., manuscript submitted). Analysis of CAI in mid-switch samples suggests that at the single cell level the switch is an "all or none" phenomenon, as there is a bimodal fluorescence profile with the positive cell having the same fluorescent profile as adult cells.

The molecular mechanism of globin gene switching has yet to be defined. The γ to β switch has been the most intensively studied, since erythroleukaemic cell lines are available which

allow the differential expression to be examined. Alongside the switching phenomenon, experiments have also focussed on the mechanism which confers tissue-specificity on globin gene expression. The recent application of the techniques of heterokaryon formation by cell fusion, DNAaseI footprinting, methylation interference and bandshift assays have allowed the interaction of trans-acting factors with globin genes to be followed. This has proven to be very instructive in analysing the pattern of globin gene expression. The concept of heterokaryon formation is dealt with below. The results of this technique and the others mentioned are dealt with in the Results and Discussion chapters of this thesis since they appeared during the timeframe of my own experiments and directly impinge on that work. However, much effort has been expended on defining the cis-acting sequences functionally active in globin expression, in particular β globin expression.

MEL (mouse erythroleukaemic cells) cells can be induced to undergo erythroid maturation which is thought to resemble closely normal erythroid differentiation, resulting in high levels of murine globin mRNA. Regulated transcription of the human β -globin gene can be obtained in MEL cells after DNA mediated gene transfer (Chao et al., 1983). Subsequent experiments showed that the sequences involved in this regulation are located both upstream and downstream of the 5'-end of the gene (Charnay et al., 1984). The presence of the downstream sequences, also shown in the chicken β -globin gene using transfection experiments (Hesse et al., 1986), have been confirmed for the human gene by experiments in transgenic

mice. Correct tissue and developmental specific expression of the genes was still obtained by using promoter deletions (Townes et al., 1985) and by using Υ - β hybrid genes (Kollias et al., 1986). The latter result in particular provided a direct link to the normal in vivo function of these sequences because it closely mimicked the situation in Hb Kenya (Ojwang et al., 1983). Patients with this condition exhibit elevated levels of foetal haemoglobin in the adult due to a deletion-fusion event between the Υ - and β -globin genes. At least part of the downstream regulatory sequences have been characterised further by showing that a fragment downstream of the β -globin gene can act as a developmental specific enhancer when placed upstream of the Υ -globin gene (Kollias et al., 1987). It was however not clear from these experiments whether the enhancer is also tissue-specific.

Evidence for the importance of the upstream sequences has been obtained by stable transfection experiments in MEL cells (Wright et al., 1984). This was possible by separation of the upstream from the downstream region in a hybrid gene construct containing the β -globin gene promoter linked to a non-inducible mouse H-2K^b gene. The complementary hybrid gene containing the H-2K^b promoter and the β -globin gene showed the importance of the downstream region. It has also been shown that the 5'-sequences prevent expression of the β -globin gene in embryonic/foetal stage K562 cells (Antoniou et al., 1987; Lin et al., 1987), but that manipulation of the 5'-end sequences can restore expression of the β -globin gene in K562 cells. More recently, an induction specific regulatory element has been localised in the 5'-end promoter and two separate

downstream tissue-specific regulatory elements, one element in the 3'-flanking region as well as one in the β -globin gene have also been identified. This has been done using hybrid β -globin /H-2K^b genes stably transfected in MEL cells (Antoniou et al., 1988).

1.5 Heterokaryons

Cell fusion systems have been used to determine whether diffusible factors present in erythroid cells are capable of activating gene expression in the nuclei of cells that do not ordinarily express these genes. However, using stable somatic cell hybrids makes it difficult to draw conclusions about trans-acting regulatory mechanisms because human chromosomes are selectively lost during passage in culture. It also requires many generations of cell growth before the hybrids can be isolated and analysed. A further disadvantage of stable cell hybrids is that they are synkaryons in which the nuclear components of one parental cell type were physically mixed with those of the other. To circumvent this problem, Blau et al (1983), prepared heterokaryons from mouse muscle cells and human amniocytes, and carried out their analysis shortly after cell fusion. Whereas prior to fusion, muscle-specific gene products could not be detected in the amniocytes, high levels of expression of human muscle-specific genes were observed in the myoblast X amniocyte heterokaryons, and they retain a full complement of both human and mouse chromosomes. Thus the activation of the human genes must occur in trans.

Baron and Maniatis (1986) have used a similar assay involving the preparation of interspecific transient

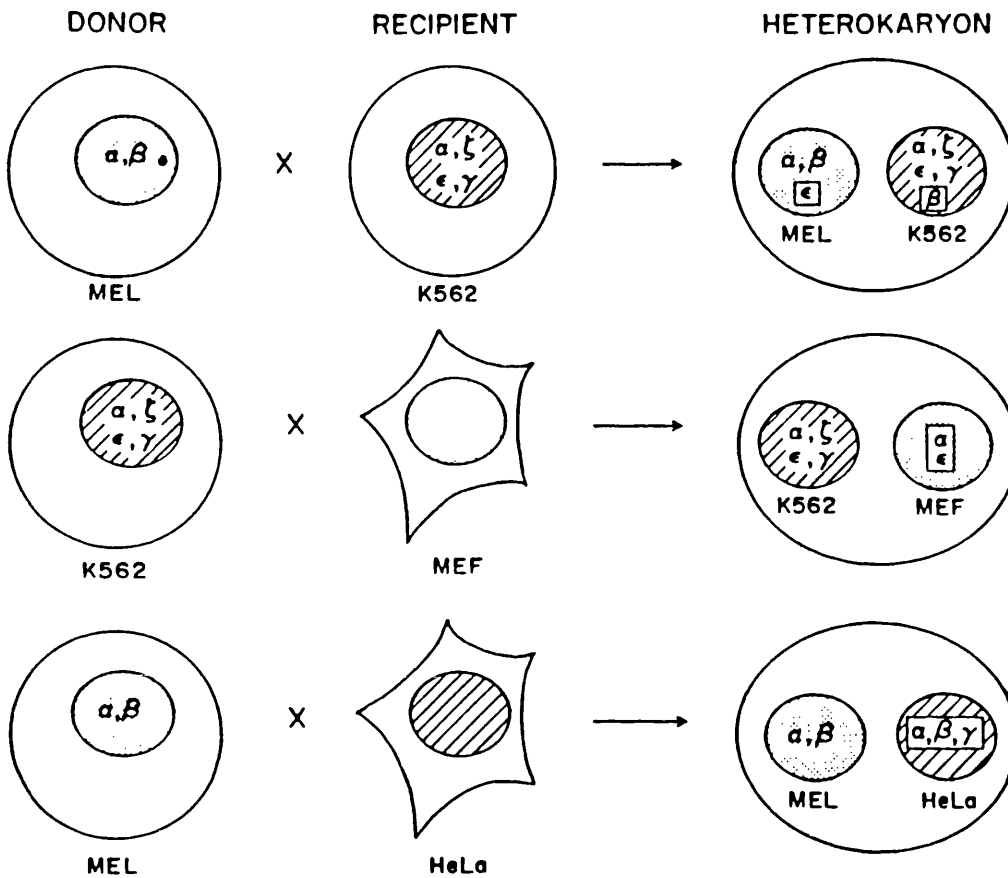


Fig. 1.8 Representation of the reprogramming of globin gene expression in transient heterokaryons as described in Baron and Maniatis (1986). Newly activated genes in the nuclei of heterokaryons have been boxed in. The recipient mouse embryo fibroblast (MEF) and human epithelial (HeLa) cells do not normally express globin genes. After fusion with an erythroid cell type, however, stage-specific globin gene activation is observed.

heterokaryons to demonstrate the presence of diffusible, trans-acting factors in erythroid cells. They observed that in general the types of globin genes expressed in the donor erythroid cell were activated (Fig. 1.8). For example, when embryonic/foetal erythroid (K562) cells were used as donor, embryonic and foetal, but not adult globin genes, were activated. Conversely, when adult erythroid (MEL) cells were used as donor, adult but not embryonic globin genes were activated. Thus, activation was not promiscuous, but reflected the phenotype of the parental erythroid cell. These experiments formed the model on which the experiments described in Chapter 5 of this thesis were carried out.

1.6 Carbonic anhydrases

Carbonic anhydrase (CA) was first isolated from erythrocytes over 50 years ago (Meldrum & Roughton, 1933; Stadie & O'Brien, 1933). The enzyme catalyses the extremely rapid and reversible hydration of CO_2 and is found in almost all organisms. It now appears that there are as many as seven genetically distinct isoenzymes, CAI-VII. Each isoform exhibits a characteristic pattern of tissue distribution. CAI is found at high levels in red blood cells, but is also present in various epithelial tissues including that of the colon. CAIII is characteristic of muscle and male rat liver, while CAII is distributed in a wide range of tissues and cell types. More recently distinct isoenzymes have been shown to exist in mitochondria (CAV), as a membrane bound form in lung and kidney (CAIV) and as a secreted salivary form (CAVI). A further isoenzyme (CAVII or CAZ) has been proposed to exist

based on the analysis of genomic clones isolated by low stringency hybridisation with a human CAII probe. Recent reviews of the CA isoenzymes are found in Fernley (1988) and Tashian (1989).

CAI, II and III are the best characterised of the CA family. All three are monomeric proteins containing 259 or 260 amino acids with a M.W. of 28-29 kd and each binding one Zn^{2+} atom. Three dimensional structures for human CAI and CAII have been determined (Notstrand et al., 1974). This makes it possible to compare residues presumed to be located within the active sites of the different CA isoenzymes (Fig. 1.9). Of the 36 residues at homologous positions, 17 (47%) are invariant. Of the other residues, all except the tiger shark red cell CA have at least one residue that seems to be unique and invariant for that isoenzyme. To what extent these residues contribute to the catalytic variation noted among the different isoenzymes is difficult to assess. For example, CAIII is a low-activity isoenzyme which is not strongly inhibited by sulphonamides (Gros & Dodgson, 1988). Conceivably, this property could be attributed to one or more of CAIII's 8 unique residues, especially the basic residues, Arg and Lys, at positions 64, 67 and 91.

Activity measurements of ^{18}O -exchange on human CAII, in which Ala and Lys have been substituted for this at position 64 by site-directed mutagenesis, indicate that His 64 contributes to the catalytic mechanism through its role in proton transfer (reviewed in Tashian, 1989). Another possibility is that the unique Phe 198 residue may modify the function of the Thr 199/Glu 106 complex which has been

CA Forms	Residue number																				
	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	
Human CA I	Y	S	H	V	H	S	F	H	N	S	F	H	N	S	F	H	N	S	F	H	N
Rabbit CA I	Y	S	H	V	H	S	F	H	N	S	F	H	N	S	F	H	N	S	F	H	N
Mouse CA I	Y	S	H	V	H	S	F	H	N	S	F	H	N	S	F	H	N	S	F	H	N
Ox CA I	Y	S	H	V	H	S	F	H	N	S	F	H	N	S	F	H	N	S	F	H	N
Horse CA I	Y	S	H	V	H	S	F	H	N	S	F	H	N	S	F	H	N	S	F	H	N
Turtle CA I	Y	S	H	V	H	S	F	H	N	S	F	H	N	S	F	H	N	S	F	H	N
Human CA II	Y	S	H	V	H	S	F	H	N	S	F	H	N	S	F	H	N	S	F	H	N
Rabbit CA II	Y	S	H	V	H	S	F	H	N	S	F	H	N	S	F	H	N	S	F	H	N
Mouse CA II	Y	S	H	V	H	S	F	H	N	S	F	H	N	S	F	H	N	S	F	H	N
Ox CA II	Y	S	H	V	H	S	F	H	N	S	F	H	N	S	F	H	N	S	F	H	N
Horse CA II	Y	S	H	V	H	S	F	H	N	S	F	H	N	S	F	H	N	S	F	H	N
Chick CA II	Y	S	H	V	H	S	F	H	N	S	F	H	N	S	F	H	N	S	F	H	N
Human CA III	Y	S	H	V	H	S	F	H	N	S	F	H	N	S	F	H	N	S	F	H	N
Mouse CA III	Y	S	H	V	H	S	F	H	N	S	F	H	N	S	F	H	N	S	F	H	N
Ox CA III	Y	S	H	V	H	S	F	H	N	S	F	H	N	S	F	H	N	S	F	H	N
Horse CA III	Y	S	H	V	H	S	F	H	N	S	F	H	N	S	F	H	N	S	F	H	N
Shark CA	X	S	X	X	X	X	X	X	X	R	V	R	R	V	R	R	V	R	R	V	R
House CA "Y"	X	S	N	V	H	Q	F	Q	V	K	Q	H	H	E	H	F	Y	L	V	G	W
Human CA VII	Y	S	N	N	H	S	V	Q	D	K	Q	H	H	E	H	V	F	L	V	G	W
Sheep CA VI	Y	S	N	N	H	T	V	N	S	K	Q	H	H	E	H	V	Y	L	V	A	Y
Human CA VI	Y	S	N	N	H	T	V	Q	G	Q	Q	H	H	E	H	V	Y	L	V	A	Y

Fig. 1.9 Diagram showing the residues postulated to occur within the active sites of animal carbonic anhydrases. Residues common to all animal sequences have been boxed. Other boxes indicate invariant and unique residues for CA isoenzymes of amniotes. The dashed box at position 64 for CAIII's indicate unique basic residues. Residues forming hydrogen-bond network to Zn-bound solvent molecule, or to the three Zn-liganded His residues (designated, Zn), in human CAI and CAII and bovine CAIII are designated by an asterisk (*). Taken from Tashian (1989).

strongly implicated in the catalytic process (Lindskog, 1983). Results from further site-directed mutagenesis studies should provide further insights into the active site mechanisms of the CA isoenzymes.

The antiquity of the active site structure is suggested by an examination of the residues found in the red cell CA of the tiger shark (Fig. 1.9), a cartilaginous fish which arose around 450 million years ago. Although the residues at positions 64-69 have not been determined, the shark CA has retained an active site structure differing little from those of the cytosolic CA isoenzymes of amniotes. However, since this represents the first structure of a non-amniotic CA, it may be premature to conclude that the active site residues of the shark CA are representative of the ancestral structure.

Examination of the structural regions of the CA proteins coded by exons provides no clues concerning the evolution of the active sites. The residues associated with the active sites (Fig. 1.9) are coded on all but one exon, and the three His residues binding the essential Zn ion are coded on exons 2 and 3. In view of the observation that exons encode various domains of proteins, it is notable that no similar correlation is evident when the exon regions of mammalian and avian CAII genes are compared with the structural or functional features of the CAII molecules.

Investigation of the genetics of the CA gene family was made possible by the cloning of cDNA for rabbit CAI (Konialis et al., 1985), human CAI (Barlow et al., 1987), human CAII (Montgomery et al., 1987), human CAIII (Lloyd et al., 1986), mouse CAI (Fraser & Curtis, 1986) and mouse CAII (Curtis,

1983). The three genes are linked on the long arm of human chromosome 8 (8q22) (Davis et al., 1987; Venta et al., 1987).

The CAI gene is particularly interesting. The "switching" on of its expression in erythroid development and its appearance early in erythroid differentiation have already been mentioned. The fact that the CAI gene is preferentially expressed in erythroid tissue, while being within a family of other genes with differing patterns of tissue distribution, adds further to the complexity of control of CAI expression. It is also clear that the development stage-specific expression of the CA genes is different. CAI is expressed very late in foetal development while, for instance, CAIII is found as early as 10 wk gestation (Lloyd et al., 1987). It is the attempt to unravel the mechanism of control of CAI expression, in particular human CAI (HCAI) expression, with which this thesis is concerned.

Thus, the aims of this project were as follows:

- i) to clone and characterise the 5'-leader region of the HCAI gene;
- ii) to map the HCAI transcription unit(s), in particular its erythroid-specific transcript;
- iii) to define the interaction of a putative erythroid-specific transcription factor with the HCAI gene;
- iv) to analyse the functional role of this factor in HCAI transcription;
- v) to examine whether or not trans-acting factors are involved in the developmental stage-specific "switching" on of HCAI expression;

vi) in the light of data published during the course of this project, (Fraser et al., 1989), to define the non-erythroid transcription unit of HCAI.

CHAPTER TWO

MATERIALS AND METHODS

2.1 Materials

Standard reagents: AR grade supplied by BDH, Fisons and Sigma.

Enzymes: Restriction enzymes from New England Biolabs, Bethesda Res. Labs. (BRL) and Anglian Biotechnology. Other enzymes used were from the same sources except SP6 RNA Polymerase, RNAase-free DNAase, S1 Nuclease and RNAases A and T₁ from Boehringer Mannheim.

Microbiology media: Bacto tryptone, yeast extract and agar from Difco.

Electrophoresis reagents: Acrylamide from BDH.
Agarose and TEMED from Sigma.
Ultra-pure Urea from BRL.

Tissue Culture: media and cultureware - sources described in Section 2.22.

Miscellaneous: dNTPs, NTPs and NAP-5 columns supplied by Pharmacia.
Yeast tRNA and spermidine from Sigma.
NA45 paper from Schleicher and Schuell.
RNAsin from Promega.
rNTPs from Boehringer Mannheim.
Radiochemicals from New England Nuclear (NEN), ³²P-γ-ATP and ³²P adCTP or Amersham, ³²P-UTP and ¹⁴C-chloramphenicol.
Nuclease-free Bovine Serum Albumin from BRL.
Guanidinium hydrochloride and thiocyanate from

Fluka.

Bacterial strains HB101: F⁻, hsd S20 (r_B⁻, m_B⁻), recA13, ara-14, proA2, lacY1, galK2, rpsL20 (Sm^r) xyl, mtl-1, supE44, λ-
LE392: hsdR⁻, hsdM⁺, supE44, supF, thi, met, lacY.
JM101: SupE, thi, δ(lac-proAB), [F', traD36, proAB, lacI^ΔZ5M15], r+m+.
Q358: hsdR⁻, SupE.

Common Buffers: 10 X TBE: 89mM Tris. HCl, 89mM Boric acid and 2mM EDTA.

1 X TE: 10mM Tris. HCl and 1mM EDTA, pH 7.4.

20 X SSC: 175.3g NaCl and 88.2g of sodium citrate / litre of water.

2.2 Methods

All solutions and (siliconised) glassware for nucleic acid work were sterilised by autoclaving at 15 p.s.i. for 20 min, unless heat labile. For RNA work the extra precaution of overnight pre-incubation at 37°C with 0.1% diethylpyrocarbonate (DEPC) was taken. Disposable gloves were worn for all experimentation. The basis for most of the methods used in this thesis can be found in Maniatis et al (1982) and Berger and Kimmel (1987). All autoradiography involving ³²P was carried out by exposure to X-ray film (Fuji) with an intensifying screen at -70°C.

2.3 Restriction enzyme digests

According to the specificity of the restriction enzyme used, as indicated by New England Biolabs, digests were performed in a solution containing 10mM Tris.HCl (pH7.5), 10mM MgCl₂ and 100µg/ml BSA together with 0, 50 or 100 mM NaCl. The only exception was SmaI digestion for which the reaction buffer contained 20 mM KCl (instead of NaCl) and Tris.HCl at pH 8.0. Incubation times were usually chosen to ensure at least a 20-fold overdigestion. Incubation temperatures were chosen as specified by New England Biolabs.

2.4 DNA modification reactions

During subcloning routines, 5'-overhangs created by restriction enzymes were filled in using the Klenow fragment of E.Coli DNA Polymerase I. Reactions were performed at 37°C for 30 min in a solution containing 10mM Tris.HCl, pH 7.5, 10mM MgCl₂, up to 2µg of DNA, 200µM of all four dNTPs and 0.25u/µl of Klenow. Reactions were terminated by heating to 65°C for 15 min.

Dephosphorylation of DNA prior to 5' labelling or, of vector DNA to prevent recircularisation during ligation was achieved using calf intestinal alkaline phosphatase (CIP). DNA was incubated with 20-30u of CIP in restriction enzyme buffer at 37°C for one hour. The reaction was terminated by heating to 65°C for 15 min and the CIP removed by a double phenol/chloroform and single chloroform extraction. The DNA was removed by precipitation from 0.3M NaOAc by 2 volumes of ethanol.

Ligations were carried out in a solution containing 50mM

Tris.HCl (pH7.4), 10mM MgCl₂, 10mM DTT, 1mM spermidine, 1mM ATP and 100µg/ml BSA. Reactions were performed at 4-8°C overnight and generally contained about a 3-fold molar excess of insert DNA to vector DNA together with 15Wu of T4 DNA ligase.

2.5 Agarose gel electrophoresis and recovery of DNA fragments

DNAs were resolved on 0.7-2.0% neutral agarose gels prepared in 20 X 20 cm flat-bed moulds. The gels were generally run at 30V overnight in TBE buffer (90mM Tris.HCl, 90mM boric acid & 2mM EDTA). Alternatively, a mini-gel of 9 X 9 cm was run at up to 120V for fast analytical results. DNA fragments were visualised by ethidium bromide staining and uv transillumination.

DNA fragments required for subcloning procedures and for use as probes were isolated from agarose gels as described in Young et al (1985). In this method, DNA is run out in a normal agarose gel. A cut is then made in the gel below the band of interest and into is placed a piece of Schleicher & Schull NA45 ion-exchange paper. Electrophoresis is then resumed so that the DNA band runs into the paper to which it binds. The paper is then transferred to an Eppendorf centrifuge tube containing 0.45ml of a 1M NaCl/50mM arginine (free base) solution which is then heated to 70°C for 30-60 min to elute the DNA. The paper is then removed and, after a single phenol/chloroform extraction, the DNA is recovered from solution by ethanol precipitation at -70°C for 15 min.

2.6 Preparation and transformation of competent E.Coli

Competent E.Coli HB101 and JM101 cells were prepared

using the CaCl_2 procedure as described by Maniatis et al (1982). 30ml of L-broth were inoculated with 0.5 ml of an overnight culture of the appropriate strain. This was incubated at 37°C with shaking until the culture reached an OD_{550} of 0.4-0.5. The culture was then chilled on ice for 10 min. The cell suspension was then centrifuged at 4000 rpm at 4°C in a Sorvall HB4 rotor. The bacterial suspension was resuspended in 15ml of an ice-cold, sterile, 50mM CaCl_2 solution and kept on ice for 15 min. The cells were then pelleted as before and resuspended in 2.5ml ice-cold, sterile, 50mM CaCl_2 .

The resuspended cells were stored on ice for at least two hours before proceeding. For transformation, up to $20\mu\text{l}$ of DNA solution (containing up to 50ng of DNA) was added to a $200\mu\text{l}$ aliquot of the cells. This mix was left on ice for 30 min and heat shocked at 37°C for 2 min before plating out. The selection media onto which the transformed cells were plated depended on the vector used, that for M13 is described below in Section 2.14. When using plasmids based on pBR322 (Bolivar et al., 1977) the cells were pre-expressed in 0.5ml L-broth at 37°C for 30 min and then plated onto L-agar plates containing $125\mu\text{g/ml}$ ampicillin (amp) and incubated overnight at 37°C . When using the Bluescript plasmid (Stratagene), after pre-expression the cells were plated onto L-agar plates containing $125\mu\text{g/ml}$ amp and 100mM IPTG. $50\mu\text{l}$ of X-gal was spread and allowed to dry on each plate prior to plating.

2.7 Large scale and rapid plasmid preparations

Bacterial plasmids were isolated from large-scale,

chloramphenicol-amplified liquid cultures by the alkaline lysis method of Birnboim and Doly (1979) as modified by Ish-Horowitz and Burke (1981) and carried out exactly as described in Maniatis et al (1982). Supercoiled plasmid DNA was purified on a CsCl-ethidium bromide gradient based on the technique of Clewell and Helinski (1972). Following centrifugation, the supercoiled DNA was recovered by puncturing the side of the centrifuge tube and drawing out the supercoil band (visualised under uv light) with a hypodermic syringe. Ethidium bromide was removed by repeated extractions with CsCl solution-saturated isopropanol. Supercoiled DNA was removed from solution by repeated ethanol precipitations then quantified and analysed for purity by spectrophotometry at 220-300nm and by gel electrophoresis. Rapid plasmid isolations were performed by the small-scale version of the above technique, also exactly as described in Maniatis et al (1982). Usually 3-4µg of plasmid DNA was recovered from 1.5ml of a saturated bacterial culture.

2.8 Purification of oligonucleotides

All the oligonucleotides used in this thesis were made on a Cyclone DNA synthesiser (Biosearch, Inc.) by Dr C.J. Taylorson of London Biotechnology Ltd., U.C.L. To cleave the oligonucleotide from the support column on which it was synthesised 1 ml of 0.88^{5g.} ammonia was passed through the column using two glass syringes attached at either end. This was repeated several times for an hour. The ammonia was then collected and heated at 65°C for 4 hours in a sealed container, to remove the protecting groups. The ammonia was removed by

lyophilization and the oligonucleotide resuspended in 1ml of TE buffer (10mM Tris.HCl, pH7.5, 1mM EDTA). The small protecting groups were removed from the oligonucleotide by passing the 1ml solution of oligonucleotide in TE down a NAP-5 (Pharmacia) desalting column. The column was pre-equilibrated in 20% ethanol. The 1ml of TE was passed down the column and this was followed by 1.5ml of 20% ethanol. The eluant was lyophilized and the resulting purified oligonucleotide resuspended in 0.5ml of sterile double distilled water (DDW). The oligonucleotide DNA was then quantified and analysed for purity by spectrophotometry at 220-300nm.

2.9 ³²P labelling of DNA

5'-end labelling of DNA was carried out using T4 polynucleotide kinase (PNK) and ³²P-γ-ATP. To 5'-label a section of plasmid, the plasmid was first digested with the appropriate restriction enzyme and the 5'-phosphate of the overhanging 5'-ends removed by treatment with CIP as in Section 2.4. Oligonucleotides to be 5'-labelled did not require this step as they were made without a 5'-phosphate group. 1μg of plasmid or 50-100ng of oligonucleotide was incubated at 37°C for one hour in a reaction containing 50mM Tris.HCl, pH7.6, 10mM MgCl₂, 1mM DTT, 1mM EDTA, 1mM spermidine, 100μCi of ³²P-γ-ATP (3000 Ci/mmol) and 1u/μl of T4 PNK. The reaction was terminated by heating to 65°C for 15 min. For oligonucleotides labelled in this manner to be used in hybridisations, the unincorporated nucleotides were removed by DE-52 cellulose chromatography (Wallace et al., 1981). The labelled oligonucleotide was diluted 10-fold with TE buffer

and loaded onto a 0.3ml DE-52 column equilibrated with TE. The column was then washed with 2ml TE buffer, the unincorporated ATP eluted with 3ml TE / 0.2M NaCl and finally the labelled oligonucleotide eluted with 1.5ml of TE / 0.5M NaCl.

Plasmid DNA or isolated restriction fragments were also labelled by either nick-translation (Rigby et al., 1977) or by random priming (Feinberg & Vogelstein, 1983). To nick-translate DNA, the reaction mix contained 200-300ng of DNA, 10mM Tris.HCl, pH8.0, 5mM MgCl₂, 1mM β-mercaptoethanol, 100 μM dATP/dGTP/dTTP, 50μCi ³²P-α-dCTP (3000Ci/mmol) and 100pg DNAase I. One minute later 10u of E.Coli DNA polymerase I was added and the mixture incubated at 14°C for one hour. In the later stages of the work in this thesis, linear DNA was labelled using the Boehringer Mannheim random primed DNA labelling kit and ³²P-α-dCTP (3000Ci/mmol). The protocol used was exactly as recommended in the manufacturer's instructions.

In both nick-translation and random-priming, unincorporated nucleotides were removed from the labelled probe as a whole by diluting the reaction mix 5-fold and spinning it through a G-50 Sephadex column (pre-equilibrated in TE) at 3000 rpm for 2 min on a bench centrifuge.

The specific activity of the labelled DNA was assessed by removing 1μl from the finished labelling reaction, diluting 10-fold and adding 1μl of the dilution to 200μl of 10% trichloroacetic acid (TCA) / 5% sodium pyrophosphate and 50μg of BSA carrier. The mix was kept on ice for 15 min and the TCA precipitate was separated from unincorporated label by filtration through a Whatman GF/C filter. After washing, the filter was dried and added to toluene/butyl-PBD for

scintillation counting.

2.10 Library plating and screening

Human genomic clones were isolated from a human genomic library kindly provided by Dr T.H. Rabbitts, Laboratory of Molecular Biology, Cambridge. This library (LeFranc et al., 1986) was constructed using DNA from an EBV-transformed B-lymphoblastoid cell line, SH. The DNA was partially digested with Sau3AI and ligated into BamHI-digested λ 2001 bacteriophage vector (Karn et al., 1984) and propagated on the E.Coli Q358 strain (Karn et al., 1980).

The aliquot of library received was first titred to determine the number of plaque forming units (pfu) / ml. An amount equivalent to 40000 pfu was added to 1ml of an overnight culture of E.Coli Q358 in 10mM MgCl₂ and 0.2% maltose. The phage were allowed to adsorb to the bacteria for 15 min at 37°C. 12ml of soft agarose (0.6% agarose, kept at 50°C to prevent setting) was then added, mixed and plated onto a pre-dried L-agar/10mM MgCl₂ plate in a 140mm Petri-dish (Sterilin). This was done for 12 separate plates giving a total of about 480,000 pfu to be screened. After setting of the agarose, the plates were inverted and incubated at 37°C for 7-9 hours depending on growth and ensuring that the plaques remained sub-confluent.

Lifts were then taken from each plate by laying a 132mm Biodyne transfer membrane (Pall) on the soft agarose and marking its position by stabbing through both filter and agar with a hot wire. A duplicate filter lift was done for each plate. After 90 sec, each filter was removed and laid, plaque

face up, on Whatman 3MM soaked in denaturing solution (0.5M NaOH, 1.5M NaCl). After 5 min, the filters were similarly laid on 3MM paper soaked in neutralising solution (3M NaOAc, pH5.5). After a further 5 min the filters were removed from the 3MM and allowed to air dry for 15 min. The filters were baked overnight at 80°C.

The filters were rinsed in 3 X SSC (10 X SSC = 1.5M NaCl / 0.3M sodium citrate) and then pre-hybridised for 4-5 hours in a solution containing 6 X SSC, 50mM sodium phosphate (pH6.5), 100µg/ml single stranded herring sperm DNA and 5 X Denhardt's solution (50 X Denhardt's = 1% BSA (Pentex), 1% Ficoll and 1% Polyvinylpyrrolidone) at a temperature determined by the base content of the probe to be hybridised to the filters. The optimum hybridisation temperature was approximated from the formula $T_d = 2^{\circ}\text{C} \times \text{no. of A:T base pairs (bp)} + 4^{\circ}\text{C} \times \text{no. of G:C bp}$ (Suggs et al., 1981), where T_d is the temperature at which $\frac{1}{2}$ of the duplexes are thermally denatured in the presence of 0.9M sodium ions. The oligonucleotide probes used to screen the library were labelled and separated from unincorporated $^{32}\text{-Y-ATP}$ as described above. The labelled oligonucleotide in TE / 0.5M NaCl was then added directly to the hybridisation solution, made up exactly as the pre-hybridisation solution. Both pre-hybridisation and hybridisation were at the same temperature, the latter being left for 12-16 hours. The filters were first rinsed 3 times with 6 X SSC at 4°C and then washed twice in 6 X SSC for 30 min at RT. The filters were now rinsed in a tetramethylammonium chloride solution (3M tetramethylammonium chloride, 50mM Tris.HCl, pH8.0, 2mM EDTA and 1mg/ml SDS) at 37°C (Wood et

al., 1985), and then washed twice for 20 min at 2-3°C above the hybridisation temperature. The filters were finally rinsed in 6 X SSC at RT before being autoradiographed at -70°C. Each plaque on the plates producing a positive signal on both primary and duplicate filters as judged by autoradiography was picked into 1ml of phage storage medium (SM) which is 100mM NaCl, 8mM MgSO₂, 50mM Tris.HCl, pH7.5 and 0.01% gelatin. Each isolate was then titred, replated on single 90mm Petri dishes (Sterilin) at 200 pfu / plate and rescreened as above. Positive plaques were again picked and screened for a third time at 50 pfu / plate so that single positive clones could be isolated.

2.11 Large scale preparation of λ bacteriophage DNA

A high-titre stock was prepared from each of the final isolates by adsorbing a sufficient number of pfu to the host bacteria to produce confluent lysis when plated out on 90mm Petri dishes. The soft L-agar overlay was smashed up in 2ml of SM and decanted into Eppendorf centrifuge tubes. 15 μ l of chloroform was added and the phage left to diffuse out for one hour at RT. The tubes were spun at 12000 rpm for 10 min and the supernatant removed and stored at 4°C.

Large scale preparation of DNA from the phage isolates was necessary to facilitate further analysis. The stocks were titred and 5 X 10⁸ pfu of each isolate were allowed to adsorb to 1 X 10¹⁰ E.Coli Q358. The same ratio of pfu to host was also used when preparing DNA of an isolate from a library made in the Charon 4A phage vector using E.Coli LE392 as host.

The phage were adsorbed for 20 min at 37°C. The mix was then added to 200ml of L-broth and 10mM MgCl₂ and incubated

with shaking at 37°C for 8-9 hours. 2ml chloroform was added and shaking continued for a further 15 min to ensure complete lysis of bacteria. 8g NaCl, 200µg DNAase and RNAase were then added and the mix kept at RT for one hour before being centrifuged in a Sorvall SS34 rotor at 7000 rpm and 4°C for 10 min to pellet bacterial debris. 10% polyethylene glycol (PEG) was added to the supernatant, dissolved very gently and kept overnight at 4°C. The solution was then centrifuged as previously, the pellet formed resuspended in 10ml of SM and extracted with 10ml of chloroform. 0.75g/ml CsCl was added to the resulting aqueous phase which was then centrifuged in 14ml polycarbonate tubes overnight at 35000 rpm and 4°C in a MSE PrepSpin centrifuge. The CsCl forms a gradient which separates intact phage from other contaminants. The intact phage was visible as a thin blue band. This was removed by puncturing the centrifuge tube and drawing out with a hypodermic syringe in a volume of about 1-2ml. This small volume containing the intact recombinant phage was dialysed overnight at 4°C in one litre of dialysis buffer (50mM Tris.HCl, pH8.0, 10mM MgCl₂ and 10mM NaCl). To each 0.5ml of phage dialysate was added 25µl 10% SDS / 40µl 250mM EDTA and 30µg of proteinase K and the mix incubated at 65°C for one hour. This was then extracted once with phenol, twice with phenol/chloroform, once with chloroform and finally precipitated from 0.3M NaOAc (pH7.0) with ethanol. The recovered phage DNA was then quantified and analysed for purity by spectrophotometry at 220-300nm and by gel electrophoresis.

2.12 Hybridisation to DNA in gel membranes

Hybridisation of labelled oligonucleotides (prepared as in Section 2.9), to plasmid and to λ phage DNA fractionated on agarose gels, was performed on dried gel membranes as described in Singer-Sam et al (1983). Following electrophoresis the gel was stained, photographed, trimmed of excess agarose and soaked in 0.5M NaOH / 0.15M NaCl for 30 min. The gel was then neutralised with 0.5M Tris.HCl (pH8.0) and 0.15M NaCl for 30 min, placed on two sheets of 3MM paper and dried at 60°C. The dried gel, resting on a single sheet of 3MM paper, was floated in DDW for one minute and the dried gel membrane was carefully peeled off the 3MM paper and immediately placed in a heat-sealed plastic bag. The DNA in the gel membrane was then hybridised, without pre-hybridisation, to the labelled oligonucleotide probe at the appropriate temperature for 6-7 hours. The hybridisation solution contained 6 X SET (1 X SET = 30mM Tris.HCl, pH8.0, 0.15M NaCl & 1mM EDTA), 0.5% NP-40 and 100 μ g/ml tRNA. The gel membrane was subsequently washed in 6 X SSC, 0.1% SDS and 0.05% sodium pyrophosphate for 15 min at the hybridisation temperature and a further 10°C above that before autoradiography at -70°C.

2.13 Rapid restriction mapping of DNA cloned in λ phage vectors

³²P labelled oligonucleotides complementary to the cohesive (cos) ends of λ phage (ON-L and ON-R) can be used to rapid restriction map DNA cloned in λ phage vectors. They are annealed to a partial digest of the recombinant and run out on an agarose gel, dried and autoradiographed, as described in

Rackwitz et al (1984).

Initially, the λ DNA to be digested was partially digested with a series of enzymes. The partial digest was carried out in a solution containing 1 μ g of λ DNA, 1 μ l of 10 X the correct buffer for the restriction enzyme used, 0.5u of the enzyme diluted in DDW and the volume adjusted to 10 μ l with DDW. This mix was incubated at 37°C. After 7 min two 2 μ l aliquots were removed and added to 1 μ l of 150mM EDTA on ice to stop further digestion.

The cos oligonucleoties, ON-L and ON-R (New England Biolabs) were 5'-end labelled as in Section 2.9. ON-L is complementary to the left cohesive terminus of λ DNA and ON-R to the right terminus. 0.5ng of labelled ON-L or ON-R was added to the aliquot of partial digest and DDW and NaCl added to make the solution 150mM NaCl. The mixture was heated to 75°C for 2 min and immediately transferred to a 45°C waterbath for 30 min to anneal. 5 μ l of gel loading buffer (36mM Tris.HCl, pH7.7, 30mM NaH₂PO₄, 60mM EDTA, 50% glycerol and 0.1% bromophenol blue) were added and the mixture was loaded onto a 0.5% agarose gel. Electrophoresis was carried out at 1.5V/cm for 24 hours in 36mM Tris.HCl, pH7.7, 30mM NaH₂PO₄ and 1mM EDTA. Relevant ³²P labelled marker tracks were also run. The gel was washed briefly and dried onto Whatman DE-81 cellulose paper at 60°C before autoradiography at -70°C.

2.14 DNA sequencing in M13 phage by the dideoxy chain-termination method

The ligation mixture of M13 vector and insert DNA was transformed into E. Coli JM101, made competent by the CaCl₂

procedure (Section 2.6) and plated with IPTG and X-gal. Under these conditions recombinant and non recombinant phage can be distinguished due to insertional inactivation of the β -galactosidase gene of M13 mp18 or mp19; recombinants appear as white plaques on the chromogenic (X-gal) substrate.

Single-stranded template was prepared from 1.5ml cultures of single white plaques picked into 1:100 dilutions of E.Coli JM101 and grown for 6-7 hours at 37°C. Cells were removed from the culture medium and the phage precipitated in 3.5% PEG 6000, 500mM NaCl for 15 min at RT. The protein coats of the particles were removed by phenol extraction and the single-stranded DNA recovered by ethanol precipitation.

Aliquots of single-stranded DNA were annealed to 2ng of appropriate primer in 10 μ l of 10mM Tris.HCl, pH8.5, 10mM MgCl₂ at 50-55°C for one hour. Nucleotide mixtures were prepared, containing the appropriate dideoxynucleotide triphosphate so that the reaction could be terminated at each of the four bases. These contained dTTP, dGTP and dCTP at 125 μ M or, if the corresponding dideoxy base was present, at 6 μ M. The dideoxy bases were present at 67 μ M for ddTTP and 33 μ M for ddATP, ddCTP and ddGTP. The annealed primer-template was divided between four tubes (2 μ l each), one for each of the nucleotide mixes (2 μ l of mix in each tube), and 2.5 μ Ci of ³⁵S- α -dATP (400 Ci/mmol) and 0.5u of Klenow fragment added to each reaction, with a final volume of 6 μ l. After 20 min at RT the reactions were chased by addition of a nucleotide mix containing each of the dNTPs at 250 μ M and given an additional 20 min incubation at RT. Samples were prepared for loading onto the 8% acrylamide sequencing gels by adding 2 μ l of formamide dye mix

(80% formamide, 10mM EDTA, 0.1% xylene cyanol and 0.1% bromphenol blue) and boiling for 3 min. After running the gels were dried and exposed to X-ray film at RT for 24-48 hours.

2.15 Isolation of total and mRNA

Total RNA was isolated from human peripheral blood (provided by the Haematology Department, UCMHMS) of patients with an elevated reticulocyte count and also from cord blood. The blood cells were washed three times in phosphate-buffered saline (PBSa; 8g/l NaCl, 2g/l KCl, 1.5g/l Na₂HPO₄, 2g/l KH₂PO₄), the buffy coat removed and the remaining cells pelleted. The cell pellet was rapidly homogenised in 4 volumes of 6M urea/3M LiCl (Auffray & Rougeon, 1980) and left at 0°C overnight to precipitate the RNA. The RNA was collected by centrifugation at 16000g for 30 min at 4°C. The pellet was resuspended in 9ml of 8M guanidium.HCl, 0.4% sodium acetate and 0.1M β-mercaptoethanol (Chirgwin et al., 1979). This solution was layered on top of 2.5ml 5.7M CsCl / 0.1M EDTA, pH8.0, in a 14ml polypropylene centrifuge tube. Following centrifugation for 18 hours at 33000 rpm and 20°C in a swing-out rotor the contents of the tube were removed to leave an RNA pellet at the bottom which was resuspended in 0.1% DEPC-treated DDW and recovered by two ethanol precipitations from 0.3M NaOAc (pH5.2).

Total RNA was recovered from cell lines by first washing the cells twice with ice-cold PBSa and then pelleting. The cell pellet was lysed directly in the same guanidinium solution as above then layered on top of a 2.5ml pad of 5M CsCl in 0.1M EDTA with a refractive index of 1.400.

Centrifugation and recovery of the RNA pellet was as above. The RNA was finally resuspended in DDW and its concentration and purity measured by spectrophotometry at 220-300nm.

Total RNA was prepared from colon tissue using guanidinium thiocyanate. Colon tissue was obtained courtesy of Dr P. Boulos, Surgery Dept., U.C.H. and the colon epithelial tissue samples removed in the UCMHMS Pathology Dept. The tissue was flash frozen in liquid nitrogen and stored at -70°C. The frozen tissue was homogenised in 8-10 volumes of 4M guanidinium thiocyanate containing 0.1M β -mercaptoethanol at pH5.0. The RNA was then pelleted by centrifugation through CsCl as above.

mRNA was purified from total RNA using oligo-dT cellulose chromatography based on the method of Aviv and Leder, 1972. Briefly, total RNA was dissolved in DDW, heated at 65°C for 5 min adjusted to binding buffer concentration (0.5M NaCl, 10mM Tris.HCl, pH7.5, 1mM EDTA, 0.1% SDS) and loaded onto an oligo-(dT) column equilibrated with the same buffer concentration. The RNA was recycled twice through the column and the column was then washed with binding buffer until the A_{260} of the eluant was <0.05. The column was then given a further wash with 0.1M NaCl, 10mM Tris.HCl, pH7.5, 1mM EDTA, 0.1% SDS and the bound RNA eluted with elution buffer (10mM Tris.HCl, pH 7.5, 1mM EDTA & 0.05% SDS). The eluted RNA was then adjusted to contain binding buffer and the binding, washing and elution procedure repeated once more. Finally, the poly (A+) mRNA was adjusted to 0.3M NaOAc, pH5.2, and precipitated with 2.5 volumes of ethanol.

2.16 Northern analysis

During the initial stages of the work in this thesis Northern analysis was carried out by the method of Lehrach et al (1977) exactly as described by Maniatis et al (1982). However, subsequently the method used was as described by Fourney et al (1988) with some modifications.

20-30µg of total RNA or 1-2µg of mRNA was resuspended in 5µl DEPC-treated DDW to which was added 25µl of an electrophoresis sample buffer made up of 0.75ml deionized formamide, 0.24ml formaldehyde, 0.1ml DEPC-treated DDW, 0.1ml glycerol, 0.08ml of a 10% bromophenol blue solution and 0.15ml 10 X MOPS buffer (0.2M 3-N-morpholinopropanesulphonic acid, 50mM NaOAc, 10mM EDTA, pH7.0). The mixture was heated at 65°C for 15-20 min and chilled on ice for 5 min. 2µl of a 1mg/ml ethidium bromide solution was added before the sample was loaded on a denaturing agarose gel. The gel consisted of 2.2g agarose, 18ml 10 X MOPS, 152.7ml DDW and 9.3ml 37% formaldehyde. Electrophoresis was for 16-18 hours in 1 X MOPS buffer at 30V.

Following electrophoresis the RNA was visualised by uv transillumination and photographed. The gel was then soaked in 10 X SSC for 2 X 20 min before transfer of the RNA to a membrane filter, either GeneScreen Plus (DuPont-NEN) or Hybond N-Plus (Amersham), by the standard capillary blot technique (Maniatis et al., 1982) in 10 X SSC. After transfer GeneScreen Plus filters were baked at 80°C for 2 hours whereas Hybond N-Plus filters were treated with 0.05M NaOH for 5 min to fix the RNA. Both types of filter were pre-hybridised and hybridised in a solution containing 6 X SSC, 5 X Denhardt's, 0.5% SDS and

100µg/ml single stranded herring sperm DNA at a temperature determined by the probe. Pre-hybridisation was for 4-5 hours and hybridisation for 14-16 hours. Filters were then washed in 1 X SSC / 0.1% SDS twice for 10 min at RT and twice in 0.5 X SSC / 0.1% SDS for 15 min at the hybridisation temperature before autoradiography at -70°C.

2.17 Primer extension analysis

Two methods were used to carry out primer extension analysis of CAI mRNA which differed only in the method of hybridisation of the primer to the mRNA. All primer extensions were done using an oligonucleotide DNA primer end-labelled with ³²P-γ-ATP (3000Ci/mmol) as in Section 2.9.

10-20µg total RNA or 2-3µg poly A(+) mRNA was co-precipitated in ethanol with 5-10ng of labelled primer and resuspended in 20µl of hybridisation buffer (40mM PIPES, pH6.4, 0.4M NaCl, 1mM EDTA, 80% deionised formamide) as in Sutton and Boothroyd (1986). The solution was heated to 80°C for 5 min then hybridised for 12-14 hours at 37°C. Following hybridisation RNA/DNA was collected after precipitation from 0.3M NaOAc, pH5.2, and 3 volumes of ethanol. The other method of hybridisation (Geliebter, 1987) involved co-precipitation of primer and RNA as above but resuspension in 20µl of annealing buffer (250mM KCl & 10mM Tris.HCl, pH8.3). The solution was heated to 80°C for 3 min and annealed at a temperature 5°C below the T_d of the primer for 45 minutes followed by ethanol precipitation as above.

The RNA/DNA collected after hybridisation was resuspended in a solution containing 50mM Tris.HCl, pH8.0, 140mM KCl, 6mM

MgCl₂, 30mM β-mercaptoethanol, 2u/μl placental RNAase inhibitor (Promega), 1mM of all 4 dNTPs and 0.5u/μl avian myeloblastosis virus reverse transcriptase (AMV-RT, Anglian Biotech). The reaction was incubated at 50°C, to minimise RNA secondary structure formation, for 1 hour. After this, EDTA was added to 15mM followed by 1μl DNAase-free RNAase (Boehringer Mannheim) and the mix incubated for 30 minutes at 37°C. The extended DNA product was collected by precipitation in 3 volumes of ethanol from 0.5M NH₄OAc and resuspended in 5μl of Maxam & Gilbert loading buffer (80% formamide, 10mM NaOH, 1mM EDTA, 0.1% xylene cyanol and 0.1% bromophenol blue) as in Maxam and Gilbert, 1980. The sample was heated to 90°C for 10 min then loaded onto a 5, 8 or 10% polyacrylamide / 50% urea (B.R.L.) denaturing gel. The gel was run for an appropriate time as indicated by the loading buffer dyes at 70W. The gel was then dried and autoradiographed at -70°C.

2.18 S1 nuclease mapping

S1 nuclease mapping was used to identify the 5'- and 3'-ends of the HCAI transcription unit based on the method of Burke (1984). Single stranded DNA probes were created from M13 DNA templates, isolated by preparative polyacrylamide gel electrophoresis and hybridised to RNA before S1 nuclease digestion.

For 5'-end mapping the HindIII-PvuII fragment containing the first exon of the HCAI gene (Fig. 3.8A) was subcloned into M13 mp18 so that the single-stranded recombinant had the coding strand. 10ng of 5' ³²P-γ-ATP labelled oligonucleotide (3'-CACCAGGAGACACCGTCGGA-5'), complementary to a region of the

first exon, was annealed to 1µg of the single-stranded recombinant in a solution containing 10mM Tris.HCl, pH7.5, 10mM MgCl₂, by heating to 70°C for 5 min and allowing to cool slowly to 37°C. The complementary strand of the M13 recombinant was then formed using the oligonucleotide as primer in the presence of 0.2u/µl Klenow enzyme and 125µM of all four dNTPs. After terminating the reaction by heating to 65°C for 15 min the solution was adjusted to 50mM NaCl and 100µg/ml BSA. The now double stranded DNA was digested with HaeIII. The digest was stopped by the addition of EDTA to 15mM and ethanol precipitated. The DNA was collected by centrifugation, resuspended in 25µl formamide dye mix and heated at 90°C for 3 min before loading on a 6.5% non denaturing polyacrylamide gel for electrophoresis at 70W until the loading buffer dyes indicated sufficient separation. One of the glass gel plates was removed and the gel covered in Saran wrap before exposure to an X-ray film plus intensifying screen for 10 min. The autoradiograph was then used to locate the required 185nt single-stranded probe band in the gel. This band was excised from the gel and the probe DNA eluted by mashing up the gel slice and incubating at 37°C overnight in a solution of 0.5M NH₄OAc, 10mM Mg(OAc)₂, 1mM EDTA, 0.1% SDS and 10µg/ml tRNA. The DNA was isolated from this solution by a series of resuspensions and ethanol precipitations as described in Maxam & Gilbert (1980).

For 3'-end mapping the HindIII-MboI fragment containing the 3'-untranslated region of the HCAI gene (Fig. 3.12A) was subcloned into M13 mp19. The complementary strand was initiated after annealing the 17 nt M13 sequencing primer

(Amersham) in the presence of 10mM Tris.HCl, pH7.5, 10mM MgCl₂, 100μM dATP / dGTP / dTTP, 0.2u/μl Klenow enzyme, 12.5μCi ³²P-α-dCTP (800Ci/mmol) and 12.5μM dCTP incubating at 37°C for 20 min. Following this, dCTP was added to 1mM and the reaction returned to 37°C for a further 20 min to complete the complementary strand synthesis. The DNA was then digested with HindIII (a unique site), ethanol precipitated, resuspended in 25μl formamide dye mix, heated to 90°C for 3 min and run out as above on a 6% non denaturing polyacrylamide gel. The 370 nt single-stranded probe was isolated as above.

For 5'-end mapping 3 X 10⁵cpm of single-stranded probe and 15μg of total human reticulocyte RNA were incubated for 3 hours at 62°C or 70°C in 15μl hybridisation buffer (0.4M NaCl, 40mM PIPES, pH 6.4, 1mM EDTA). For 3'-end mapping 1.4 X 10⁵cpm of probe was hybridised to 9μg of the total RNA for 1 hour at 50°C or 58°C. The rest of the procedure was exactly as described in Burke (1984) except that the S1 nuclease (Boehringer-Mannheim) digestion was carried out for 2 hours at 20°C and the products resolved by electrophoresis on a 7% polyacrylamide/50% urea gel.

2.19 RNA sequencing

This technique (Geleibter, 1987) allows the direct sequencing of a single mRNA species in poly (A+) mRNA whose cDNA has not been fully cloned, using a sequence-specific oligonucleotide primer.

The primer was 5'-end labelled as in Section 2.9 and the reaction terminated by heating to 65°C for 15 min. 1μl of this reaction containing 2.5ng primer was added to 11μg poly (A+)

mRNA in 250mM KCl, 10mM Tris.HCl, pH8.3, to a volume of 12 μ l. The mix was heated to 80°C for 3 minutes and annealed at 5°C below the T_d for the primer. 2 μ l of this mixture was then added to 3.5 μ l reverse transcription buffer (24mM Tris.HCl, pH 8.3, 16mM MgCl₂, 10mM DTT, 0.8mM dATP/dCTP/dTTP, 1.5mM dGTP, 0.4u placental RNAase inhibitor and 4.5u AMV-RT) and 1 μ l of either 1mM ddATP / ddCTP / ddGTP / ddTTP depending on which base is to be sequenced. The mix was then incubated at 50°C for 50 min after which 2.2 μ l of formamide stop mix was added to each sequencing reaction, the sample heated to 95°C for 3 min then loaded onto a 8% polyacrylamide/50% urea sequencing gel. After electrophoresis the gel was dried and autoradiographed at -70°C.

2.20 RNAase protection analysis

The detection of low-levels of specific RNA transcripts was achieved by the technique of RNAase mapping using single-stranded RNA probes generated from SP6 vectors hybridised to complementary mRNA species. The hybrids were then digested with specific RNAases and the characteristic protected fragment resolved by gel electrophoresis (Melton et al., 1984).

DNA fragments were cloned in pSP6 vectors (obtained from Boehringer Mannheim) in such an orientation that when used as templates for SP6 RNA polymerase they produced single-stranded RNA transcripts complementary to the specific mRNA to be detected (Fig. 5.3 and 5.4). The SP6 vector containing the cloned insert was then linearised by digestion with an appropriate restriction enzyme (at a unique site not between

the SP6 promoter and the insert) phenol/chloroform and chloroform extracted then resuspended in DEPC-treated DDW after ethanol precipitation. 1µg of linearised vector was then added to the transcription reaction containing 10mM DTT, 5mM spermidine, 40mM Tris.HCl, pH7.5, 6mM MgCl₂, 10mM NaCl, 60u placental RNAase inhibitor, 0.75mM ATP / CTP / GTP, 37.5µM UTP, 0.5u/µl SP6 RNA polymerase and 35-50µCi ³²P-α-UTP at either 410 or 3000Ci/mmol. The reaction mix was incubated at 40°C for one hour before the addition of 1µl RNAase-free DNAase and incubation at 37°C for 25 min to remove the DNA template. The reaction mix was adjusted to 5mM Tris.HCl, pH7.5 / 5mM EDTA before phenol/chloroform extraction. The organic phase was re-extracted in 1% SDS / 10mMTris.HCl, pH7.5 / 10mM EDTA and both aqueous phases pooled and precipitated from 0.3M NaOAc, pH5.2, with ethanol, the RNA probe then being resuspended in DEPC-treated DDW.

The RNA probe was co-precipitated with up to 80µg of total RNA and resuspended in 30µl of hybridisation buffer (as in Section 2.18), heated to 85°C for 5 min before hybridisation at 45°C for 12-16 hours. After hybridisation samples were digested for 1 hour at 30-34°C (optimised for different probes) with 300µl RNAase solution (40µg/ml RNAase A, 2µl/ml RNAase T, both Boehringer Mannheim, 10mM Tris.HCl, pH7.5, 5mM EDTA, 300mM NaCl). The RNAase solution was boiled for 10 min before use to remove any residual double-stranded nuclease activity. After digestion 30µl of 10% SDS and 75µg of Proteinase K (Sigma) was added to the reaction mix and incubated at 37°C for 20 min to terminate RNAase activity. Each sample was then extracted twice with phenol/chloroform, ethanol precipitated

with 10µg tRNA carrier, resuspended in formamide dye mix, heated to 90°C for 5 min and loaded onto a 5% polyacrylamide / 50% urea gel. After electrophoresis the gel was dried and autoradiographed at -70°C.

2.21 Analysis of CAI protein by SDS PAGE, isoelectric focusing and Western blotting

The protein analysis was carried out on a whole cell lysate of mouse erythroleukaemic cells (MEL C88, Section 2.22). The MEL C88 cells were washed twice in PBSa and resuspended at 2.5×10^7 cells/ml of sonication buffer (10mM Tris.HCl, pH7.6, 1mM MgCl₂, 20mM KCl and 10% glycerol), as described in Conscience & Meier (1980) plus 1% NP-40. The cells were then sonicated and lysis confirmed by phase-contrast microscopy. The lysate was centrifuged at 20000g for 30 min at 4°C and the supernatant removed and stored at -70°C.

Protein was analysed by SDS PAGE by the method of Laemmli (1970). The gel was prepared using 12.5ml acrylamide stock (30% acrylamide, 0.8% N,N'-methylene bisacrylamide), 7.5ml 1.5M Tris.HCl (pH8.8), 0.4% SDS and DDW to 28.5ml in the running gel, and 2.6ml acrylamide stock, 3.3ml 0.5M Tris.HCl (pH 6.8), 0.4% SDS and 6.7ml DDW in the stacking gel. 1ml 1% ammonium persulphate and 35µl TEMED. Samples were prepared by boiling for 2 min in an equal volume of SDS sample buffer (10% SDS, 10% β-mercaptoethanol, 50% sucrose, 0.005% bromophenol blue and 0.125M Tris.HCl, pH 6.8. Electrophoresis was for 5 hours at 32mA in a running buffer containing 25mM Tris.HCl (pH 8.5), 192mM glycine and 0.1% SDS. The part of the gel not to be blotted was stained in Coomassie Blue R250 and destained in

30% methanol / 10% glacial acetic acid.

Isoelectric focusing was done in a gel made up of 2.67ml acrylamide stock, 0.3ml pH 5-7 ampholines (Pharmacia), 0.9ml pH 7-9 ampholines, 0.13ml pH 3.5-10 ampholines and 20.3ml DDW, polymerised with 0.67ml 1% ammonium persulphate and 20 μ l TEMED. Samples were loaded on small pieces of Whatman 3MM paper soaked in cell lysate and laid on the gel. Electrophoresis was at 300V and 1000V for 1 hour followed by 1500V for 1½ hours on a horizontal cooled plate. At each end of the gel was a 1cm wide strip of 3MM paper, dipped in 1M NaOH at the cathode and 2M H₃PO₄ at the anode.

The resolved proteins on both type of gel were Western blotted according to the technique of Towbin et al (1979). The proteins were transferred to a nitrocellulose filter (Schleicher and Schull, 0.22 μ m) by blotting overnight at 100mA on a Biorad Transblot apparatus in 25mM Tris.HCl (pH7.5), 192 mM glycine and 20% methanol. After transfer, the nitrocellulose filter was blocked by gentle agitation in 3% gelatin (Biorad) in TBS buffer (20mM Tris.HCl, pH7.5, 500mM NaCl) for 1 hour, followed by 2 X 5 min washes in TTBS (TBS buffer containing 0.05% Tween-20, Biorad). The filter was then treated with a solution containing rabbit anti-HCAI antibody and 1% gelatin / TTBS for 1 hour at RT. The rabbit anti-HCAI was a gift from Dr. Y. Edwards, U.C.L., having been purified by Sepharose-Protein A resin. The antibody solution was removed and the filter washed twice in TTBS for 5 min. A second antibody solution was then added containing goat anti-rabbit Ig conjugated with horse radish peroxidase and incubated for 1 hour. Following two more washes in TTBS and

one in TBS the filter was developed. 60mg of 4-chloronaphthol (Biorad) was dissolved in 20ml ice-cold methanol and 60µl ice-cold hydrogen peroxide was added to 100ml of TBS at RT. The 2 solutions were mixed and added to the nitrocellulose filter. Colour was developed for no longer than 45 min after which the filter was rinsed in DDW and photographed.

2.22 Tissue culture and cell lines

Tissue culture practice was standard as described in Freshney, 1983. Media (Gibco-BRL) were either reconstituted from powder and sterilised by filtration through a 0.22µm filter or, bought ready-made. Cells were passaged in 175cm² culture flasks, either Falcon (Becton-Dickinson) or Nunc (Gibco-BRL), at 37°C and 5% CO₂. As far as possible cells were maintained in logarithmic growth phase at 2-4 X 10⁵ cells/ml. Total cell and viable cell counts were done using a Neubauer haemocytometer and Trypan Blue (Sigma) staining. A wide variety of cell lines were used in the work in this thesis. All media were supplemented with 10% (FCS) foetal calf serum (Flow), 100 international u/ml penicillin, 100µg/ml streptomycin, 2.5µg/ml amphotericin B and 2mM glutamine. HeLa cells are a line of human cervical carcinoma cells (Gey et al., 1952). K562 cells are human erythroleukaemic cells with an embryonic/foetal phenotype (Lozzio & Lozzio, 1975). K562 SAI cells are a semi-adherent subclone of K562 (Spandidos, 1984). HEL (92.1.7) cells are human erythroleukaemic cells (Martin & Papayannopoulou, 1982). F412B2 cells are a mouse erythroleukaemic line (Spandidos & Paul, 1982). Finally, SW480 cells are a human colorectal adenocarcinoma cell line

(Leibovitz et al., 1976). All of these lines were maintained in Dulbecco's minimal essential (DMEM).

MEL C88 (APRT-) cells are a mouse erythroleukaemic cell line (Deisseroth & Hendrick, 1978). They were maintained in α -MEM supplemented with 50 μ g/ml diaminopurine. HUT-78 are a human T-lymphocyte cell line (Gootenberg et al., 1981) and HL-60 a human monocytic cell line (Collins et al., 1977) and both were maintained in RPMI-1640, the latter with 10% heat-inactivated FCS. A-498 (American Type Culture Collection - human tumour bank, cell line 44) and Caki-1 (Fogh and Trempe, 1985) are both human renal carcinoma cell lines. A498 were maintained in α -MEM and Caki-1 in McCoy's 5A medium. LLC-RK1 are a non-transformed rabbit kidney cell line (Hull et al., 1965) which were maintained in Medium 199 without a CO₂ atmosphere.

2.23 Cell fusion

Transient heterokaryons were formed by fusing MEL C88 with K562 SAI cells and MEL C88 with HeLa cells. The method used was on that of Baron and Maniatis (1986). In these experiments the cells were adhered to tissue culture plates at high density and treated with lectin to promote cell to cell contact. The cells were then fused using PEG and incubated for 48 hours before harvesting and isolation of total cellular RNA.

The cell fusions were carried out on 100mm Corning tissue culture plates (Bibby). The plates were pre-treated with 10 μ g/ml of poly-L-lysine-HBr (Sigma, M.W. 150-300,000) in PBSa for two hours, then rinsed in DDW. For each plate 5 X 10⁷ MEL

C88 cells and 1×10^7 K562 or HeLa cells were required. The K562 SAI cells were previously induced by culturing with $20\mu\text{M}$ haemin (Rutherford et al., 1979). The 5×10^7 MEL C88 cells were resuspended in 4ml of Iscove's DMEM (IDMEM) / 10% FCS and the 1×10^7 K562 SAI or HeLa in 1ml of IDMEM / 10% FCS and disaggregated by 2 gentle passages with a 19G needle. The lectin, phytohaemagglutinin-P (Wellcome) was added at $50\mu\text{g/ml}$ and the cells added to the rinsed culture plates. The plates were incubated for 90 min at which point virtually 100% of cells were adherent. The culture medium was aspirated and the cells washed with 10ml of serum-free medium. 2ml of Koch-Light PEG 1000 (50% in freshly prepared serum-free IDMEM, adjusted to pH7.4) was then applied to cells for 60 sec. The cells were washed twice, very gently, with 10ml of serum-free IDMEM and finally 10ml of IDMEM / 10% FCS added and the plates incubated for 48 hours at 37°C and 5% CO_2 . Total cellular RNA was isolated using Guanidinium-HCl as in Section 2.15.

2.24 Transfection of tissue culture cells

The method used for transfection of F412B2 cells was a variation of that described in Rosenthal (1987). 1.5×10^6 F412B2 cells in IDMEM / 10% FCS were plated on 100mm Corning tissue culture dishes about 20-22 hours before transfection. For each plate the CaPO_4 / DNA precipitate was formed by first mixing $387\mu\text{l}$ of DDW, $500\mu\text{l}$ of 2 X Hebes buffered saline (2 X HBS; 50mM Hebes, 280mM NaCl and 1.5mM Na_2HPO_4 , adjusted to pH7.03) and $50\mu\text{l}$ of DNA in TE buffer (up to $40\mu\text{g}$ of supercoiled plasmid DNA/plate). This mix was vortexed briefly and then $63\mu\text{l}$ of a freshly prepared solution of 2M CaCl_2 added.

A fine precipitate was formed, checked under the microscope and left for 30 min at RT. The precipitate was then added dropwise to the cells in their medium and incubated for 24 hours. The medium and precipitate were then removed and 20ml of fresh medium added for a further incubation of 24 hours. Following this, the cells were removed from the plates with 2mM EDTA in PBSa and resuspended in 1.5ml PBSa for preparation of cell lysate (see below).

HeLa cells were most efficiently transfected using a variation of the above method but also including a glycerol shock. 3×10^6 HeLa cells in DMEM / 10% FCS were plated in 75cm² tissue culture flasks 20-22 hours before transfection. The CaPO₄/DNA was added as above and left in contact with the cells for 6 hours. The cells were washed once with serum-free DMEM then 5ml of 15% glycerol in 1 X HBS was added for 2 min. The cells were then washed twice with PBSa and once with serum-free DMEM. 30ml of DMEM / 10% FCS was then added and incubated for 48 hours. The cells were then removed and resuspended as above.

Plasmids used in transfections were supercoil-purified by double banding in CsCl gradients and checked by agarose gel electrophoresis.

2.25 CAT and β -galactosidase assays

To assay for transfected CAT (chloramphenicol acetyltransferase) and β -galactosidase activity, lysates were made from the transfected cells. The cells, having been resuspended in 1.5ml of PBSa after transfection, were then centrifuged for 20 seconds at 4000rpm and resuspended in 200 μ l

0.25M Tris.HCl, pH7.5. The cells were lysed by 3 cycles of freeze-thawing and then centrifuged for 10 min at 4°C and 12000rpm. The supernatant was then removed and stored at -20°C.

β-galactosidase assays were performed essentially as described by Herbomel et al (1984). 1ml of 60mM Na₂HPO₄, 40mM NaH₂PO₄, 10mM KCl, 1mM MgCl₂ and 50mM β-mercaptoethanol was added to 30-50μl of cell lysate followed by 0.2ml of an o-nitrophenylgalactoside solution (2mg/ml in 60mM Na₂HPO₄/40mM NaH₂PO₄). The mix was incubated at 37°C for 1 hour and the reaction stopped by adding 0.5ml 1M Na₂CO₃. The OD₄₂₀ was then measured and expressed as OD₄₂₀ X 100 for the total volume of the cell extract. For a particular transfection, the efficiency was assessed by β-galactosidase activity. An equivalent amount of β-galactosidase activity from each plate was then taken and assayed for CAT activity. CAT assays were performed essentially as described by Gorman et al (1982). Variable volumes of cell extract up to 70μl were added to 70μl 0.25M Tris.HCl, pH7.6, with 0.6μCi ¹⁴C-chloramphenicol (CM, NEN-Du Pont) and the total volume made to 140μl with DDW. The mix was incubated at 37°C for 5 min, 20μl of 5mM Acetyl CoA added and the incubation continued for 1 hour. The assay reaction was stopped by adding 1ml cold ethyl acetate to extract the CM. After vortexing, the organic phase was transferred to a fresh Eppendorf centrifuge tube and evaporated off under a stream of nitrogen. The CM residue was then resuspended in 30μl of ethyl acetate and spotted onto a silica gel thin layer chromatography (tlc) plate. This was then placed into a TLC tank containing 95:5

chloroform/methanol for 90 min. The tlc plate was then air-dried and autoradiographed at 4°C for 6-10 hours.

2.26 Protein preparation for bandshift assays

Two methods were used to prepare protein extracts for use in bandshift assays. One involved the isolation of nuclei from tissue culture cells and subsequent preparation of a nuclear protein extract. By the other method a whole cell protein extract was made from cell lines.

Nuclear extracts were prepared from cell lines as follows. Cells were first washed twice in PBSa and then twice with lysis buffer (10mM Hepes, pH7.9, 10mM NaCl, 3mM MgCl₂ and 0.05% NP-40) on ice. Cellular lysis was confirmed by Trypan Blue staining. The lysate was resuspended in 10ml of the lysis buffer without NP-40 and twice sedimented through 10ml of 30% sucrose dissolved in the same buffer by centrifugation in a Sorvall HB4 rotor at 25000rpm and 4°C for 5 min. The nuclear pellet obtained was resuspended in 3ml of nuclear lysis buffer (20mM Hepes, pH7.9, 1.5mM MgCl₂, 25% glycerol, 0.42M NaCl, 0.2mM EDTA, 0.5mM EDTA and 0.5mM phenylmethylsulphonylfluoride, PMSF) and stirred gently at 0°C for 30 min (Wildeman et al., 1984). The nuclear lysate was centrifuged at 25000g and 4°C for 20 min, the supernatant collected and then precipitated by the addition of 0.33g/ml of ammonium sulphate with gentle stirring for 30 minutes at 0°C. The precipitate was collected by centrifugation at 25000g and 4°C for 20 min then resuspended in 1ml of dialysis buffer (20mM Hepes, pH7.9, 20mM KCl, 1mM MgCl₂, 17% glycerol, 2mM DTT and 0.5mM PMSF). The resuspension was dialysed against 100ml of

the same buffer overnight at 4°C. Aliquots were frozen in liquid nitrogen and stored at -70°C. Protein concentration was estimated using the Biorad protein assay.

Whole cell extracts were prepared by modification of the method described in Dale et al (1989). Frozen cell pellets were made of total volume 0.2ml containing 2-3 X 10⁷ cells. 1ml of ice-cold extraction buffer (10mM Hepes, pH7.9, 0.4M NaCl, 1.5mM MgCl₂, 0.1mM EGTA, 0.5mM DTT, 0.5 mM PMSF and 5% glycerol) was added to a single pellet and the cells lysed as they thawed by pipetting in and out until a homogeneous mix was formed. This mixture was then centrifuged at 29000rpm and 4°C for 25 min on a Beckman SW50.1 rotor. The supernatant was removed, half was frozen in aliquots and the other half was loaded onto a NAP-5 desalting column. The NAP-5 column was pre-equilibrated with 5ml of ice-cold binding buffer which is exactly the same as extraction buffer except that it has only 50mM NaCl. The column was then eluted with 1ml of binding buffer, the eluant aliquoted and frozen in liquid nitrogen for storage at -70°C.

2.27 Bandshift assays

Bandshift assays were performed to study the binding of potential regulatory proteins to specific DNA sequences. ³²P-labelled DNA probes were used to study the binding. The probes were either fragments of cloned DNA or double-stranded oligonucleotides.

Cloned DNA probes were generated as described in Section 4.1. Double stranded oligonucleotides were made by annealing complementary single-stranded oligonucleotides designed so as

to leave four nucleotide 5'-overhangs. 50ng of the double-stranded oligonucleotides were then 5'-end labelled as in Section 2.9 and the 5'-overhangs filled in using Klenow enzyme in the presence of 250 μ M of all four dNTPs. The labelled probes were purified by electrophoresis on a 10% polyacrylamide gel.

For bandshift assays using nuclear protein extracts the following method was adopted. 50-100 μ g of nuclear protein was pre-incubated with 5 μ g poly(dI-dC).poly(dI-dC) (Sigma) in a final volume of 40 μ l of a binding buffer (50mM KCl, 1mM EDTA, 10mM Tris.HCl, pH8.0, 1mM DTT, 5mM MgCl₂, 12.5% glycerol, 0.1% Triton X-100, Sigma) on ice for 30 min. 20000cpm of labelled DNA was added and incubated for a further 30 min on ice. Samples were electrophoresed on a non-denaturing 5% polyacrylamide gel in 0.5 X TBE for 2 hours at 150V. The gel was then dried and autoradiographed at -70°C.

Bandshift assays using whole cell extracts were carried out essentially as described in Dale et al (1989). 10 μ l of extract was pre-incubated with 5 μ g of poly(dI-dC).poly(dI-dC) for 15 min at RT. 10 μ l of another binding buffer (10mM HEPES, pH7.9, 50mM NaCl, 1.5mM MgCl₂, 0.1mM EGTA, 0.5mM DTT, 0.5mM PMSF and 5% glycerol), 10 μ g of BSA, 10-20000cpm of probe, Ficoll to a final concentration of 4%, and competitor DNA, if used, were then added. The volume was adjusted to 40 μ l with DDW. The mixture was incubated for a further 15 min at RT. The samples were then electrophoresed on a 5% non denaturing polyacrylamide gel in 0.5 X TBE at 150V for 2 hours. The gel was then dried for 30 minutes and autoradiographed at -70°C.

CHAPTER THREE

ISOLATION AND CHARACTERISATION OF THE 5' END OF THE HUMAN CARBONIC ANHYDRASE I GENE

In order to study directly the structure and regulation of the HCAI gene, the first step was to obtain HCAI genomic clones. Initially a fragment of the HCAI cDNA (Barlow et al., 1987) was used to screen a human genomic library in λ 2001 (LeFranc et al., 1986). Two recombinants (HGCAI 5.1 & HGCAI 2.1) were isolated by this procedure. Sequence analysis of these clones showed them to contain exons 2-7 (J.H. Barlow, personal communication) as shown in Fig. 3.13. The work in this chapter describes how clones containing the first exon which contains protein coding sequence (λ 104 & H24) and subsequently clones containing the transcriptional start site and promoter region of the HCAI gene (λ 201-204) were isolated and characterised.

3.1 Isolation and mapping of λ 104 and H24 clones

Initial attempts to isolate genomic clones 5' to HGCAI 5.1 and 2.1 using a 5' fragment of the HCAI cDNA were unsuccessful. It was then decided to screen the human genomic library with an oligonucleotide complementary to the first exon of the HCAI gene containing protein coding sequence. The oligonucleotide used was a 21mer (3'-GGTCTGACCCCTATACTACTG-5') designated oligo #3, complementary to the region from codon 3 (proline) to codon 9 (aspartate), see Fig. 3.3C. The sequence was obtained from the HCAI cDNA (Barlow et al., 1987). About

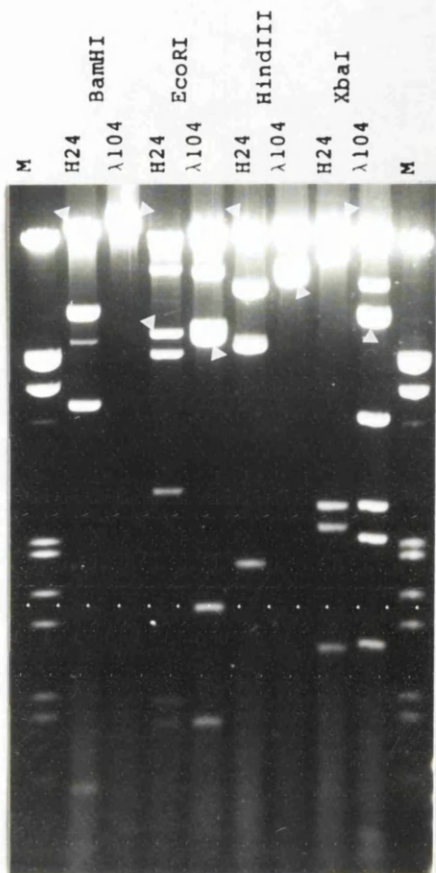
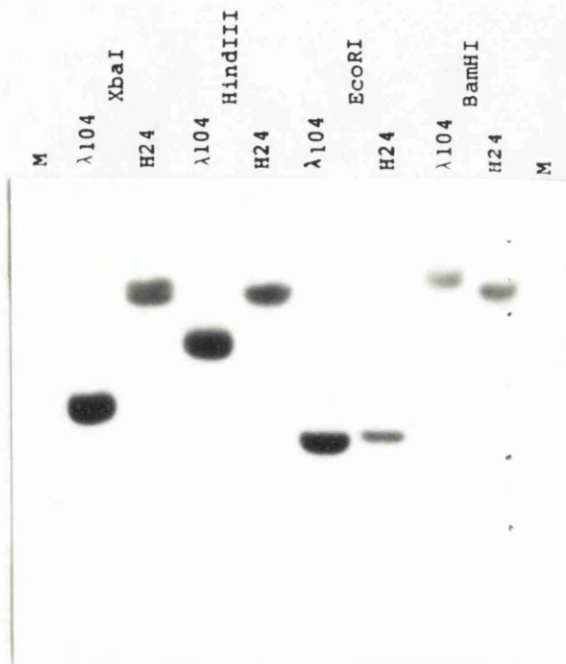
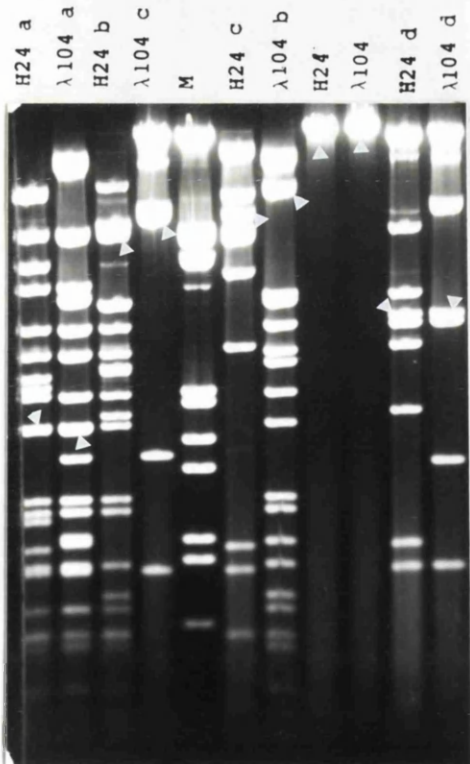
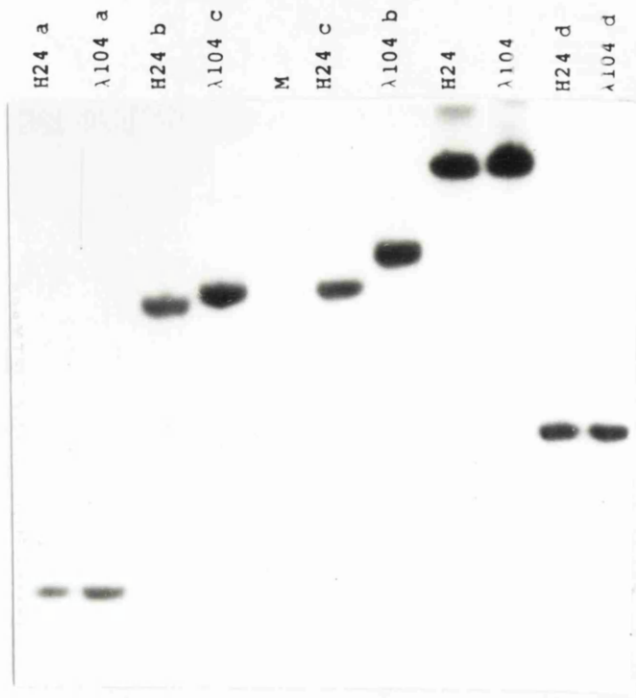
Fig. 3.1 Restriction enzyme digests of H24 and λ 104 DNA and autoradiographs of these digests probed with ^{32}P labelled oligo #3.

A BamHI, EcoRI, HindIII and XbaI digests of λ 104 and H24 DNA electrophoresed on a 0.7% agarose gel. The bands marked with Δ are those which hybridise to the ^{32}P labelled oligo #3. The marker track, M, is a HindIII-EcoRI digest of λ DNA.

B The agarose gel in Fig. 3.1A was dried down and probed with ^{32}P labelled oligo #3. The restriction fragments which hybridise to the labelled oligo contain exon 1c of the HCAI gene.

C EcoRI/PstI (a), PstI/HindIII (b), EcoRI/ BamHI (c) and (d) EcoRI/HindIII digests of λ 104 and H24 DNA electrophoresed on a 0.7% agarose gel. The bands marked with Δ are those which hybridise to the ^{32}P labelled oligo #3. The marker track, M, is as in Fig. 3.1A.

D The agarose gel in Fig. 3.1C was dried down and probed with ^{32}P labelled oligo #3. The restriction fragments which hybridise to the labelled oligo contain exon 1c of the HCAI gene.

A**B****C****D**

500,000 pfu of the human genomic library was screened with ³²P-labelled oligo #3. Four possible positive clones were picked and replated for a secondary screening. One definite positive clone emerged and was further plated at low titre and screened so that a single positive plaque could be isolated. Large scale DNA preparation of this clone, designated λ104, was then carried out to facilitate further analysis. DNA was also prepared from another clone, designated H24. This clone was a gift from Dr R.E.Tashian, Univ. of Michigan, USA. It was isolated from a human genomic library made in the Charon 4A phage vector during the cloning of the human CAII gene (Venta et al., 1984).

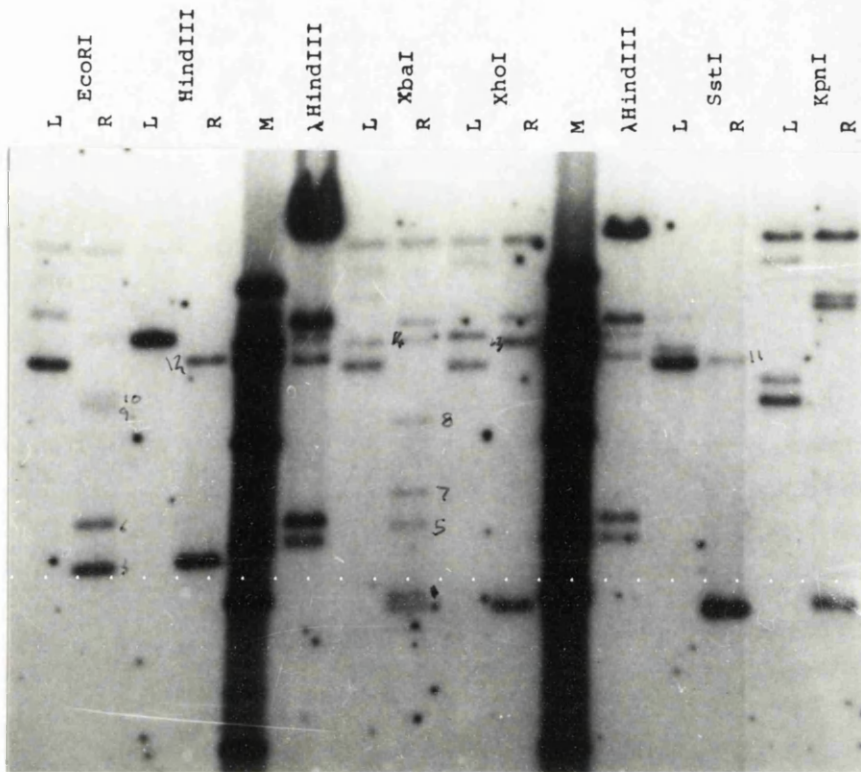
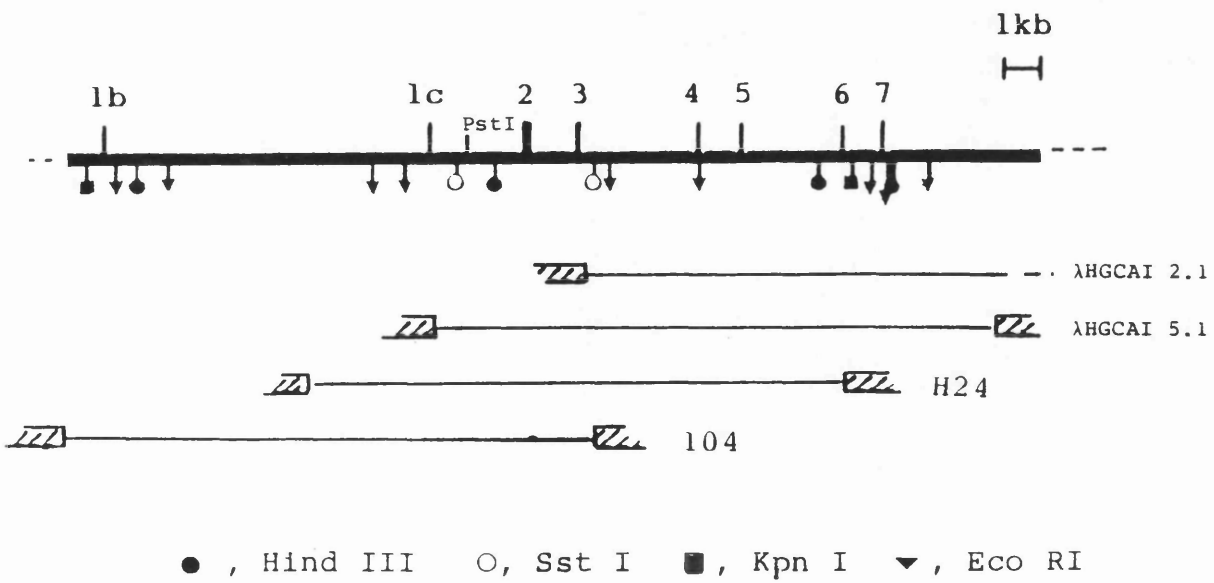
The next objective was to map the clones and identify a fragment containing the start of the coding sequence which could be subcloned and sequenced with the premise that the promoter region of the gene would be continuous with it.

Dr Tashian had provided a restriction enzyme site map of the H24 clone. The restriction map of λ104 was determined by single and double digests resolved by agarose gel electrophoresis using a series of restriction enzymes with hexanucleotide recognition sequences (Fig. 3.1A and 3.1C). The technique of "cos" mapping was also used to order the restriction fragments of the λ104 clone. "Cos" mapping allows a sequential reading of restriction sites from a single gel. This was achieved by partially digesting λ104 DNA and then annealing a ³²P end-labelled oligonucleotide complementary to either the left or right cohesive ("cos") end of λ phage (Rackwitz et al., 1984). The partial digest should produce fragments of varying length, from the largest, equivalent to

Fig. 3.2A "cos" mapping gel of λ 104 DNA

The ^{32}P labelled "cos" oligos complementary to left (L) and right (R) arms of λ DNA were hybridised to λ 104 DNA, partially digested with the restriction enzymes shown. The restriction fragments were size fractionated by electrophoresis on a 0.5% agarose gel. The marker track, M, is a mixture of ^{32}P labelled fragments of HindIII, EcoRI/HindIII and SalI digests of λ DNA.

Fig. 3.2B A map of the restriction enzyme sites in the cloned inserts of λ 104, H24 and λ HGCAI 5.1 DNA.

A**B**

the whole clone to smaller ones equivalent to sites of increasing distance from either the left or right end depending on the oligonucleotide used. When this is done with a series of restriction enzymes and run on the same gel the order in which the restriction sites come from a single end in the clone can be directly read from bottom to top of the autoradiograph of the dried gel. A "cos" map of λ 104 DNA is shown in Fig. 3.2A. To determine which restriction fragment had the protein coding sequence, gels with single and double digests of H24 and λ 104 DNA were dried and hybridised to ^{32}P labelled oligo #3. Fig. 3.1B and 3.1D show autoradiographs of gels so treated. It can be seen that in several cases for the same restriction enzyme(s) that oligo #3 hybridises to different size fragments. This is not just a consequence of the clones being in different λ phage vectors but also that H24 and λ 104 contain different regions of genomic DNA, as is seen from aligning their sequences (Fig. 3.2B). H24 extends 3' as far as exon 7 but λ 104 only extends downstream as far as exon 3 but much more 5'. The 1.8 kb EcoRI-PstI fragment shown in Fig. 3.1D (lanes 1 & 2) was the smallest hybridising to oligo #3 and was thus chosen for more detailed analysis by DNA sequencing.

3.2 Nucleotide sequence analysis of λ 104 and H24 DNA

Sequence information around the first exon of the protein coding region of λ 104 was obtained by subcloning the 1.8 kb EcoRI-PstI fragment into M13 and subsequent dideoxy chain-termination nucleotide sequencing. The fragment was isolated from an agarose gel by elution onto NA45 paper and then subcloned into double-stranded M13 mp18 and mp19 DNA. The M13

Fig.3.3A Sequencing strategy for the first exon of the HCAI protein coding region in λ 104 and H24 DNA

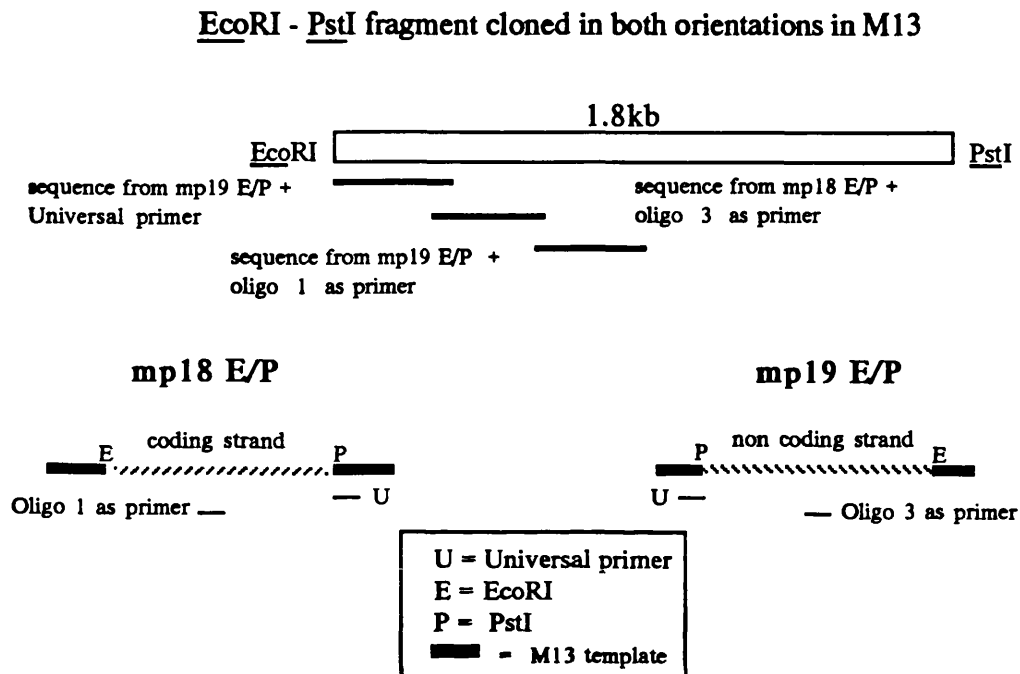


Fig. 3.3B The nucleotide sequence obtained 5' and 3' to the first exon of the protein coding region of λ 104 and H24

CAAGCAGGGATTTAAGGCATCTCCTGCATGCATGCACAGTTGC

AGTTAGTTATTCCAGGTATTATTTTTGTTTTTCAG[▼]AAAAAGAAA

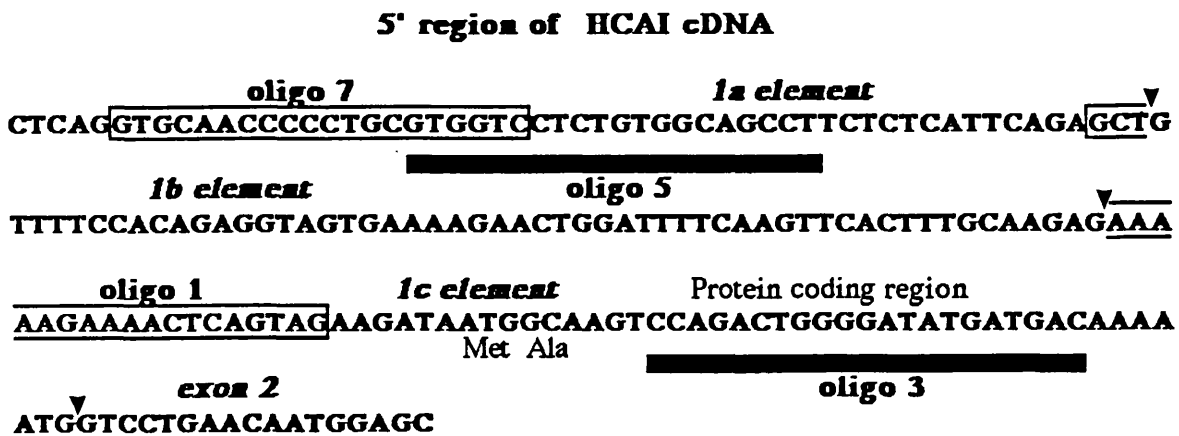
5' untranslated region Protein coding region

ACTCAGTAGAAGATAATGGCAAGTCCAGACTGGGGATATGATG

Met

ACAAAAT[▼]GGTAACACTTC

Fig. 3.3C HCAI cDNA sequence showing the origin of the oligonucleotides used as primers. The solid black lines represent oligonucleotides complementary to the coding strand of HCAI cDNA



DNA had previously been digested at unique polylinker EcoRI and PstI sites and dephosphorylated to prevent recircularisation during ligation. Single-stranded M13 DNA was prepared from clear plaques and the presence of an insert determined by agarose gel electrophoresis. The sequencing strategy then employed is shown in Fig. 3.3A. The single-stranded M13 recombinants containing either the coding strand, mp18, or the non-coding strand, mp19, were both sequenced from either end using the Universal Sequencing Primer (Amersham). The mp18 recombinant was also sequenced using oligo #3, complementary to the coding strand, as primer to read sequence 5' to the first protein coding exon. The mp19 recombinant was also sequenced using, as a primer, an oligonucleotide designated oligo #1 (5'-GCTAAAAGAAAAGTCTAG-3' - sequence from HCAI cDNA) which is complementary to the non-coding strand. This allowed the 3'-end of the exon into intron sequence to be read. Fig. 3.3A shows that the region of interest proved to be closer to the EcoRI end of the cloned fragment.

Figure 3.3B shows the sequence obtained 5' and 3' to the first protein coding exon. An exon-intron junction interrupts codon 12 (glycine) where the first G of the codon is followed by the invariant GT dinucleotide of the intronic 5' splice site (Padgett et al., 1986).

To explain the 5'-end of the exon it is first necessary to examine the 5' cDNA structure (Fig. 3.3C). Sequence analysis of a number of cDNA clones showed that three sequence elements can contribute to the 5' leader (N.Lowe, personal communication). By far the most common type of cDNA contains

elements 1a and 1c and a second rare type contains elements 1a, 1b and 1c. When this is compared to sequence obtained from λ 104 it can be seen that the genomic clone contains 1c but that, immediately 5', the sequence diverges. The first two nucleotides 5' to 1c are AG the invariant dinucleotide pair of intronic 3' splice sites (Padgett et al., 1986). This observation, plus the fact that the 1a element (and the 1b element) of the cDNA is not continuous with 1c in λ 104, suggested that the 5'-untranslated region of the HCAI gene as transcribed in reticulocytes contains at least one intron. To confirm that situation was not just a phenomenon of the λ 104 clone, the 1.8 kb EcoRI-PstI fragment from H24 was also subcloned into M13mp18 and mp19 then sequenced as above. The sequence obtained was identical to that for λ 104.

An oligonucleotide specific to the 1a cDNA element was made, as this region was now presumed to be adjacent to the 5' promoter of the gene. The oligonucleotide sequence was (3'-CACCCAGGAGACACCGTCGGA-5') and designated oligo #5 (Fig. 3.3C). To confirm that 1a really was part of the HCAI gene transcript and not a cloning artefact, oligo #5 was ^{32}P end-labelled and hybridised to RNA from two different patients with high reticulocyte counts (suffering pyruvate kinase deficiency) and RNA from the umbilical cord of a healthy newborn baby. The oligo hybridised to the same sized transcript, approx. 1.5 kb, in all three samples (Fig. 3.4A). To determine if λ 104 contained the 1a element, a dried down gel of λ 104 DNA digests was hybridised to ^{32}P -labelled oligo #5. No hybridisation was detected.

An oligonucleotide specific for element 1b of the cDNA

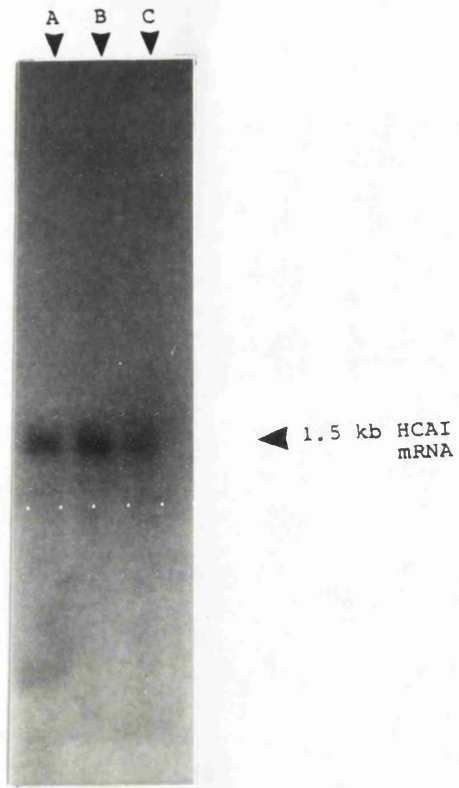
Fig. 3.4A Northern analysis of erythroid RNAs using ³²P
labelled oligo #5

Each track contains 20µg of total human RNA. Tracks A and B are reticulocyte RNA from two different patients with reticulocytosis due to pyruvate kinase deficiency. Track C contains cord blood RNA from a healthy newborn baby. The RNAs were probed with ³²P end-labelled oligo #5, specific for exon 1a.

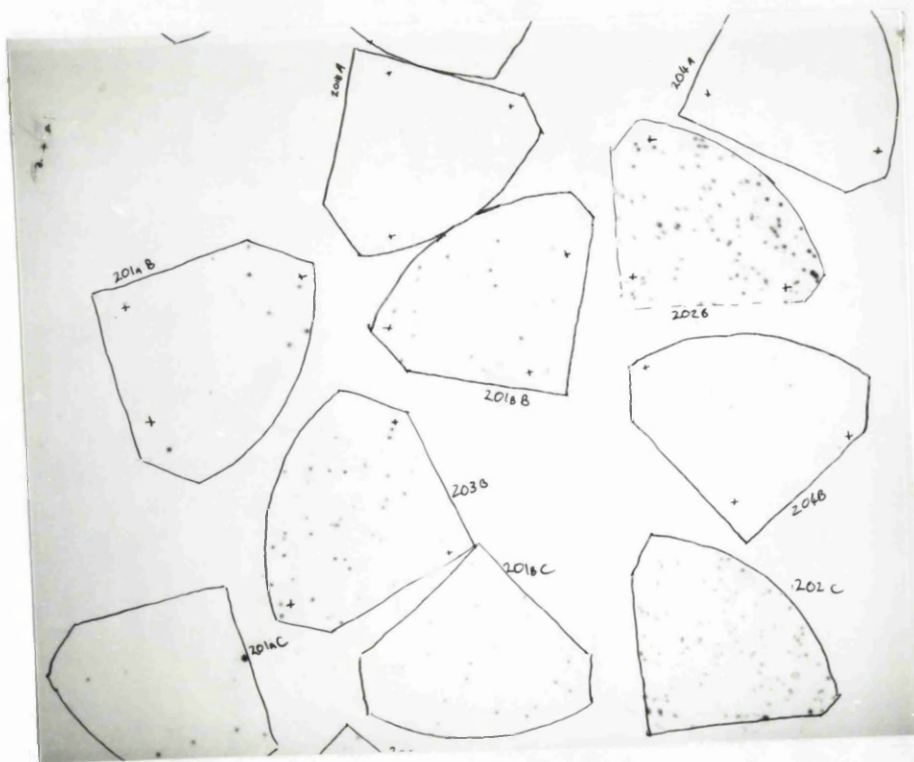
Fig. 3.4B Tertiary screen of clones isolated from the λ2001
human genomic library probed with ³²P labelled
oligo #5

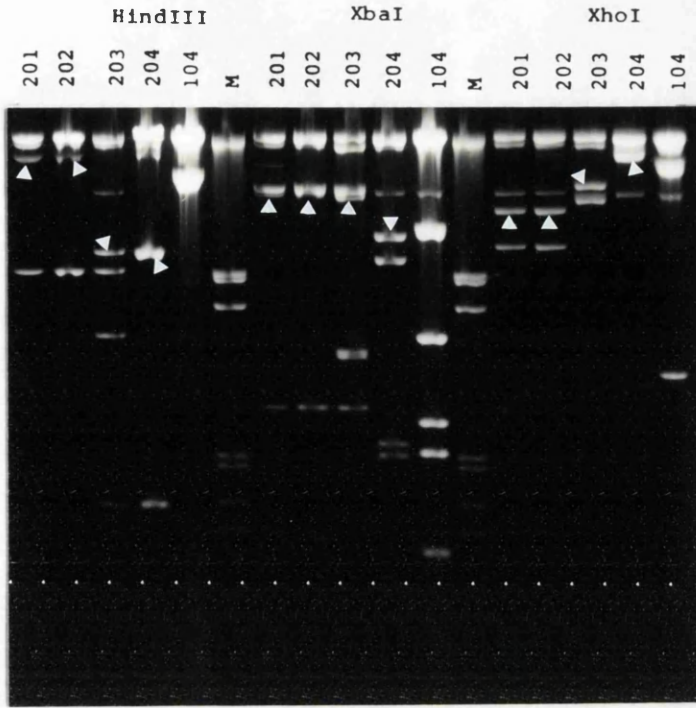
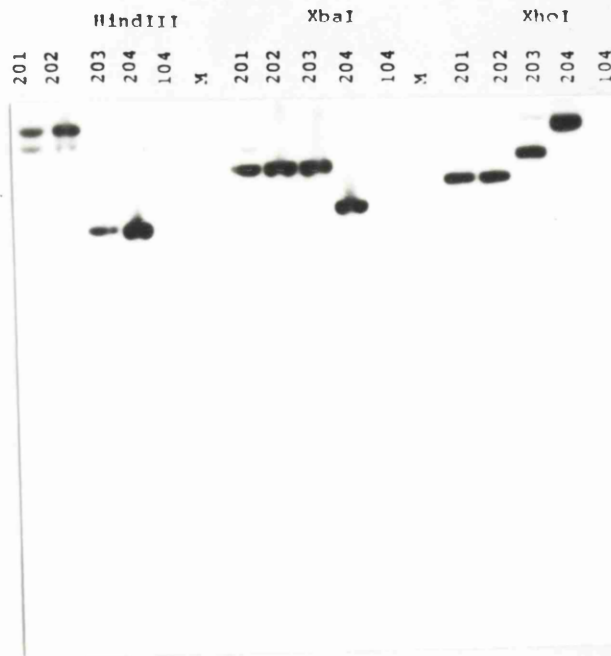
A tertiary screen of the clones λ201-204 isolated from the λ2001 library. This screening allowed the isolation of single discrete plaques of each clone.

A

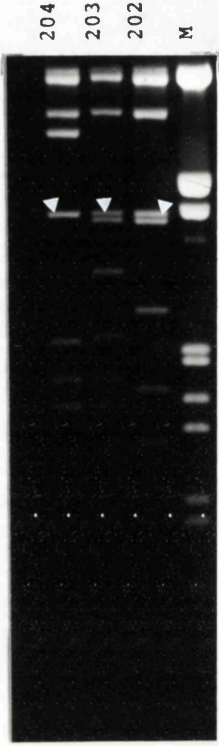


B



A**B**

C



D



was synthesised. It was ^{32}P -labelled and hybridised to λ 104 DNA digests. This 1b oligo hybridises to a KpnI-EcoRI fragment at the 5'-end of the λ 104 clone (J.C.Sowden, personal communication). No functional significance has been assigned to the very rare appearance of 1b in the HCAI gene transcript.

3.3 Isolation, mapping and nucleotide sequencing of λ 201-204 clones

^{32}P -labelled oligo #5 was used to screen the λ 2001 human genomic library. Following secondary and tertiary screenings 4 single positive clones were isolated (Fig. 3.4B). These clones were designated λ 201-204. Following the same procedures as in Section 3.1 restriction maps of the clones were drawn up using data from single and double restriction enzyme digests and "cos" mapping with the same enzymes (Fig. 3.5A, 3.5C & 3.6A). The restriction fragment pattern of λ 201 and λ 202 showed them to be identical. The gels containing digested DNA from the clones were dried and hybridised with ^{32}P -labelled oligo #5 (Fig. 3.5B & 3.5D). This allowed restriction fragments containing the 1a element to be identified and one to be selected for further analysis. The fragment selected was a 4 kb SstI-HindIII fragment present in all four clones which was purified from an agarose gel containing λ 204 DNA digested with those enzymes. The fragment was subcloned into Bluescript (KS+) plasmid which had been digested with SstI and HindIII then dephosphorylated with CIP to prevent recircularisation during ligation. The recombinant plasmid formed, designated pBSc204SH was analysed to confirm its identity and then prepared on a large scale for further analysis. Single and

Fig. 3.6A "cos" mapping gel of λ 200 series clones 202 & 203

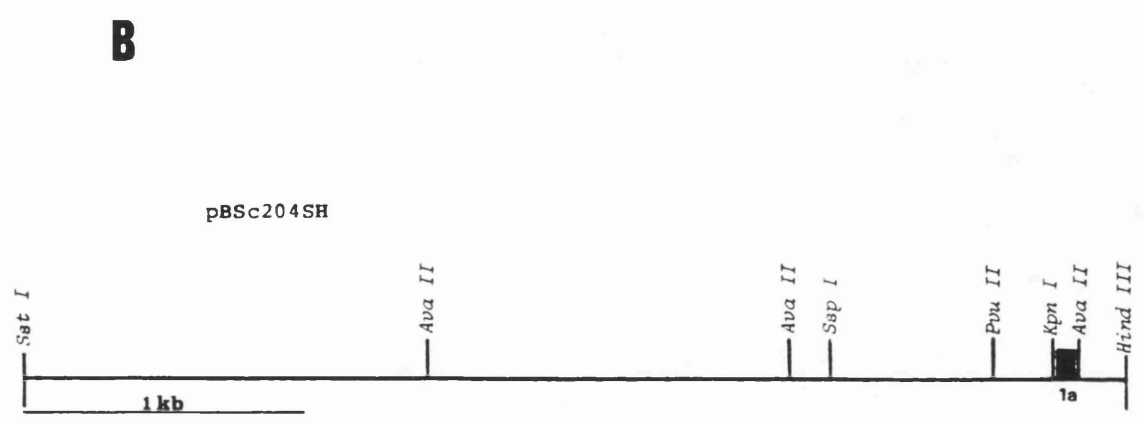
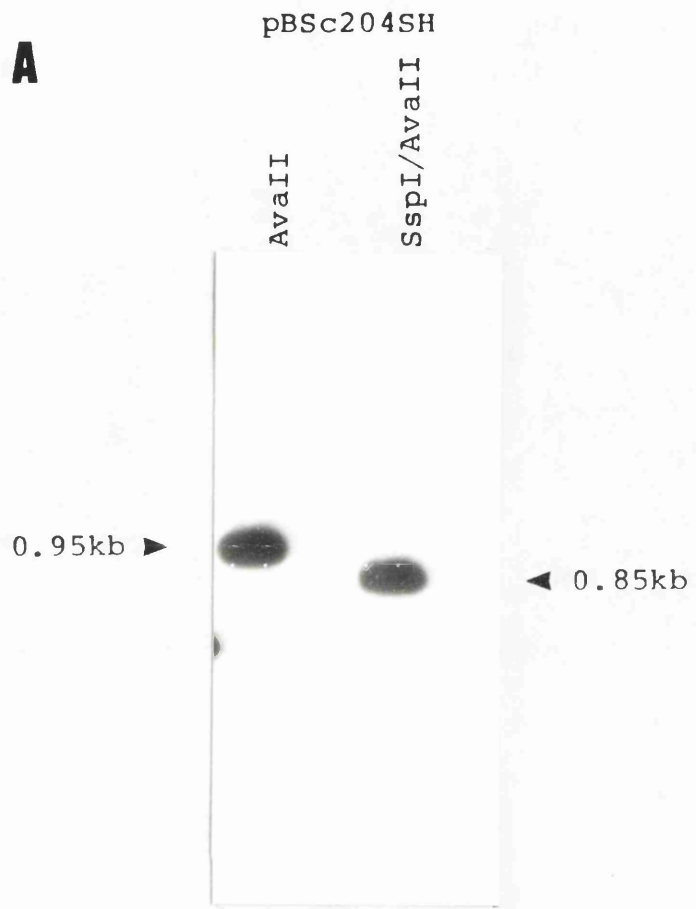
In the gel shown, a ^{32}P end-labelled "cos" oligo complementary to the left (L) arm of λ DNA was hybridised to λ 202 and λ 203 DNA and the ^{32}P end-labelled "cos" oligo complementary to the right (R) arm of λ DNA hybridised to λ 202 DNA. The λ DNA was partially digested with the restriction enzymes shown and the fragments resolved by electrophoresis on a 0.5% agarose gel. The marker track, M, is as in Fig. 3.2A.

Fig. 3.6B A map of the restriction enzyme sites in the cloned inserts of λ 201-204 DNA

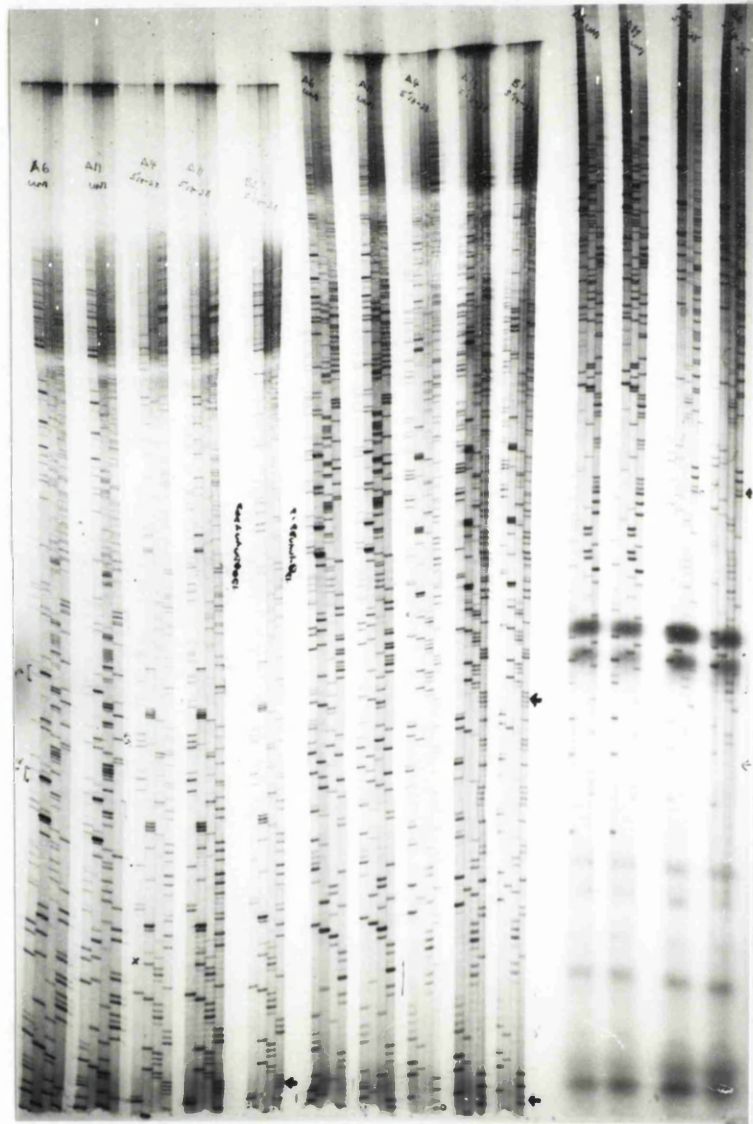
Fig. 3.7A Probing of pBSc204SH restriction enzyme digests with ³²P labelled oligo #7

Restriction enzyme digests of pBSc204SH were electrophoresed on a 0.8% agarose gel and then the gel dried. This was probed with ³²P end-labelled oligo #7. A 0.95kb fragment hybridised in the AvaII digest and a 0.85kb fragment in the SspI/AvaII digest.

Fig. 3.7B Restriction map of the SstI/HindIII insert in pBSc204SH



B



double restriction enzyme digests of the plasmid were carried out to map the insert. These gels were then probed with ³²P-labelled oligonucleotide specific for 1a to position the element within the insert. The oligo used for this, oligo #7, had the sequence (5'-GTGCAACCCCCTGCGTGGTC-3') and was complementary to the non-coding strand of the cDNA. The reason this oligonucleotide was used instead of oligo #5 was because the AvaII sites within the insert were quite useful for mapping purposes. Oligo #5 has the whole AvaII recognition site (5' GGTCC 3') within it and so could not efficiently hybridise to AvaII digested DNA. The 1a element was located within the insert as shown in Fig. 3.7A and 3.7B.

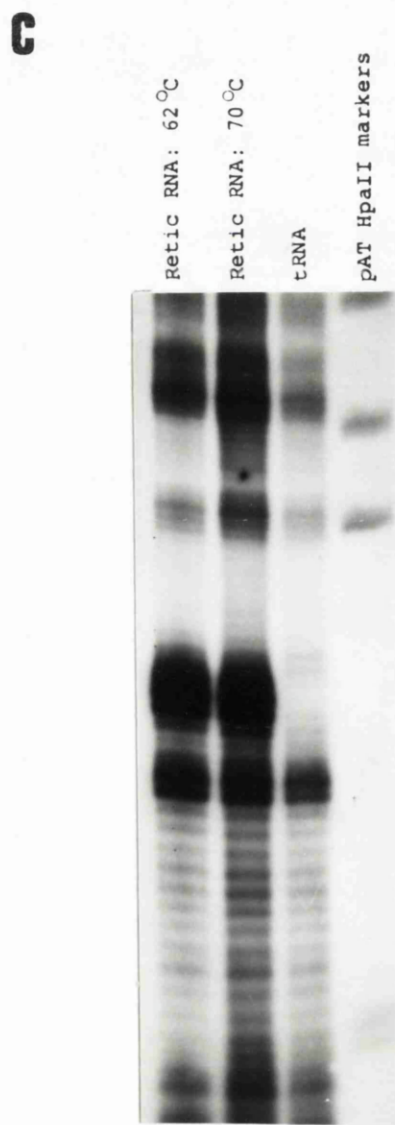
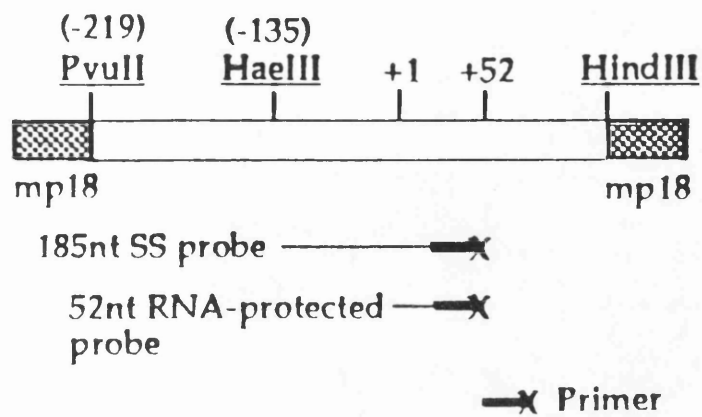
The strategy applied to obtain the nucleotide sequence of the genomic DNA flanking the 1a element is shown in Fig. 3.8A. The PvuII-HindIII fragment was subcloned into HincII-HindIII-digested and dephosphorylated M13 mp18, the single-stranded recombinant containing the coding strand, and sequenced using oligo #5 and the Universal Sequencing Primer. The same fragment was also subcloned into similarly treated M13 mp19, the single stranded recombinant containing the non-coding strand, and sequenced using the Universal primer and oligo #7. The SspI-PvuII fragment was subcloned into SmaI digested and dephosphorylated mp19 in both orientations and sequenced from either end using the Universal primer. The 1.3 kb KpnI fragment, isolated from a digest of λ 204, was subcloned into KpnI digested and dephosphoylated M13 mp19 in both orientations and sequenced from either end using the Universal primer. An example of a typical sequencing gel of these recombinants is shown in Fig. 3.8B.

3.4 Mapping of the transcription start site of HCAI

Initial scrutiny of the nucleotide sequence obtained (Fig. 3.11) shows that all of the 1a element found in the cDNA sequence is present in the HCAI genomic sequence. From this it could be presumed that the sequence 5' to this must contain the HCAI gene leader sequence containing various elements involved in the transcriptional regulation of the gene. In order to identify these elements it was important to establish the exact location of the transcription start site. This was achieved by a combination of S1 mapping and primer extension analysis on adult human reticulocyte RNA.

An end-labelled single-stranded DNA probe was generated from the M13 mp18 PvuII-HindIII template using oligo #5 as primer (Section 2.18) as shown in Fig. 3.9A. After isolation from a 6.5% polyacrylamide gel, labelled single-stranded probe was hybridised with 15µg of total human reticulocyte RNA for 3 hours at either 62°C or 70°C followed by S1 nuclease digestion. The protected fragments were resolved on a 7% polyacrylamide / 50% urea gel. The results of this experiment are shown in Fig. 3.9B. The 185 nt single-stranded probe is shown in lane 1. Lanes 2 & 3 show the S1 nuclease protected fragments after hybridisation at 62°C and 70°C respectively. A 52 bp protected fragment is indicated in both tracks. The size of 52 bp is measured relative to ³²P end-labelled pAT HpaII markers run alongside. It was precisely determined by overexposure of the gel (Fig. 3.9C). This shows a single bp step ladder due to "nibbling" of the RNA/DNA hybrid by the S1 nuclease (Hentschel et al., 1980) and allows the position of each bp between the

A S1 map of HCAI 5' end



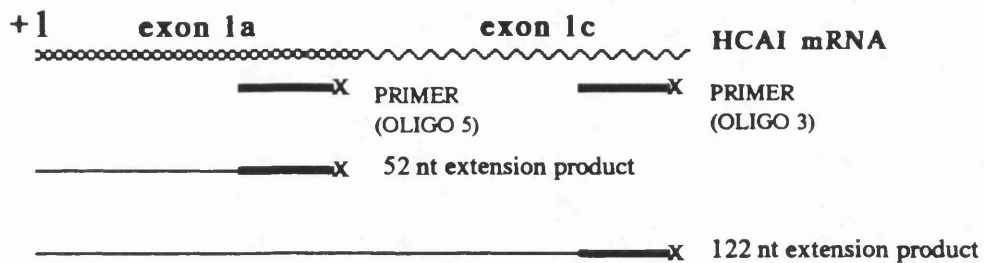
67 and 34 bp fragments of the pAT HpaII markers to be seen. No protection is visible in lane 4 which has 15µg of tRNA hybridised with the single-stranded probe under the same conditions as for lane 2. The tRNA acts as a control to indicate any non-specific protection.

The S1 mapping suggests that the transcription start site is at a G residue, 67 bp from the 3'-end of the 1a element.

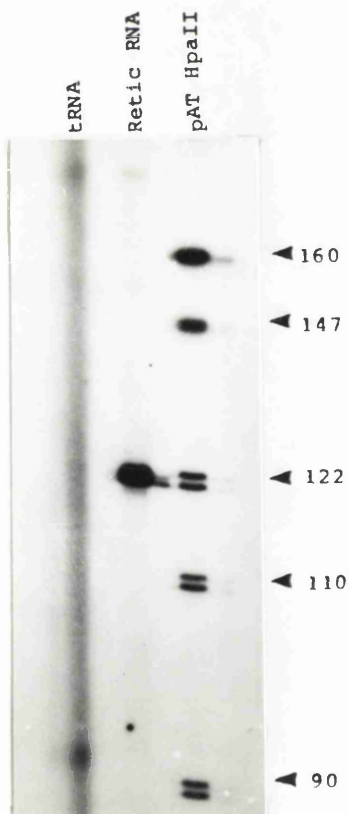
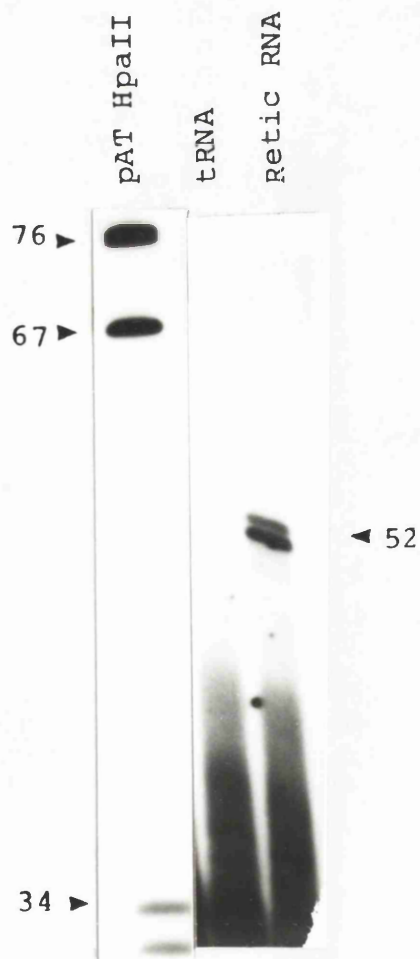
In order to confirm the placing of the initiation site, the transcript was also mapped by primer extension. Two different primers were used, oligo #5 which lies within the 1a element and oligo #3 which lies within the 1c element (Fig. 3.10A). The ³²P end-labelled oligonucleotides were annealed to 10-20µg of total human reticulocyte RNA and extended to form a cDNA/RNA hybrid with reverse transcriptase in the presence of all four dNTPs as described in Section 2.17. The extension products were resolved by electrophoresis on denaturing polyacrylamide gels. The results are shown in Fig. 3.10B and 3.10C. Oligo #3 gives a 122 nt extension product. From the 5'-end of oligo #3 to the 5'-end of 1c is 54 nt which, since in the vast majority of transcripts 1a and 1c are continuous, means that 1a is 68 nt long. This correlates exactly with the S1 nuclease mapping data. Oligo #5 gives a 52 nt extension product which again confirms the transcription start site to be the G residue, accordingly designated +1 in Fig. 3.11.

In the rest of this thesis all numbering of base positions in the 5'-end of the HCAI gene will be relative to this position, with all upstream sequences carrying a (-) prefix.

The structure of the 5'-leader of the HCAI gene is now

A

(x denotes ^{32}P labelling)

B**C**

apparent (Fig. 3.11). It is divided into what are now referred to as exons 1a and 1c. We now know that these are separated by a 36.5 kb intron (Fig. 3.13B) which contains the 1b element which very infrequently appears in the HCAI transcript. The 201-204 clones do not overlap the λ 104 clone therefore further cloning work was carried to fill the gap with overlapping clones as shown in Figure 3.13B (N.Lowe, personal communication).

3.5 Nucleotide sequence analysis of the HCAI gene 5' flanking region

Having defined the transcription start site of the HCAI gene, it was then possible to make a detailed analysis of the region upstream for consensus sequences of cis-acting elements known to be involved in the transcriptional regulation of other eukaryotic genes (as discussed in Section 1.1). The sequences identified are shown in Fig. 3.11.

A TATA box motif is present at -28 [5'-CATAAG-3']. The consensus sequence for TATA, in those eukaryotic genes which have such a motif, is [5'-TATA^A/_TA-3'] (Brethnach & Chambon, 1981). However, in the TATA box of the erythroid-specific globin genes, the first T residue is usually a C e.g. in the human β globin gene at -32 [5'-CATAAA-3'] (Konkel et al., 1979), in the mouse β^{maj} globin gene at -32 [5'-CATATA-3'] (Lawn et al., 1980) or in the human θ globin gene [5'-CATATA-3'] (Leung et al., 1987).

Another motif in the 5'-flanking region of eukaryotic genes is the CCAAT box. This is based on the sequence [5'-CCAAT-3'] but variations within this are possible and the


```

-360 TGAGTCAAATCGACATCAAGATATAAAGGAATCAGTGACTAAATATATTTTCATATATGGTATTTTATTGATTATTGGTCTTGACCTAGTATGG
      TaqI
      AP-1
      GFI(A)
-260 AGGCCTGGCTAGAGGCTGGTCAGTTTCCTCTCTTGAGCAGCTGATTAATCCACACCCCAACCACCTCCCTTATCAGGTTCTCACACTCTGGGGCCACT
      'CACCC'
      GFI(B)
      PvuII
      SphI GFI(C)
-160 ATGTACCCACTCTAATCACCCACAGGGCCAGACATCAGACAATTAAGGACAGGGCCCGTCCCGCCCAAGCCCGCCAAAATTATGCCAAATTATCAAATAATTAT
      RsaI HaeIII HaeII +1
      'CACCC' 'TATA'
- 60 TCAACCTAGCTAACCCCAACCCTTTTGGCTGTACATAAGCTGCCCATTCGCCCTCCAGGCTGTGGTACCCTCCTCAGGTGCAACCCCTGCGTGGTCCT
      DdeI

```

Fig. 3.11 Diagram of the 5'-region flanking the transcription start site. Restriction enzyme sites mentioned in the text and consensus sequences for trans-acting factor binding sites are shown.

region from -90 to -60 has three potential CCAAT boxes (Fig. 3.11). Covering one of these sequences at -81 is an OTF-1, octamer binding factor, motif [5'-ATGCAAAT-3'] (Fletcher et al., 1987). OTF-1 is a ubiquitous factor which has been shown to enhance transcription when binding to its recognition site in eukaryotic promoters (Sive et al., 1986). Another factor recognises the same sequence, OTF-2, but this has been shown to be lymphoid-specific (Scheidreit et al., 1987). Very close upstream of the OTF-1 consensus site (at -93) is an Sp-1 binding site consensus sequence [5'-CCCGCC-3']. The most common Sp-1 consensus is [5'-CCGCC-3'] (Briggs et al., 1986). However, the sequence at -93 of the HCAI gene has been shown to bind Sp-1 from HeLa cell extracts when present in the 5'-flanking region of the HSV-1 intermediate early gene 3 (Jones & Tjian, 1985). Binding of the ubiquitous Sp-1 to the 5'-flanking region of genes has been shown to upregulate their transcription (Kadonaga et al., 1988).

The 5'-flanking region of HCAI also contains two putative CACCC boxes, at -209 [5'-CACCC-3'] and -44 [5'-CACCC-3']. It has been suggested that CACCC box of globin genes binds an erythroid-specific factor (Mantovani et al., 1988) but this factor has now been shown to be ubiquitous (de Boer et al., 1988; Plumb et al., 1989; Schule et al., 1988a). At -324 there is an AP-1 factor binding consensus sequence [5'-GTGACTAA-3'] (Lee et al., 1987). Most viral and cellular genes that are upregulated by AP-1 binding, such as human metallothionein II_A and the SV40 early promoter, are also induced by treating cells with phorbol ester tumour promoters such as TPA (12-O-tetradecanoyl-13-acetate) as described in Lee et al (1987).

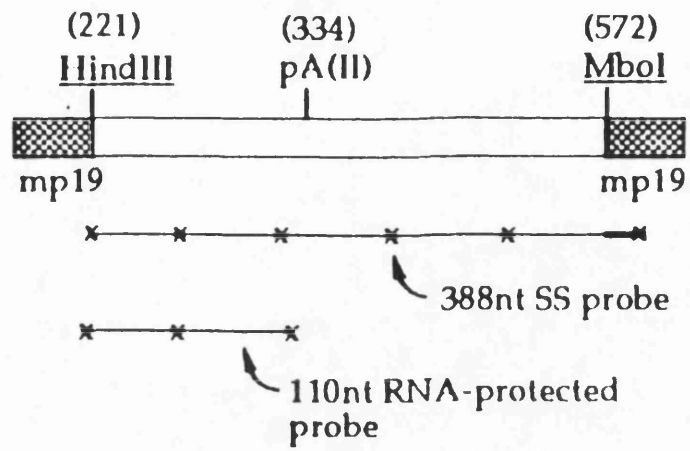
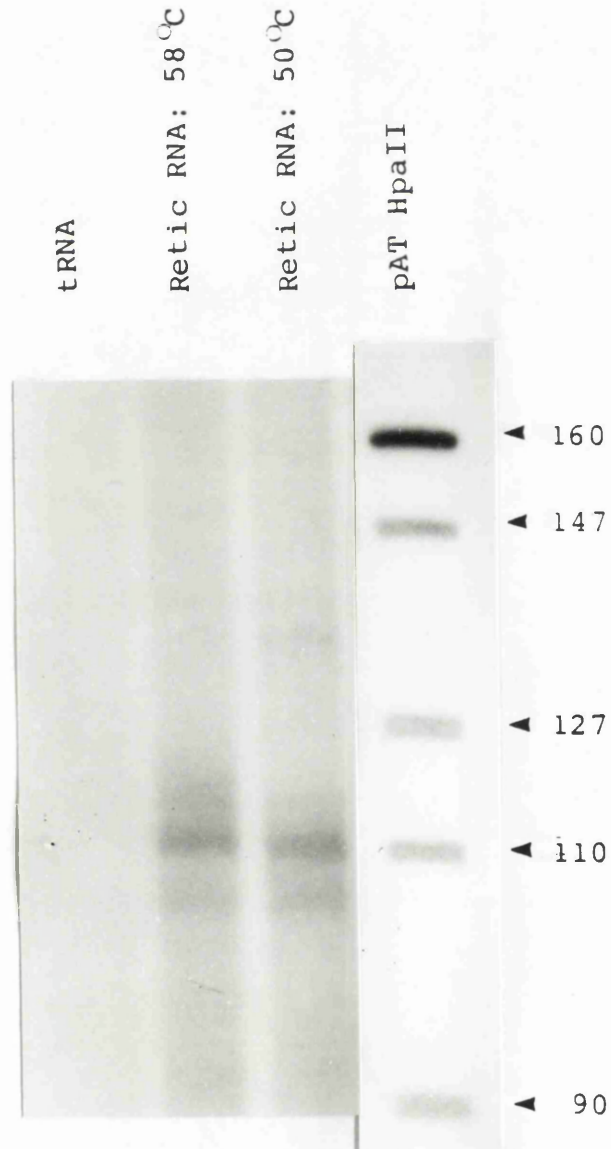
As described previously (Section 1.6) the HCAI gene is highly expressed in erythroid cells but in few other tissues apart from gastrointestinal mucosa and the some epithelial tissue. Its high level of expression in erythroid cells would suggest that the regulation of its transcription would have some common features with erythroid-specific genes such as the globins. At the time when the HCAI gene sequence became available, data was beginning to be published suggesting the presence of an erythroid-specific binding factor recognising a particular consensus sequence within the 5'-flanking regions of the chicken α^D -globin gene (Kemper et al., 1987) and chicken adult β -globin gene (Evans et al., 1988). In both of these cases, the consensus was [5'-^A/_TGATA^A/_G-3']. Based on this, a potential erythroid factor binding site was identified at -190 of the HCAI gene, [5'-CTTATCA-3'] or [5'-TGATAAG-3'] in the opposite orientation. Further work on the erythroid-specific factor has revealed that it has a role in transcriptional regulation of erythroid expression (Evans et al., 1988, deBoer et al., 1988 & Plumb et al., 1989) and recognises as consensus either [5'-GATAAG-3'] or [5'-CTTATC-3'] (Plumb et al., 1989) or [5'-^A/_CPy^T/_AATC^A/_TPy-3'] / [5'-Pu^T/_AGAT^A/_TPu^T/_G-3'] (Wall et al., 1988). The latter consensus allowed the identification of 2 other potential erythroid factor binding sites, [5'-TGATTAT-3'] at -290 and [5'-CTAATCA-3'] at -149 of the HCAI gene.

3.6 Mapping of the 3' end of the HCAI transcription unit and nucleotide sequence analysis

Having defined the 5' transcription start S1 nuclease mapping was used to define the 3' maturation point and site of

A

S1 map of HCAI 3' end

**B**

poly(A) addition. A uniformly labelled single-stranded DNA probe was generated from an M13 template which consisted of a HindIII-MboI fragment containing the 3'-untranslated region of the HCAI gene (Fig. 3.13B) subcloned into M13 mp19. The single-stranded recombinant has the anti-sense DNA strand from which the probe was generated using the Universal sequencing primer, as shown in Fig. 3.12A (and described in Section 2.18). After isolation from a 6% polyacrylamide gel, labelled probe was hybridised to 9µg of total human reticulocyte RNA for 1 hour at 50°C or 58°C followed by S1 nuclease digestion. Lower temperatures of hybridisation were used than for the previous S1 mapping experiment due to increased AT content of the probe. The protected fragments were resolved on a 7% polyacrylamide / 50% urea gel. The results of this experiment are shown in Fig. 3.12B. The 388 nt single-stranded probe is shown in Lane 1. Lanes 2 and 3 show the S1 nuclease protected fragments after hybridisation at 50°C and 58°C respectively. A 110 bp protected fragment is indicated in both tracks. No protection is seen in lane 4 which shows the probe hybridised with tRNA as a control.

The HCAI gene has 2 polyadenylation sites. The most proximal to the coding sequence is described in Barlow et al., 1987. It is designated pA(I) (Fig. 3.13A) and was identified from cDNA cloning. However, of 8 cDNA clones isolated containing the 3'-end, only one used this polyadenylation site, the others using a more distal site (N.Lowe, personal communication). The latter is the site mapped above and designated pA(II) (Fig. 3.13A). This site is 334 bp from the end of the final protein coding codon and 18 bp downstream

Fig. 3.13A Diagram of the region flanking the 3' end of the HCAI gene.

The two polyadenylation signal and addition sites are shown, as well as the factor binding consensus sequences.

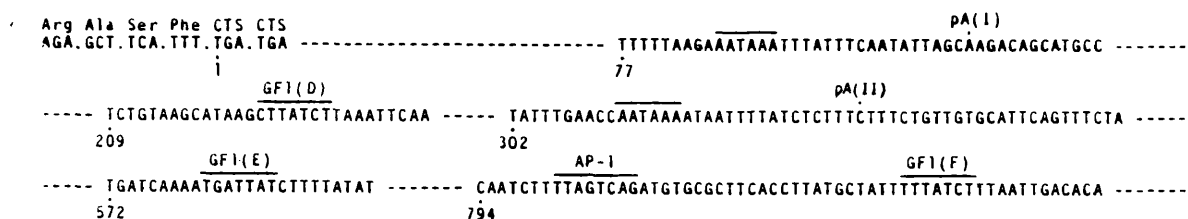
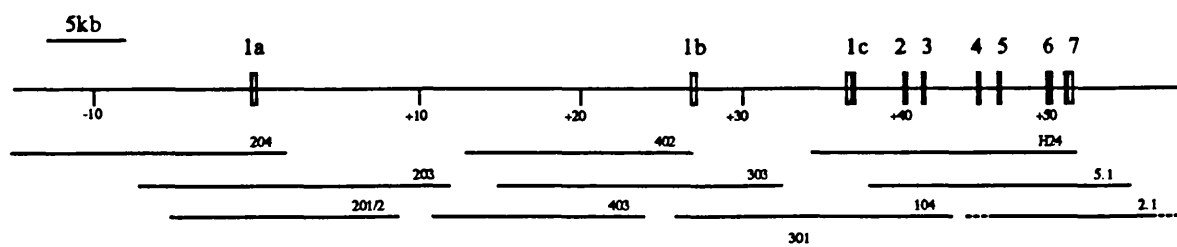


Fig.3.13B Diagram of the human carbonic anhydrase I gene.

The scale is numbered in kb from the transcription start site. Exons are indicated by black bars (coding sequence) or open boxes (non coding sequence) and are not drawn to scale due to their small size. The annotated lines below indicate the extent of the cloned regions in the various lambda recombinants isolated.



from an AATAAA hexamer that acts as a polyadenylation signal in most non-histone eukaryotic mRNAs (Birnstiel et al., 1985).

Scrutiny of the DNA sequence 3' to pA(II) revealed further consensus sequences for factor binding. In order to position sequences at the 3'-end of the gene the first base of the termination codon at the end of the protein coding sequence is numbered 1. Therefore, pA(I) is 109 and pA(II) is 334. There is an AP-1 factor binding consensus at 801 [5'-TTAGTCAG-3'] ([5'-CTGACTAA-3'] in the opposite orientation). There are also two potential erythroid-specific factor binding sites. One is at 581 where there are in fact two possible sites on either strand overlapping one another [5'-TGATTATCT-3'] and [5'-AGATAATCA-3']. The other site is at 833 [5'-TTTATCT-3']. Inspection of the sequence between the two polyadenylation sites revealed a further possible erythroid factor binding site at 223 [5'-CTTATCT-3'].

Fig. 3.13B shows the regions discussed above within the framework of the complex structure of the HCAI gene.

This chapter has described the cloning and characterisation of the areas of the HCAI gene containing the transcription and translation start sites. The structure of the 5'-end of the gene is complicated by the presence of a 36.5 kb intron within the 5'-untranslated region. The 5' and 3' limits of HCAI mRNA have been mapped by S1 nuclease and primer extension analysis. Various consensus sequences recognised by DNA binding factors, which have been shown to play a direct role in transcriptional upregulation of other

genes, have been identified. Of particular interest, in view of the erythroid upregulation of HCAI expression, is the presence of potential erythroid-specific factor binding sites. These will be investigated in the next chapter.

CHAPTER FOUR

INVESTIGATION OF THE BINDING OF AN ERYTHROID-SPECIFIC FACTOR TO REGIONS FLANKING THE HCAI GENE AND ITS ROLE AS A POSSIBLE TRANSCRIPTION FACTOR

Having identified sequences around the HCAI gene that might bind an erythroid-specific factor the next objective was to discover whether they would. This was done using bandshift (also called gel retardation) assays (Fried & Crothers, 1981). This technique works by first mixing a labelled DNA probe with a nuclear or whole cell extract. The mixture, which includes poly (dI.dC).poly(dI.dC) to bind protein that otherwise would non specifically bind to the probe DNA, is then run on a native 5% polyacrylamide gel. Probe DNA which has bound protein has a retarded mobility in the gel compared to free DNA and this is visualised by autoradiography.

4.1 Origin and preparation of the DNA probes used in bandshift assays

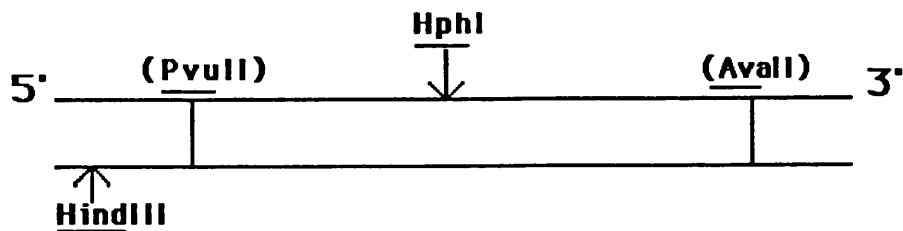
Two types of probe were used in the bandshift assays. These were either ³²P end-labelled double stranded oligonucleotides or similarly labelled fragments of cloned DNA. All the consensus sequences noted in Sections 3.5 and 3.6 were investigated. The three sites at the 5'-end of the HCAI gene were labelled (A), (B) and (C) (Fig. 3.11) and the three sites at the 3'-end (D), (E) and (F). Double-stranded oligonucleotides were synthesised covering all these sites except (D). The anti-sense strand of these oligonucleotides is

Fig. 4.1 DNA probes used in bandshift assays

A The oligonucleotides used in bandshifts. These were made as complementary single-stranded sequences and annealed before use. The anti-sense strand of the double stranded DNA is shown below. The underlined sequences are consensus binding sites.

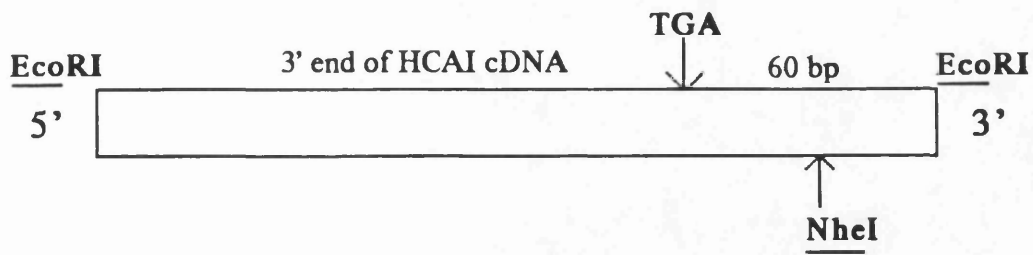
ag2 (-197) 5'-GATCCGGGCAACTGATAAGGATTCCCAGATC-3' (-167) Ma, globin
 CACCC (-219) 5'-CTGATTAAATCCACACCCCA-3' (-196) HCAI
 A (-300) 5'-GTATTTTATIGATTATTGTGCTG-3' (-277) HCAI
 B (-199) 5'-ACCACTTCCCITATCAGGTTCTC-3' (-177) HCAI
 C (-155) 5'-CCCACTCTAATCACCACAGGGCCA-3' (-132) HCAI
 E (572) 5'-TGATCAAATGATTATCITTTATAT-3' (596) HCAI
 F (808) 5'-CTATTTTITATCITTAATTGACACA-3' (831) HCAI

B Diagram of 5'-HphI fragment and the NheI-EcoRI fragment containing GF-1 site (D)



Sequence of the 5'-HphI fragment

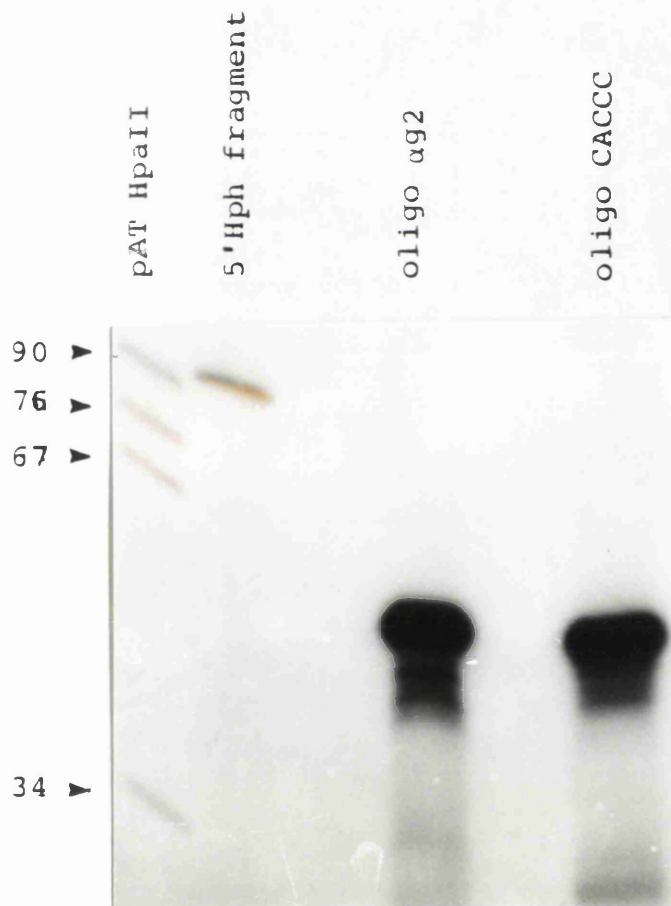
5'-ACTATAGCTTAAGGACGTCGGGCTGATTAAATCCACACCCCAACCACTTCCITATCAGGTT
 CTCACACTCTGGGGCCACTATGTACCC-3'



Sequence of the NheI-EcoRI fragment containing GF-1 site (D)

5'-CTAGCAAATAGTAATCTGTAAGCATAAGCTTATCITAAATTCAAGTTTAGTTTGAGG-3'

C An example of a polyacrylamide gel for the purification of ³²P labelled probes used in bandshift assays



shown in Fig. 4.1A. Two other double-stranded oligonucleotides were also used, the "CACCC" oligo, spanning -219 to -196 of the HCAI gene, and oligo ag2 which was a gift from Dr.M.Plumb, Beatson Institute, Glasgow, which covered from -197 to -167 of the mouse α 1-globin gene (Plumb et al., 1989). A labelled DNA fragment containing site (D) was generated by excision of an EcoRI fragment (Fig. 4.1B) from a 3' HCAI cDNA clone. The fragment was eluted onto NA45 paper then purified by phenol/chloroform and chloroform extractions followed by ethanol precipitation. The purified fragment was then dephosphorylated with CIP, kinased with ^{32}P - γ -ATP and T4 PNK and finally digested with NheI. This produced two fragments, a large one and a much smaller 60 bp one, which were resolved by electrophoresis on a 8% polyacrylamide gel. The 60 bp NheI-EcoRI band was then eluted and purified as in Section 2.18. The sequence of this fragment is shown in Fig. 4.1B.

The first bandshift experiments were carried out using a fragment of cloned DNA from the 5'-end of the HCAI gene containing the (B) site. This was prepared from a plasmid containing the 258 bp PvuII-AvaII fragment from -219 to +38 of the HCAI gene subcloned into the SmaI site of a Bluescript plasmid (KS+). The plasmid was first digested at a unique HindIII site, dephosphorylated with CIP and then kinased as above. Secondary digestion with HphI produced the probe to be used (Fig. 4.1B) which was purified from a 8% polyacrylamide gel as above. This fragment is referred to as the "5'-Hph fragment". Its sequence is shown in Fig. 4.1B.

4.2 Bandshift assays using nuclear protein extracts

Initially bandshift assays were carried out on protein extracts of the isolated nuclei of tissue culture cells as described in Sections 2.26 and 2.27. Fig. 4.2A shows the results of a bandshift on nuclear protein extracts from the HEL and MEL C88 erythroid cell lines using the "5' Hph fragment" as probe. Lane 1 shows the probe on its own without any protein. Lane 2 shows the effect of adding HEL cell nuclear protein. Two retarded bands are seen. When 100ng of unlabelled ag2 oligo was added, both the retarded bands disappear (lane 3). This is a competition effect. The two retarded bands are due to protein binding to the 5' Hph fragment which can be specifically competed by excess ag2 oligo. The only sequence similarity between the 5' Hph fragment and the ag2 oligo is [5'-CTGATAAGG-3'], in the ag2 oligo, and [5'-CCTTATCAG-3'] (on the complementary strand [5'-CTGATAAGG-3']) in the fragment. When the HEL nuclear protein extract is incubated with the labelled 5' Hph fragment and excess unlabelled "CACCC" oligo, the two retarded bands are not affected (lane 4) confirming the specificity of the ag2 oligo's competition effect. When MEL C88 nuclear protein extract was used in place of HEL the same results were obtained as shown in lanes 5-7. This experiment suggested that a protein factor or factors present in the nuclei of erythroid cells would bind to the sequence [5'-CTGATAAGG-3'].

To confirm the erythroid-specificity of this binding a bandshift was done using the 5' Hph fragment and nuclear protein extracts from HEL and HeLa, a non-erythroid cell line. The result is shown in Fig. 4.2B. The 5' Hph fragment produces the two retarded bands with HEL nuclear extract and no

Fig. 4.2 Bandshift assays using nuclear protein extracts

A Bandshift on nuclear protein extracts from the HEL and MEL C88 erythroid cell lines using ^{32}P labelled 5'-Hph fragment as probe and unlabelled oligos ag2 and CACCC as competitors. The two DNA-protein complexes seen are defined as I and II.

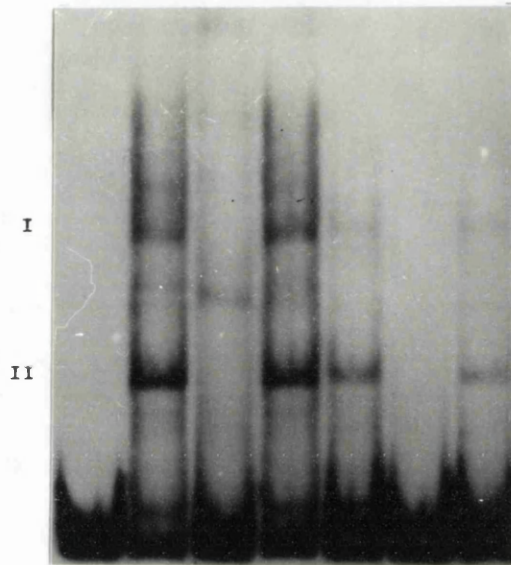
B Bandshift on nuclear protein extracts from HEL and HeLa cells using ^{32}P labelled 5'-Hph fragment as probe.

C Bandshift on nuclear protein extracts from four different erythroid cell lines using ^{32}P labelled oligo ag2 as probe.

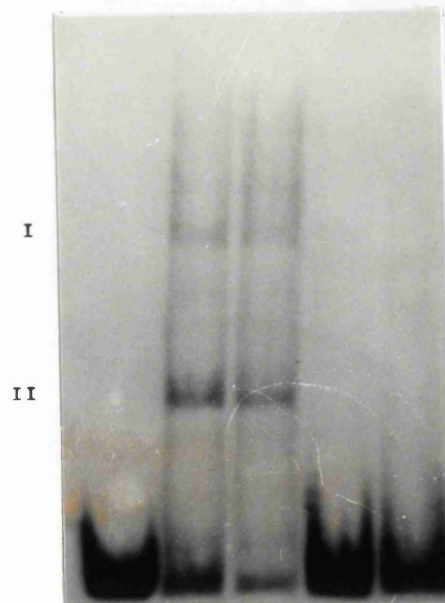
D Bandshift on HEL cell nuclear protein extract using ^{32}P labelled oligo B in the presence of increasing quantities of unlabelled oligo ag2.

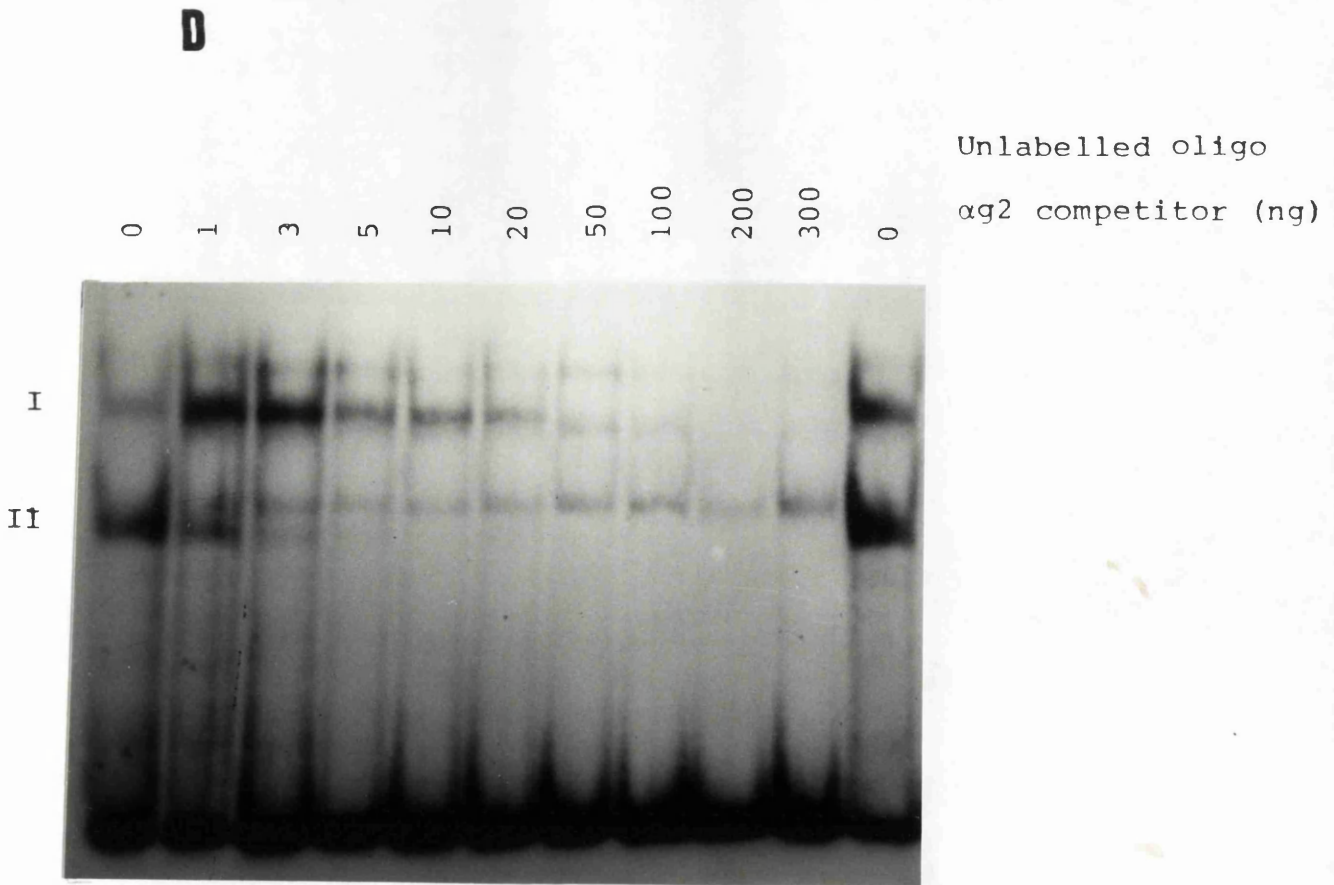
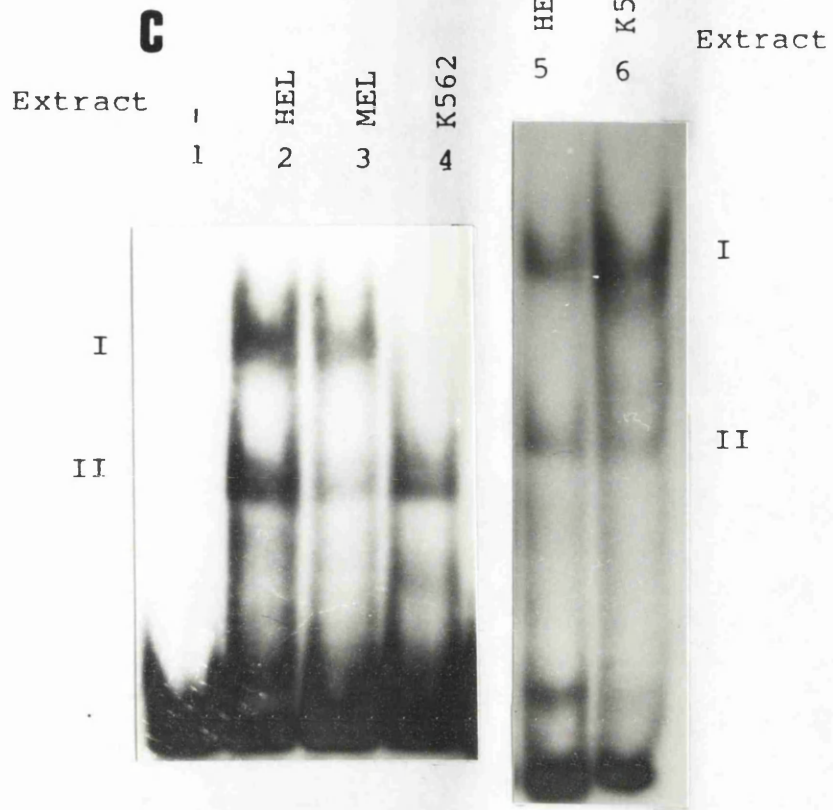
A

		HEL		MEL C88			Extract
		α g2	CACCC	α g2		CACCC	Oligonucleotide Competitor
1	1	3	4	5	6	7	
1	2	3	4	5	6	7	

**B**

		HEL		HeLa		Amount of
		5 μ l	10 μ l	5 μ l	10 μ l	Extract
0	1	2	3	4	5	
0	1	2	3	4	5	





retardation with the HeLa nuclear extract. Thus the protein(s) which binds to the fragment is not present in HeLa cells which are known to contain other ubiquitous, as opposed to cell-specific, DNA binding factors such as SP-1 (Jones & Tjian, 1985), AP-1 (Lee et al., 1987) and CACCC-box binding factor (Plumb et al., 1989).

Figure 4.2C shows a composite of bandshifts with ³²P end-labelled oligo ag2 as probe and nuclear protein extracts from 4 erythroid cell lines. HEL (lanes 2 & 5) and MEL C88 (lane 3) produce the two band pattern as before. The two bands will now be referred to as complexes I and II where I refers to the lower mobility DNA-protein complex and II refers to the higher mobility DNA-protein complex. Thus, both complexes are seen in the K562 SAI extract (lane 6) but only complex II in the extract from K562 cells grown in suspension (lane 4). While initially of interest, the presence of only complex II in the nuclear extract from the suspension grown K562 cells was almost certainly an artefact of the nuclear protein extraction procedure. When a similar experiment is performed with whole cell extract, complex I is present (discussed below). Of greater significance is the fact that complexes I and II are present in the extracts from HEL, MEL C88 and K562 SAI. This points to the fact that the DNA binding factor(s) concerned is present at different developmental stages. Primarily on the basis of globin expression, the erythroid cell lines can be said to be at different stages of development. The K562 SAI cells express human embryonic and foetal globins (Spandidos, 1984, Rutherford et al., 1979 & Benz et al., 1980). HEL cells seem to have a foetal/adult phenotype expressing α and

globin genes (Papayannopoulou et al., 1987) and MEL C88 cells express adult globin genes (Wright et al., 1983).

Correspondingly HEL cells express HCAI whereas both types of K562 cell do not (Fig. 5.2B) and the MEL C88 cells express the mouse CAI gene (Fig. 5.1B).

Complexes I and II appear to be present in similar amounts, based on the intensity of the two bands on the autoradiographs, although there usually seems to be slightly more of II than I. Therefore, it was interesting to see whether or not the complexes had different binding affinities to the DNA recognition sequence. This was done by performing bandshift assays with HEL cell nuclear extract and ³²P end-labelled oligo B in the present of increasing quantities of unlabelled ag2 oligo. The result of this experiment is shown in Fig. 4.2D. After the addition of 5ng of unlabelled ag2 oligo the complex II is virtually absent whereas the removal of complex I requires the addition of 50ng of ag2 oligo. Complex I would therefore seem to have a much higher affinity than complex II.

The band which appears in the absence of complex II is only associated with the use of oligo B as a probe. It may be due to the fact that, when the erythroid factor is binding, it excludes another factor which can only bind when the erythroid factor is competed off. DNAase I footprinting analysis with competitor and in extracts from other cell types show no other binding near site (B) apart from at the -209 CACCC-box (J.C.Sowden, personal communication). However, bandshifts using labelled oligo B and erythroid nuclear extract in the presence of excess unlabelled ag2 oligo and CACCC oligo still

show the appearance of this band (data not shown).

However, the results produced using nuclear protein extracts were not completely satisfactory. This was because of the difficulty of reproducing results from different extracts of the same cell lines. Often anomalous banding patterns were found. Initially, this was thought to be a consequence of the condition of tissue cultured cells when harvested. The cells were then closely monitored and maintained in the logarithmic growth phase before harvesting, isolation of nuclei and protein extraction. Cultures were also seeded with cells that had undergone very few passages. However, neither of these changes improved the reproducibility of bandshifts using the extracts produced. Another explanation was that the problem originated in the protein extraction procedure. This involved the isolation of nuclei, their lysis with the subsequent ammonium sulphate precipitation and isolation of the nuclear proteins as described in Wildeman et al (1984). It is possible that the isolation of nuclei at pH 7.9 and in the solution described in Section 2.26, although yielding nuclei that looked perfectly rounded under the microscope, in fact produced nuclei with highly decondensed chromatin (Chevaillier & Philippe, 1973). This process may have caused modifications in DNA-binding proteins. Using 33% ammonium sulphate to precipitate the nuclear proteins may also have contributed to the variability of different preparations since it did not take account of differing concentrations of nuclear proteins to be precipitated. Another likely possibility is proteolysis during the extended preparative procedure. Therefore, a different method was adopted whereby protein was extracted

from whole cells and without any precipitation step.

4.3 Bandshift assays using whole cell extracts

The method for whole cell extract preparation was based on the method of Dale et al (1989) and a personal communication from Dr.A.M. Ali Imam, I.C.R.F., London, as detailed in Section 2.26. It simply involved a whole cell lysis followed by centrifugation and removal of the supernatant which was then desalted before flash freezing. This method required only a fraction of the time of the nuclear protein extraction, involved much fewer manipulations and, most significantly, proved to be completely reproducible irrespective of cell line and growth conditions. Using these extracts the erythroid-specific and stage-specific nature of the DNA binding factor were again investigated as well as binding to all of its potential consensus sequences flanking the HCAI gene.

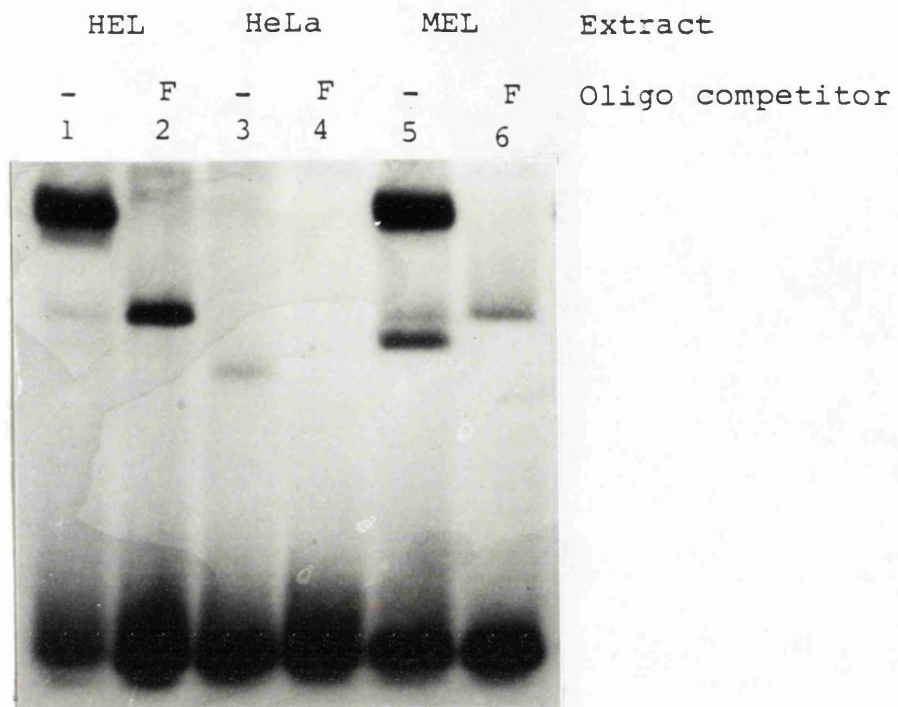
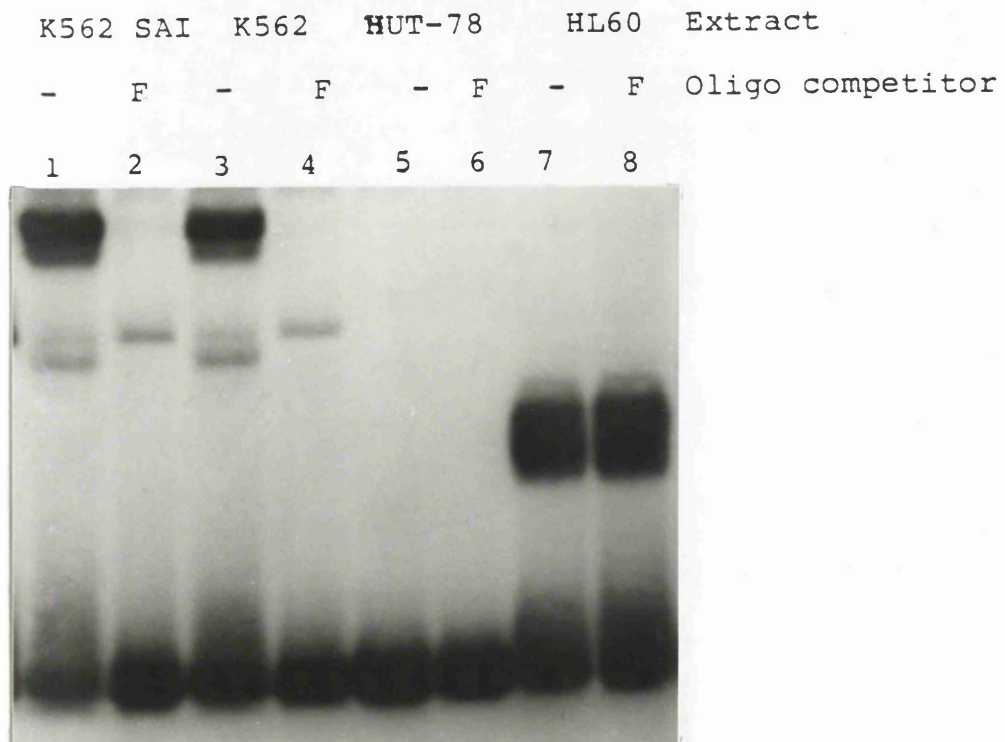
Figure 4.3A shows a bandshift using HEL, HeLa and MEL C88 whole cell extracts with ^{32}P labelled oligo B as probe and excess unlabelled oligo F as a competitor. Lanes 1 & 5 with HEL and MEL respectively show the dramatic difference between this method and the previous one. Two complexes are again seen but in this case there is vastly more complex I than complex II, both of which are competed with excess unlabelled oligo F (which contains an erythroid factor binding site, discussed below) as shown in lanes 2 and 6 (the artefactual band present in competition of oligo B is again seen). Using probes labelled to a similar degree as those used with the nuclear protein preparation a much greater proportion is retarded in

Fig. 4.3 Bandshift assays using whole cell extracts

A Bandshift assay on whole cell extracts from HEL, HeLa and MEL C88 cells with ^{32}P labelled oligo B as probe and excess unlabelled oligo F as a competitor.

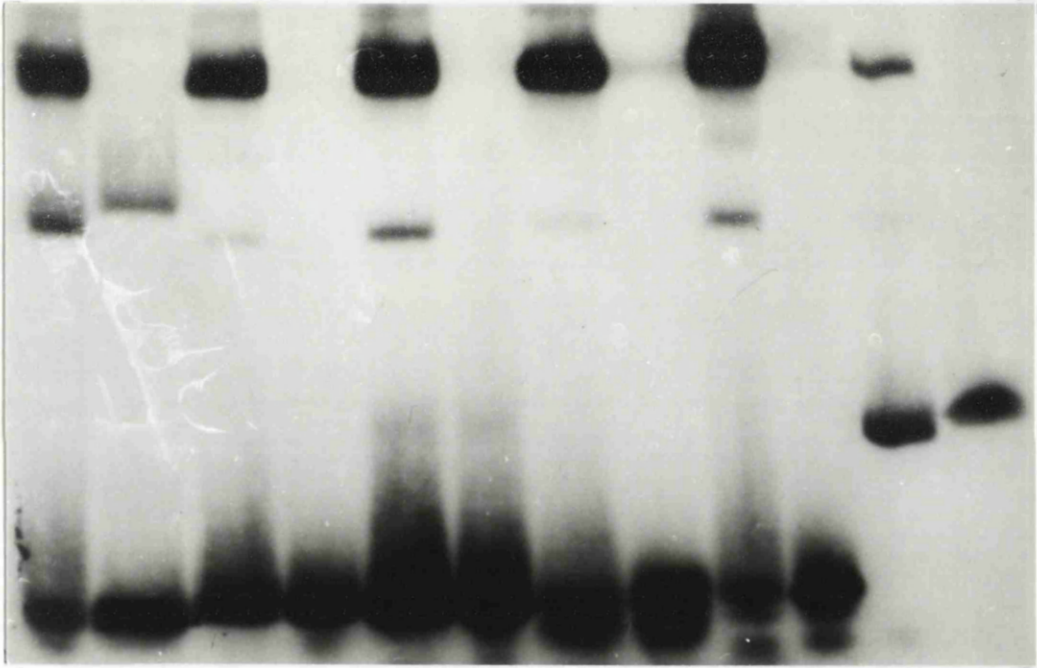
B Bandshift assay on whole cell extracts from K562 and K562 SAI cells and also the non-erythroid haemopoietic cells HL60 and HUT-78. The probe used was ^{32}P labelled oligo B as probe and excess unlabelled oligo F as a competitor.

C Bandshift assay on whole cell extracts from MEL C88 cells with ^{32}P labelled oligos A, B, C, E & F and the NdeI-EcoRI site (D)-containing fragment as probes. The unlabelled competitors used were excess oligo ag2 or oligo B.

A**B**

C

Probe		oligo B		oligo A		oligo C		oligo E		oligo F		NdeI-EcoRI (D) fragment	
oligo	competitor	-	ag2	-	B	-	B	-	B	-	B	-	B
1	2	3	4	5	6	7	8	9	10	11	12		



DNA-protein complex. These observations allow two assertions to be made about the bandshift results obtained with nuclear protein extract: firstly, the similar levels of complex I and II may be due to complex II being a degradation product of complex I whose proportion is increased during the harsher and more laborious protein preparation; secondly, for the same reason, the higher proportion of probe retarded in the whole cell extract maybe due to the binding factor being closer to its native conformation and hence more readily able to bind. It should also be noted that performing the bandshift assay using the whole cell extract technique does not involve the addition of dye-containing loading buffer before gel electrophoresis since 4% Ficoll is present in the binding mix itself and this is sufficient to pull it down to the bottom of the wells. The latter method seems an improvement since it produces sharp straight bands as opposed to more smeary bands with more complex at the edges than in the middle.

Close scrutiny of lane 3 in Fig. 4.3A shows the presence of a retarded band in the HeLa extract. This was not seen in bandshifts assays of the HeLa nuclear protein extract, probably because of the lower sensitivity of that technique. The fact that this band is competed with excess unlabelled oligo F shows that this is a specific interaction with the factor binding site on oligo B. The factor in HeLa would appear to be present in very low abundance or else have a very weak affinity for the binding site. It appears to produce only a single complex which has a different mobility from the 2 complexes present in erythroid cells. Since this mobility difference is seen using the same probe in HeLa and erythroid

cell extracts it must be due to a factor present in HeLa being smaller than that forming either of the erythroid-specific complexes. The HeLa factor may be a modified version of the erythroid form or a completely different factor which recognises a the same or a similar consensus sequence.

Figure 4.3B (lanes 1-4) shows that both complex I and II are present in K562 SAI and suspension-grown K562 cells. This is a reproducible result which points to the fact that the presence of only complex II in the nuclear extract of the suspension K562 cells (Fig. 4.2C, lane 4) was an inconsistency. The result again confirms the presence of the DNA binding factor across the embryonic to foetal to adult developmental stages of the erythroid cell lineage as represented by K562, HEL and MEL C88 cells.

To confirm that the factor was not present in other haemopoietic cells other than erythroid ones, whole cell extracts from lymphoid lineage cells (the human T cell line, HUT-78) and monocyte/macrophage lineage cells (the HL60 line) were also tested by bandshift assay. Neither cell line produces the complexes I and II seen with erythroid cells (Fig. 4.3B, lanes 5-8). However, a different large retarded band is seen in HL60 cells and not competed with oligo F, so it would appear not to be specific for the erythroid factor consensus sequence. It would appear that this complex formed in HL60 is completely independent of the sequence of the labelled probe since other unrelated ³²P end-labelled oligonucleotides also produce this large band (data not shown).

There are 6 potential erythroid factor binding sites

around the HCAI gene. Figure 4.3C shows an experiment investigating their ability to bind the erythroid factor present in a MEL C88 whole cell extract. Sites (A), (B), (C), (E) and (F) are present in ³²P end-labelled double stranded oligonucleotides (Fig. 4.1A) and site (D) in a ³²P end-labelled 60 bp DNA fragment (Fig. 4.1B). All of the labelled probes show the characteristic DNA-protein complexes I and II which are specifically competed by excess unlabelled oligo B except for labelled oligo B itself which was specifically competed by excess unlabelled oligo F. The DNA fragment containing site (D) produces a weaker signal because it is labelled to a lower specific activity than the oligonucleotide probes. This experiment carried out using whole HEL cell extract yields precisely the same result (data not shown). The significant result of this experiment is that all the sites bind yet they have variable consensus sequences as underlined in Fig. 4.2A & 4.2B. The immediately common motif is the presence of a core -ATC- or -GAT- trinucleotide (depending on orientation). Based purely on these 6 sites a hexanucleotide consensus sequence can be deduced: [5'-T^A/_TATC^A/_T-3'] or, in the opposite orientation [5'-T^T/_AGAT^A/_TA-3'].

4.4 Summary of the results of the bandshift assays

From the bandshift assays using nuclear protein and whole cell extracts the following results can be summarised:

- i) there is an erythroid-specific DNA binding factor;
- ii) it is present at each developmental stage of the erythroid cell lineage;
- iii) around the HCAI gene it binds at 6 sites which have a

consensus sequence [5'-T^T/AATC^A/T-3'];

- iv) a factor present in HeLa cells binds to the same sequence.

As mentioned in Section 3.4, since the research in this thesis was begun other investigators have established the presence of an erythroid-specific DNA binding factor (to be discussed later). The results of their work are consistent with the properties described here. Recently, a cDNA encoding a factor having these properties has been cloned and designated GF-1 (Tsai et al., 1989). GF-1 would appear to be the major DNA binding protein specific to the erythroid lineage. It can be assumed therefore that the erythroid-specific DNA binding factor binding to the sites around the HCAI gene is GF-1 and will be referred to as such hereafter.

4.5 The effect of TPA on HCAI expression in HEL cells

HEL cells express the HCAI gene (Fig. 4.4 & 5.2A). When HEL cells are treated with the phorbol ester 12-O-tetradecanoyl-phorbol-13-acetate (TPA) at 10^{-6} M they undergo morphological, functional and biochemical changes so that they exhibit a macrophage-like phenotype (Papayannopoulou et al., 1983). Apart from the drastic morphology changes, the cells greatly enhance their phagocytic ability and acquire receptors for binding and degradation of chemically modified lipoproteins. It has been shown previously that the biosynthesis of CAI protein in TPA-treated HEL cells is reduced in comparison to control cultures (Villevall et al., 1985). Following on from this the effect on HCAI mRNA levels

Fig. 4.4 Photomicrographs of untreated and TPA-treated HEL cells

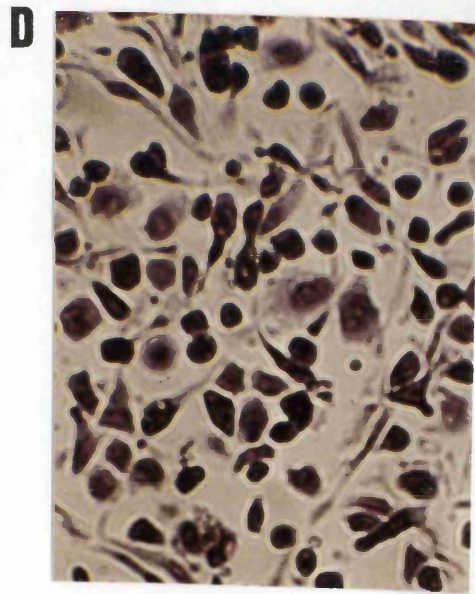
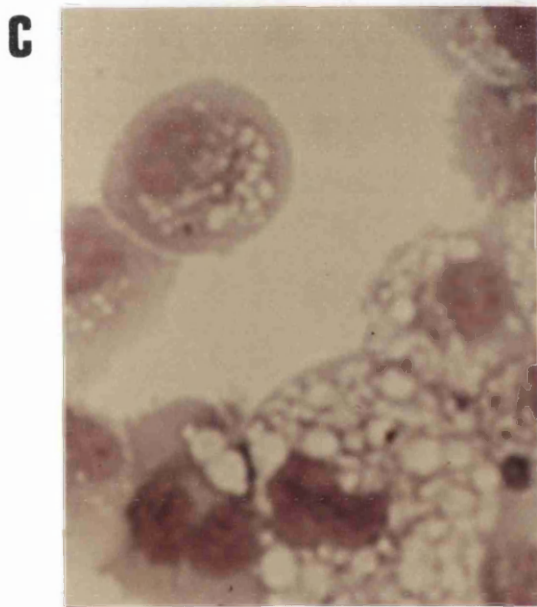
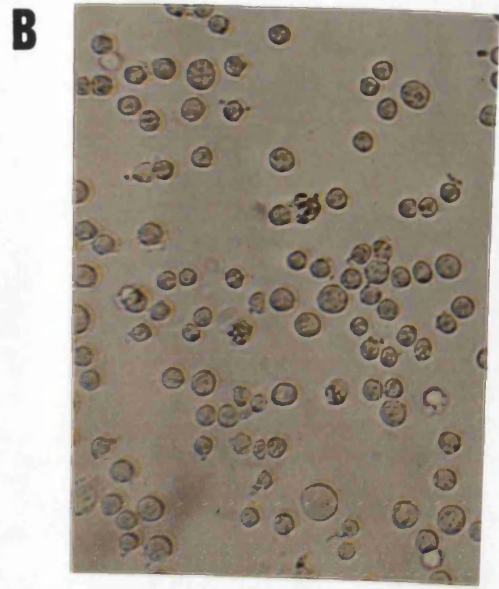
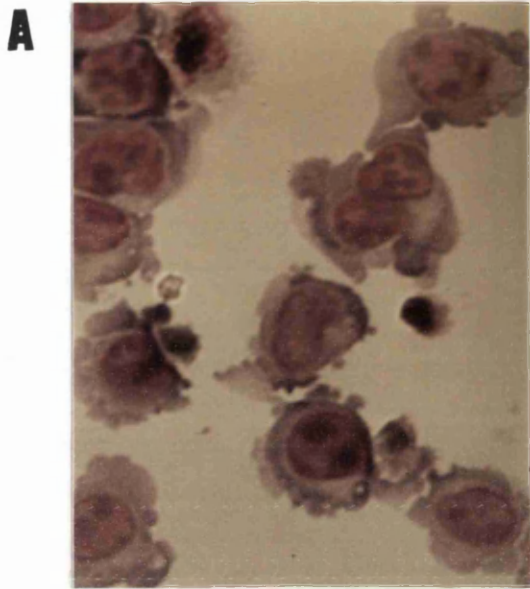
A Giemsa-stained untreated HEL cells at high magnification

B Giemsa-stained untreated HEL cells at low magnification

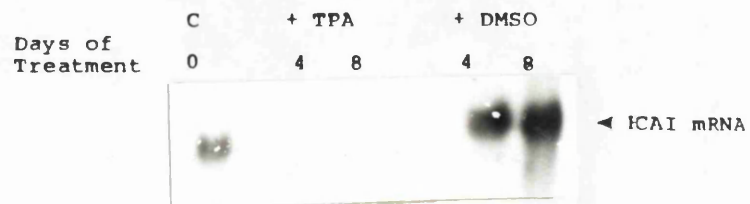
C Giemsa-stained TPA-treated HEL cells at high magnification

D Giemsa-stained TPA-treated HEL cells at low magnification

E Northern analysis of equivalent amounts of total RNA from control and TPA-treated HEL cells probed with ^{32}P labelled HCAI cDNA. The RNA from DMSO-treated HEL cells was necessary as a control since the TPA had been dissolved in DMSO.



E



in TPA-treated HEL cells was investigated. HEL cells also have the GF-1 factor. A correlation between the effect of TPA treatment on HCAI mRNA levels and GF-1 binding as assayed by bandshift was also examined. If such a correlation existed it would provide indirect evidence of a direct role for GF-1 in HCAI transcription.

HEL cells were treated with TPA at 10^{-6} M for 4 or 8 days. That the TPA induced the HEL cells to undergo a macrophage-like shift was assessed by staining for the characteristic morphological changes anticipated (Papayannopoulou et al., 1983). When stained with Giemsa and viewed under high magnification, the TPA-treated cells (Fig. 4.4C) show large differences compared to untreated HEL cells (Fig. 4.4A). They show a dramatic decrease in nuclear/cytoplasmic ratio, a loss in cytoplasmic basophilia and a large increase in cytoplasmic vacuolisation. When viewed under lower magnification the change is still apparent. The control HEL cells appear as round, free cells; adherent cells and cell aggregates are rare (Fig. 4.4B). TPA-treated cells however, adhere firmly and stretch their cytoplasm in a variety of ways so that cells are seen having a thin spindle-like or dendritic appearance or very flat cells with indistinct cytoplasmic borders (Fig. 4.4D).

Total RNA was prepared from untreated HEL cells, HEL treated with TPA (10^{-6} M) for 4 or 8 days and HEL treated with 0.01% DMSO for 4 or 8 days. The latter was necessary as a control because the TPA was dissolved in DMSO. The amount of TPA that was added equivalent to 10^{-6} M made the solution 0.01% DMSO. 30 μ g of each RNA was then Northern blotted and probed

with ^{32}P - α -dCTP labelled HCAI cDNA. The result of this is shown in Fig. 4.4E. Compared to the control cultures (lanes 1, 4 & 5), the steady-state level of HCAI mRNA has been greatly reduced in the TPA-treated HEL cells, both at 4 and 8 days (lanes 2 & 3). As determined by scanning densitometry on the LKB Ultrosan XL, the decrease in HCAI mRNA levels is 7-8 fold. This is a much more dramatic decrease than seen for HCAI protein (Villevall et al., 1985) but that is not surprising considering the protein is likely to be more stable than the mRNA transcript and have a much longer half-life.

Having seen a large fall in the HCAI mRNA level, whole cell extracts were made from the TPA-treated and control cultures to examine the effect on GF-1 binding to ^{32}P -labelled oligo E in a bandshift assay. However, no change was seen in GF-1 binding in the TPA-treated cell extracts (data not shown). This infers that, even though GF-1 is present and binding, that HCAI expression can still be heavily down-regulated. As mentioned previously (and discussed further in the next Sections), GF-1 is believed to be a transcriptional regulator of erythroid expression (Evans et al., 1988; deBoer et al., 1988; Plumb et al., 1989; Tsai et al., 1989). However, though it may be necessary for the expression of genes from erythroid-specific promoters, GF-1 would appear not to be sufficient in itself. This applies to the HCAI gene, since GF-1 is present in K562 cells which do not express HCAI and is apparently unaffected in TPA-treated HEL cells. However, this may be an over-simplification since presence as determined by bandshift assay does not give any clue as to subtle modifications e.g. phosphorylation/dephosphorylation or

glycosylation, which may affect the biological activity of the factor and yet not its ability to bind DNA. In the case of the TPA treatment, there is also the slightly less likely possibility that GF-1 has an inordinately long half-life so even if it is affected by TPA treatment 8 days is not sufficient to see it.

4.6 A transfection system to assay transcriptional activity

Having established the presence of GF-1 binding sites in the 5'- and 3'-flanking regions the HCAI gene, but not shown any direct evidence for a role for GF-1 in the regulation of HCAI expression, a more direct approach was sought to establish this role. Ideally, the cloned gene would be introduced into a homologous cell type but this has great drawbacks: the difficulty of distinguishing the expression of the cloned gene from that of endogenous genes; in the case of HCAI, the gene is far too large to be introduced into cells by normally available methods and the only homologous cell type available (HEL, the human erythroid cell line expressing HCAI, is very recalcitrant to transfection. A solution to these problems was to link the promoter of the cloned gene to an easily distinguishable and assayable reporter gene in a plasmid. The reporter gene used was chloramphenicol acetyltransferase (CAT). CAT is responsible for the chloramphenicol resistance of many different prokaryotic species. It makes an ideal reporter gene because there is no endogenous eukaryotic analogue and there is an easy sensitive and reproducible assay for its activity (Gorman et al., 1982a).

Before promoter constructs could be assayed, it was necessary to find a suitable cell line into which they could be introduced and a reliable method of introduction by transfection. The constructs were to be assayed after transient expression rather than from stably transfected cell lines. This was thought to be easier because it requires less time as it eliminates the time consuming procedure of isolating cloned cell lines. This method also obviates the clonal variability caused by integration into different genomic locations and with differing frequencies. It also requires a more efficient method of transfection, since the number of stably transfected clones available for assay can be increased by allowing them to grow longer whereas the transiently transfected cells are a fixed number. The cells chosen for transfection were MEL cells since they are erythroid, have GF-1 and express mouse CAI so should therefore express HCAI constructs.

Firstly, attempts were made to transfect the MEL cell C88 subclone with the plasmid pIGA-101 (gift from Dr D.S. Latchman, U.C.L.). This plasmid contained the promoter regions of the Herpes Simplex Virus 1 (HSV-1) intermediate early genes 22 and 47 fused with the CAT gene. HSV-1 promoters are known to drive reporter gene expression in erythroid cells (Dr J. Frampton, Beatson Inst., Glasgow, personal communication). It was not possible to use the calcium phosphate / DNA precipitate method since the cells are non-adherent. Transfection using a Biorad Gene Pulser electroporator, following both the manufacturer's recommendations and the method of Spandidos (1987) produced no CAT expression. Another

type of MEL cell F412B2, was tried. Several subclones of F412B2 were tested for transfection efficiency. The one which proved useful for transfection was a gift from Dr J. Frampton. This type of MEL cells are adherent on tissue culture surfaces and can be efficiently transfected using a variation of the standard calcium phosphate / DNA precipitate technique (Section 2.24). However, it is important that the passage number of the cells to be transfected be kept as low as possible, up to a maximum of 10 passages from source before transfection efficiency falls.

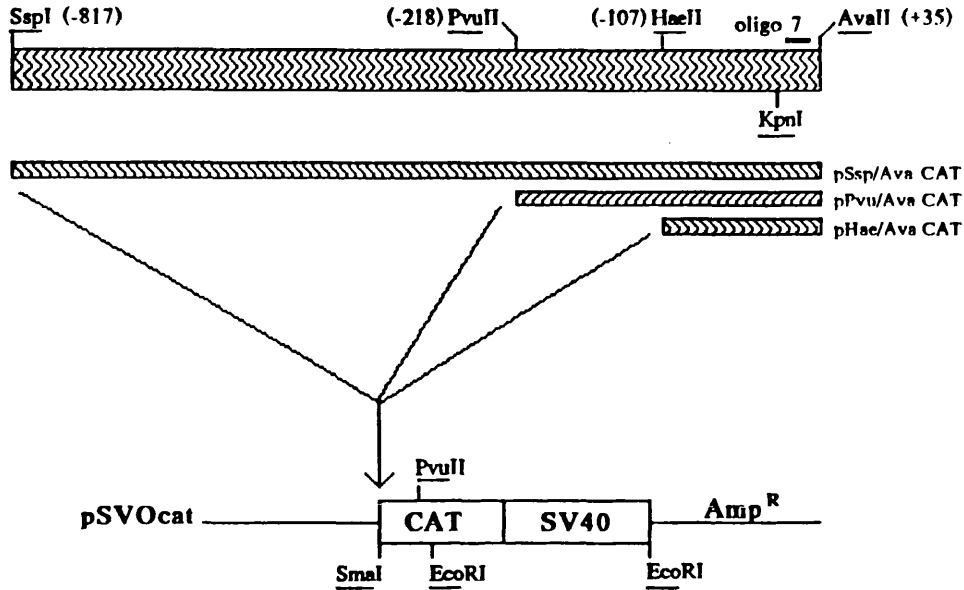
4.7 HCAI constructs to study transcriptional activity

Promoter/CAT gene constructs could now be made for transfection into the F412B2 MEL cells. 5'-flanking regions of the HCAI gene were subcloned, in the correct 5' to 3' orientation, into pSV0cat. This vector consists of the β -lactamase gene and origin of replication from pBR322 together with the coding sequence and some 5'-untranslated sequence of the CAT gene from the E.Coli transposon Tn9, followed by segments of SV40 DNA containing the small t antigen intron and the early region polyadenylation site (Gorman et al., 1982a). The CAT sequence has no preceding transcriptional promoter. Therefore, the unique SmaI site, just upstream of the CAT gene, was used to subclone the HCAI gene fragments.

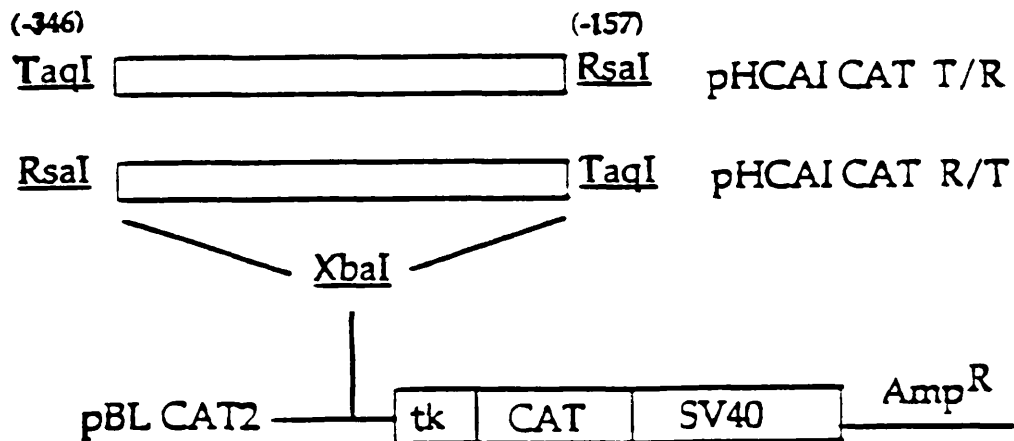
Three fragments were subcloned into the SmaI site (Fig. 4.5A): a SspI / AvaII fragment (-817 to +35 in HCAI 5'-flanking region); a PvuII / AvaII fragment (-218 to +35) and a HaeII / AvaII fragment (-107 to +35). These fragments were chosen to see the effect of going from a construct containing

Fig. 4.5 Diagrams of HCAI promoter-CAT gene constructs

A Diagram of the fragments and vector used to make HCAI promoter constructs in pSV0cat.



B Diagram showing the TaqI-RsaI fragment of HCAI (-348 to -157) inserted (in either orientation) into the unique XbaI site of pBLCAT2. This site is upstream of a minimal thymidine kinase (tk) promoter fused to the CAT reporter gene, the SV40 small t intron and polyadenylation site.



all of the 5' GF-1 binding sites to a situation where none of them were present. To begin the subcloning procedure, the SspI / AvaII and PvuII / AvaII fragments were first isolated on NA45 paper from digests of pBSc204SH. The purified fragments were blunt-ended with Klenow and ligated in pSV0cat, that had previously been digested with SmaI and dephosphorylated with CIP. After transformation of E.Coli HB101 cells and selection of ampicillin-L agar plates, rapid plasmid preparation of positive recombinants were size screened on an agarose gel, using supercoiled pSV0cat for comparison.

Recombinants with inserts were further screened for correct orientation (i.e. the 3' AvaII end being adjacent to the CAT gene) by digestion with PvuII for the SspI / AvaII construct and KpnI / EcoRI for the PvuII / AvaII construct. To make the HaeII / AvaII-CAT gene construct was less straightforward. It was necessary to completely digest the isolated PvuII / AvaII fragment with HaeII, blunt end the digestion products with Klenow and then ligate into SmaI digested and dephosphorylated pSV0cat. Two possible inserts, PvuII / HaeII or HaeII / AvaII, could be produced by this procedure. Therefore, filter lifts were taken from the positive recombinant colonies, after transformation and ampicillin-selection, in order to identify the desired recombinants. The filters were denatured on 0.5M NaOH/1.5M NaCl for 10 minutes and neutralised for 5 min on 3M NaOAc, pH 5.5, followed by baking at 80°C for 4 hours. The filters were prehybridised at 58°C in 6 X SSC, 5 X Denhardt's, 50mM Na₂HPO₄, pH 6.5 & 100µg/ml denatured salmon sperm DNA before hybridisation overnight, under the same conditions, with oligo

#7. The filters were then washed for 2 X 15 min, at 48°C before autoradiography at -70°C overnight. Oligo #7 hybridises to a region within the HaeII / AvaII fragment only, so it identifies only HaeII / AvaII recombinants. Rapid plasmid preparations of these were made and correctly-orientated clones identified by KpnI / EcoRI digestion.

Large scale preparations of each of the three constructs, and pSV0cat, were caesium chloride banded and supercoil purified, then transfected into F412B2 MEL cells. CAT expression was detected at a low level in the cells transfected with the three constructs but, it was at a similar level to that found in cells transfected with the pSV0cat vector (Fig. 4.6A). pSV0cat has no eukaryotic promoter, yet there appears to be some recognition of cryptic promoter sequences by the F412B2 MEL cells. It is likely that these sequences lie within the pBR322 region of pSV0cat (Langner, 1986). The fact that the expression due only to the vector was at the same level as that for the 5'-flanking region / CAT gene fusion constructs precluded the further use of the latter as viable tools to study the transcriptional regulation of the HCAI gene.

Another strategy had to be considered. Two factors needed to be borne in mind: firstly, on the evidence of the 5'-flanking region/CAT gene fusion constructs, the putative HCAI gene promoter region does not seem to function very well in MEL cell transfections and; secondly, the initial point of interest is whether the GF-1 binding sites affect transcription. The method adopted was to insert a fragment of the HCAI 5' flanking region upstream of a functional minimal

eukaryotic promoter driving CAT gene expression. This construct would be introduced into MEL cells and a non-erythroid cell line to check for any induction in the level of CAT expression in the erythroid cells.

The fragment to be subcloned was the TaqI / RsaI fragment at -348 to -157 in the 5'-flanking region. It contains the GF-1 binding sites (A) and (B). The vector into which it was inserted was pBLCAT2 (Luckow & Schutz, 1987), shown in Fig. 4.5B. This is the same as pSV0cat except, fused to the CAT gene is the fragment of the HSV-I thymidine kinase (tk) gene promoter spanning from -105 to +51 upstream of which is a polylinker region.

The TaqI / RsaI fragment was prepared by first isolating the SspI / AvaII (-815 to +35) as above. This fragment was then digested with TaqI and RsaI. The digested products were blunt ended with Klenow and then ligated into pBLCAT2 vector that had been digested with XbaI (a unique site upstream of the tk promoter), also blunt-ended with Klenow and then dephosphorylated. After transformation of E.Coli HB101 recombinants were selected on ampicillin-L-agar. To identify those recombinants which contained the TaqI / RsaI fragment, filter lifts from the positive colonies were taken and screened with a ³²P end-labelled oligonucleotide as above. The oligonucleotide used was the coding / anti-sense strand of the double-stranded oligo B, since site (B) is within the TaqI / RsaI fragment. Rapid plasmid preparations from positively hybridising colonies were then further screened by HinfI / HindIII and HinfI / XhoI double digests compared with similarly digested vector, to allow the orientation of the

insert to be determined. Recombinants possessing the TaqI / RsaI fragment in both orientations were selected. The recombinant with the TaqI / RsaI fragment 5' to 3' was designated pHCAICAT T/R and that with the fragment 3' to 5' designated pHCAICAT R/T (Fig. 4.5B).

4.8 Transient expression studies of pHCAICAT T/R & R/T in MEL and HeLa cells

Equivalent amounts of caesium chloride-banded and supercoil-purified preparations of pHCAICAT T/R, pHCAICAT R/T and pBLCAT2 were introduced into F412B2 MEL cells and HeLa cells as described in Section 2.24. The HeLa cell transfection was most efficient when incorporating a glycerol shock step after the cells had been exposed to the calcium phosphate/DNA precipitate. Each transfection was in fact a co-transfection of the CAT construct and ^{an} independent reference gene on another plasmid to allow for the transfection efficiency to be normalised between different plates of cells. This other plasmid contained a β -galactosidase gene driven by the HSV-1 intermediate early gene 4 promoter (a gift from Dr J. Frampton). It was well expressed in both cell types transfected and easily assayed as described in Section 2.25.

Having co-transfected each of the CAT plasmids and the β -galactosidase reference plasmid into F412B2 cells, the cells were harvested, lysates formed and assayed for β -galactosidase expression. β -galactosidase activity was measured for each individual plate cell lysate. A volume of each cell lysate having an equivalent level of β -galactosidase activity was then assayed for CAT activity. The CAT activity measured was

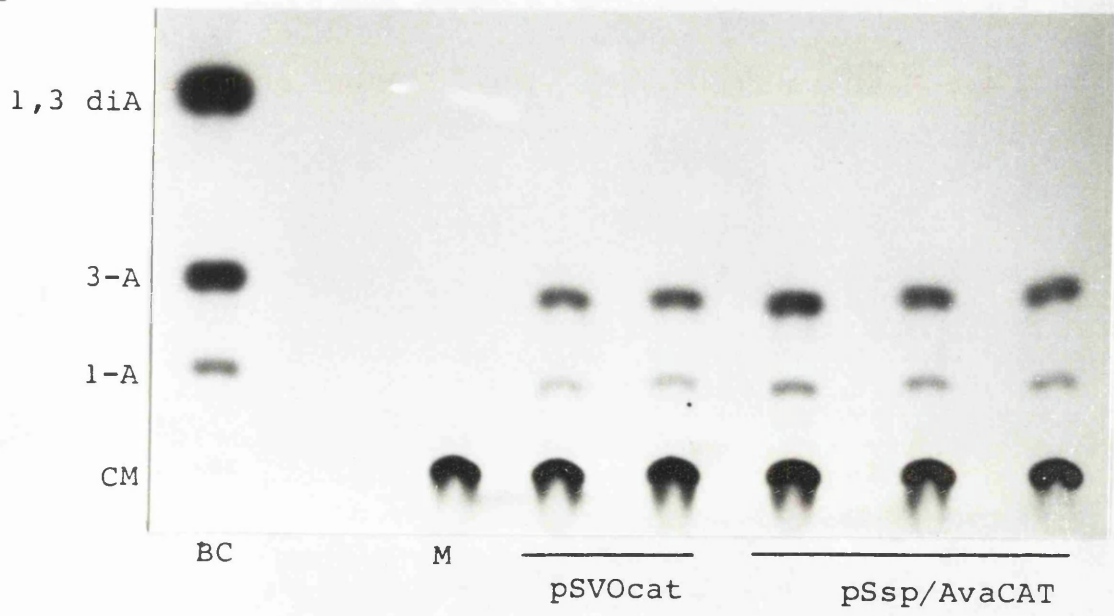
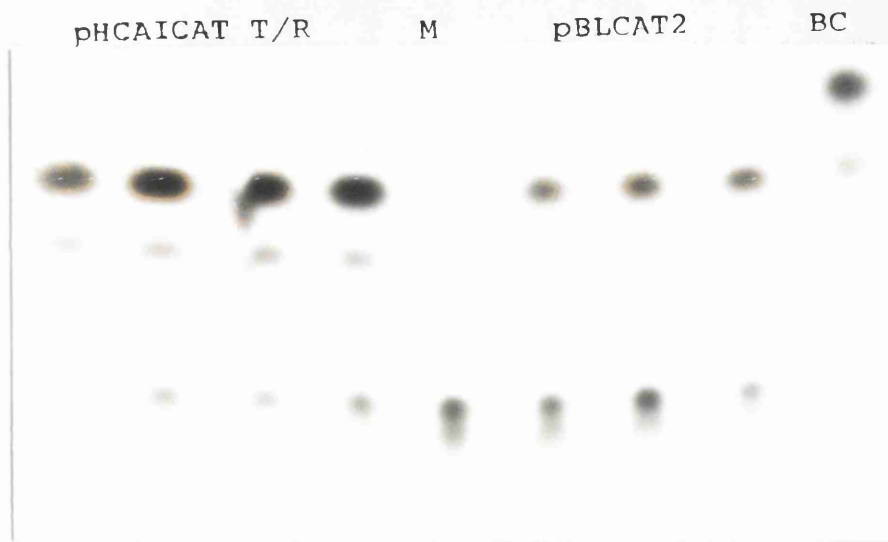
Fig. 4.6 Results from CAT assays of HCAI promoter constructs transfected into HeLa and F412B2 MEL cells.

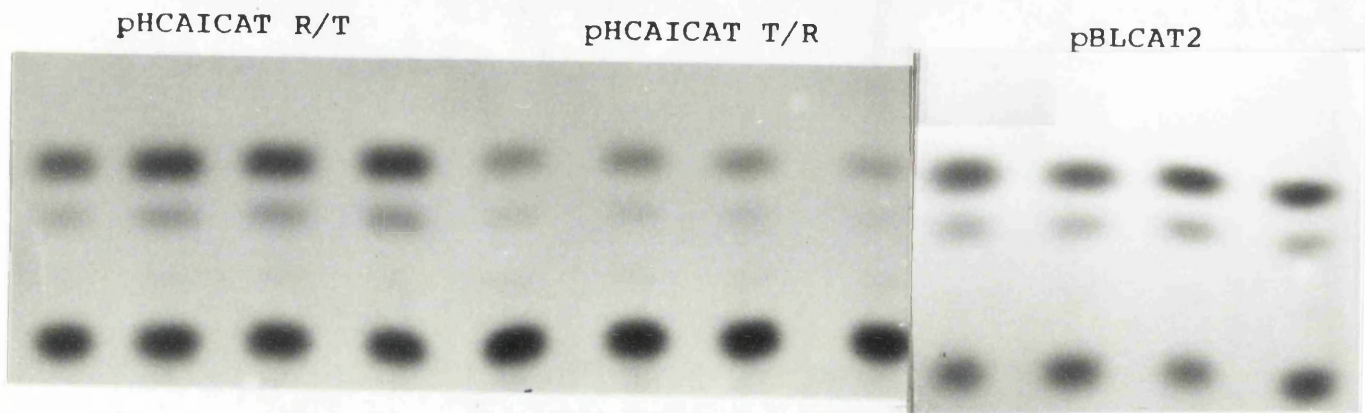
A Diagram showing a CAT assay of extracts from F412B2 MEL cells transfected with pSsp/AvaCAT and pSV0cat. The autoradiograph was exposed for 4 days. CM = chloramphenicol. 1-A, 3-A & 1,3 diA = 1-acetyl-, 3-acetyl- and 1,3 diacetyl-chloramphenicol respectively. BC = bacterial control, an E.Coli HB101 extract expressing pRSVcat and M = mock transfection control.

B CAT assay of extracts from F412B2 MEL cells transfected with pBLCAT2 and pHCAICAT T/R. The volumes of extract in each assay express an equivalent level of β -galactosidase activity thus normalising for different transfection efficiencies.

C CAT assay of extracts from HeLa cells transfected with pBLCAT2, pHCAICAT T/R and pHCAICAT R/T.

D The results of the CAT assays on normalised volume of extracts of the transfected cells were analysed by scanning densitometry of autoradiographs of the tlc plates. The data derived from each are shown relative to pBLCAT2 in both HeLa and F412B2 MEL cells.

A**B**

C**D**

	HeLa	F4
pBL CAT2	1.0	1.0
pHCAI CAT T/R	0.5	2.8
pHCAI CAT R/T	0.9	2.5

thus normalised for different transfection efficiencies.

CAT inactivates chloramphenicol by the formation of mono- and di-acetylated derivatives. The CAT assay involves incubating cell lysates with ^{14}C -chloramphenicol (CM) and resolving acetylated products by thin-layer chromatography (tlc). Fig. 4.6B is an autoradiograph of a tlc-plate with CAT assays of a pBLCAT2 and pHCAICAT T/R F412B2 cell transfection showing CM and its forms. It shows the degree of CAT activity in extracts from different plates of F412B2 cells transfected with pHCAICAT T/R (lanes 1-4), with pBLCAT2 (lanes 6-8), with TE buffer, as a control, (lane 5) and in an E.Coli HB101 extract expressing pRSVcat (lane 9). pRSVcat is similar to pSV0cat except that the expression of the CAT gene is driven by the Rous Sarcoma Virus 5' long terminal repeat which is a very strong promoter (Gorman et al., 1982b).

Figure 4.6B clearly shows a higher level of CAT activity from pHCAICAT T/R over pBLCAT2 vector. Assessed by scanning densitometry of the 3-acetyl CM form in each track, the activity is 2.8 fold greater (Fig. 4.6D). This can be assumed to be due to a higher level of expression of the CAT gene on the pHCAICAT T/R plasmid. The only difference between the two plasmids is the presence of a TaqI / RsaI fragment containing the two GF-1 binding sites. It would therefore seem that the presence of erythroid-specific DNA binding factor sites upstream of the minimal tk promoter allow the upregulation of transcription from this promoter in an erythroid environment. A similar result is obtained with the pHCAICAT R/T plasmid (data not shown) when there is 2.5 fold increase in CAT activity.

The same experiment was then repeated in HeLa cells. This

was essential for two reasons: firstly, it was necessary to show that the effect seen was erythroid-specific. It may have been that the nature of the HCAI-sequence containing plasmids was such that they were more suitable for transcription if, for instance, the presence of the insert had altered their conformation to be so; secondly, as well as two GF-1 binding sites, the TagI / RsaI fragment contains consensus sequences for AP-1 and/or CACCC-box factor. Using HeLa cells allows a distinction to be made between the latter two and GF-1 since HeLa cells contain AP-1 (Lee et al., 1987) and CACCC-box (deBoer et al., 1988; Plumb et al., 1989; Schule et al., 1988) but not GF-1. MEL cells contain GF-1 as well as CACCC-box factor (deBoer et al., 1988; Plumb et al., 1989) and AP-1 (Hirai et al., 1989).

The results of the HeLa transfection are shown in Fig. 4.6C and 4.6D. The CAT expression of the HCAI-sequence containing plasmids, relative to pBLCAT2 as 1.0, is 0.9 in the case of pHCAICAT R/T and 0.5 for pHCAICAT T/R. Therefore, no increase in transcription of the CAT gene is seen in HeLa cells. If anything there is some decrease in expression. This would suggest that it is the presence of the GF-1 binding sites which confers the increased expression of the pHCAICAT plasmids compared to the control plasmid, in MEL cells and that GF-1 is a positive regulator of erythroid-specific transcription.

This chapter has shown that at least 6 DNA sequences, based on the consensus [5'-T^T/_AATC^A/_T-3'] and flanking the HCAI gene, bind an erythroid-specific factor. This factor,

designated GF-1, is present at all developmental stages of the erythroid lineage. A different DNA binding factor, in HeLa cells, appears to recognise the GF-1 binding consensus sequence. It has also been shown that TPA treatment of HEL cells decreases HCAI expression without manifestly altering GF-1. However, the presence of GF-1 binding sites does appear to increase expression from a eukaryotic promoter in erythroid cells and not in non-erythroid cells.

CHAPTER FIVE

STUDIES ON THE TRANS-ACTIVATION OF ERYTHROID GENE EXPRESSION IN TRANSIENT HETEROKARYONS

The previous chapter showed that an erythroid-specific factor, GF-1, binds to the HCAI gene. However, GF-1 is present and shows the same pattern in bandshift assays, in K562 cells (not expressing HCAI) as in HEL cells (which do express HCAI). As well as this, GF-1 seems unchanged in TPA-treated HEL cells as compared to normally grown HEL cells, when in the former the level of HCAI mRNA has dropped 7-8 fold. These observations suggest that although GF-1 may have a role in the regulation of the HCAI gene in erythroid tissue, other factors are also involved. In particular, the presence of GF-1 at all the developmental stages of the erythroid lineage makes it likely that another trans-acting factor (or factors) regulates the stage-specific activation of the HCAI gene at the foetal-adult interface. To test this hypothesis, cell fusion studies were carried out using transient heterokaryons. The aim was to determine whether diffusible factors present in erythroid cells were capable of activating the expression of genes in the nuclei of cells that do not ordinarily express these genes.

5.1 Assay of cell lines for absence or presence of CAI and confirmation of stage-specific activation of HCAI mRNA

The cell lines to be used in the fusion studies were MEL C88, K562 SAI (haemin-induced, see Section 5.2) and HeLa

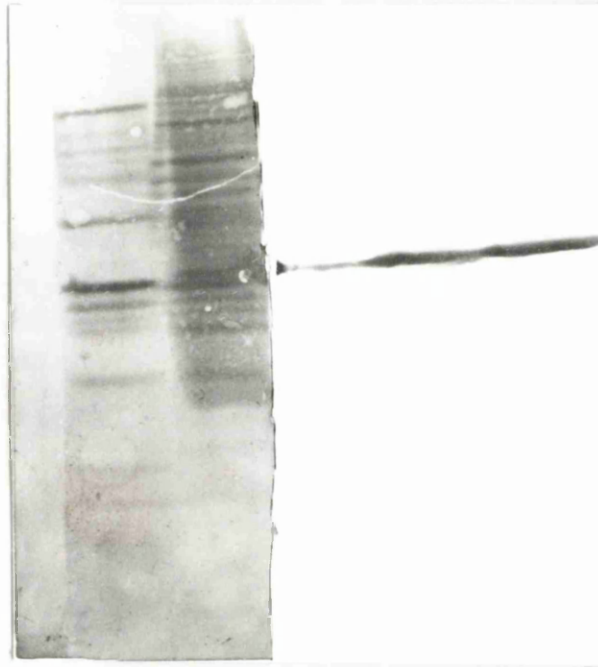
Fig. 5.1 Western blot analyses of MCAI protein in MEL C88 cells.

A SDS PAGE gel electrophoresis of MEL C88 cell lysate. After running, the gel was Western blotted onto a filter. Lanes 1 and 2 were Coomassie Blue stained whereas lanes 3-5 were probed with anti CAI antibody. This binding was visualised by horseradish peroxidase Colour Reagent. M = M.W. marker proteins and L = MEL C88 cell lysate.

B Three similar amounts of MEL C88 cell lysate were resolved on an isoelectric focusing gel under conditions optimised to separate the CA isoenzymes. The gel was Western blotted, probed with anti-CAI antibody and the binding visualised as above. The mouse CAI, having run to the anode, cross-reacted with the antibody. There was also a very slight cross-reaction with CAII, which had run to the cathode.

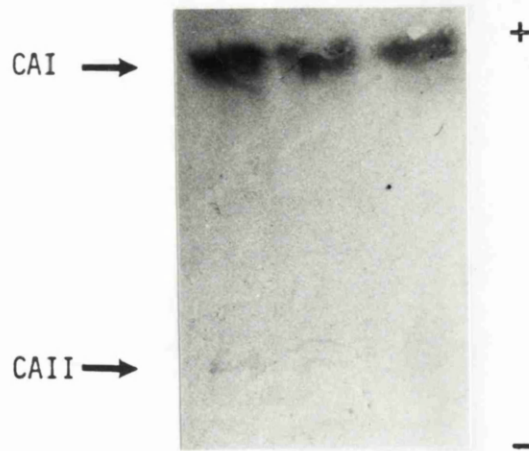
A

Coomassie
Blue Stained Anti-HCAI Ab probed
M L M L L



B

MEL C88 cell lysate



cells. For the reasons discussed in Section 5.2, it was first necessary to show that the MEL C88 subclone expressed mouse CAI and that neither K562 SAI (haemin-induced) nor HeLa cells expressed HCAI. It was also necessary to confirm at the mRNA level that HCAI is expressed in a stage-specific manner.

The presence of the mouse CAI gene product as distinct from the other CA isoenzymes, particularly CAII, was confirmed by antibody probing of Western blots of resolved MEL C88 cell lysate. The antibody used to probe for mouse CAI was against HCAI and had been raised in rabbit (gift from Dr Y. Edwards, U.C.L.). In order to show that this cross-reacted with mouse CA protein, an SDS PAGE gel of MEL C88 cell lysate was run, Western blotted and probed with the anti-HCAI antibody. The filter was then probed with goat anti-rabbit Ig conjugated with horse radish peroxidase (HRP). Finally, a HRP colour reagent was used to visualise antibody binding. The result is seen in Fig. 5.1A. The two concentrations of MEL C88 lysate (lanes 4 and 5) show a single band binding equivalent to a single band in the molecular weight (M.W.) marker proteins track. Lane 2 shows a sample of lysate equivalent to lane 5 and lane 1 has the same M.W. markers as in lane 3. However, although lanes 1 and 2 were Western blotted onto the same filter with lanes 3, 4 and 5 they were not probed with antibody but Coomassie Blue stained instead. The staining procedure shrunk the filter slightly, but it was still possible to line up lanes 1 and 2 against 3, 4 and 5, which shows that the band in lane 3 is equivalent to the CA, 29 kd, M.W. marker (lane 1) which is in turn equivalent to the band present in the MEL cell lysate tracks. This showed that the

Fig. 5.2A Northern analysis to detect HCAI expression in various cell lines

Northern analysis on total RNA from K562 SAI and K562 suspension cells (both haemin-induced), HeLa cells and HEL cells. The RNAs were probed with ^{32}P labelled HCAI cDNA.

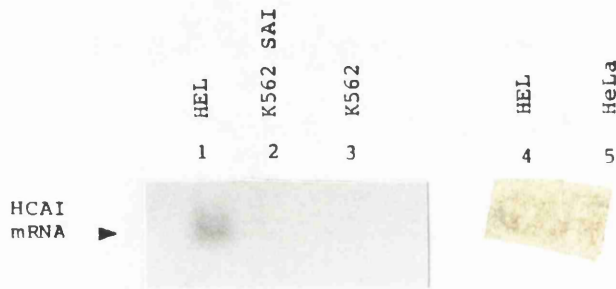
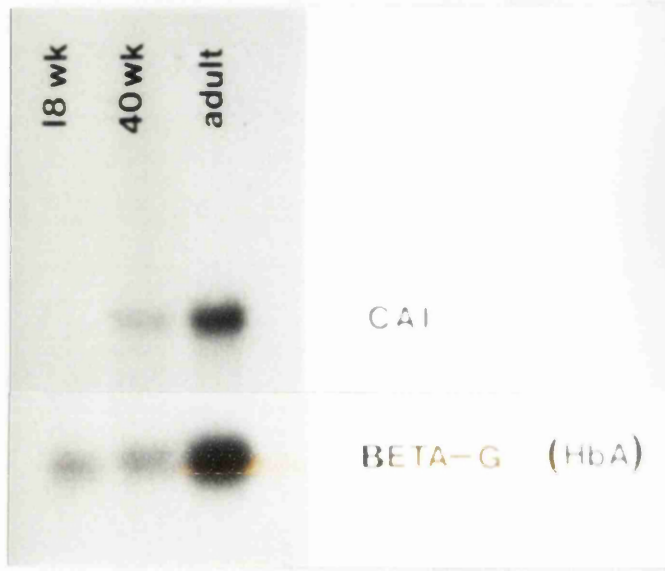
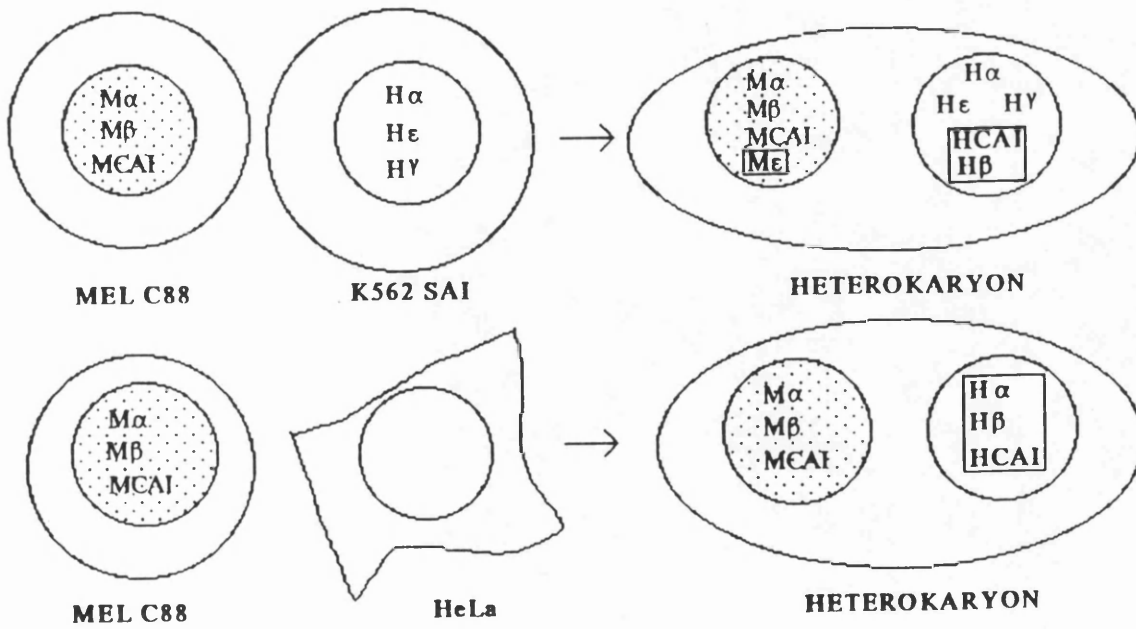


Fig. 5.2B Northern analysis of the developmental expression of HCAI and globin mRNA

Northern analysis on total RNA from cord bloods of babies born at 18 wk, 37 wk and 40 wk gestation as well as an adult sample. The RNAs were probed with ^{32}P labelled HCAI cDNA and human β -globin cDNA.

Fig. 5.2C Schematic representation of cell fusion experiments

The diagram illustrates the potential reprogramming of gene expression in the transient heterokaryons formed by cell fusion. Genes that would be newly activated in the nuclei of heterokaryons have been boxed in. M = mouse, H = human, α = α -globin, β = β -globin, ϵ = ϵ -globin and γ = γ -globin.

B**C**

rabbit anti-HCAI antibody cross-reacts with mouse CA. However, as CAI, II and III all have the same M.W. further analysis was required to show CAI specificity.

The isoenzymes CAI, II and III can be separated on the basis of their different isoelectric points by the technique of isoelectric focusing (IEF). Therefore, the MEL C88 cell lysate was resolved by IEF under conditions optimal to separate the CA isoenzymes (H. Isenberg, U.C.L., personal communication). Under these conditions CAI runs to the anode, CAII to the cathode and the two isoforms of CAIII in between (Tweedie & Edwards, 1989). The result of MEL cell lysate IEF, blotting and probing as above is shown in Fig. 5.1B. Mouse CAI was present, having run to anode and cross-reacted with the antibody. There was also a very slight cross-reaction with CAII, which had run to the cathode.

Northern analysis was carried out on total RNA from K562 SAI and K562 suspension cells (both haemin-induced), HeLa cells and HEL cells. The probe used was ³²P labelled HCAI cDNA. The results are shown in Fig. 5.2A. HCAI is expressed in HEL cells (lanes 1 and 4) but is not present in K562 SAI, K562 or HeLa (lanes 2, 3 and 5).

Northern analysis was also used to confirm the stage-specific activation of HCAI in erythroid tissue. Total RNA, from cord bloods of babies born at 18 wk, 37 wk and 40 wk gestation as well as an adult sample, was probed with HCAI cDNA and also with human β -globin cDNA (gift from Dr F. Grosveld, N.I.M.R., London). The results are shown in Fig. 5.2B. The β -globin gene shows the expected pattern of developmental expression (Fig. 1.4). β -globin gene transcripts

were present as early as 18 wk gestation (lane 1) but at low levels compared to their abundance in the adult sample. HCAI mRNA was not detectable at 18 wk but was apparent in the 37 wk, 40 wk and adult samples. This reflects the situation of CAI protein content in erythroid cells which is not detectable before 36 wk but thereafter rises to adult levels (Dr D. Linch, U.C.M.H.M.S., personal communication).

5.2 Transient heterokaryon formation

The cell fusion studies first involved forming transient heterokaryons between MEL C88 cells and K562 cells (initially the suspension-grown subclone of K562 was used, but all the results described refer to heterokaryons with K562 SAI cells as these proved more amenable to fusion). As shown above, MEL C88 cells express the mouse CAI gene whereas K562 SAI cells (haemin-induced) do not express HCAI. It is also shown above that in normal development of the human erythroid lineage, HCAI expression is not characteristic of an embryonic/foetal phenotype but rather of the adult. The adult phenotype MEL cells do not express the mouse embryonic globin, designated ϵ -globin (Hansen et al., 1982). The embryonic/foetal phenotype K562 SAI cells however, constitutively express the human ϵ -globin gene and 3- to 5- fold stimulation of synthesis of this globin gene is observed upon exposure to haemin (Rutherford et al., 1979).

The idea behind forming heterokaryons between MEL and K562 cells is to see whether the embryonic / foetal phenotype of the K562 can activate embryonic / foetal type gene expression in the adult phenotype of the MEL and vice versa.

This is illustrated schematically in Fig. 5.2C. In other words, after fusion, to detect mouse ϵ -globin transcripts and human CAI transcripts. In a previous investigation the activation of mouse ϵ -globin in MEL X K562 heterokaryons has been shown (Baron & Maniatis, 1986). In the work described here the activation of mouse ϵ -globin was sought to show that a viable cell fusion system had been created to allow the investigation of HCAI activation.

The fact that the heterokaryons are formed between cell lines from different species make the definitive detection of gene activation easier. This is because there are sufficient nucleotide sequence differences between the mouse and human ϵ -globin and CAI genes to allow highly specific nucleic acid probes to be constructed.

The MEL C88 X K562 SAI (haemin-induced) and MEL C88 X HeLa fusions were carried out as detailed in Section 2.23, based upon the method developed in Baron & Maniatis, 1986. The cells to be fused were first adhered at high density to tissue culture plates that had been treated with poly-L-lysine. The cells were treated with the lectin phytohaemmagglutinin-P to promote cell-to-cell contact. They were then fused using PEG 1000 and incubated for 48 hours before harvesting and isolation of total cellular RNA. Successful fusion between cells was initially judged by Giemsa staining and phase-contrast microscopy.

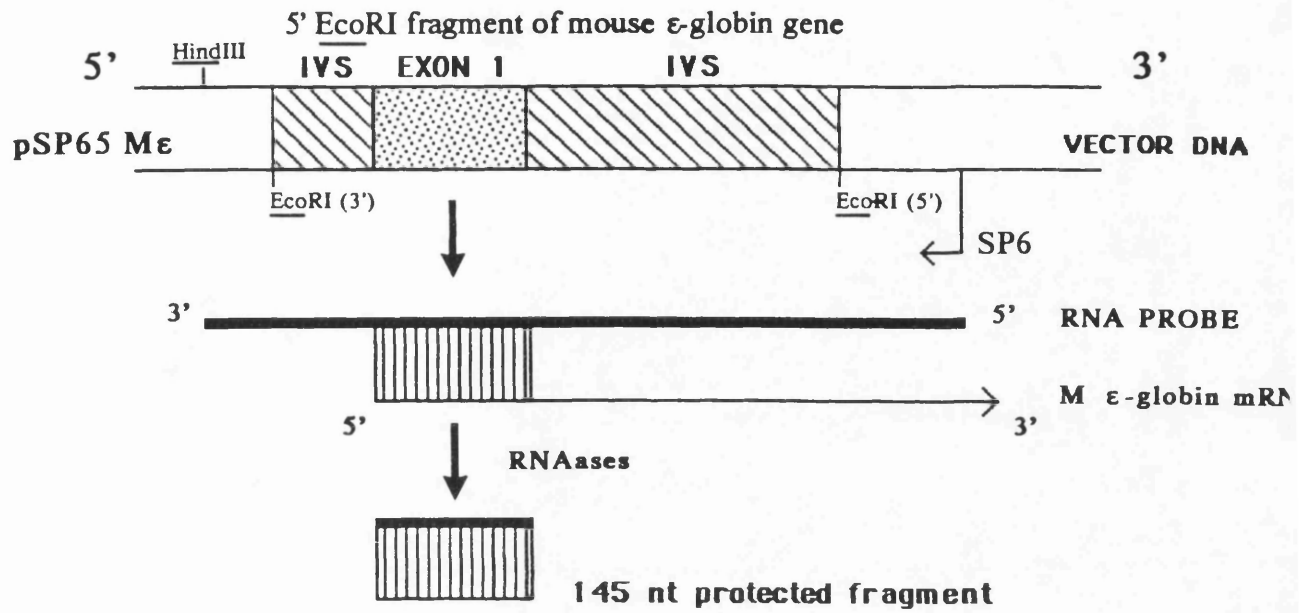
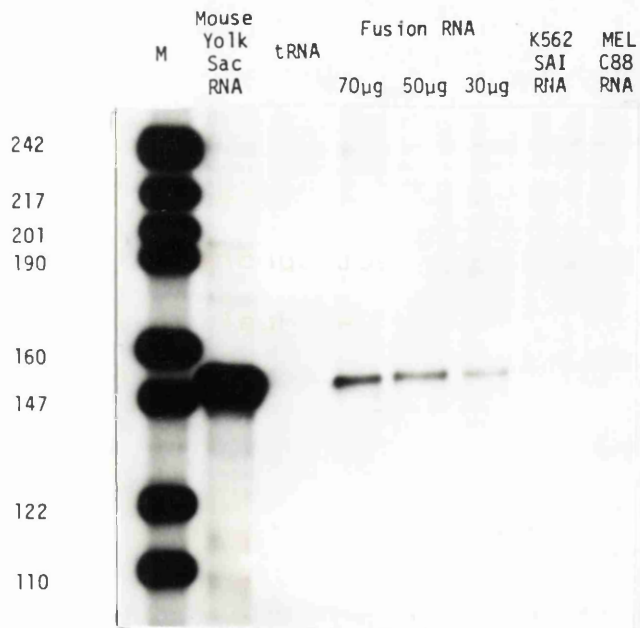
5.3 Analysis of mouse ϵ -globin gene activation

To determine whether the previously non-expressed embryonic globin gene was activated in the mouse nuclei of the

Fig. 5.3 Detection of mouse ϵ -globin transcripts in transient heterokaryons

A Diagram illustrates the SP6-derived probe, pSP65Me, used to detect mouse ϵ -globin transcripts. The mouse ϵ -globin probe was prepared by transcription from a SP6 plasmid containing exon 1 of the mouse ϵ -globin gene cloned in the anti-sense orientation. The uniformly labelled RNA probe was hybridised with various RNAs followed by RNAase digestion.

B RNAase protection analysis of mouse ϵ -globin transcripts. pSP65Me DNA was linearised by HindIII digestion before transcription in the presence of SP6 polymerase and ^{32}P - α -UTP. 2×10^6 cpm of pSP65Me RNA probe was hybridised at 45°C overnight to various cellular RNAs. The resulting RNA-RNA hybrids were treated with RNAase and the final digestion products analysed by electrophoresis on a 5% denaturing polyacrylamide gel. M = pAT HpaII markers and the other labelling refers to the RNAs hybridised with the probe.

A**B**

K562 X MEL heterokaryons, the total cellular RNA was analysed by RNAase protection. Since no metabolic selection was used to eliminate unfused human and mouse cells, the RNA was derived from a population of cells that included unfused cells, homokaryons and heterokaryons. This posed a serious problem for detection because, on average, as judged by phase-contrast microscopy and Giemsa staining, only 5% of cells in any one plate were heterokaryons. This meant that the activated ϵ -globin transcript had to be detected against a very high background of other RNAs. The problem was even more acute in the case of the HCAI transcript, as discussed below. The requirement for high sensitivity of detection eventually dictated the use of uniformly ^{32}P labelled single-stranded RNA probes. Earlier attempts to detect activated mouse ϵ -globin mRNA using a ^{32}P end-labelled oligonucleotide, made specific to mouse ϵ -globin cDNA sequence, in primer extension analysis, were not successful. Primer extension with the same oligonucleotide, in the presence of ^{32}P - α -dCTP (to be incorporated in the extension product) also failed.

The generation of the uniformly labelled single-stranded RNA probe, hybridisation, RNAase protection and subsequent resolution of protected fragments was carried out as described in Section 2.20 based on the original method of Melton et al (1984). The probe used for mouse ϵ -globin detection was prepared by transcription from a SP6 plasmid containing exon 1 of the mouse ϵ -globin gene cloned in the anti-sense orientation. This plasmid, designated pSP65Me, is shown schematically in Fig. 5.3A. It was constructed by Baron & Maniatis (1986), and was a gift from Dr E. Whitelaw, Oxford.

Probe was generated by transcription from linearised pSP65Me in the presence of SP6 polymerase and 35 μ Ci of 410Ci/mmol ³²P- α -UTP. 2 X 10⁶ cpm of the probe was hybridised with total cellular RNA. Resulting RNA-RNA hybrids were treated with RNAase and the final digestion products analysed by electrophoresis on a 5% denaturing polyacrylamide gel.

The result of this experiment is shown in Fig. 5.3B. Lane 1 shows the result of using the pSP65Me probe in RNAase mapping of total RNA from a 10 day mouse embryo yolk sac (a gift from Dr N. Dillon, N.I.M.R., London). The mouse yolk sac has high levels of mouse ϵ -globin gene expression and hence a strong protected fragment of 145 nt is seen. This size corresponds to that expected for accurately initiated transcripts from the mouse ϵ -globin gene using this probe (Baron & Maniatis, 1986). Total RNA from haemin-induced K562 SAI and MEL C88 cells is analysed in lanes 6 and 7 respectively. No protected fragment is seen since neither express mouse ϵ -globin. Lane 2 shows a tRNA control for non-specific protection, of which there is none. The total RNA from fused mixtures of MEL X K562 cells is analysed in lanes 3-5. A protected fragment at 145 nt, equivalent to that seen in yolk sac RNA, is present in each track, at decreasing intensity with decreasing amount of RNA analysed. Therefore, transient heterokaryon formation between the MEL C88 and haemin-induced K562 SAI cells has resulted in trans-activation of the previously dormant mouse ϵ -globin gene. It has been shown previously that this process is not an artefact of PEG treatment (Baron & Maniatis, 1986). Furthermore, no mouse ϵ -globin was detected with RNA from a mixture of MEL and K562

cells treated precisely as above except without the PEG fusion step (data not shown).

5.4 Construction of a HCAI-specific probe for analysis of cell fusion RNA

Having established that trans-activation of dormant gene expression was occurring in the system, it was then possible to investigate for the presence of HCAI transcripts. This was more problematical than the detection of ϵ -globin since the level of expression of HCAI in erythroid cells is very much lower than that of globin genes.

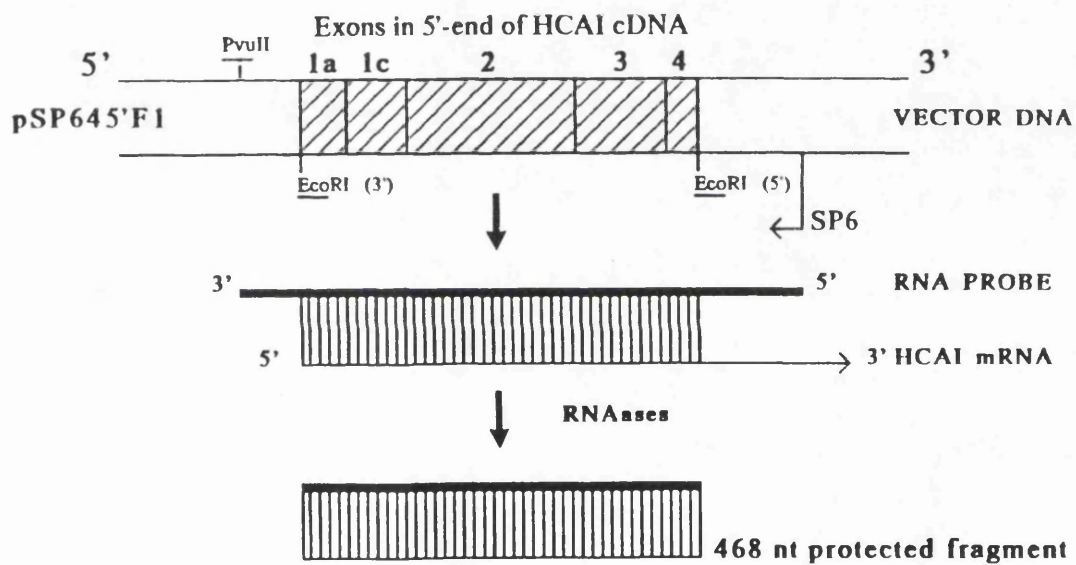
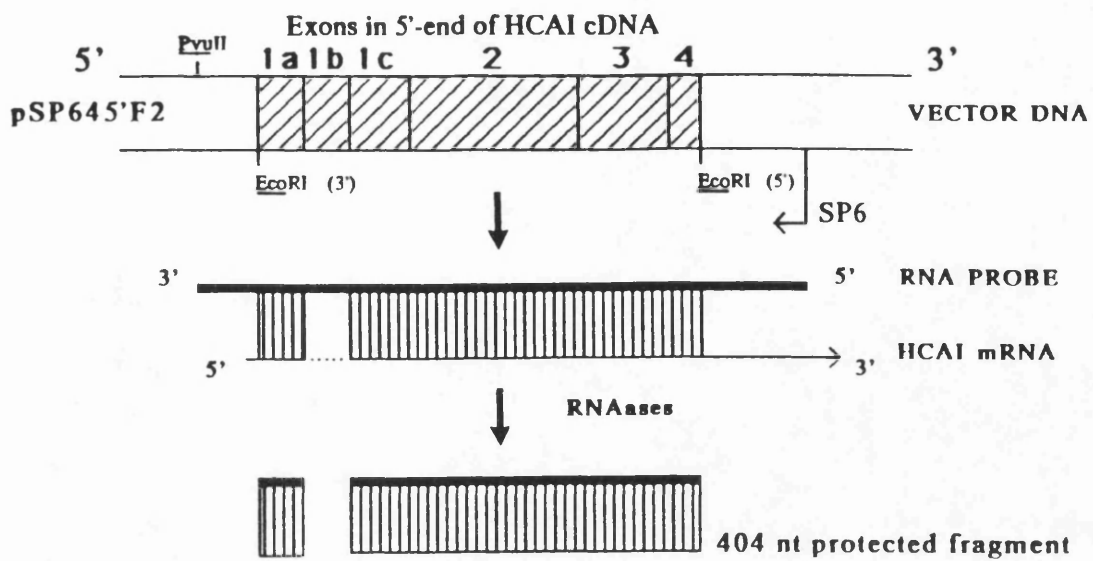
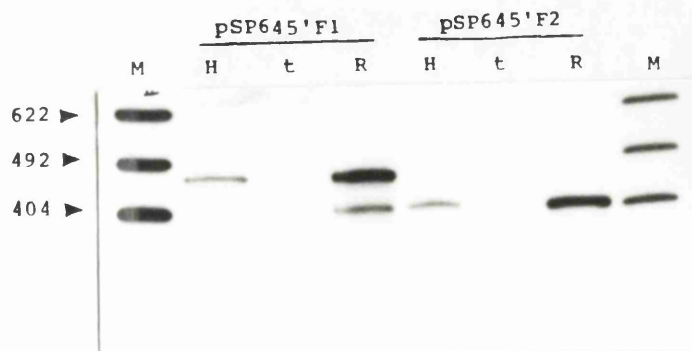
It would have been desirable to use a probe covering the transcription start site, as in pSP65Me however, the structure of the 5'-untranslated region made this impractical. Exon 1a is only 68 bp long. The PvuII-HindIII fragment containing exon 1a (see Fig. 3.9A) was subcloned in anti-sense orientation into pSP64. However, the probe generated (labelled with high specific activity ^{32}P - α -UTP), produced a 68 bp protected fragment which was not an efficient detector of HCAI mRNA in total RNA from normal adult human reticulocytes (data not shown). On this basis, the same probe could not be used to detect much lower levels of HCAI mRNA, if present, in the fused cell RNA.

Instead, a more suitable probe was constructed from the 5'-end of HCAI cDNA. As mentioned in Section 3.2, and shown in Fig. 3.3C, two types of 5'-HCAI cDNA are found. The most abundant contains most of exon 1a and exon 1c. A second, very rare type, contains the same piece of exon 1a, the 1b element and exon 1c. The EcoRI fragments from the cDNAs containing

Fig. 5.4 Detection of HCAI transcripts in transient heterokaryons

A Diagram illustrates the SP6-derived probes used to detect HCAI transcripts. These probes contain the two different forms of 5'-HCAI cDNA. The EcoRI fragments from the cDNAs containing each of these regions were subcloned in anti-sense orientation into pSP64. The plasmid containing exon 1a and exon 1c was termed pSP645'F1 and that containing exon 1a, the 1b element and exon 1c designated pSP645'F2.

B RNAase protection analysis of HEL cell and total human reticulocyte to assess the usefulness of pSP645'F1 and pSP645'F2. Both plasmids were linearised by PvuII digestion and a uniformly labelled RNA probe generated from each by transcription in the presence of SP6 polymerase and ³²P-α-UTP. Each probe was hybridised with HEL cell, reticulocyte RNA and yeast tRNA as a control, followed by RNAase digestion and resolution by denaturing polyacrylamide gel electrophoresis.

A**B**

these regions were subcloned in anti-sense orientation into pSP64 (Fig. 5.4A). The plasmid containing 1a and 1c only, was termed pSP645'F1 and the other plasmid designated pSP645'F2. To assess the usefulness of the two probes, they were used in RNAase protection assays of total RNA from HEL cells and adult human reticulocytes. The result is shown in Fig. 5.4B.

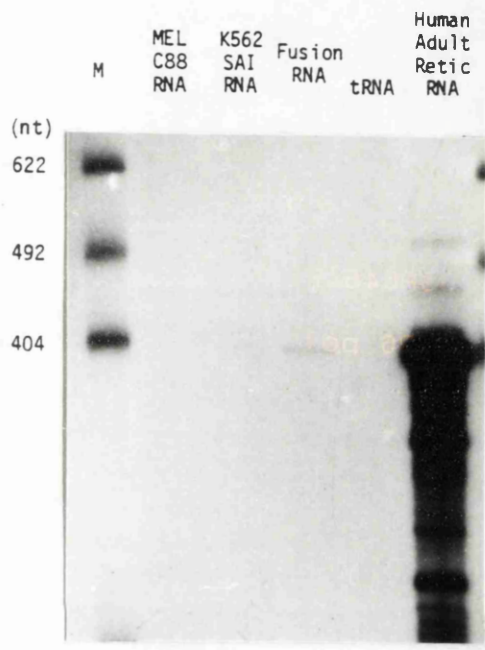
pSP645'F1 produced a protected fragment of 468 nt with both HEL and reticulocyte RNA (lanes 1 and 3) and pSP645'F2 a protected fragment of 404 nt with HEL and reticulocyte RNA (lanes 4 and 6). The pSP645'F2 protected fragment is 404 nt because the 1b element is not present in the vast majority of HCAI mRNA transcripts and therefore, RNAase protection of the probe only extends to the 5'-end of exon 1c. Both probes produced strong protected fragments however, in the tRNA controls, weak artefactual bands were visible at sizes close to that of the genuine protected fragments (lanes 2 and 5). This was of significance because it would seriously interfere with definitive identification of the presence of low levels of HCAI mRNA in the fused cell RNA. Less artefactual bands were visible with the pSP645'F2 probe (lane 5) than with the other probe. Therefore, pSP645'F2 was used to analyse the fused cell RNA.

5.5 Analysis of HCAI gene activation

Probe was generated from linearised pSP645'F2 in the presence of SP6 polymerase and 50 μ Ci of 3000Ci/mmol 32 P- α -UTP. 1 X 10⁶ cpm of probe / hybridisation was then used in an RNAase protection assay, the result of which is shown in Fig. 5.5. Lane 5 showed a strong protected fragment in human adult

Fig. 5.5 Detection of HCAI transcripts in transient heterokaryons

RNAase protection analysis of HCAI transcripts. The probe was generated from pSP645'F2, linearised by PvuII digestion, in the presence of SP6 polymerase and 50 μ Ci of 3000Ci/mmol ³²P- α -UTP. 1 X 10⁶ cpm of probe was hybridised at 45°C overnight to various cellular RNAs. The resulting RNA-RNA hybrids were treated with RNAase and the final digestion products analysed by electrophoresis on a 5% denaturing polyacrylamide gel.



reticulocyte RNA as expected at 404 nt. Neither MEL C88 or haemin-induced K562 SAI cell RNA had any HCAI transcripts (lanes 1 and 2). No artefactual bands were seen in the tRNA control (lane 4). Two factors aided the removal of the artefactual bands: less probe was added than for the test experiment (Fig. 5.4B) but that which was added was labelled to a much higher specific activity and the temperature of RNAase digestion was increased to 34°C from 30°C and continued for an extra 15 min. Analysis of total RNA from the fused mixture of MEL X K562 cells is shown in lane 3. A single protected fragment of 404 nt, equivalent to that in the reticulocyte RNA, was present. When RNA from a mixture of MEL C88 and haemin-induced K562 SAI cells which had been treated as the fused cells, except without any PEG 1000 treatment, no HCAI transcripts were detected (data not shown). Therefore, the formation of MEL X K562 transient heterokaryons had resulted in the activation of the previously dormant HCAI gene in the human cell nuclei.

Transient heterokaryons were also formed (as judged by phase-contrast microscopy and Giemsa staining) between MEL C88 and HeLa cells. However, analysis of total cellular RNA from several different sets of fused mixtures of these cells failed to show activation of HCAI in the nuclei of the non-erythroid HeLa cells.

This chapter has demonstrated that MEL C88 cells express the mouse CAI gene whereas haemin-induced K562 SAI cells do not express the HCAI gene. The HCAI mRNA has also been shown to be expressed in a developmental stage-specific manner.

Transient heterokaryons have been made between MEL C88

cells and haemin-induced K562 SAI cells and between MEL C88 cells and HeLa cells. In the MEL C88 X K562 SAI heterokaryons mouse ϵ -globin and human CAI gene expression has been activated. This is believed to be due to diffusible stage-specific trans-acting factors.

CHAPTER SIX

TRANSCRIPT ANALYSIS OF RABBIT CAI AND FURTHER TRANSCRIPT ANALYSIS OF HCAI, REVEALING A SECOND, NON-ERYTHROID PROMOTER

The work so far described in this thesis has focussed on the characterisation and regulation of the human CAI gene, particularly with regard to its expression in erythroid tissue. During the course of this work I also examined two other aspects of CAI expression: (i) the transcript of the rabbit CAI (RCAI) gene and (ii) the non-erythroid-specific transcript from the HCAI gene.

6.1 Completing the nucleotide sequence of the protein-coding region of the rabbit CAI cDNA

Previous work in this laboratory led to the isolation of a cDNA clone for RCAI (Konialis et al., 1985). The amino acid sequence deduced from this clone began at residue 25 through to the end of the protein coding region. The cDNA contained a further 360 bp of 3'-untranslated sequence. I decided to further characterise the transcript by primer extension analysis and RNA sequencing. These two techniques work on the same principle. From the RCAI cDNA sequence already available an oligonucleotide primer was synthesised, complementary to RCAI mRNA. The primer was ³²P end-labelled, annealed to poly (A+) rabbit reticulocyte RNA and the cDNA strand synthesised using reverse transcriptase. The latter step was carried out solely with dNTPs present for primer extension analysis, and for direct transcript RNA sequencing, in the presence of a

Fig. 6.1 Primer extension analysis of RCAI mRNA

A Diagram showing the RCAI 21mer primer in the 5'-end of the RCAI cDNA as deduced in Konialis et al (1985). The heavy black line represents the 21mer (3'-CACTTTGTACTGTGGAGAGAC-5') which is complementary to the coding strand.

B Diagram showing the primer and extension product of primer extension analysis of the 5' end of RCAI mRNA.

C 3µg of poly (A+) rabbit reticulocyte RNA was co-precipitated with 2-3ng of the ³²P end-labelled RCAI 21mer primer and resuspended in 20µl of 250mM KCl and 10mM Tris.HCl (pH8.3). The solution was heated to 80°C for 3 min and then the primer annealed at 57°C for 45 min. The RNA/DNA hybrid was then reverse transcribed in the presence of AMV-RT for one hour at 50°C. Following RNAase digestion the extension product was resolved on an 8% polyacrylamide / 50% urea.

mixture of ddNTPs and dNTPS.

The primer synthesised was a 21mer with sequence (3' CACTTTGTA CTGTGGAGAGAC 5'), complementary to the mRNA coding for amino acid residues 38-44 (Fig. 6.1A). The primer, 5'-end labelled with ^{32}P - γ -ATP, was then used in primer extension analysis of 3 μg of poly (A+) rabbit reticulocyte RNA, as described in Section 2.17. The primer and its extended product are shown diagrammatically in Fig. 6.1B and the autoradiograph of the actual gel in Fig. 6.1C. The extended product is about 232 nt long. This allows the distance between the transcription start site and the translation start site to be calculated. The coding sequence, including the initiation methionine codon, contains 45 amino acid residues up to the end of the primer. This is coded for by 135 nt hence, the 5'-untranslated region is 97 nt long.

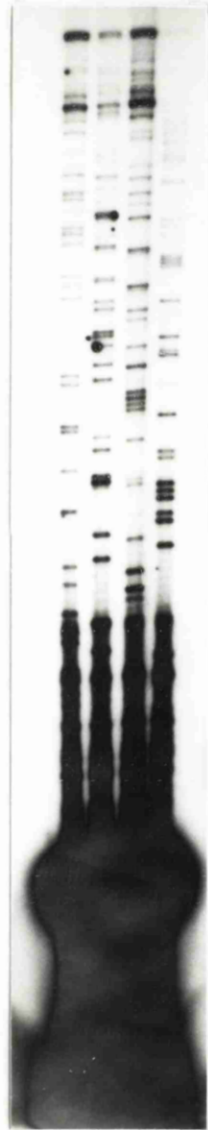
It is possible to compare this with the 5'-untranslated region of the erythroid-specific transcript of the two CAI genes that have been cloned i.e. HCAI, as described in this thesis and mouse CAI (MCAI) (Fraser et al., 1989). HCAI mRNA has a 5'-untranslated region (i.e. exon 1a + part of exon 1c) of 92 nt and MCAI, one of 87 nt. Therefore, the length of the 5'-untranslated region of the CAI gene in erythroid tissue is quite well conserved across species.

Using the same primer and 10-12 μg of poly (A+) rabbit reticulocyte RNA, sequence of the cDNA extended product was obtained by the method of direct transcript-"RNA sequencing"-as described in Section 2.19. The autoradiograph of the RNA sequencing gel is shown in Fig. 6.2. Sequence was obtained which covered the previously uncloned protein coding sequence.

Fig. 6.2 An RNA sequencing gel

Using the RCAI 21mer primer and 10-12 μ g of poly (A+) rabbit reticulocyte RNA, the sequence of the cDNA extended product was obtained by the method of direct transcript -"RNA sequencing" based on Geliebter (1987). This diagram shows an example of an RNA sequencing gel.

A C G T



← ³²P Labelled
Primer

The three amino acid residues marked with an asterisk are different from the RCAI protein sequence given in Hewett-Emmett et al (1984). In the latter residues 10, 11 and 24 are glutamate, methionine and aspartate respectively, whereas they are glutamine, asparagine and asparagine as derived from the transcript sequencing. Of this region only residue 7 is thought to play a role in the active site of CAI and this is conserved as a tyrosine. Comparison of the nucleotide sequence and protein sequence for the first 25 amino acids of RCAI with MCAI and HCAI shows strong conservation (Fig. 6.3A).

6.2 The RCAI cDNA is homologous to a novel transcript in kidney cells

As mentioned previously, there are thought to be at least 7 different CA genes. CAI, II and III are all found in the cytoplasm but the other gene products are thought to be non-cytoplasmic. In particular a membrane-bound form has been shown to present in the human kidney (Wistrand et al., 1984). This human kidney enzyme (CAIV) has been purified by differential centrifugation and affinity chromatography. It has an apparent subunit molecular mass of 68-70 kd. The amino acid content of the CAIV isoenzyme suggests it is different from the other CAs (ibid.). However, this is not necessarily significant, since if CAIV were to contain, for instance, a large transmembrane domain, this would distort its overall amino acid content relative to CAI. Of more significance could be the fact that when amino acid residues are compared only CAI and CAIV do not contain tryptophan residues. However, preliminary studies (ibid.) suggest that the antigenic profile

Fig. 6.3A Comparison of the nucleotide sequence and protein sequence over the first 25 amino acids of RCAI with MCAI and HCAI. The asterisks on the nucleotide sequence denote mismatches whereas the asterisks on the RCAI protein sequence denote deviations from that previously published for RCAI in Hewett-Emmett et al (1984).

Fig. 6.3B Northern analysis for CAI homologous transcripts in kidney cell lines. 30µg of total RNA from two human renal carcinoma cell lines, A498 and Caki-1, HEL cells and adult human reticulocytes were probed with ³²P HCAI cDNA. 30µg of total RNA from an untransformed rabbit kidney cell line, LLC-RK1 and 2µg of poly (A+) rabbit reticulocyte RNA were probed with ³²P labelled RCAI cDNA.

A

```

                    5                               10                               15
HCAI  ATG GCA AGT CCA GAC TGG GGA TAT GAT GAC AAA AAT GGT CCT GAA CAA>

RCAI  ATG GCA AGT tct GAC TGG GcA TAT GAT GgC cAA AAT GGT CCT GAg CAc>
          * *                *                * *                * *

MCAI  ATG GCA AGT gCA GAC TGG GGA TAT Gga agC gAA AAT GGT CCT GAc CAA>
          *                ** ** *                *

                    20                               25
HCAI  TGG AGC AAG CTG TAT CCC ATT GCC AAT GGA

RCAI  TGG AGC AAG CTG TAc CCC ATT GCg AAT GGA
          *                *

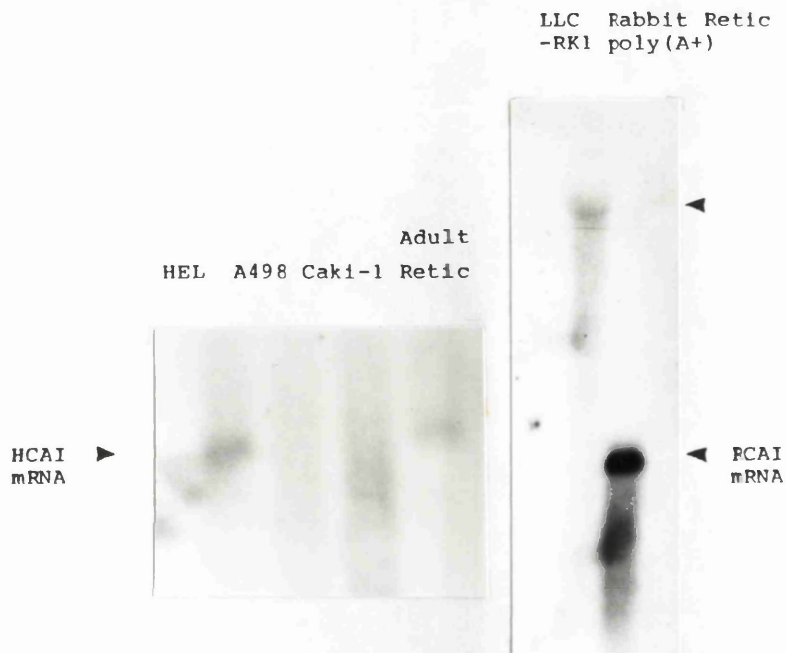
MCAI  TGG AGC AAG CTG TAT CCC ATT GCC AAT Ggt
          *
  
```

```

                    1                               5                               10
HCAI  Ala Ser Pro Asp Trp Gly Tyr Asp Asp Lys Asn Gly Pro
MCAI  Ala Ser Ala Asp Trp Gly Tyr Gly Ser Glu Asn Gly Pro
RCAI  Ala Ser Ser Asp Trp Ala Tyr Asp Gly Gln Asn Gly Pro

                    15                               20                               25
HCAI  Glu Gln Trp Ser Lys Leu Tyr Pro Ile Ala Asn Gly
MCAI  Asp Gln Trp Ser Lys Leu Tyr Pro Ile Ala Asn Gly
RCAI  Glu His Trp Ser Lys Leu Tyr Pro Ile Ala Asn Gly
          *
  
```

B



of CAIV and CAI are different.

It was decided to check whether or not there might be any homology at the nucleotide level of CAI with the kidney CAIV. This was first approached by using the HCAI cDNA (Barlow et al., 1987) to screen total RNA isolated from two human kidney transformed cell lines, A498 (American Type Culture Collection-human tumour bank, cell line 44) and Caki-1 (Fogh and Trempe, 1985). These are both renal carcinoma lines. Northern analysis of RNA from the kidney cell lines, HEL cell and adult human reticulocyte RNA are shown in Fig. 6.3B. The HCAI cDNA probe hybridises to the HCAI mRNA in the HEL and reticulocyte RNA (lanes 1 and 4) but does not hybridise to any transcript in the two kidney RNAs at the stringency used (lanes 2 and 3).

Since the RCAI cDNA was available it was used to screen a rabbit kidney cell line. The cell line used was LLC-RK1 (Hull et al., 1965) which was immortalised but not transformed. Total RNA from this line was probed with a section of pRCAI (Konialis et al., 1985). The probe was a 2.5 kb fragment isolated from a PvuI-AvaII digest of pRCAI and contained the RCAI cDNA. As a control, poly (A+) rabbit reticulocyte RNA was run alongside the LLC-RK1 RNA. The result of the probing is seen in Fig. 6.3B. The RCAI cDNA hybridises to an approx. 1.5 kb HCAI mRNA transcript in the reticulocyte RNA (lane 6). However, it also hybridises to a much larger transcript in the LLC-RK1 RNA (lane 5). The LLC-RK1 RNA does not contain the 1.5 kb mRNA found in the reticulocyte RNA.

Using the 28S and 18S ribosomal RNAs as markers, the size of the rabbit kidney transcript is about 4.9 kb. This size of

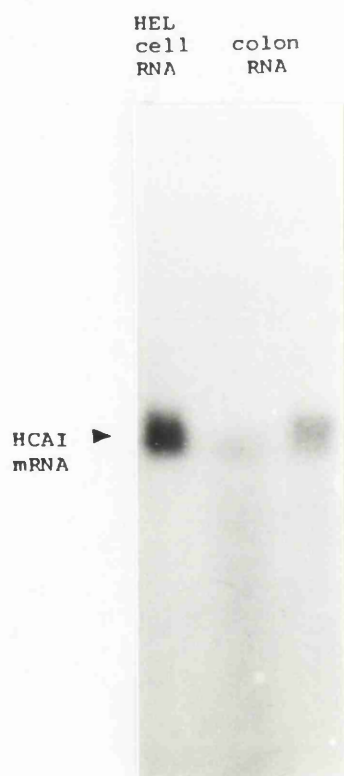
transcript will code for a large protein, much larger than the 28-29 kd cytoplasmic CA isoenzymes. Using the size estimate of CAIV, and neglecting the contribution of carbohydrate residues, this would require a coding region of 636 amino acids to code for a 70 kd protein. However, only just over 1.9 kb of transcript is required to code for 636 amino acids which still leaves a further 3 kb of the unidentified LLC-RK1 transcript unaccounted for. This "extra" amount would seem inordinately large to be due to 5'- and 3'-untranslated regions and poly (A) tail. The fact that the transcript is present in kidney and not in reticulocyte RNA suggests some measure of specificity. The possibility exists that the size estimate for CAIV is quite wrong. It is also possible that the LLC-RK1 transcript which hybridises to RCAI cDNA is in fact not CAIV. Further work is required to settle this question. The homology of the cDNA to the transcript may well be sufficient to allow clones to be isolated from a rabbit kidney cDNA library. These clones could be sequenced to definitively characterise them and the protein for which they code.

6.3 Human colon expresses HCAI: Northern and primer extension analysis

Another facet of CAI gene transcription is that it would seem to be under the control of two separate promoters, one erythroid-specific and the other active in the colon, if not more generalised. This has come to light from recent work on the MCAI gene (Fraser et al., 1989). Using primer extension analysis and RNA sequencing of the 5'-end of the transcript, the MCAI gene has been shown to have two quite distinct start

Fig. 6.4 Northern analysis on colon RNA for HCAI mRNA

30µg of total RNAs from colon tissue and HEL cells were probed with ³²P labelled HCAI cDNA.



sites. The structure of the 5'-end of the MCAI gene is also similar to that described in this thesis for the HCAI 5'-end. In the way that the HCAI exons 1a and 1c are separated by a 36.5 kb intron, the exons containing the transcription and translation start sites are separated by an intron of >10 kb in the MCAI gene (ibid.). Therefore, it seemed likely that the HCAI gene would also have the same type of promoter structure.

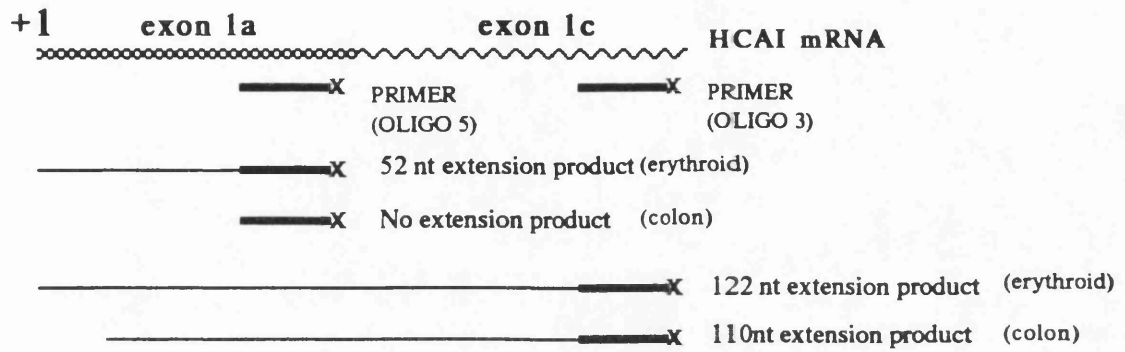
The first step in investigating the non-erythroid transcript of the HCAI gene was to find a non-erythroid tissue which expressed HCAI. Histochemical studies had previously suggested that CAI protein was present in intestinal and corneal epithelium, vascular endothelium and eye lens as well as erythroid tissue (Spicer et al., 1979). The work on MCAI had used RNA isolated from whole mouse large intestine. It was not feasible to obtain a whole large intestine from human from which undegraded RNA could be prepared. A colorectal cell line, SW 480 (Leibovitz et al., 1976) was investigated for the presence of HCAI RNA but proved negative. Fortunately, sections of gut could be acquired as a by-product of intestinal resectioning operations. These were kindly provided by Dr P. Boulos, Surgery, U.C.H. RNA was only prepared from areas of intestinal epithelium from the proximal and distal colon and caecum.

Total colon RNA was then probed with ³²P labelled HCAI cDNA in Northern analysis (Fig. 6.4). The HCAI cDNA hybridises to the HCAI mRNA present in HEL cells (lane 1). It also hybridises to a similar size transcript in colon RNA (lanes 2 & 3). The colon RNA in lane 2 is degraded, hence the weak smeary band. Thus, it would appear that HCAI mRNA is present

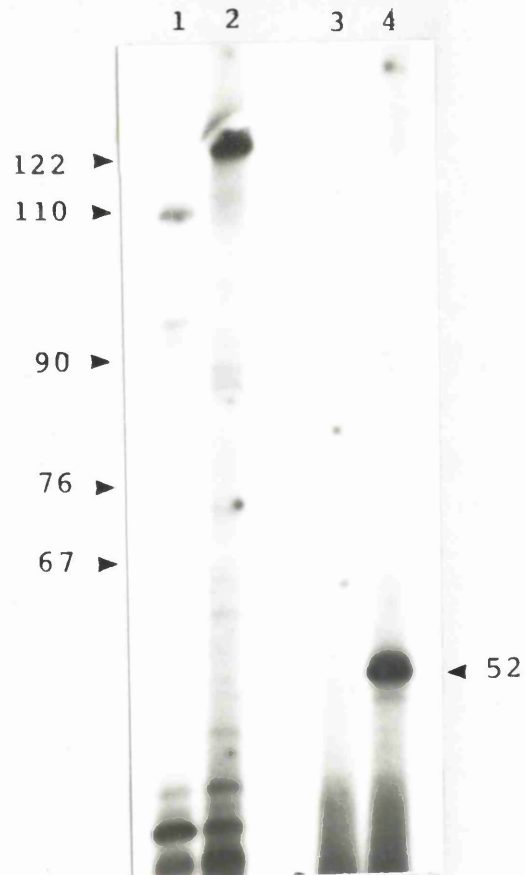
Fig. 6.5 Primer extension analysis of HCAI mRNA in colon tissue

A Diagram showing the primers and extension products of primer extension analysis of the 5'-end of HCAI mRNA in both colon and erythroid tissue.

B 20µg of either total human colon or human reticulocyte RNA was co-precipitated with 2-3ng of either ³²P end-labelled oligo #3 or ³²P end-labelled oligo #5 and resuspended in 20µl of 250mM KCl and 10mM Tris.HCl (pH8.3). The solution was heated to 80°C for 3 min and then both primers annealed at 57°C for 45 min. The RNA/DNA hybrids formed were then reverse transcribed in the presence of AMV-RT for one hour at 50°C. Following RNAase digestion the extension products were resolved on a 10% denaturing polyacrylamide gel.

A

(x denotes ^{32}P labelling)

B

- 1 Colon RNA primed with oligo 3
- 2 Retic RNA primed with oligo 3
- 3 Colon RNA primed with oligo 5
- 4 Retic RNA primed with oligo 5

in colon tissue. A similar experiment probing caecum RNA produced no hybridising transcript (data not shown). Therefore, CAI expression is localised within the large intestine. Earlier studies on CA in the gastrointestinal tract of guinea-pig (Carter and Parsons, 1971), using protein gel electrophoresis and activity staining, identified a specific distribution of CAI and CAII. CAII was only present in the stomach and proximal colon, whereas CAI was found in the caecum and both proximal and distal colon but not in the stomach. Both were present in the small intestine but at low levels. The absence of HCAI mRNA in human caecum is at variance with the situation in guinea-pig.

The fact that Northern analysis shows two transcripts of similar size in colon and erythroid tissue, does not give any information as to different start sites of transcription. This was obtained from primer extension analysis of human colon and reticulocyte RNA. Two oligonucleotide primers were used, as in Section 3.4. Oligo #5 which lies within exon 1a and oligo #3 which lies within exon 1c (Fig. 6.5A). The rationale behind this is based on the MCAI situation (Fraser et al., 1989). The mouse equivalent to exon 1c is continuous with the colon/non-erythroid start site. If the same arrangement is found in the HCAI gene then the 1c primer should result in an extended product in colon RNA but the 1a primer should not, since exon 1a would not be present in that transcript.

The result of the primer extension analysis is shown in Fig. 6.5B. Using the 1a primer a correctly initiated 52 nt product is found in reticulocyte RNA (lane 4) but no extension product is present in colon RNA (lane 3). The 1c primer

Fig. 6.6A Sequence of the region flanking the postulated HCAI gene transcription start site in colon tissue

Fig. 6.6B Comparison of the HCAI and MCAI colon promoter region

Fig. 6.6C Proposed structure for the 5'-region of the HCAI gene. The scale is numbered in kb from the transcription start site. Exons are indicated by black bars (coding sequence) or open boxes (non coding sequence).

A

ATAATTCTCTTTTACAATATAAGAAAATTAAGCAATGAAA
 CTAACATAGCCCTTGTAGATTTTTTTTACAACACCTTTTTT
 TTAGATATGTGTA^tCTTCCIGATAAGCAGAGATGATGAAAT
 AATGGCCTATTA^gAAAAGCAAATAAGTTTCTATAAAAACGCC
 CAAGCAGGGATTTAAGGCATCTCCTGCATGCACAGTTGCA
 GTTAGTTATTCCAGGTATTATTTTTGTTTTTCAGAAAAAGA
 of erythroid HCAI mRNA
 AAAAACTCAGTAGAAGATAATG

B

HCAI ATAATTCTCTTTTACAATATAAGAAAATTAAGCAATGAAACTAACATAGCCCTTGTAGATTTT>

MCAI ATATTT^tCa^gtTTTTAtAtTAcAA^tAaCA^tATAAAaCAATcAtgCTgACcTA-CCCc^aaTAgA-ca^t>
 * ** *

HCAI TTTACAACACCTTTTTTTTAGATATGTGTA^tCTTCCIGATAAGCAGAGATGATGAAATAATGGC>

MCAI TTTACAATACCaTTTTaTTtATgTGTaTACTTCCGTATgAGCAaAGtTGATGAAAcAgTGaa>
 *

HCAI CTATTA^gAAAAGCAAATAAGTTTCTATAAAAACGCCCAAGCAGGGATTTAAGGCATCTCCTGCAT>
 (•+1)

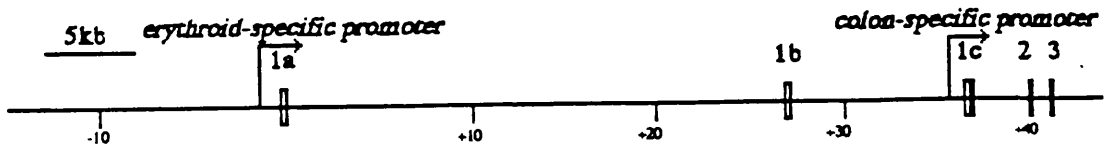
MCAI CTATTAcAcGgAAAcAAGcTTCTAT--AAAaGCCCA^gtGgtGtcAcTTAAGatcTCTgCTGCAg>
 *

HCAI GCACAGTTGCAGTTAGTTATTCCAGGTATTATTTTTGTTTTTCAGAAAAAGAAAAA^gACTCAGTA>

MCAI GtgtAagTGCAGTTAGTcATTtCAcaTATcATTTcTcTTTaCAG--gAA^gtAcAaAcCTaAaTA>
 *** ** *

HCAI GAAGATAATG>

MCAI agAGAA^gAATG>
 ** *

C

produces the expected 122 nt extension in reticulocyte RNA (lane 2). However, in colon RNA the 1c primer results in a 110 nt extension product (lane 1). Therefore, the HCAI gene uses two different start sites, one in erythroid tissue and one in colon. Working on the assumption that a similar situation holds in the HCAI gene as the MCAI gene, the transcription start site for the HCAI colon transcript will be 80 bp upstream of the ATG initiation codon and designated +1 (Fig. 6.6A).

6.4 HCAI and MCAI: nucleotide sequence comparison of colon promoter region

The region upstream of exon 1c, previously thought to be intron, is now obviously of considerable interest since it apparently contains another promoter region. The sequence of some of this area, up to 150 bp beyond the start site, had been obtained previously during the sequence of exon 1c. This is described in Section 3.2 and the sequencing strategy illustrated in Fig. 3.3A. The HCAI sequence is compared to that published for the MCAI colon promoter region (Fraser et al., 1989) in Fig. 6.6B. The alignment was carried out using the "best-fit" provided by the IBI/Pustell DNA sequence analysis software from International Biotechnologies, Inc. There is a very strong homology between the HCAI and MCAI sequences, both in their 5'-untranslated regions and into their putative promoter regions. Both have a strong consensus TATA box within the recognised limiting distance from the start site. The HCAI TATA box consensus [5'-TATAAA-3'] is at -31. However, there do not seem to be any other general

promoter elements such as an Sp1 site or a CCAAT box though, curiously, there is a potential GF-1 binding site [5'-TGATAAG-3'] at -81. However, since in all probability, there is no GF-1 factor in colon tissue this should be of no functional significance.

The nucleotide comparison would seem to confirm that there is a strong case for believing the promoter arrangement of the HCAI and the MCAI genes are similar. The proposed structure for the 5'-region of the HCAI gene is shown in Fig. 6.6C. The best characterised example of a gene with two promoters is human porphobilinogen deaminase (PBGD) (Chretien et al., 1988). This gene has an erythroid-specific promoter and a housekeeping promoter. In this case the promoters are only 3 kb apart and the erythroid one is closest to the protein coding region. However, the two promoters of the PBGD lead to two transcripts which code for two slightly different isoenzymes. The alcohol dehydrogenase gene of Drosophila melanogaster is the only gene that has the same structural arrangement of promoters and 5'-non-coding exons as HCAI and MCAI (Benyajati et al., 1983). Further work is required to confirm the sequence of the 5'-untranslated region of the HCAI colon mRNA transcript. This can be done by RNA sequencing of poly (A+) colon RNA using ³²P labelled oligo #3 as primer.

The work in this chapter has firstly, completed the nucleotide sequence for the protein coding region of the RCAI mRNA and then, shown that it is homologous to a large transcript in rabbit kidney cells that may code for the membrane-bound form of CA known to be in kidney. Secondly, the

presence of HCAI mRNA in proximal and distal colon tissue has been confirmed. Furthermore, it would seem that, as in the case of MCAI, the HCAI colon mRNA is transcribed from a different promoter to erythroid HCAI mRNA. The promoter active in colon is continuous with exon 1c of the HCAI gene whereas the erythroid promoter is a further 36.5 kb upstream.

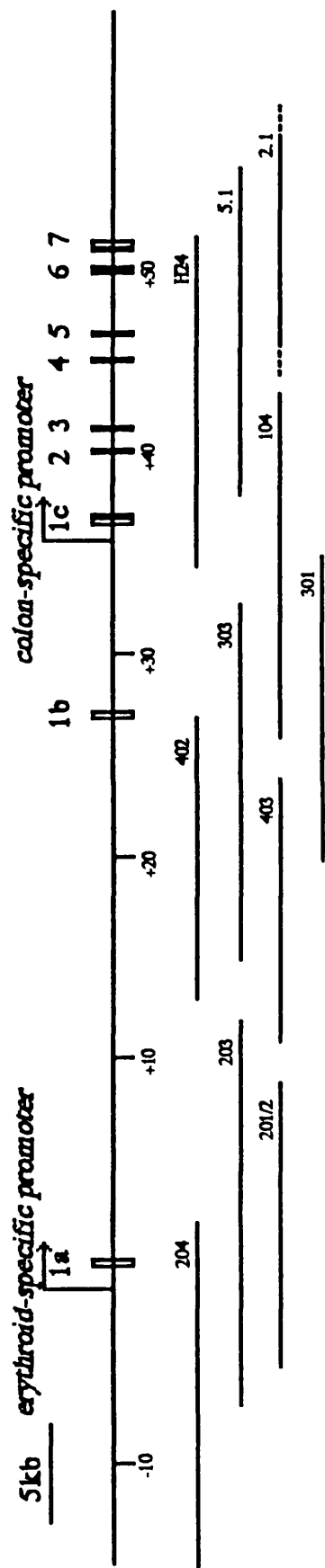


Fig. 7.1 Structure of the HCAI gene

CHAPTER SEVEN

DISCUSSION

Initially, this thesis has sought to clone and characterise the 5'-end of the HCAI gene and then to map the limits of HCAI mRNA by S1 nuclease mapping and primer extension analysis. It has also investigated the regulation of transcription of the HCAI gene by examining the role of trans-acting factors which are required for erythroid- and developmental stage-specificity. This chapter will now seek to address the significance of these results, particularly in the context of other work that has recently appeared on the transcription of erythroid genes, notably the globins.

7.1 Structure of the HCAI gene and transcription unit(s)

The structure of the HCAI gene as presently known is shown in Fig. 7.1. The results of the cloning and characterisation of the λ 104, H24 and λ 201-204 clones, followed by subsequent cloning to join exon 1c to exon 1a (by Nick Lowe in this laboratory) have placed the upstream HCAI promoter 36.5 kb from the beginning of the protein coding region. This promoter produces an erythroid-specific transcript. The 5'-end of this transcript has two alternate forms. The most common (6 out of 7 cDNAs cloned, N.Lowe, personal communication) has exon 1a and exon 1c separated by the 1b element. The 54 bp 1b element maps to a separate region in the genomic clones of the gene, between exons 1a and 1c. It is not obvious why it appears in some HCAI transcripts. It has

several characteristics of an intron (Fig. 3.3C), in that it has 5' and 3' splice site consensus sequences, a polypyrimidine tract at the 3'-end and an appropriately positioned potential "A-residue" for a branchpoint. However, it is probably too short for efficient splicing (Wieringa et al., 1984). It is also possible that the sequence flanking the 3'-end of the 1b element may be responsible for the rarity of its inclusion in the HCAI transcript. The adjacent region 3' to 1b starts with a GC, rather than the virtually invariant GT dinucleotide in 5' splice sites (Padgett et al., 1986). Unless a more suitable splice site were to exist in the region between 1b and 1c, any failure of processing at this point would produce a mRNA over 10 kb, rather than 1.4-1.5 kb normally.

Other genes have also been shown to process introns in their 5'-untranslated region. The gene which most resembles CAI, in terms of possession of an "optional exon", is the hydroxymethyl glutaryl CoA synthase gene which gives rise to a small 5'-untranslated exon (59 bp) in up to 50% of its transcripts. In this case it has been suggested that the exon has some functional role on the basis that this situation is conserved between hamster and humans (Gil et al., 1987). Mouse CAI gene transcripts are not reported to contain any "optional" exons (Fraser and Curtis, 1986; Fraser et al., 1989).

The 1b element is probably a curiosity so that essentially the region between 1a and 1c can be regarded as intron. The MCAI gene has also been shown to contain a large intron within its 5'-untranslated region (Fraser et al.,

1989). Therefore, this intron would seem to be a conserved feature of CAI gene structure. Apart from this, the exon/intron structure of the CAI gene is the same as the other CA genes that have been sequenced. The HCAI exons 1c-7 are equivalent to exons 1-7 in HCAII (Venta, 1985) and HCAIII (Lloyd, 1987). The boundaries of exons 1c-7 lie in the same positions within the protein as those in human CAIII (Lloyd et al., 1987) and chicken CAII (Yoshihara et al., 1987). They differ from mouse CAII where the junction between exons 4 and 5 interrupts a glycine codon 14 bases 5' to its position in CAI (Venta et al., 1987).

The intron between exons 1a and 1c would appear to be only present in HCAI as transcribed in erythroid tissue, since the work on colon tissue has shown that HCAI has a non erythroid promoter continuous with exon 1c. The question to be asked is why has the HCAI gene another, erythroid-specific, promoter 36.5 kb upstream of the rest of the gene? The erythroid promoter must override the downstream promoter, at least in erythroid tissue. This is seen from primer extension using the 1c primer (oligo #3) on reticulocyte RNA, which shows only transcripts initiated from the upstream, erythroid, promoter (Fig. 6.5B, lane 2). It might be then that the upstream promoter is more active than that used in colon. Alternatively, the different 5'-untranslated regions might confer differing levels of mRNA stability or translatability. However, the existing evidence suggests that there is no remarkable difference in the levels of CAI protein in red blood cells and colon epithelial cells (Spicer et al., 1979; Carter & Parsons, 1971). Another explanation for the 5'-end

structure of the HCAI gene is that the erythroid promoter and exon 1a could have been acquired from an erythroid-specific gene by duplication and translocation. It is also possible that the promoter and exon 1a could belong to an unidentified erythroid-specific gene located with the intron. However, no other example of this form of a gene within a gene has been found. In addition, if an oligonucleotide specific to the erythroid exon 1a is used to probe reticulocyte RNA, only the CAI mRNA is detected (Fig. 3.4A). There is no detectable transcript from an erythroid-specific "gene within a gene".

The actual size of the intron between exons 1a and 1c is quite intriguing. It is 10 times as big as any of the other introns in the CAI gene. It is not quite on the same scale as the first intron of the human c-abl gene which is at least 200 kb long (Bernards et al., 1987). However, similar questions must apply to how the transcripts from very large genes are processed. The length of the CAI erythroid-specific transcription unit is over 50 kb. If primary transcription from this gene is continuous it is not obvious how the correct donor and acceptor sites are identified in such a very long transcript. The possibility may exist of discontinuous transcription of genes with very large introns for instance, jumping from DNA encoding exon 1a to DNA encoding exon 1c. Looping of chromatin might facilitate the bringing together of these regions resulting in a smaller primary transcribed intron such that the donor and acceptor splice sites in the primary RNA transcript would be the only available splice sites.

At the 3'-end of the gene, two different polyadenylation

sites are used. The proximal site is 109 bp downstream from the first base of the termination codon and the distal one 334 bp. On the basis of the 3' cDNA clones isolated and Northern analysis (N.Lowe, personal communication), usage of the distal site is much greater than the proximal one. Mouse CAI cDNA is similar in that the cDNA sequence reported (Fraser & Curtis, 1986) contains an unused polyadenylation signal 60 bp downstream from the translation termination codon, while the utilised signal gives rise to a 320 bp untranslated sequence.

7.2 Several GF-1 binding sites flank the HCAI gene erythroid-specific transcription unit

The bandshift assays have shown that there are six sites flanking the erythroid-specific HCAI transcription unit (Fig. 3.11 & 3.13A) which bind an erythroid-specific factor. This factor appears to be the major DNA binding protein specific to the erythroid lineage. A cDNA encoding this factor has recently been cloned and designated GF-1 (Tsai et al., 1989) and this nomenclature is followed in this thesis. Previous to this, and during the time that that the work in this thesis was being carried out, other investigators have reported the presence of an erythroid-specific DNA binding factor. These reports have centred almost exclusively on the globin genes.

The initial investigation of erythroid-specific DNA binding was done on the chicken globin genes (Kemper et al., 1987; Evans et al., 1988). The α -like and β -like clusters of the chicken globin genes are shown in Fig. 7.2. The chicken adult β -(β^A) globin gene was shown to have a DNAaseI hypersensitive domain at its 5'-end, stretching from -40 to

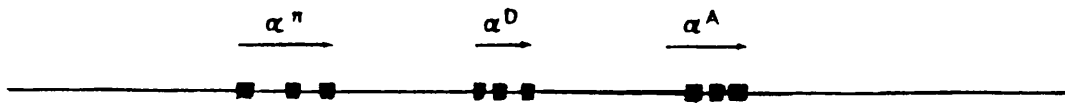


Fig. 7.2A Schematic diagram of the three chicken α -globin genes with the direction of transcription indicated by the arrows. Closed boxes indicate exons of the genes.

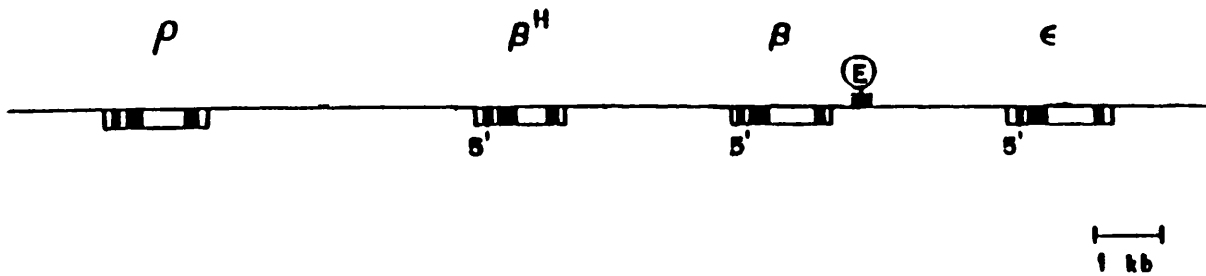


Fig. 7.2B Diagram of the organisation of the chicken β -like globin genes and the position of the β -globin enhancer. The boxes indicate the globin genes, the filled-in regions representing exons. E represents the 184bp β -globin enhancer (Choi & Engel, 1986).

-240 bp. This domain was present in chromatin from 9 day old embryos onwards when the β^A -globin gene is expressed, but absent in 5 day old embryos in which the gene is not expressed (McGhee et al., 1981).

The adult α -globin genes, α^A and α^D , are expressed in the primitive cells of 5 day old embryos as well as in the definitive cells of older embryos (Schalekamp, 1983). Despite this variance from β^A -globin, it was assumed that the globin genes would still have common regulatory elements. The α^D -globin gene has a DNAaseI hypersensitive domain which extends from -130 to +80 relative to the transcription start site (Kemper et al., 1987). DNAaseI footprinting revealed a protected region of as much as 35 bp within this domain. At lowest protein concentration, a core sequence of [5'-AAGATAAGG-3'] was protected. At higher concentrations of protein, protection extends to adjacent sequences on either side, showing that multiple identical or different factors bind in this region. This was confirmed by bandshift assays using a DNA fragment covering this region and showing multiple protein-DNA complexes. An identical [5'-AAGATAAG-3'] sequence was also found at -200 in the chicken β^{rho} -globin gene that had the same properties in bandshift assays as its counterpart from the α^D -globin gene.

From these initial observations, further investigation revealed the presence of an erythroid-specific DNA binding factor by concentrating on the 3'-flanking region of the chicken β^A -globin gene. This region, as well as its hypersensitive domain, contains a strong tissue-specific enhancer (Choi & Engel, 1986). This area was examined in

transient expression studies on 10 day old embryonic chicken erythrocytes. The cells were transfected with plasmids carrying the β^A -globin promoter at the 5'-end of the CAT reporter gene and the β^A -globin enhancer at the 3'-end. The enhancer was constructed from various synthetic oligonucleotides that could be independently altered to vary the parental sequence (Evans et al., 1988). A 49 bp region of the enhancer was thus identified as playing the most critical role in its function. Further analysis of this region using bandshift and footprinting assays, identified two sequences which bound a factor in 10 day old embryonic chicken erythrocyte extracts. These sequences were [5'-AGATAAA-3'] and [5'-TGATAGC-3']. Comparing these results to those in other extracts from heart tissue, the chicken pre B lymphoblast cell line, MSB, chicken lens and HeLa cells showed that the factor binding to these sequences was specific to erythroid tissue. Aligning these two β^A -globin sequences to those found to bind erythroid nuclear factors in the β^{rho} - and α^D -globin gene promoters suggested the consensus sequence [5'-^A/_TGATA^A/_G-3'] (ibid.). The factor binding was designated Eryf1.

Analysis of the human β -globin gene promoter and 3' enhancer also revealed the binding of an erythroid-specific factor. Four sites for erythroid factor binding were identified in the 3' β -globin enhancer in a similar fashion to above, using DNAaseI footprinting, methylation interference and bandshift assay (Wall et al., 1988). Comparison of the sequences at these four sites produced a consensus binding sequence of [5'-^A/_CPy^T/_AATC^A/_TPy-3'] for the factor, designated by NF-E1 by these authors. Fig. 7.3A shows these sites

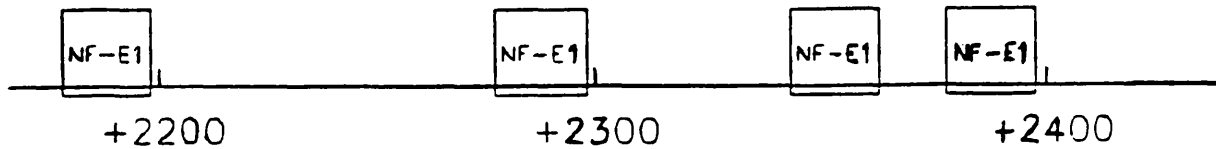


Fig. 7.3A Schematic diagram showing the four NF-E1 binding sites in the human β -globin 3' enhancer relative to the transcription start site of the gene, which is taken as +1.

```

MEFVALGGPDAGSPTP-F ----- PDEAG 22
*D*PG**ALGTSE*L*Q*VDSALVSS*SDST

AFL ----- GLGGGERTEA ----- 35
G*FSSGPE**DAASSSTSPNAATAAASALAY

----- GLLASYPPSGRVSLVP ----- 52
YREAEAYRHSPVFQV**LLNSMEGI*GGSPY

---WA-DTGTLTGPQWVPPATQHEPPH ----- 75
AS**YGKTA*YPAST*C*SHEDA*SQALEDO

----- YLELLQPPRGSP ----- P 88
EGKSNNTF*DT*KTE*L**DLLTLGTALPAS

HPSSG ----- PLL ----- PLSSGP ----- 102
L*VT*SAYGGADFPS*FFSPTGS*****AAYS

----- PPCEARE[C*V*N*CGATATPLWRR 123
SPKFHGS LPLA*****

DGTGHYLCNACGLYHRLNGQNRPLIRPKKRL 154
*R*****KH*****M

LVSKR]AGTV[C*SN*CTSTTTLWRRSPHGD PVC 185
I*****Q[T*****T*****NAS*****

NACGLYYKLHQVNRPLTMRKDCIQTRNR]KVS 216
*****F*****A*

SKGKKRRPP ----- GGGNPSATAGGGAPMGGGG 244
G*****K*GSNLA*A*AAEGP***FHVVA*SS

DPSMPPPPPPPPAA-APPQSDALY-ALGPVVL 273
SSGNCGEVASGL*LGTAGTAN*QG*****

SG ----- HFLPF ----- GNSGGFFGGGAGGYTA 296
**PVS*LM**PGPLL*SPTTS*PT*PAPT*S

PPGLSPQI ----- 304
STSVIAPLSS

```

Fig. 7.3B Diagram of the Eryf1 sequence, using the single letter amino acid code with numbering at the right. Beneath it is an alignment of the GF-1 amino acid sequence (Tsai et al., 1989). An asterisk represents an amino acid identity; dashes signify gaps in the Eryf1 sequence. The repeated elements are indicated by brackets; each repeat contains two Cys-x-x-Cys motifs that are underlined (taken from Evans & Felsenfeld, 1989).

relative to the transcription start site of the gene, which is taken as +1. Similar characterisation of the human β -globin promoter region revealed the presence of two other NF-E1 binding sites at -200 and -120 (deBoer et al., 1988). However, the site at -120 has a very low affinity for binding to the erythroid factor and it would seem like that only the site at -200 is a genuine binding site. Its sequence is [5'-AGATATATCT-3']. This contains two overlapping consensus binding sites, one on each strand. Further *in vitro* binding analysis showed that the mouse α - and β^{maj} -globin genes and the erythroid-specific promoter of the human PBGD gene also bind an erythroid-specific factor (Plumb et al., 1989).

The first conclusion to be drawn from these results, and those for HCAI, is that the factor described is specific to erythroid tissue. It is not present in other tissues in the same form (a factor from HeLa does bind to the same consensus site, discussed below). It is possible that the protein is completely absent from these cells or that its binding to DNA is suppressed through modification or association with an inhibitor, as for NF- κ B (Sen & Baltimore, 1986).

Another important point which emerges from the comparison of this data with that described in this thesis for HCAI is that the erythroid-specific DNA binding factor is present at all developmental stages. This is illustrated in Fig. 4.3A and 4.3B. The factor binds to ^{32}P -labelled oligo B in K562 cells which have an embryonic/foetal phenotype, HEL cells, which have a foetal/adult phenotype and MEL cells, which have an adult phenotype. The Eryf1 binding activity is present in embryonic and adult erythroid nuclear extracts. Likewise, NF-

E1 and that factor described by Plumb et al (1989) is also present in both K562 and MEL cell nuclear extracts. As discussed below, the sequence specificity and binding pattern produced suggest that NF-E1, Eryf1 and the other activities described, are the same DNA binding factor. The fact that this factor is present throughout development suggests that it is not primarily responsible for the stage-specific expression of either HCAI or the globin genes. It may well be necessary for this process but not the major regulator. However, the possibility is not totally excluded that some subtle change in the factor takes place, which does not affect its DNA binding activity but, alters any interaction it may make in transcriptional activation.

On the basis of the binding studies described above various consensus binding sites have been proposed:

- 1) Evans et al., 1988 (Eryf1) $A/TGATA/G$
- 2) Wall et al., 1988 (NF-E1) $A/G^A/TGATA/T^A/G^A/C$
- 3) Plumb et al., 1989 GAT A A G
- 4) 6 binding sites around the HCAI gene $A/TGATA/T A$

The sequences can be written in either orientation. From the comparison of the sequences above it can be seen that the core of the consensus is probably $A/TGATA/T$, though only the GAT is invariant. However, the consensus cannot simply be this trinucleotide since by itself it occurs too widely in the genome to be specific (even if it did bind the factor, the amount of factor present would have to be very large to bind to all the available GAT sequences). There is no evidence to suggest that other sequences either side of the consensus site aid the binding of the erythroid-specific factor.

As mentioned in Sections 4.2 and 4.3, the method used to prepare the cell/nuclear extract for bandshift assay has a great influence on the banding pattern produced. Prior to actually isolating the factor, the pattern of these bandshifts can yield some information as to the nature of the binding activity involved. Using the probes for the sequences flanking HCAI, two complexes (I and II) were seen in the bandshift assays of the erythroid extracts. The complexes were present in roughly equal amounts when using nuclear extracts but the lower mobility complex I was much more prominent in whole cell extracts. A similar pattern to the latter was obtained from embryonic or adult chicken whole blood extracts (Evans et al., 1988) where the extract was made from a crude nuclear preparation and the proteins extracted from chromatin in 0.42M NaCl.

The results of the bandshifts assays on the human β -globin NF-E1 sites are complicated by the fact that the probes used contained more than one binding consensus. The exception to this is the oligonucleotide designated D (Wall et al., 1988). This probe does not bind any factor in non erythroid cell extracts. When used on erythroid extracts it does produce a two band pattern in which the lower mobility complex usually predominates over the higher mobility complex. However, in the K562 nuclear extract used the two complexes are present in equal amounts. Plumb et al (1989) also showed the presence of two complexes in crude nuclear extracts from erythroid cells. They also observed that after fractionation of the crude protein extract by non specific DNA-cellulose affinity chromatography, there is a tendency to remove the lower

mobility complex.

Two different explanations are possible as to why two complexes are formed. It could be that it is due to an artefact of the extract preparation procedure. The isolation of nuclei, ammonium sulphate precipitation and proteolysis, all could have an effect on the banding pattern. However, since two complexes are so frequently seen they may not be artefacts. An obvious interpretation then would be that the erythroid-specific DNA binding factor is a dimer which is in equilibrium with its monomeric form. However, as discussed below, the erythroid-specific DNA binding factor has been isolated (EF1, Perkins et al., 1989; Eryf1, Evans & Felsenfeld, 1989; EF-1, Barnhart et al., 1989; GF-1, Tsai et al., 1989) and a cDNA encoding it has been cloned (GF-1, Tsai et al., 1989; Eryf1, Evans & Felsenfeld, 1989). This has allowed some direct study of the factor itself.

Perkins et al (1989) have isolated chicken EF1 and shown it to have a MW of 37-39 kd. They have also suggested that the two complexes formed in bandshifts with the purified factor, termed EF1 and EF1', may have a dimer-monomer relationship which is mediated by phosphorylation. Eryf1, also from chicken, has a MW of about 38 kd. Using purified Eryf1 in bandshifts, two complexes can still be found. The presence of the higher mobility complex accumulates during purification and extended storage. It is postulated that this is due to proteolytic degradation of the Eryf1 factor (Evans & Felsenfeld, 1989). Consistent with this idea is that isolation of a 22 kd protein co-purified with 38 kd Eryf1. The 22 kd protein forms the higher mobility complex in bandshifts with

an Eryf1 binding site probe.

A further interpretation can be drawn from studies on purified MEL cell GF-1 (Tsai et al., 1989). These authors conclude that GF-1 DNA binding activity is associated with a single chain 50 kd polypeptide and that proteolytic fragments of 38 kd and 28 kd retain DNA binding capacity. The 38 kd and 20 kd polypeptides arise during purification. When intact GF-1 is used in bandshift assays under conditions in which DNA probe is limited, an additional low mobility complex is observed. This was seen when using a probe with a single binding site and it was inferred that this additional complex was due to either a dimer of GF-1 bound to a single DNA molecule or perhaps a dimer of protein-DNA complexes. EF-1, also isolated from MEL cells, was made up of two polypeptides, 18 kd and 19 kd (Barnhart et al., 1989). Two possibilities might explain this latter result. There may be multiple forms of the erythroid-specific DNA binding factor which perhaps recognise slightly different sequences or, more likely, the factor is susceptible to proteolysis which would explain the range of sizes attributed to it. The effect of post-translational modifications in different cell lines and in tissues from different species should also not be discounted. The two complexes formed with whole erythroid cell extracts and HCAI probes may then be explained by proteolysis although the possibility monomer-dimer equilibrium still exists. The latter possibility^{of} should be resolved by studies on the purified mammalian factors.

7.3 What is the role of GF-1?

The erythroid-specific factor, GF-1 (and its other designations), has been constantly referred to as a DNA binding protein. The key question is: does it also have a role in the regulation of transcription of erythroid-specific genes? This would seem very likely since GF-1 binding sites are conserved throughout the promoters and enhancers of erythroid-specific genes. The commonest way in which this possibility can be investigated is by transfection studies. Constructs with or without GF-1 binding sites and, including a reporter gene driven by a minimal promoter, can be transfected into erythroid cells and the effect on the levels of transcription of the reporter gene measured.

deBoer et al (1988) used the stable transfection of β -globin promoter constructs into MEL cells to look at the influence of the binding sites on transcription. This was not a systematic study of the role of the GF-1/NF-E1 solely. Work from this laboratory had previously shown that an important sequence, necessary for the increase of β -globin transcription after DMSO induction, lies at -150. The transfection experiments showed that for the -150 sequence to be effective, it required the presence of either the -200 or the -120 GF-1/NF-E1 binding sites. These experiments suggest that a single NF-E1 site attached to a β -globin minimal promoter (i.e. containing a single CACCC, CCAAT and TATA box) is not sufficient to allow transcription with or without DMSO induction.

Further stable transfection experiments were carried out with various mouse α -globin gene promoter sequences attached to a reporter gene (Plumb et al., 1989). These experiments

focussed on the difference between the levels of transcription in erythroid as opposed to non erythroid cells, in this case fibroblasts. The results suggested that the presence of a GF-1 binding site (referred to as a GATAAG motif by the authors) allowed a 12-fold higher level of transcription in MEL cells over fibroblasts. An important point, arising from these experiments and the previous ones, with the human β -globin promoter, are that the position of the binding site relative to transcription start site and other binding sites may be of importance. For instance, a fragment of the mouse α -globin promoter containing the GF-1 binding site was taken out of context and placed beside a region of the promoter, containing only a TATA box preceding the transcription start site. This construct was not transcribed at any higher level after transfection into erythroid cells compared to fibroblasts (ibid.). Similarly, no effect is seen if up to 5 GF-1 sites are placed upstream in the same construct.

More specific experiments are possible with the cDNA coding for GF-1. Tsai et al (1989) cotransfected non-erythroid cells (COS or NIH 3T3) with a minimal promoter containing multiple GF-1 binding sites, GF-1 cDNA in an expression vector and reporter constructs. A "modest" (no specific data given) increase in transcription from the minimal promoter is observed in the presence of expressed GF-1.

The experiments described in this thesis (Section 4.8) investigate the role of GF-1 in transcription by inserting two GF-1 binding sites, (A) and (B) (Fig. 3.11) upstream of a minimal tk promoter driving a CAT reporter gene. The presence of the sites, in either orientation, produced 2.8 fold greater

CAT activity than a control construct without the sites in transiently transfected MEL cells. No increase in activity from the constructs containing the GF-1 binding sites was seen after a similar transient transfection of non-erythroid, HeLa cells. As with the data from Plumb et al (1989), this would suggest that the binding of GF-1 stimulates transcription of erythroid-specific genes. The 12 fold increase seen with the mouse α -globin transfectants suggests GF-1 acts as an upregulator however, the more "modest" increases in transcriptional activity noted here and by Tsai et al (1989) are at variance with this. It may be that GF-1 is a necessary component of the transcription complex in erythroid cells but this does not require GF-1 to upregulate transcription by itself. Further investigation is required to judge whether GF-1, in combination with other factors, just supports or positively upregulates transcription in erythroid tissue.

The presence of three GF-1 binding sites, (D), (E) and (F), in the 3'-flanking region of the HCAI gene is intriguing. The 3' enhancer of human β -globin has four GF-1/NF-E1 binding sites (Wall et al., 1988), the 3' enhancer of chicken α^A -globin has three sites (Evans et al., 1988) and the 3' enhancer of chicken β -globin has two erythroid-specific factor binding sites (ibid.). It would therefore seem a good prospect for there to be enhancer activity associated with the 3'-flanking region of the HCAI gene. It is tempting to say this would also be erythroid-specific. However, there is no reason to suppose that HCAI's colon / non-erythroid transcription unit has a different 3'-flanking region. Therefore, in the absence of GF-1 from non erythroid cells, the putative enhancer should not

function and this transcription unit would not be as efficiently transcribed in the non-erythroid context.

The isolation of cDNAs for the mouse erythroid-specific DNA binding factor, GF-1, and the chicken equivalent, Eryf1, allow a comparison of the amino acid sequence of the two (Fig. 7.3B). There is a central region in which 105 of the 120 amino acids (i.e. 88%) are identical between the two proteins. If conservative substitutions are included, the two sequences are 94% conserved. Within this region are two highly similar repeated sequences, each of which contains a zinc finger motif. This motif is known to be part of the DNA binding domain of other transcription factors (Evans & Hollenberg, 1988). In both GF-1 and Eryf1, the zinc finger motifs are of the "two-pairs-of-cysteine-residues" form rather than the two cysteine and two histidine type. Outside the central domain, the amino and carboxy regions vary considerably suggesting that, whatever their function, they have less fastidious sequence requirements.

It is interesting to speculate whether the presence of GF-1 protein in erythroid cells is of fundamental importance to the erythroid commitment of these cells. As mentioned previously, it would seem unlikely that GF-1 is the major regulator of "switching" of erythroid gene expression. However, it is present at all stages of development and differentiation so far examined. It may well be that it can be found directly after the point when a pluripotent haemopoietic stem cell becomes committed to the erythroid lineage. The classification of stem cell populations has so far eluded investigators, it may well be that GF-1 is actually the

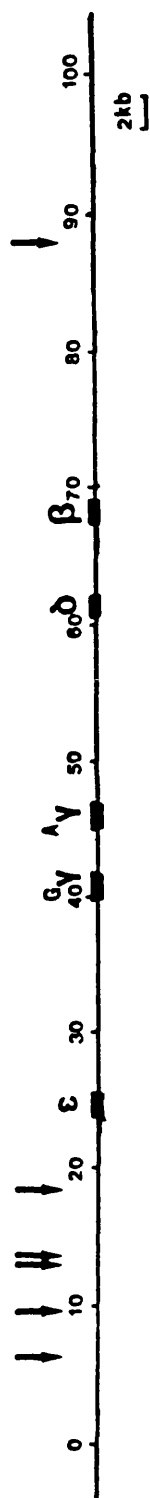


Fig. 7.4 The human β -globin-like gene locus with distances in kb. The arrows indicate DNAase I hypersensitive sites.

erythroid determining factor in the way that MyoD, myogenin, myd and myf-5 appear to be myogenic determining factors (Section 1.2).

A possible site of action of GF-1 apart from in the proximal promoter and 3' enhancer sections around erythroid-specific genes may be in the so-called in dominant control region. This has so far only been shown around the human β -like globin gene locus. This comprises 5 DNAaseI hypersensitive sites 5' to the ϵ -globin and one DNAaseI hypersensitive 3' to the β -globin gene (Fig. 7.4). These DNAaseI sites have been mapped and shown to be erythroid-specific and present when any one of the globin genes is expressed (Grosveld et al., 1987). A "minilocus" was constructed that contained the 5'-and 3'-flanking regions of the human β -globin locus (encompassing the DNAaseI hypersensitive sites) and the β -globin gene. This construct was then used to make transgenic mice. The minilocus was expressed tissue-specifically in the transgenic mice, at a level directly related to its copy number yet independent of its position of integration in the mouse genome. Furthermore, the expression per minilocus copy was the same in each mouse and as high as that of the endogenous mouse β -globin gene (op.cit.). These results indicated that the DNA regions flanking the human β -globin locus contain dominant regulatory sequences that specify position-independent expression. These regions presumably normally activate the complete human multigene β -globin locus.

Further work has shown that the presence of only the 5' DNAaseI hypersensitive sites is necessary for the dominant

control region effect. It has also been shown that high level expression of heterologous genes can be obtained, using the same method as used for β -globin (Blom van Assendelft, 1989). The region required to be incorporated into vectors to exercise the dominant control effect in erythroid cells has also been reduced from over 30 kb to 6.5 kb (Talbot et al., 1989).

It has been reported (Tsai et al., 1989) that one of the DNAaseI hypersensitive sites of human β -globin has a GF-1 binding site. The other sites remain to be investigated. It may well turn out that GF-1 binding is intimately connected with the dominant control region and its effect. This makes the availability of the GF-1 cDNA very useful. It has been proposed, for instance, that the dominant control region could be used to overexpress a cDNA encoding some commercially valuable gene product (e.g. a clotting factor like Factor VIII) in erythroid cells. If the cells could be cotransfected with an expression vector for GF-1, this may provide an even greater boost to the formation of the valuable gene product.

The key to understanding the role of GF-1, as that of the other trans-acting factors, will be in their interaction (i.e. protein-protein) with one another. Overlapping or superimposed binding sites for multiple factors can result in different positive and negative factors competing for sites. Synergistic effects between different factor binding sites has already been observed (Schule et al., 1988b) and shown to be dependent on strict spacing between adjacent binding elements. The interaction between various transcription factors e.g. fos and jun, has also been noted (reviewed in Curran & Franza, 1988).

Further exploration of this avenue of research will surely yield a more complex knowledge of the direct role of trans-acting factors in transcription.

All the evidence points to GF-1 being the major DNA binding protein of the erythroid cell lineage. However, this thesis has provided evidence that non-erythroid, HeLa, cells contain a "GF-1 like" factor in that it specifically binds a GF-1 recognition site (Fig. 4.3A, lane 3). This HeLa factor seems to be present at very low abundance or else it has a very weak affinity for the binding site. It is present as a single complex which has a higher mobility than either of the two complexes with GF-1. It is therefore smaller than the erythroid GF-1, of which it may be a modified form. It could also be a totally unrelated factor which coincidentally also recognises a similar consensus binding site to GF-1. This would not be unique among DNA binding factors or even proven transcription factors. The transcription factor ATF (Hurst & Jones, 1987 - which is now thought to be the same as CREB, Gonzalez et al., 1989) has almost exactly the same recognition site as AP-1 (Lee et al., 1987). The AP-1 consensus binding site is [5'-^C/_GTGACT^C/_AA-3'] whereas the ATF differs only by one nucleotide at position 6, [5'-^C/_GTGACG^C/_AA-3'].

7.4 Further remarks on transcription of the HCAI gene

Constructs, made up of regions of the HCAI promoter and the CAT reporter gene, produced only a low level of CAT activity after transfection into MEL cells (Section 4.7). The level of CAT activity detected was at a similar level to that found in MEL cells transfected with the vector, pSV0cat, minus

the HCAI sequences. The CAT activity produced by the vector alone is probably due to some recognition of cryptic promoter sequences that lie within the pBR322 region of pSV0cat (Langner, 1986). That the HCAI promoter region is unable to drive efficient reporter gene expression in erythroid cells may be because an essential cis-acting element or elements has not been included in the construct. However, up to -817 bp beyond the transcription start site is present in the larger construct and this would usually be sufficient to include many of the cis-acting elements necessary for function (e.g. Myers et al., 1986). It could be that a necessary enhancer region has been omitted or that expression from the HCAI promoter is enhancer dependent. This is not without precedent. For example, when the human globin genes were introduced into COS cells, the β -globin promoter was almost totally dependent upon the effect of the SV40 enhancer whereas the α -globin promoter functioned independently (Humphries et al., 1982). It is highly speculative but, in vivo the enhancer activity for HCAI gene expression might be provided by the possible 3' enhancer proposed above.

HEL cells constitutively express the HCAI gene. However, when treated with the phorbol ester, TPA, the steady-state level of HCAI mRNA falls by 7-8 fold (Fig. 4.4E). The TPA treatment changes the erythroid phenotype of HEL cells to that resembling cells of a macrophage-like phenotype (Papayannopoulou et al., 1983). Therefore, it would seem reasonable that it would be accompanied by a downregulation of the erythroid-specific HCAI transcription unit. Bandshift assays were then carried out on TPA-treated HEL cell extracts

and control cell extracts to see if the TPA had caused any change to GF-1. No difference was observed in GF-1 binding between the TPA-treated and control HEL cell cultures. It is possible that TPA treatment may not have affected GF-1's DNA binding activity while modifying its functional domain by, for example, phosphorylation. However, if as far as can be observed by bandshift assay, it is not affected, this would suggest that although GF-1 may be necessary for the expression of genes from erythroid-specific promoters, it is not sufficient in itself.

Another point of interest is that exposure to TPA decreased the level of HCAI mRNA. The HCAI gene has two AP-1 consensus binding sites, one at -324 in the 5'-flanking region and the other at 801 in the 3'-flanking region. Most viral and cellular genes that respond to AP-1 are also induced by TPA (Angel et al., 1987; Lee et al., 1987). If the AP-1 sites are accessible in the HCAI gene it is surprising then that TPA treatment downregulates HCAI expression. TPA treatment of the human promyelocytic cell line, HL60, results in induction of HCAII mRNA (Shapiro et al., 1989). This suggests different responses between the HCAI and HCAII genes to TPA and emphasises the more general contrast between the regulation of their transcription. MEL cells when induced to differentiate by DMSO also show a differential response between MCAI and MCAII expression. The CAI mRNA is decreased whereas the level of CAII mRNA is increased (Fraser & Curtis, 1987).

This is a suitable juncture to reemphasise the position of the HCAI gene within the CA gene family. HCAI, II and III have been mapped to the long arm of chromosome 8 (Davis et

al., 1987; Nakai et al., 1987) and recent pulse-field electrophoresis data shows that all three genes lie within 200 kb of each other (N.Lowe, personal communication). Since, as described in Section 1.6, these three genes have quite different patterns of tissue-specific and developmental stage-expression, their proximity on chromosome 8 suggests that quite specific elements must be present close to each gene which determine these patterns. It has yet to be shown, but presumably some^{cf} the other uncharacterised CA genes are also present in the same locus. This makes the study of the erythroid- and stage-specific nature of HCAI expression particularly interesting in the context of all the effort expended on understanding globin expression. Since the latter, while being stage-specific is in the context of a completely erythroid-specific locus, whereas CAI is within a locus of differing tissue-specificities.

7.5 Trans-acting factors confer stage-specificity on erythroid HCAI gene expression

The results of the transient heterokaryons created between erythroid cells at different stages in development show the trans-activation of the mouse ϵ -globin and human CAI genes. By identifying the trans-activation of the mouse ϵ -globin gene, it was confirmed that viable heterokaryons had been formed as described previously by Baron and Maniatis (1987). The system was then used to show trans-activation of the HCAI gene in the K562 cell nuclei.

It has previously been proposed that a generalised system of cellular repression can account for the fact that only a

fraction of the potential repertoire of a cell's genes are ever expressed (Weintraub, 1985, and references therein). This repression was thought to be irreversible and stably propagated during replication. However, Blau et al (1983; 1985) then showed that this state is not irreversible but could be modified by diffusible factors present in cytoplasm. They showed that the fusion of multinucleated muscle cells with a variety of non-muscle cell types, to form stable heterokaryons, resulted in the activation of muscle-specific genes in the non-muscle nuclei. The parental nuclei remained distinct and did not fuse therefore, the reprogramming of the differentiated state occurred by trans-activation. This trans-activation of tissue-specific gene expression was also shown to occur in the absence of DNA replication (Chiu & Blau, 1984).

The experiments in this thesis, as well as those of Baron and Maniatis (1987) have taken a similar approach in examining the role of trans-acting factors in the developmental regulation of erythroid gene expression. However, unlike the muscle syncytia used by Blau et al the erythroid cell lines used were not naturally multinucleated. Therefore, the resulting heterokaryons were not stable but transient.

From the Baron and Maniatis studies (illustrated in Fig. 1.8), several conclusions emerged:

(i) erythroid cells contain stage- and possibly globin gene-specific trans-acting factors which function in the developmental regulation of globin genes;

(ii) these factors are active both in erythroid and in non-erythroid cells, whether primary or established;

(iii) the presence of the appropriate trans-acting factors, while necessary, may not in all cases be sufficient for full activation of globin genes. The mouse embryonic α -like (ζ) globin gene was not activated in any of the heterokaryons examined, even though the parental human erythroid (K562) cell expressed high levels of ζ -globin mRNA. This gene is apparently tightly repressed, perhaps in a chromosomal configuration which prevents its interaction with positive trans-acting factors;

(iv) globin genes in non expressing differentiated cells are not irreversibly repressed.

Several conclusions in addition to these can be reached from the work described here. Most importantly, the stage-specific expression of the HCAI gene has been shown to be controlled by trans-acting factors. No activation of the HCAI gene was detected in transient heterokaryons between MEL C88 and the non-erythroid transformed cell line, HeLa. Therefore, the stage-specific trans-acting factors contained in MEL are inadequate to activate HCAI expression in HeLa. Perhaps the HCAI gene is in an inaccessible conformation in the HeLa genome. This situation is not without precedent. It has been found that heterokaryons between murine muscle cell lines and HeLa cells do not express human muscle-specific genes unless the HeLa cells are pre-treated with 5-azacytidine (Chiu & Blau, 1985). This would suggest that for a given cell type some genes are more readily activated than others and likewise some genes can only be activated in certain cell types.

The nature of the stage-specific, trans-acting factors, whether they are positive activators or derepressors, cannot

be illuminated by the experiments described here. As already discussed, GF-1, an erythroid-specific DNA binding protein with an apparent role in erythroid-specific transcriptional regulation, has been identified. However, GF-1 would not appear to be the stage-specific factor(s) referred to here since it is present at each stage of erythroid development (Section 4.4). It remains to be shown whether the stage-specific trans-acting factors which activate globin genes are the same that activate HCAI or whether they are gene-specific and/or tissue-specific.

It is interesting to consider these results in the context of other data on the developmental regulation of chicken β^A - and ϵ -globin genes from Choi & Engel (1988). These investigators looked at the role of the chicken erythroid-specific enhancer located in the intergenic region between the β^A - and ϵ -globin genes (Fig. 7.2B). They did this by analysis of transcripts produced in transfected primitive and definitive erythroid cells from plasmid constructs of deletion and substitution mutants of regions around the β^A -globin and ϵ -globin genes. The results showed that the β -globin enhancer is required in cis for transcriptional activation of both the adult β^A - and embryonic ϵ -globin genes. They also showed that the region of DNA containing the whole β^A -globin gene locus is both necessary and sufficient for correct developmental regulation of this gene. A further conclusion from these experiments was that the enhancer was not the sole genetic element responsible for differential embryonic and adult β -globin gene regulation: the ϵ -globin gene required a cis-linked β^A -globin gene and the enhancer for proper developmental

regulation. It was shown that sequences from -112 to -20 of the β^A -globin were responsible both for this ϵ -globin "cis-suppression" effect and for preferred β^A -globin gene transcription in definitive erythroid cells. These observations suggest that for these genes, erythroid-specificity is defined by one genetic element (the enhancer) that is physically separate from the element responsible for erythroid cell temporal control of stage-specific β -like globin gene expression (designated the stage selector element, SSE).

This idea raises many questions. For instance, is the SSE a positive element activated only in definitive erythroid cells, or a negative cis regulatory element repressed by trans-acting factors present only in primitive cells? Will any enhancer suffice to promote stage-specificity with the SSE? These results have significance as regards the discovery in this thesis that the HCAI gene is under the control of stage-specific trans-acting factors. Trans-acting factors need cis-elements to bind to. Therefore, following on this model described for chicken β -like globin genes, there may well be an equivalent to the SSE within the HCAI gene. This could confer a stage-specificity on erythroid HCAI gene expression different from that seen in the other adjoining genes of the CA locus. If such an element exists in the HCAI gene it would be intriguing to see if the non-erythroid promoter of HCAI also responded to similar stage-specific factors or if they are tissue-specific as well.

7.6 Further work

From the results presented in this thesis several avenues of research warrant further analysis.

1) Expression assays are required on erythroid cells transfected with reporter gene constructs containing various segments of the HCAI gene 3'-flanking region. These should detect the presence of any 3' enhancer activity. These studies could be extended to non erythroid cells to determine if any enhancer activity found was tissue-specific. A stable transfection system is probably the best way to define this since the 5'-flanking regions of the HCAI gene do not seem to be very effective in transient expressions assays.

2) Extensive DNAaseI footprinting studies of the 5'- and 3'-flanking regions of the HCAI erythroid transcription unit are required to define whether all the consensus binding sites do actually bind factors.

3) Reporter gene-minimal promoter constructs containing various regions 5' and 3' to the HCAI erythroid transcription unit could be transfected into erythroid cells to check on the synergism between different factor binding sites and the trans-acting factors binding to them.

4) The HCAI colon / non-erythroid transcription unit, 5' and 3', needs to be defined by S1 nuclease mapping.

5) The regions flanking this transcription unit need to be analysed by in vitro binding assays i.e. DNAaseI footprinting

and bandshifts.

6) Attempts could be made to define the cis-elements binding the stage-specific trans-acting factors responsible for HCAI expression late in foetal development. One way of doing this would be to capitalise on the heterokaryon technique already worked out. Stable transfectants of K562 SAI cells could be made containing various regions of the HCAI gene attached to a reporter gene. Transfectants selected would have to be non expressing. These could then be fused with MEL C88 cells and the activation of reporter gene expression measured. This could be carried out even more sensitively than with SP6 RNAase mapping by using polymerase chain reaction amplification instead. This would hopefully allow a region or regions binding stage-specific factors to be identified. More detailed analysis may then yield sufficient information on sequence-specificity for the isolation of these factors by DNA-affinity chromatography or the cloning of their cDNAs from expression libraries.

REFERENCES

- Allfrey, V.G. (1977) in Chromatin and Chromosome Structure
(eds) Li, H.J. & Echardet, R. Academic Press, New York.
pp 167-191.
- Amara, S.G., Evans, R.M. & Rosenfeld, M.G. (1984) Mol. Cell.
Biol. 4, 2151-2160.
- Angel, P., Imagawa, M., Chiu, R., Stein, B., Imbra, R.J.,
Rahmsdorf, H.J., Jonat, C., Herrlich, P. & Karin, M.
(1987) Cell 49, 729-739.
- Antoniou, M., deBoer, E., & Grosveld, F. (1987) Nucl. Acids
Res. 15, 1886.
- Antoniou, M., deBoer, E., Habets, G. & Grosveld, F. (1988)
EMBO J. 7, 377-384.
- Arndt, K. & Fink, G. (1986) Proc. Natl. Acad. Sci. USA 83,
8516-8520.
- Auffray, C. & Rougeon, F. (1980) Eur. J. Biochem. 107, 303-
314.
- Aviv, H. & Leder, P. (1972) Proc. Natl. Acad. Sci. USA 69,
1408-1412.
- Axelrad, A.A., McLeod, D.L., Shreeve, M.M. & Heath, D.S.
(1974) in Haemopoiesis in culture (ed) Robinson, W.A.
U.S. Dept. of Health, Washington D.C. pp 74-80.
- Barlow, J.H., Lowe, N., Edwards, Y.H. & Butterworth, P.H.W.
(1987) Nucl. Acids Res. 15, 2386.
- Barnhart, K.M., Kim, C.G. & Sheffrey, M. (1989) Mol. Cell.
Biol. 9, 2606-2614.
- Baron, M.H. & Maniatis, T. (1986) Cell 46, 591-602.
- Benyajati, C., Spoerel, N., Haymerle, H. & Ashburner, M.

- (1983) *Cell* 33, 125-133.
- Benz, E.J., Murnane, M.J., Tonkonow, B.L., Berman, B.W.,
Mazur, E.M., Cavallisco, C., Jenko, T., Snyder, E.L.,
Forget, B.G. & Hoffman, R. (1980) *Proc. Natl. Acad. Sci.*
USA 77, 3509-3513.
- Berger, S.L. & Kimmel, A.R. (eds) (1987) *Methods in
Enzymology: Guide to Molecular Cloning Techniques*,
Academic Press, Inc.
- Bernards, A., Rubin, C.M., Westbrook, C.A., Paskind, M. &
Baltimore, D. (1987) *Mol. Cell. Biol.* 7, 3231-3236.
- Bird, A.P. (1986) *Nature* 321, 209-213.
- Birnboim, H.C. & Doly, J. (1979) *Nucl. Acids Res.* 7, 1513-
1523.
- Birnstiel, M.L., Busslinger, M. & Strub, K. (1985) *Cell* 41,
349-359.
- Blau, H.M., Chiu, C.P. & Webster, C. (1983) *Cell* 32, 1171-
1180.
- Blau, H.M., Pavlath, G.K., Hardeman, E.C., Chiu, C.P.,
Silberstein, L., Webster, S.G., Miller, S.C. & Webster,
C. (1985) *Science* 230, 758-766.
- Blom van Assendelft, G., Hanscombe, O., Grosveld, F. &
Greaves, D.R. (1989) *Cell* 56, 969-977.
- Bohmann, D., Bos, T.J., Adman, A., Nishimura, T., Vogt, P.K. &
Tjian, R. (1987) *Science* 238, 1386-1392.
- Bolivar, F., Rodriguez, R.L., Greene, P.J., Betlach, M.C.,
Heynecker, H.L., Boyer, H.W., Crosa, J.H. & Falkow, S.
(1977) *Gene* 2, 95-113.
- Bos, T.J., Bohmann, D., Tsuchie, H., Tjian, R. & Vogt, P.K.
(1988) *Cell* 52, 705-712.

- Boyer, S.H., Siegel, S. & Noyes, A.N. (1983) *Dev. Biol.* 97, 250-253.
- Braun, T., Buschhausen-Denker, G., Bober, E., Tannich, E. & Arnold, H.H. (1989) *EMBO J.* 8, 701-709.
- Breathnach, R. & Chambon, P.A. (1981) *Ann. Rev. Biochem.* 50, 349-383.
- Briggs, M.R., Kadonaga, J.T., Bell, S.P. & Tjian, R. (1986) *Science* 234, 47-52.
- Burke, J.F. (1984) *Gene* 30, 63-68.
- Buskin, J.N. & Hauschka, S.D. (1989) *Mol. Cell. Biol.* 9, 2627-2640.
- Carter, M.J. & Parsons, D.S. (1971) *J. Physiol.* 215, 71-94.
- Chao, M.V., Mellon, P., Charnay, P., Maniatis, T. & Axel, R. (1983) *Cell* 32, 483-493.
- Charnay, P., Treisman, R., Mellon, P., Chao, M.V., Axel, R. & Maniatis, T. (1984) *Cell* 38, 251-263.
- Chevallier, P. & Philippe, M. (1973) *Exp. Cell Res.* 82, 1-14.
- Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. & Rutter, W.J. (1979) *Biochemistry* 18, 5294-5299.
- Chiu, C.P. & Blau, H.M. (1984) *Cell* 37, 879-887.
- Chiu, C.P. & Blau, H.M. (1985) *Cell* 40, 417-424.
- Choi, O.R. & Engel, J.D. (1986) *Nature* 323, 731-734.
- Choi, O.R. & Engel, J.D. (1988) *Cell* 55, 17-26.
- Chretien, S., Dubart, A., Beaupain, D., Raich, N., Grandchamp, B., Rosa, J., Goossens, M. & Romeo, P.H. (1988) *Proc. Natl. Acad. Sci. USA* 85, 6-10.
- Clewell, D.B. & Helinski, D.R. (1972) *J. Bacteriol.* 110, 1135-1141.
- Clissold, P.M., Arnstein, H.R.V. & Chesterton, C.J. (1977)

Cell 11, 353-361.

Collins, S.J., Gallo, R.C. & Gallagher, R.E. (1977) Nature
270, 347-349.

Conscience, J.F. & Meier, W. (1980) Exp. Cell Res. 125, 111-
119.

Curran, T. & Franza Jr., B.R. (1988) Cell 55, 395-397.

Curtis, P.J. (1983) J. Biol. Chem. 258, 4459-4463.

Dale, T.C., Ali Imam, A.M., Kerr, I.M. & Stark, G.R. (1989)
Proc. Natl. Acad. Sci. USA 86, 1203-1207.

Dang, C.V., McGuire, M., Buckmire, M. & Lee, W.M.F. (1989)
Nature 337, 664-666.

Darnell, J.E. (1982) Nature 297, 365-371.

Davis, M.B., West, L.F., Barlow, J.H., Butterworth, P.H.W.,
Lloyd, J.C. & Edwards, Y.H. (1987) Somat. Cell Mol.
Genet. 13, 173-178.

Davis, R.L., Weintraub, H. & Lassar, A.B. (1987) Cell 51, 987-
1000.

deBoer, E., Antoniou, M., Mignotte, V., Wall, L. & Grosveld,
F. (1988) EMBO J. 7, 4203-4212.

Deisseroth, A. & Hendrick, D. (1978) Cell 15, 55-63.

Dexter, T.M. (1979) Clin. Haematol. 8, 453-458.

Dierks, P., van Ooyen, A., Cochran, M.D., Dobkin, C., Reiser,
J. & Weissman, C. (1983) Cell 32, 695-706.

Dorn, A., Bollekens, J., Staub, A., Benoist, C. & Mathis, D.
(1987) Cell 50, 863-872.

Dynan, W.S. & Tjian, R. (1985) Nature 316, 774-778.

Dynan, W.S., Sazer, S., Tjian, R. & Schmike, R.T. (1986)
Nature 319, 246-248.

Eisen, H., Bach, R. & Emery, R. (1977) Proc. Natl. Acad. Sci.

USA 74, 3898-3902.

Emerson, B.M., Lewis, C.D. & Felsenfeld, G. (1985) Cell 41, 21-30.

Emeson, R.B., Hedjran, F., Yeakley, J.M., Guise, J.W. & Rosenfeld, M.G. (1989) Nature 341, 76-80.

Evans, R.M. & Hollenberg, S.M. (1988) Cell 52, 1-3.

Evans, T., Reitman, M. & Felsenfeld, G. (1988) Proc. Natl. Acad. Sci. USA 85, 5976-5980.

Evans, T. & Felsenfeld, G. (1989) Cell 58, 877-885.

Feinberg, A.P. & Vogelstein, B. (1983) Anal. Biochem. 132, 6-13.

Fernley, R.T. (1988) Trends Biochem. Sci. 13, 356-359.

Fitcher, J.H., Foon, K.A. & Martin, J.C. (1981) N. Engl. J. Med. 305, 17-25.

Fletcher, C., Heintz, N. & Roeder, R.G. (1987) Cell 51, 773-781.

Fogh, J. & Trempe, G. (1975) in Human Tumour Cells in vitro (ed) Fogh, J. Plenum Press, New York. pp 115-159.

Fourney, R.M., Miyakoshi, J., Day, R.S. & Paterson, M.C. (1988) BRL Focus 10, 5-7.

Fraser, P.J. & Curtis, P.J. (1986) J. Mol. Evol. 23, 294-299.

Fraser, P.J. & Curtis, P.J. (1987) Genes Dev. 1, 885-861.

Fraser, P., Cummings, P. & Curtis, P.J. (1989) Mol. Cell. Biol. 9, 3308-3313.

Freshney, R.I. (1983) Culture of animal cells - a manual of basic techniques, Alan R. Liss Inc., New York.

Fried, M. & Crothers, D.M. (1981) Nucl. Acids Res. 9, 6505-6525.

Friend, C., Scher, W., Holland, J.G. & Sato, T. (1971) Proc.

- Natl. Acad. Sci. USA 68, 378-382.
- Fukuda, M., Fukuda, M.N., Papayannopoulou, T. & Hakamori, S.
(1980) Proc. Natl. Acad. Sci. USA 77, 3474-3478.
- Gahmberg, C.G., Jokinen, M. & Anderson, L.C. (1978) Blood 52,
379-387.
- Garcia, J., Wu, F. & Gaynor, R. (1987) Nucl. Acids Res. 15,
8367-8385.
- Geliebter, J. (1987) BRL Focus 9, 5-8.
- Gey, G.O., Coffman, W.D. & Kubicek, M.T. (1952) Cancer Res.
12, 264-265.
- Gil, G., Smith, J.R., Goldstein, J.L. & Brown, M.S. (1987)
Proc. Natl. Acad. Sci. USA 84, 1863-1866.
- Gonzales, G.A., Yamamoto, K.K., Fischer, W.H., Karr, D.,
Menzel, P., Biggs III, W., Vale, W.V. & Montminy, M.R.
(1989) Nature 337, 749-752.
- Gootenberg, S.E., Ruscetti, F.W., Mier, J.W., Gazdar, A. &
Gallo, R.C. (1981) J. Exp. Med. 154, 1403-1418.
- Gorman, C.M., Moffat, L.F. & Howard, B.H. (1982a) Mol. Cell.
Biol. 2, 1044-1051.
- Gorman, C.M., Merlino, G.T., Willingham, M.C., Pastan, I. &
Howard, B.H. (1982b) Proc. Natl. Acad. Sci. USA 79, 6777-
6781.
- Graves, B.J., Johnson, P.F. & McKnight, S.L. (1986) Cell 44,
565-576.
- Gros, G. & Dodgson, S.J. (1988) Ann. Rev. Physiol. 50, 669-
694.
- Gross, D.S. & Garrard, W.T. (1988) Ann. Rev. Biochem. 57, 159-
197.
- Grosveld, F., Blom van Asendelft, G., Greaves, D.R. & Kollias,

- G. (1987) *Cell* 51, 975-985.
- Hansen, J.N., Konkel, D. & Leder, P. (1982) *J. Biol. Chem.* 257, 1048-1052.
- Hart, R.P., McDevitt, M.A. & Nevins, J.R. (1985) *Cell* 43, 677-683.
- Harvey, R.P., Robins, A.J. & Wells, J.R.E. (1982) *Nucl. Acids Res.* 10, 7851-7863.
- Hashimoto, C. & Steitz, J.A. (1986) *Cell* 45, 581-591.
- Heintz, N. & Roeder, R.G. (1984) *Proc. Natl. Acad. Sci. USA* 81, 2713-2717.
- Hentschel, C.C., Irminger, J.C., Bucher, P. & Birnstiel, M.L. (1980) *Nature* 285, 147-151.
- Herbomel, P., Bourachot, B. & Yaniv, M. (1984) *Cell* 39, 653-662.
- Hesse, J.E., Nickol, J.M., Leiber, M.R. & Felsenfeld, G. (1986) *Proc. Natl. Acad. Sci. USA* 83, 4312-4316.
- Hewett-Emmett, D., Hopkins, P.J., Tashian, R.E. & Czelusniak, J. (1984) *Ann N.Y. Acad. Sci. USA* 429, 338-358.
- Hirai, S.I., Ryseck, R.P., Mehta, F., Bravo, R. & Yaniv, M. (1989) *EMBO J.* 8, 1433-1439.
- Hofer, E. & Darnell, J.E. (1981) *Cell* 23, 585-593.
- Huehns, E.R. & Farooqui, A.M. (1975) *Nature* 254, 335-337.
- Huehns, E.R., Flynn, F.V., Butler, E.A. & Beaven, G.H. (1961) *Nature* 189, 490-497.
- Hull, R.N., Dwyer, A.C., Cherry, W.R. & Tritch, O.J. (1965) *Proc. Soc. Exp. Biol. Med.* 118, 1054-1059.
- Humphries, R.K., Ley, T., Turner, P., Moulton, A.D. & Nienhuis, A.W. (1982) *Cell* 30, 173-183.
- Hurst, H.C. & Jones, N.C. (1987) *Genes Dev.* 1, 1132-1146.

- Ingraham, H.A., Chen, R., Mangalam, H.J., Elsholtz, H.P., Flynn, S.E., Lin, C.R., Simmons, D.M., Swanson, L. & Rosenfeld, M.G. (1988) *Cell* 55, 519-529.
- Ingram, V.M. (1972) *Nature* 235, 338-339.
- Ish-Horowicz, D. & Burke, J.F. (1981) *Nucl. Acids Res.* 9, 2989-2998.
- Johnson, P.F., Landshulz, W.H., Graves, B.J. & McKnight, S.L. (1987) *Genes Dev.* 1, 133-146.
- Jones, K.A. & Tjian, R. (1985) *Nature* 317, 179-182.
- Jones, K.A., Yamamoto, K.R. & Tjian, R. (1985) *Cell* 42, 559-572.
- Jones, K.A., Kadonaga, J.T., Rosenfeld, P.J., Kelly, T.J. & Tjian, R. (1987) *Cell* 48, 79-89.
- Kadonaga, J.T. & Tjian, R. (1986) *Proc. Natl. Acad. Sci. USA* 83, 5889-5893.
- Kadonaga, J.T., Courey, A.J., Ladika, J., Tjian, R. (1988) *Science* 242, 1566-1570.
- Karn, J.S., Brenner, S., Barnett, L. & Cesareni, G. (1980) *Proc. Natl. Acad. Sci. USA* 77, 5172-5176.
- Karn, J., Matthes, H.W.D., Gait, M.J. & Brenner, S. (1984) *Gene* 32, 217-224.
- Katinka, M., Vasseur, M., Montreau, N., Yaniv, M. & Blangy, P. (1981) *Nature* 290, 720-722.
- Kemper, B., Jackson, P.D. & Felsenfeld, G. (1987) *Mol. Cell. Biol.* 7, 2059-2069.
- Ko, H.S., Fast, P., McBride, W. & Staudt, L.M. (1988) *Cell* 55, 135-144.
- Kollias, G., Wrighton, N., Hurst, J. & Grosveld, F. (1986) *Cell* 46, 89-94.

- Kollias, G., Hurst, J., deBoer, E. & Grosveld, F. (1987) *Nucl. Acids Res.* 15, 5739-5747.
- Konialis, C.P., Barlow, J.H. & Butterworth, P.H.W. (1985) *Proc. Natl. Acad. Sci. USA* 82, 663-667.
- Konkel, D.A., Maizel Jr., J.V. & Leder, P. (1979) *Cell* 18, 865-873.
- Kornblihtt, A.R., Vibe-Pedersen, K. & Baralle, F.E. (1984) *Nucl. Acids Res.* 12, 5853-5867.
- Kryske, M.H., Piette, J. & Yaniv, M. (1987) *Nature* 328, 254-257.
- Laemmli, U.K. (1970) *Nature* 227, 680-685.
- Landschulz, W.H., Johnson, P.F. & McKnight, S.L. (1988) *Science* 240, 1759-1764.
- Langner, R.D. (1986) *Proc. Natl. Acad. Sci. USA* 83, 1598-1601.
- Lassar, A.B., Buskin, J.N., Lockshon, D., Davis, R.L., Apone, S., Hauschka, S.D. & Weintraub, H. (1989) *Cell* 58, 823-831.
- La Thangue, N.B. & Rigby, P.W.J. (1988) in *Transcription and Splicing* (eds) Hames, B.D. & Glover, D.M. IRL Press, Oxford. pp 1-42.
- Lawn, R., Efstratidis, A., O'Connell, C. & Maniatis, T. (1980) *Cell* 21, 647-651.
- Lee, W., Mitchell, P. & Tjian, R. (1987) *Cell* 49, 741-752.
- LeFranc, M.P., Forster, A., Baer, R., Stinson, M.A. & Rabbitts, T.H. (1986) *Cell* 45, 237-246.
- Lehrach, H., Diamond, D., Wozney, J.M. & Boedtker, H. (1977) *Biochemistry* 16, 4743-4751.
- Leibovitz, A., Stinson, J.C., McCombs III, W.B., McCoy, C.E., Mazur, K.C. & Mabry, N.D. (1976) *Cancer Res.* 36, 4562-

4569.

- Leung, S., Proudfoot, N.J. & Whitelaw, E. (1987) *Nature* **329**, 551-554.
- Lin, H.J., Anagnou, N.P., Rutherford, T.R., Shimada, T. & Nienhuis, A. (1987) *J. Clin. Invest.* **80**, 374-380.
- Lindskog, S. (1983) in *Zinc Enzymes* (ed) Spiro, T.G. John Wiley, New York. pp 79-121.
- Lloyd, J., McMillan, S., Hopkinson, D.A. & Edwards, Y.H. (1986) *Gene* **41**, 233-239.
- Lloyd, J., Brownson, C., Tweedie, S., Charlton, J. & Edwards, Y.H. (1987) *Genes Dev.* **1**, 594-602.
- Lozzio, C.B. & Lozzio, B.B. (1975) *Blood* **45**, 321-334.
- Luckow, B. & Schutz, G. (1987) *Nucl. Acids Res.* **15**, 5490.
- McGhee, J.D., Wood, W.I., Dolan, M., Engel, J.D. & Felsenfeld, G. (1981) *Cell* **27**, 45-55.
- McLauchlan, J., Gaffney, D., Whitton, J.L. & Clements, J.B. (1985) *Nucl. Acids Res.* **13**, 1347-1368.
- Maclean, N., Gregory, S.P. & Flavell, R.A. (eds) (1983) *Eukaryotic Genes: Structure, Activity and Regulation*. Butterworths, London.
- Maniatis, T., Fritsch, E.F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, New York.
- Maniatis, T., Goodbourn, S. & Fischer, J.A. (1987) *Science* **236**, 1237-1245.
- Mantovani, R., Margaretti, N., Nicolis, S., Giglioni, B., Comi, P., Cappellini, N., Bertaro, M.T., Caligaris-Cappio, F. & Ottolenghi, S. (1988) *Nucl. Acids Res.* **16**, 4299-4313.

- Martin, P.J. & Papayannopoulou, T. (1982) *Science* 216, 1233-1235.
- Mather, E.L., Nelson, K.J., Haimovich, J. & Perry, R.P. (1984) *Cell* 36, 329-338.
- Matsui, T., Segall, J., Weil, P.A. & Roeder, R.G. (1980) *J. Biol. Chem.* 255, 11992-11996.
- Maxam, A.M. & Gilbert, W. (1980) *Methods in Enzymology* 65, 499-560.
- Meldrum, N.V. & Roughton, F.J.W. (1933) *J. Physiol.* 80, 113-142.
- Melton, D.A., Krieg, P.A., Rebagliati, M.R., Maniatis, T., Zinn, K. & Green, M.R. (1984) *Nucl. Acids Res.* 12, 7035-7056.
- Melton, D.W., McEwan, C., McKie, A.B. & Reid, A.M. (1986) *Cell* 44, 319-328.
- Miller, J., McLachlan, D. & Klug, A. (1985) *EMBO J.* 4, 1609-1614.
- Mizushima-Sugano, J. & Roeder, R.G. (1986) *Proc. Natl. Acad. Sci. USA* 83, 8511-8515.
- Montgomery, J.C., Venta, P.J., Tashian, R.E. & Hewett-Emmett, D. (1987) *Nucl. Acids Res.* 15, 4687.
- Moore, C.L. & Sharp, P.A. (1985) *Cell* 41, 845-855.
- Morgan, W.D., Williams, G.T., Morimoto, R.I., Green, J., Kingston, R.E. & Tjian, R. (1987) *Mol. Cell. Biol.* 7, 1129-1138.
- Myers, R.M., Tilly, K. & Maniatis, T. (1986) *Science* 232, 613-618.
- Nakai, H., Byers, M.G., Venta, P.J., Tashian, R.E. & Shows, T.B. (1987) *Cytogenet. Cell Genet.* 44, 234-235.

- Nienhuis, A.W. & Maniatis, T. (1987) in *The Molecular Basis of Basis of Blood Diseases* (eds) Stamatoyannopoulos, G., Nienhuis, A.W., Leder, P. & Majerus, P.W. Saunders, Philadelphia. pp 28-65.
- Nomiyama, H., Fromental, C., Xiao, J.H. & Chambon, P. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 7881-7885.
- Notstrand, B., Vaara, I. & Kannan, K.K. (1974) in *Isoenzymes: Molecular Structure*, vol. I (ed) Market, C.L. Academic Press, New York.
- O'Hare, P. & Goding, C.R. (1988) *Cell* **52**, 435-445.
- Ojwang, P.J., Nakatsuji, T., Gardiner, M.B., Reese, A.L., Gilman, J.G. & Huisman, T.H.J. (1983) *Haemoglobin* **7**, 115-123.
- Ondek, B., Shepard, A. & Herr, W. (1987) *EMBO J.* **6**, 1017-1025.
- Padgett, R.A., Grabowski, P.J., Konarska, M.M., Seiler, S. & Sharp, P.A. (1986) *Ann. Rev. Biochem.* **55**, 1119-1150.
- Papayannopoulou, T., Nakamoto, B., Yokochi, T., Chait, A. & Kannagi, R. (1983) *Blood* **62**, 832-845.
- Papayannopoulou, T., Nakamoto, B., Kurachi, S. & Nelson, R. (1987) *Blood* **70**, 1764-1772.
- Perkins, N.D., Nicolas, R.H., Plumb, M.A. & Goodwin, G.H. (1989) *Nucl. Acids Res.* **17**, 1299-1314.
- Picard, D. & Schaffner, W. (1984) *Nature* **307**, 80-82.
- Piette, J. & Yaniv, M. (1987) *EMBO J.* **6**, 1331-1337.
- Pinney, D.F., Pearson-White, S.H., Konieczny, S.F., Latham, K.E. & Emerson, C.P. (1988) *Cell* **53**, 781-793.
- Plumb, M., Frampton, J., Wainwright, H., Walker, M., MacLeod, K., Goodwin, G. & Harrison, P. (1989) *Nucl. Acids Res.* **17**, 73-92.

- Proudfoot, N.J. & Brownlee, G.G. (1976) *Nature* 263, 211-214.
- Ptashne, M. (1988) *Nature* 335, 683-689.
- Quesenberry, P. & Levitt, L. (1979) *N. Engl. J. Med.* 301, 755-760.
- Rackwitz, H.R., Zehetner, G., Frischauf, A.M. & Lehrach, H. (1984) *Gene* 30, 195-200.
- Rigby, P.W.J., Dieckmann, M., Rhodes, C. & Berg, P. (1977) *J. Mol. Biol.* 113, 237-251.
- Robinson, J., Sieff, C., Delia, D., Edwards, P.A.W. & Greaves, M.F. (1981) *Nature* 289, 68-71.
- Rosenthal, N. (1987) in *Methods in Enzymology: Guide to Molecular Cloning Techniques* (eds) Berger, S.L. & Kimmel, A.R. Academic Press, Inc. pp 704-720.
- Ruiz-Opazo, N., Weinberger, J. & Nadal-Ginard, B. (1985) *Nature* 315, 67-70.
- Rutherford, T., Clegg, J. & Weatherall, D. (1979) *Nature* 280, 164-165.
- Sazer, S. & Schmike, R. (1986) *J. Biol. Chem.* 261, 4685-4690.
- Schalekamp, M. (1983) in *Regulation of haemoglobin biosynthesis* (ed) Goldwasser, E. Elsevier Science Publishing Inc., New York pp 323-332.
- Scheidreit, C., Heguy, A. & Roeder, R.G. (1987) *Cell* 51, 783-793.
- Schroeder, W.A., Huisman, T.H.J., Shelton, J.R., Shelton, J.B., Kleihauer, E.F., Dozy, A.M. & Robberson, B. (1968) *Proc. Natl. Acad. Sci. USA* 60, 537-544.
- Schule, R., Muller, M., Otsuka-Murakani, H. & Renkawitz, R. (1988a) *Nature* 332, 87-90.
- Schule, R., Muller, M., Kaltschmidt, C. & Renkawitz, R.

- (1988b) *Science* 242, 1418-1420.
- Sen, R. & Baltimore, D. (1986) *Cell* 47, 921-928.
- Shapiro, L.H., Venta, P.J., Yu, Y.S.L. & Tashian, R.E. (1989)
FEBS Lett. 249, 307-310.
- Shaw, G. & Kamen, R. (1986) *Cell* 46, 659-667.
- Singer-Sam, J., Simmer, R.L., Keith, D.H., Shively, L.,
Teplitz, M., Itakura, K., Gartler, S.M. & Riggs, A.D.
(1983) *Proc. Natl. Acad. Sci. USA* 80, 802-806.
- Singh, H., LeBowitz, J.H., Baldwin, A.S. & Sharp, P.A. (1988)
Cell 52, 415-423.
- Sive, H.L., Heintz, N. & Roeder, R.G. (1986) *Mol. Cell. Biol.*
6, 3329-3340.
- Spandidos, D.A. & Paul, J. (1982) *EMBO J.* 1, 15-20.
- Spandidos, D.A. (1984) *Mol. Biol. Med.* 2, 167-175.
- Spandidos, D.A. (1987) *Gene Anal. Tech.* 4, 50-56.
- Spicer, S.S., Steward, P.J. & Tashian, R.E. (1979) *J.*
Histochem. Cytochem. 27, 820-831.
- Stadie, W.C. & O'Brien, H. (1933) *J. Biol. Chem.* 103, 521-529.
- Staudt, L.M., Singh, H., Sen, R., Wirth, T., Sharp, P.A. &
Baltimore, D. (1986) *Nature* 323, 640-643.
- Struhl, K., (1987) *Cell* 50, 841-846.
- Sturm, R., Das, G., Herr, W. (1988) *Genes Dev.* 2, 1582-1599.
- Suggs, S.V., Wallace, R.B., Hirose, T., Kawashima, E.H. &
Itakura, K. (1981) *Proc. Natl. Acad. Sci. USA* 78, 6613-
6617.
- Sutton, R.E. & Boothroyd, J.C. (1986) *Cell* 47, 527-535.
- Talbot, D., Collis, P., Antoniou, M., Vidal, M., Grosveld, F.
& Greaves, D.R. (1989) *Nature* 338, 352-355.
- Tapscott, S.J., Paris, R.L., Thayer, M.J., Cheng, P.F.,

- Weintraub, H. & Lassar, A.B. (1988) *Science* 242, 405-411.
- Tashian, R.E. (1989) *BioEssays* 10, 186-192.
- Towbin, H., Staehlin, T. & Gordon, J. (1979) *Proc. Natl. Acad. Sci. USA* 76, 4350-4354.
- Townes, T.M., Lingrel, J.B., Chen, H.Y., Brinster, R.C. & Palmiter, R.D. (1985) *EMBO J.* 4, 1715-1723.
- Tsai, S.F., Martin, D.I.K., Zon, L.I., D'Andrea, A.D., Wong, G.G. & Orkin, S.H. (1989) ^{Nature} 339, 446-451.
- Turner, R. & Tjian, R. (1989) *Science* 243, 1689-1694.
- Tweedie, S. & Edwards, Y. (1989) *Biochem. Genet.* 27, 17-30.
- Veldman, G.M., Lupton, S. & Kamen, R. (1985) *Mol. Cell. Biol.* 5, 649-658.
- Venta, P.J., Montgomery, J.C., Wiebauer, K., Hewett-Emmett, D. & Tashian, R.E. (1984) *Ann. N.Y. Acad. Sci.* 429, 309-323.
- Venta, P.J., Montgomery, J.C., Hewett-Emmett, D. & Tashian, R.E. (1985) *J. Biol. Chem.* 260, 12130-12135.
- Venta, P.J., Montgomery, J.C. & Tashian, R.E. (1987) in *Isoenzymes: Current topics in biological and medical research* (eds) Razzatti, M.C., Scandalios, J.G. & Whitt, G.S. Alan R. Liss Inc., New York. pp 58-72.
- Villeva, J.L., Testa, U., Vinci, G., Tonthat, H., Bettaieb, A., Titeux, M., Cramer, P., Edelman, L., Rochant, H., Breton-Gorius, J. & Vainchenker, W. (1985) *Blood* 66, 1162-1170.
- Wall, L., deBoer, E. & Grosveld, F. (1988) *Genes Dev.* 2, 1089-1100.
- Wallace, B.R., Johnson, M.J., Hirose, T., Miyake, T., Kawashima, E. & Itakura, K. (1981) *Nucl. Acids Res.* 9, 879-894.

- Weatherall, D.J. & Clegg, J.B. (1981) *The Thalassaemia Syndromes*. Blackwell Scientific, Oxford.
- Weintraub, H. (1985) *Cell* 42, 705-711.
- Weintraub, H., Tapscott, S., Davis, R., Thayer, M., Adam, M., Lassar, A. & Miller, A.D. (1989) *Proc. Natl. Acad. Sci. USA* 86, 5434-5438.
- Wieringa, B., Hofer, E. & Weissmann, C. (1984) *Cell* 37, 915-925.
- Wildeman, A.G., Sassone-Corsi, P., Grundstrom, T., Zenke, M. & Chambon, P. (1984) *EMBO J.* 3, 3129-3133.
- Wirth, T., Staudt, L. & Baltimore, D. (1987) *Nature* 329, 174-178.
- Wistrand, P.J. (1984) *Ann. N.Y. Acad. Sci. USA* 429, 195-206.
- Wood, W.G. (1976) *Br. Med. Bull.* 32, 282-287.
- Wood, W.G., Clegg, J.B. & Weatherall, D.J. (1979) *Br. J. Haematol.* 43, 509-512.
- Wood, W.J., Gitschier, J., Lasky, L.A. & Lawn, R.M. (1985) *Proc. Natl. Acad. Sci. USA* 82, 1585-1588.
- Wright, C.V.E., Cho, K.W.Y., Oliver, G. & DeRobertis, E.M. (1989) *Trends Biochem. Sci.* 14, 52-56.
- Wright, S., deBoer, E., Grosveld, F.G. & Flavell, R.A. (1983) *Nature* 305, 333-336.
- Wright, S., Rosenthal, A., Flavell, R. & Grosveld, F. (1984) *Cell* 38, 265-273.
- Wright, W.E., Sassoon, D.A. & Lin V.K. (1989) *Cell* 56, 607-617.
- Yen, T.J., Machlin, P.S. & Cleveland, D.W. (1988) *Nature* 334, 580-585.
- Yoshihara, C.M., Lee, J.D. & Dodgson, J.B. (1987) *Nucl. Acids*

Res. 15, 753-770.

Young, R.A., Bloom, B.R., Grosskinsky, C.M., Iwany, J.,
Thomas, D. & Davies, R.W. (1985) Proc. Natl. Acad. Sci.
USA 82, 2583-2587.

Multiple GF-1 binding sites flank the erythroid specific transcription unit of the human carbonic anhydrase I gene

Hugh J.M. Brady, Jane C. Sowden, Mina Edwards, Nicholas Lowe and Peter H.W. Butterworth

Department of Biochemistry, University College London, Gower Street, London WC1E 6BT, England

Received 14 September 1989

Six potential GF-1 sites which bind an erythroid factor are present in the 5' and 3' regions flanking the erythroid-specific transcription unit of the human carbonic anhydrase I (HCAI) gene. When two of these sites are placed upstream of a minimal eukaryotic promoter they confer up-regulated expression in erythroid over non-erythroid cells. The presence of the erythroid factor in TPA-treated HEL cells in which the level of HCAI transcript has greatly decreased and in non-HCAI-expressing K562 cells suggests that in these cases the presence of the factor is not sufficient for HCAI expression.

GF-1; Erythroid specific transcription factor; Human carbonic anhydrase I; *Trans*-acting protein

1. INTRODUCTION

Certain conserved DNA sequence elements to which transcription factors bind are necessary for the expression of most eukaryotic genes by RNA polymerase II; other elements, binding specific *trans*-acting protein factors have been shown to determine cell-specific expression [1]. In erythroid cells, the promoters of globin genes contain the conserved 'TATA' or 'CATA', 'CAAT' and 'CACCC' sequence cassettes and recent work has identified a sequence motif 'GATAAG' (or closely related variants thereof) which binds an erythroid-specific protein [2–5]. This sequence element is conserved across species and is found in either orientation in the regulatory regions of erythroid-specific genes. A cDNA encoding this factor has recently been cloned and designated GF-1 [6]. This paper defines the erythroid-specific transcription unit of the HCAI gene, the expression of which is characteristic of erythroid cells of the adult phenotype [7], and examines the binding of the erythroid-specific factor to sequences flanking it.

2. MATERIALS AND METHODS

2.1. Transcription unit mapping

Total human reticulocyte RNA was prepared by the guanidinium hydrochloride/caesium chloride method [8]. Primer extension analysis [9] of the 5'-end of HCAI mRNA used a single-stranded DNA oligonucleotide primer (3'-CACCAGGACAGACCGTCGGA-5') complementary to a sequence in the 5'-leader region of the HCAI gene (from +33 to +52). The primer was 5'-end labelled with

T4 polynucleotide kinase and [γ -³²P]ATP and hybridised to 25 μ g of RNA followed by extension with reverse transcriptase.

S₁-nuclease mapping used single-stranded DNA probes generated from M13 templates [10]. For 5'-end mapping, a *PvuII-HindIII* fragment containing the first HCAI exon was subcloned into M13mp18 and the complementary strand synthesised using the above primer. For 3'-end mapping, a *HindIII-MboI* fragment from the 3'-untranslated region (221–572 bp downstream from the stop codon) was subcloned into M13mp19 and synthesis of the complementary strand initiated using the 17mer Amersham sequencing primer. 3 \times 10⁵ cpm of 5'-end probe was hybridised to 15 μ g total human reticulocyte RNA at 62 or 70°C for 3 h. 1.4 \times 10⁵ cpm of 3'-end probe was hybridised to 9 μ g of RNA for 1 h at 50 or 58°C. S₁-nuclease digestion was carried out for 2 h at 20°C.

2.2. Cell lines and tissue culture

The erythroid cell lines used were K562 [11], K562-SA1 [12], HEL (92.1.7) [13] and mouse erythroleukemic (MEL) cells F412B2 (TK⁻) [14]. K562 have an embryonic/foetal phenotype, the others an adult phenotype. All were grown in Dulbecco's MEM (DMEM) with 10% foetal calf serum (Gibco) plus penicillin (100 U/ml), streptomycin (100 μ g/ml) and amphotericin B (2.5 μ g/ml). Another MEL cell line C88 (APRT⁻) [15] was also used and grown in α -MEM supplemented with 10% foetal calf serum and 50 μ g/ml diaminopurine. HeLa cells were grown in DMEM as above. HL-60 (myeloid) [16] and HUT-78 (lymphoid) [17] were grown in RPMI 1640 plus 10% foetal calf serum with antibiotics and amphotericin B as above. All media were supplemented with 2 mM glutamine.

HEL cells were induced to undergo a macrophage-like shift by treatment with 12-*O*-tetradecanoyl-phorbol-13-acetate (TPA) [18]. TPA was dissolved in dimethylsulphoxide (DMSO) and used at 10⁻⁶ M for 4–8 days. Control cultures contained equivalent amounts of DMSO (0.01%).

2.3. Protein preparation

Whole cell extracts used for gel retardation were prepared by modification of the method of Dale et al. [19]: frozen cell pellets were made of total volume 0.2 ml containing 2–3 \times 10⁷ cells. 1.0 ml ice-cold extraction Buffer A (10 mM Hepes, pH 7.9, 0.4 M NaCl, 1.5 mM MgCl₂, 0.1 mM EGTA, 0.5 mM DTT, 0.5 mM PMSF and 5% glycerol) was added to a single pellet for lysis. The lysate was cen-

Correspondence address: H.J.M. Brady, Department of Biochemistry, University College London, Gower Street, London WC1E 6BT, England

trifuged at $100000 \times g$ at 4°C for 15 min. The supernatant was desalted with Buffer B (same as Buffer A but with 50 mM NaCl) on a NAP-5 column (Pharmacia) and stored at -70°C .

Nuclear proteins for footprinting were prepared from 10^8 cells which were washed twice in phosphate-buffered saline and twice in Buffer I (0.05% Nonidet P-40, 10 mM Hepes pH 7.9, 10 mM NaCl, 3 mM MgCl_2). The lysate was resuspended in 10 ml Buffer II (10 mM Hepes, pH 7.9, 10 mM NaCl, 3 mM MgCl_2) and sedimented twice through 10 ml 30% sucrose in Buffer II at $1000 \times g$ at 4°C for 5 min. Nuclei were resuspended in 3 ml Buffer III (20 mM Hepes pH 7.9, 0.42 M NaCl, 1.5 mM MgCl_2 , 0.2 mM EDTA, 0.5 mM DTT, 0.5 mM PMSF, 25% glycerol) and the protein extract prepared as described by Wildeman [20] apart from the addition of 0.5 mM PMSF to the buffers used. Typically $5\text{--}10 \text{ mg}\cdot\text{ml}^{-1}$ protein was obtained from 10^8 cells and stored in aliquots at -70°C .

2.4. Oligonucleotides and probe preparation

The oligonucleotide ' $\alpha\text{g}2$ ' derived from mouse α_1 -globin was a gift from Dr M. Plumb (Beatson Institute, Glasgow). Other oligonucleotides (antisense strand shown below) were made as complementary single-stranded sequences and annealed before use:

Oligo A: 5'-GTATTTTTATTGATTATTGTGCTG-3'
 Oligo B: 5'-ACCACTTCCCTTATCAGGTTCTC-3'
 Oligo C: 5'-CCCCTCTAATCACACAGGGCCA-3'
 Oligo E: 5'-TGATCAAATGATTATCTTTTATAT-3'
 Oligo F: 5'-CTATTTTATCTTAATTGACACA-3'
 Oligo $\alpha\text{g}2$: 5'-GATCCGGCAACTGATAAGGATTCCC-
 AGATC-3'
 Oligo CACCC: 5'-CTGATTAATCCACACCCCA-3'

The oligonucleotides were 5'-end labelled as above and 5'-overhangs were filled in using excess dNTPs and Klenow fragment and purified by electrophoresis on a 10% polyacrylamide gel.

Fragment 'D', a 57 bp *NheI-EcoRI* fragment (195–251 bp downstream from the 'stop' codon) which lies between the two polyadenylation sites, was dephosphorylated using calf intestinal phosphatase and ^{32}P end-labelled as above.

2.5. Gel retardation assay

Gel retardation assays using whole cell extracts were carried out essentially as described by Dale et al. [19]. $10 \mu\text{l}$ of extract was preincubated with $1 \mu\text{l}$ of $5 \text{ mg}\cdot\text{ml}^{-1}$ poly(dI-dC)·poly(dI-dC) for 15 min at 20°C . Additional components were added to a final concentration of $0.5 \times$ Buffer B, 2% Ficoll (w/v), $0.25 \text{ mg}\cdot\text{ml}^{-1}$ BSA, $10\text{--}20000$ cpm end-labelled DNA and 100 ng competitor DNA where indicated in a final volume of $40 \mu\text{l}$. The mixture was incubated for a further 15 min at 20°C . Samples were electrophoresed on a 5% nondenaturing polyacrylamide gel in $0.5 \times$ TBE (89 mM Tris, 89 mM boric acid and 5 mM EDTA) at 150 V for 2 h.

2.6. Footprinting analysis

The 255 bp *PvuII-AvaII* fragment (–219 to +35) was subcloned into the *SmaI* site of Bluescript plasmid (KS+, Stratagene). Both strands were labelled for DNase I footprinting of the promoter region of HCAI: the coding and noncoding strands were 5'-end labelled at the polylinker *HindIII* site and the *DdeI* site at +14, respectively, and fragments were purified after secondary digestion with *HaeII* (–107) for the coding strand and with *PstI* (in the polylinker) for the noncoding strand. Markers were prepared by Maxam-Gilbert sequencing of the 5'-end labelled fragments.

Nuclear protein ($50\text{--}100 \mu\text{g}$) was preincubated with $1 \mu\text{g}$ poly(dI-dC)·poly(dI-dC) in $40 \mu\text{l}$ binding buffer (50 mM KCl, 5 mM MgCl_2 , 1 mM EDTA, 10 mM Tris-HCl pH 8.0, 1 mM DTT, 12.5% glycerol and 0.1% Triton X-100) at 4°C for 30 min. Labelled fragments (20000 cpm) were added and incubated at 4°C for 30 min. DNase I digestion at $0.5 \mu\text{g}\cdot\text{ml}^{-1}$, in the presence of protein and $0.01 \mu\text{g}\cdot\text{ml}^{-1}$ in the absence of protein, was at 20°C for 2 min, followed by the addition of 0.1 vol. of 'stop' solution (1 mM EDTA, 10% SDS, $1 \text{ mg}\cdot\text{ml}^{-1}$ tRNA). DNA was purified by organic extraction and resolved on an 8% denaturing polyacrylamide gel.

2.7. Transfection

F412B2 MEL cells and HeLa cells were transfected using calcium phosphate/DNA precipitation as described by Rosenthal [21]. 1.5×10^6 F412B2 cells were plated on 100 mm Corning tissue culture dishes (Bibby) 20–22 h before transfection whereas 3×10^6 HeLa cells were plated on 75 cm^2 Falcon tissue culture flasks (Becton and Dickinson) at the equivalent time. The precipitate was left in contact with F412B2 cells for 24 h and with HeLa cells for 6 h, followed by glycerol shock for 2 min. Both types of cell were left for 48 h after adding the precipitate before harvesting. Cell lysates were then made by 3 cycles of freeze-thawing.

2.8. Chloramphenicolacetyl transferase (CAT) and β -galactosidase assays

β -Galactosidase assays were performed exactly as described by Herbomel et al. [22]. Equivalent amounts of β -galactosidase activity for each transfected plate or flask were then assayed for CAT activity exactly as described by Gorman et al. [23].

2.9. Northern analysis

RNA was separated and transferred onto Gene Screen Plus [24]. HCAI mRNA was detected by hybridising the filter with 1×10^3 cpm HCAI cDNA [25] labelled with $[\gamma\text{-}^{32}\text{P}]\text{dCTP}$ using random primers. Hybridisation and washing conditions were as recommended for Gene Screen Plus.

3. RESULTS AND DISCUSSION

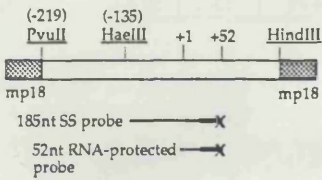
S_1 -mapping and primer extension studies have defined the position of the 5'-end of the transcription unit (fig.1A,B). S_1 -mapping has also identified the polyadenylation site, pA(II) at the 3'-end of the most abundant HCAI mRNA species (fig.1C) which lies 225 bp downstream from an alternative (yet rarely used) site of 3'-end maturation, pA(I), previously described from an analysis of cDNA clones [25]. Consensus sequences for the binding of general transcription factors are apparent (fig.1D). At –28 there is a globin-like 'CATA' motif [26] and three potential 'CAAT' box sequences [26] between –60 and –90. The flanking sequences also contain consensus binding sites for characterised transcription factors: for the 'CACCC'-binding factor [27] at –209 and –47; for AP-1 [28] at –324 and 801 bp downstream from the end of the protein-coding sequence; for Sp1 [29] at –93 and Oct-1 [30] at –81. Based on previously reported consensus sequences [3–5] for the binding of an erythroid-specific transcription factor, GF-1, six potential sites are found flanking the HCAI gene: sites A, B and C at –290, –190 and –149, respectively, and sites D, E and F located 223 bp, 581 bp and 833 bp downstream from the 'stop' codon. Site D lies between the two polyadenylation sites and site E has the sequence motif in two orientations.

Gel retardation assays show that all six GATAAG-like sequences flanking the HCAI transcription unit bind the same erythroid-specific protein (fig.2). Double-stranded oligonucleotides (23- or 24-mers) containing sites A, B, C, E and F and a 57 bp fragment containing site D were used. Each gives rise to a banding pattern containing a more abundant upper band

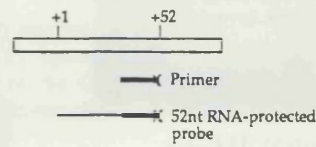
and a much less abundant lower band when incubated with a protein extract from erythroid (MEL) cells. In each case, competition using Oligo-B or Oligo- α 2 (in which the only common sequence is a GATAAG-like

motif) shows binding to be specific to the GATAAG-like motif (lanes 15–26). Protein extracts from erythroid cell lines regardless of developmental phenotype (K562, K562-SA1, MEL and HEL, lanes 1,

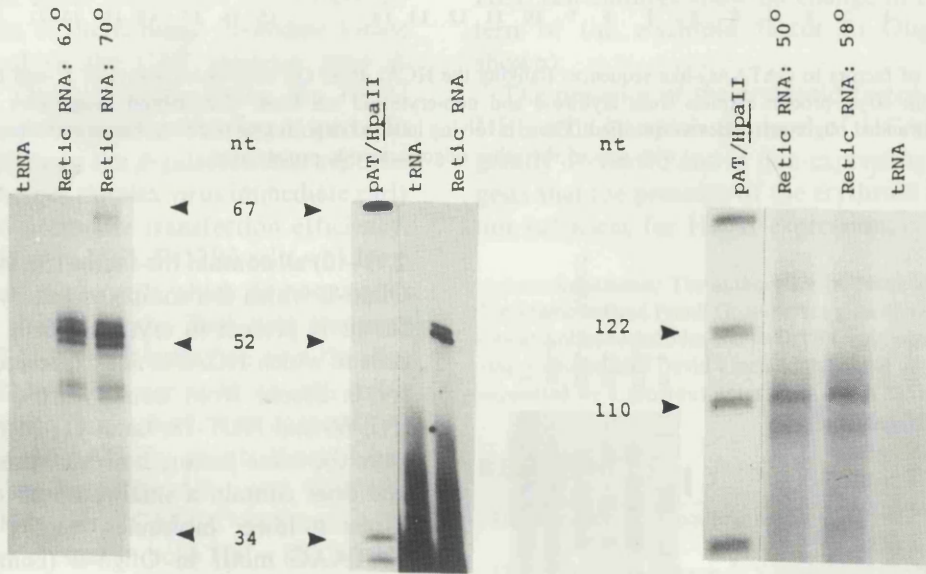
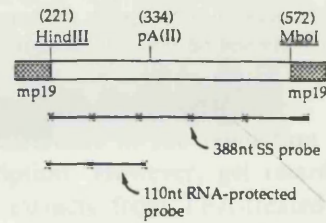
A. S1 map of HCAI 5' end



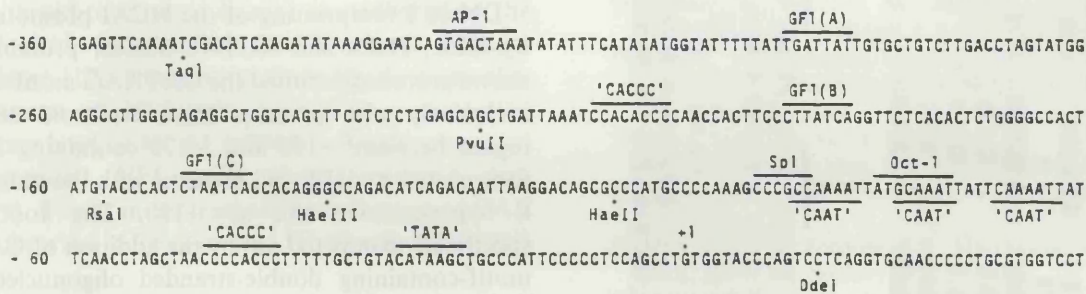
B. Primer extension analysis of HCAI 5' end



C. S1 map of HCAI 3' end



D. (i) 5'-region flanking the transcription start site



(ii) 3' region flanking the translation stop signal

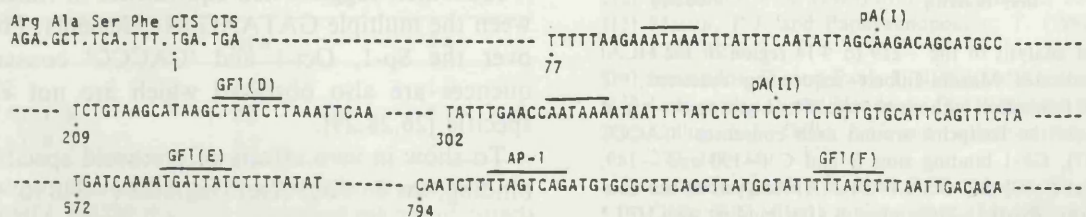


Fig.1. The HCAI transcription unit. (A) S₁-mapping and (B) primer extension analysis defining the transcription start site. (C) S₁-mapping of the most 3'-polyadenylation site. (D) DNA sequences flanking the 5'- and 3'-end of the HCAI transcription unit showing relevant restriction endonuclease cleavage sites and consensus sequences for binding of ubiquitous and cell-type specific transcription factors and for 3'-end maturation.

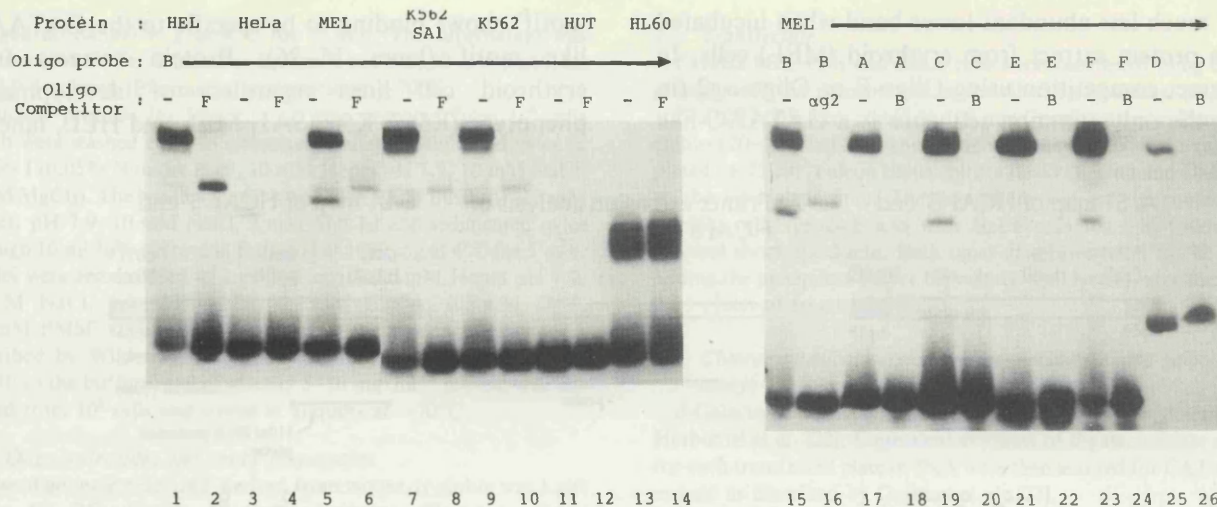


Fig.2. The binding of factors to GATAAG-like sequences flanking the HCAI gene. Gel retardation assays of 5'-end labelled double-stranded oligonucleotides with 50 µg protein extracts from erythroid and non-erythroid cell lines. Competition assays were performed with 150 ng unlabelled double-stranded oligonucleotides as specified. Those involving labelled oligo-B give rise to a characteristic band which does not occur with any of the other oligonucleotide probes used.

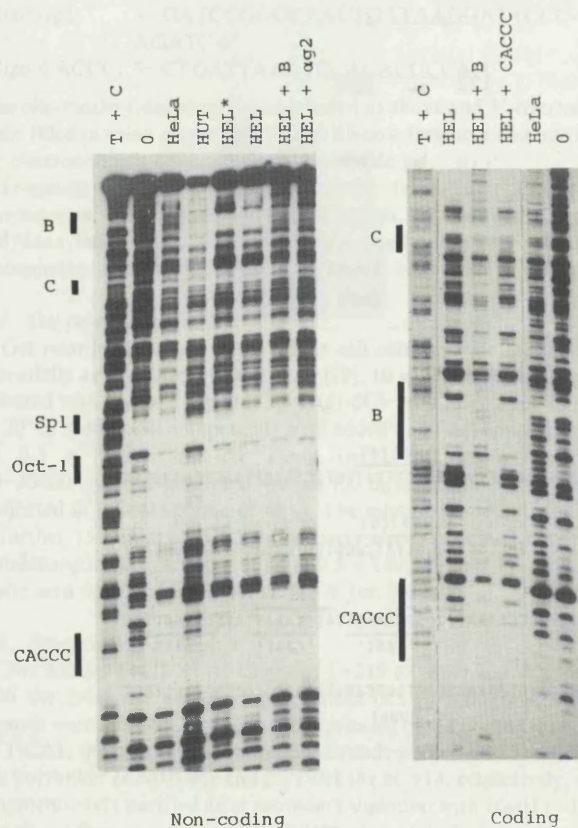


Fig.3. Footprint analysis of the -219 to +14 region of the HCAI gene. T+C indicates Maxam-Gilbert sequencing reactions; '0' indicates DNase I digestion without protein. The boxes to the left of the panels denote the footprint around each consensus: CACCC (-209 and -47), GF-1 binding sites B and C (-190 and -149, respectively), Sp1 (-93) and Oct-1 (-81). (A) Analysis of the non-coding strand from *PvuII* (-219) to (+14) after binding with HEL* (50 µg) and HEL (100 µg), HeLa (100 µg) or HUT-78 (100 µg) nuclear extracts. Competition of the footprint over GF-1 (site B) in HEL extracts was by the addition of 200 ng double-stranded oligonucleotide B or αg2. (B) Analysis of the coding strand from

2, 5-10) all contain the factor (forming complexes with Oligo-B which are competed out by Oligo-F); thus the factor is present in erythroid cells even in embryonic cells in which HCAI is not expressed (see [6]). The factor is absent from non-erythroid haemopoietic cells (HL-60 and HUT-78, lanes 11-14). HeLa cells do not have the same factor; however, this non-erythroid cell line does contain a small amount of a protein which forms a lower molecular weight complex with the GATAAG motif in Oligo-B (competed by Oligo-F, lanes 3 and 4). Comparing the six binding sites with the other published consensus sequences [3-5] suggests a core recognition site of 3'-Py-A-A-T-C-A-5'.

DNase I footprinting of the HCAI promoter region by HEL, HeLa and HUT-78 nuclear proteins (fig.3) shows protection around the GATAAG motif at Site B, exclusively with proteins from erythroid cells. The region between -193 and -179 containing Site B is footprinted by HEL cell proteins with the induction of a hypersensitive site at -180. The footprint is specifically competed out by the addition of GATAAG motif-containing double-stranded oligonucleotides B and αg2 but not the 'CACCC' oligonucleotide. No footprint is evident on either DNA strand for Site C at -150 which suggests non-equivalence in function between the multiple GATAAG-like elements. Footprints over the Sp-1, Oct-1 and 'CACCC' consensus sequences are also observed which are not erythroid specific [26,28,29].

To show in vivo effects of erythroid specific factor binding, the 5' *TaqI-RsaI* fragment (-348 to -157) of

PvuII (-219) to *HaeIII* (-107) after binding with protein extracts from HEL (100 µg) and HeLa (100 µg); competitor for footprint at site B was 200 ng Oligo-B.

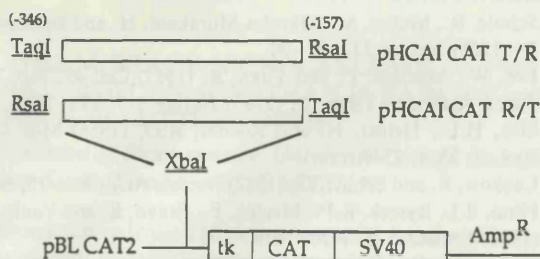


Fig.4. *TaqI-RsaI* fragment of HCAI (-348 to -157) inserted (in either orientation) into the unique *XbaI* site of pBLCAT2, upstream of a minimal thymidine kinase (tk) promoter fused to the CAT reporter gene, the SV40 small t intron and polyadenylation site.

HCAI was placed in either orientation into an expression vector upstream of the minimal thymidine kinase promoter [31] fused to the CAT reporter gene illustrated in fig.4. Constructs containing the HCAI fragment, or vector alone, were cotransfected into cells with a plasmid containing the β -galactosidase reporter gene driven by the herpes simplex virus immediate early gene 4 promoter to normalise transfection efficiency. The transfected cells were MEL F412B2 cells which express mouse CAI and HeLa cells which do not express CAI. The plasmid containing the *TaqI-RsaI* fragment of HCAI shows a 2.5–2.8-fold induction over the control plasmid in MEL cells but not in HeLa cells (table 1). This fragment which flanks the 5'-end of the HCAI gene contains two GATAAG-like motifs (Sites A and B), and consensus sequences for AP-1 and 'CACCC' binding proteins. However, 'CACCC'-box and AP-1 binding proteins are present in both HeLa [27,28] and MEL cells [4,32].

HEL cells constitutively express HCAI. When treated with the phorbol ester TPA, a shift takes place from erythroid to myeloid lineage as evidenced by the morphological, biochemical and functional changes they undergo [17]. Northern analysis (fig.5) shows that the treatment of HEL cells with TPA reduces the steady-state level of HCAI mRNA 7–8-fold compared with untreated HEL cells (from scanning densitometry). This is in contrast to the induction of CAII mRNA observed in TPA-treated HL60 cells [33] which

Table 1

Effect of an HCAI 5'-flanking region on minimal promoter function

	HeLa	F4
pBL CAT2	1.0	1.0
pHCAI CAT T/R	0.5	2.8
pHCAI CAT R/T	0.9	2.5

Each construct (illustrated in fig.4) was transfected separately into HeLa and MEL F412B2 (F4) cells. Normalised volumes of extracts from transfected cells (see section 2) were assayed for CAT activity and subsequently analysed by scanning densitometry. The data derived from each construct are given relative to pBL-CAT2 in each cell line

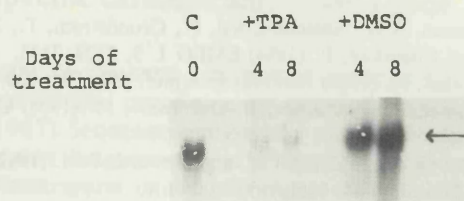


Fig.5. Northern analysis of equivalent amount of total RNA from control and TPA-treated HEL cells probed with 32 P-labelled HCAI cDNA.

indicates a difference in the regulation of CAI and CAII transcription. However, gel retardation assays with protein extracts from TPA-treated and control HEL cell cultures show no change in the binding pattern of the erythroid factor to Oligo-B (data not shown).

The presence of the erythroid factor in TPA-treated HEL cells in which the level of HCAI transcript has greatly decreased and in non-expressing K562 cells suggests that the presence of the erythroid factor (GF-1) is not sufficient for HCAI expression.

Acknowledgements: The authors are indebted to Drs Mark Plumb, Jon Frampton and Frank Grosveld for gifts of various cell lines, vectors and oligonucleotides and to Drs Yvonne Edwards, Ali Imam, Irving Johnston and David Linch for help and advice. This work was supported by a generous grant from the Wellcome Trust.

REFERENCES

- [1] Maniatis, T., Goodbourn, S. and Fischer, J.A. (1987) *Science* 236, 1237–1245.
- [2] Kemper, B., Jackson, P.V. and Felsenfeld, G. (1987) *Mol. Cell. Biol.* 7, 2059–2069.
- [3] Evans, T., Reitman, M. and Felsenfeld, G. (1988) *Proc. Natl. Acad. Sci. USA* 85, 5976–5980.
- [4] DeBoer, E., Antonio, M., Mignotte, V., Wall, L. and Grosveld, F. (1988) *EMBO J.* 7, 4203–4212.
- [5] Plumb, M., Frampton, J., Wainwright, H., Walker, M., MacLoed, K., Goodwin, G. and Harrison, P. (1989) *Nucleic Acids Res.* 17, 73–92.
- [6] Tsai, S.-F., Martin, D.I.K., Zon, L.I., D'Andrea, A.D., Wong, G.G. and Orkin, S.H. (1989) *Nature* 339, 446–451.
- [7] Boyer, S.H., Siegel, S. and Noyes, A.N. (1983) *Dev. Biol.* 97, 250–253.
- [8] Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) *Biochemistry* 18, 5294–5299.
- [9] Sutton, R.E. and Boothroyd, J.C. (1986) *Cell* 47, 527–535.
- [10] Burke, J.F. (1984) *Gene* 30, 63–68.
- [11] Lozzio, C.B. and Lozzio, B.B. (1975) *Blood* 45, 321–334.
- [12] Spandidos, P.A. (1984) *Mol. Biol. Med.* 2, 167–175.
- [13] Martin, P.J. and Papayannopoulou, T. (1982) *Science* 216, 1233–1235.
- [14] Spandidos, D. and Paul, J. (1982) *EMBO J.* 1, 15–20.
- [15] Deisseroth, A. and Hendrick, D. (1978) *Cell* 15, 55–63.
- [16] Collins, S.J., Gallo, R.C. and Gallagher, R.E. (1977) *Nature* 270, 347–349.
- [17] Gootenberg, S.E., Ruscetti, F.W., Mier, J.W., Gazdar, A. and Gallo, R.C. (1981) *J. Exp. Med.* 154, 1403–1418.
- [18] Papayannopoulou, T., Nakamoto, B., Yokochi, T., Chait, A. and Kannagi, R. (1983) *Blood* 62, 832–845.
- [19] Dale, T., Ali Imam, A.M., Kerr, I.M. and Stark, G.R. (1989) *Proc. Natl. Acad. Sci. USA* 86, 1203–1207.

- [20] Wildeman, A.G., Sassone-Corsi, P., Grundstrom, T., Zenke, M. and Chambon, P. (1984) *EMBO J.* 3, 3129-3133.
- [21] Rosenthal, N. (1987) *Methods Enzymol.* 152, 704-720.
- [22] Herbomel, P., Bourachot, B. and Yaniv, M. (1984) *Cell* 39, 653-662.
- [23] Gorman, C.M., Moffat, L.F. and Howard, B.M. (1982) *Mol. Cell. Biol.* 2, 1044-1051.
- [24] Fournay, R.M., Miyokoshi, J., Day, R.S. and Paterson, M.C. (1988) *BRL Focus* 10, 5-7.
- [25] Barlow, J.H., Lowe, N., Edwards, Y.E. and Butterworth, P.H.W. (1987) *Nucleic Acids Res.* 15, 2386.
- [26] Proudfoot, N.J., Shander, M.H.M., Manley, J.L., Gefter, M.L. and Maniatis, T. (1980) *Science* 209, 1329-1336.
- [27] Schule, R., Muller, M., Oksuka-Murakani, H. and Renkawitz, R. (1988) *Nature* 332, 87-90.
- [28] Lee, W., Mitchell, P. and Tjian, R. (1987) *Cell* 49, 741-752.
- [29] Jones, K.H. and Tjian, R. (1985) *Nature* 317, 179-182.
- [30] Sive, H.L., Heintz, N. and Roeder, R.G. (1986) *Mol. Cell. Biol.* 6, 3329-3340.
- [31] Luckow, B. and Schutz, G. (1987) *Nucleic Acids Res.* 15, 5490.
- [32] Hirai, S.I., Ryseck, R.P., Mechta, F., Bravo, R. and Yaniv, M. (1989) *EMBO J.* 8, 1433-1439.
- [33] Shapiro, L.H., Venta, P.J., Ya-Shion, L.Yu. and Tashian, R.E. (1989) *FEBS Lett.* 249, 307-310.

An unusual 5'-leader in the human erythroid-specific carbonic anhydrase I gene

HUGH J. M. BRADY, NICHOLAS LOWE,
JANE C. SOWDEN, JONATHAN H. BARLOW and
PETER H. W. BUTTERWORTH

Department of Biochemistry, University College London,
Gower Street, London WC1E 6BT, U.K.

Carbonic anhydrase I (CAI) is one of a family of at least five very closely related isoenzymes whose major function is the hydration of carbon dioxide. The genes for three of these isoenzymes have been localized to, and are probably tightly linked on, the long arm of human chromosome 8 (see Davis *et al.*, 1987). This family of genes is particularly interesting because the duplicated CA isoforms, which appear to have maintained linkage for more than 300 million years, exhibit very different patterns of tissue-specific expression (reviewed by Tashian & Hewett-Emmett, 1984). CAII is found in a wide variety of tissues, CAIII is muscle-specific (although it is found in male rat liver), there is a kidney-specific, membrane-bound species (CAIV) and a mitochondrial species (CAV). CAI is restricted to the erythrocyte and its expression is activated at a late stage of human fetal development, in the last few weeks (2-4) before birth.

Abbreviations used: CA, carbonic anhydrase; nt, nucleotide.

Initially we screened a human reticulocyte cDNA library for recombinants containing CAI coding sequence (Barlow *et al.*, 1987). Sequence analysis of a number of cDNA clones showed that the processing of primary gene transcripts of the CAI transcription unit was complex (Fig. 1). The simplest interpretation of the data is that three sequence elements can contribute to the 5'-leader which are shown as 1a (69 nt), 1b (54 nt) and 1c (23 nt). One type of cDNA contains elements 1a + 1c (92 nt) and a second type contains all three elements (1a + 1b + 1c: 146 nt). The shorter of the two sequences is the dominant species. Northern analysis on total human reticulocyte RNA using oligonucleotide probes made to sequence elements 1a and 1b produced no detectable signal with the 1b probe suggesting that CAI mRNAs containing this element constitute less than 5% of the total. Alternative polyadenylation sites (see Fig. 1) are also used although at present there is no data to relate one type of 5'-leader to the use of a particular polyadenylation site and the physiological significance of these alternative processing events is not known.

A λ 2001 human genomic library (courtesy of Dr T. H. Rabbitts) was screened with oligonucleotides made to the three 5'-leader sequence elements. From the recombinants obtained, the 1b leader fragment has been mapped to a posi-

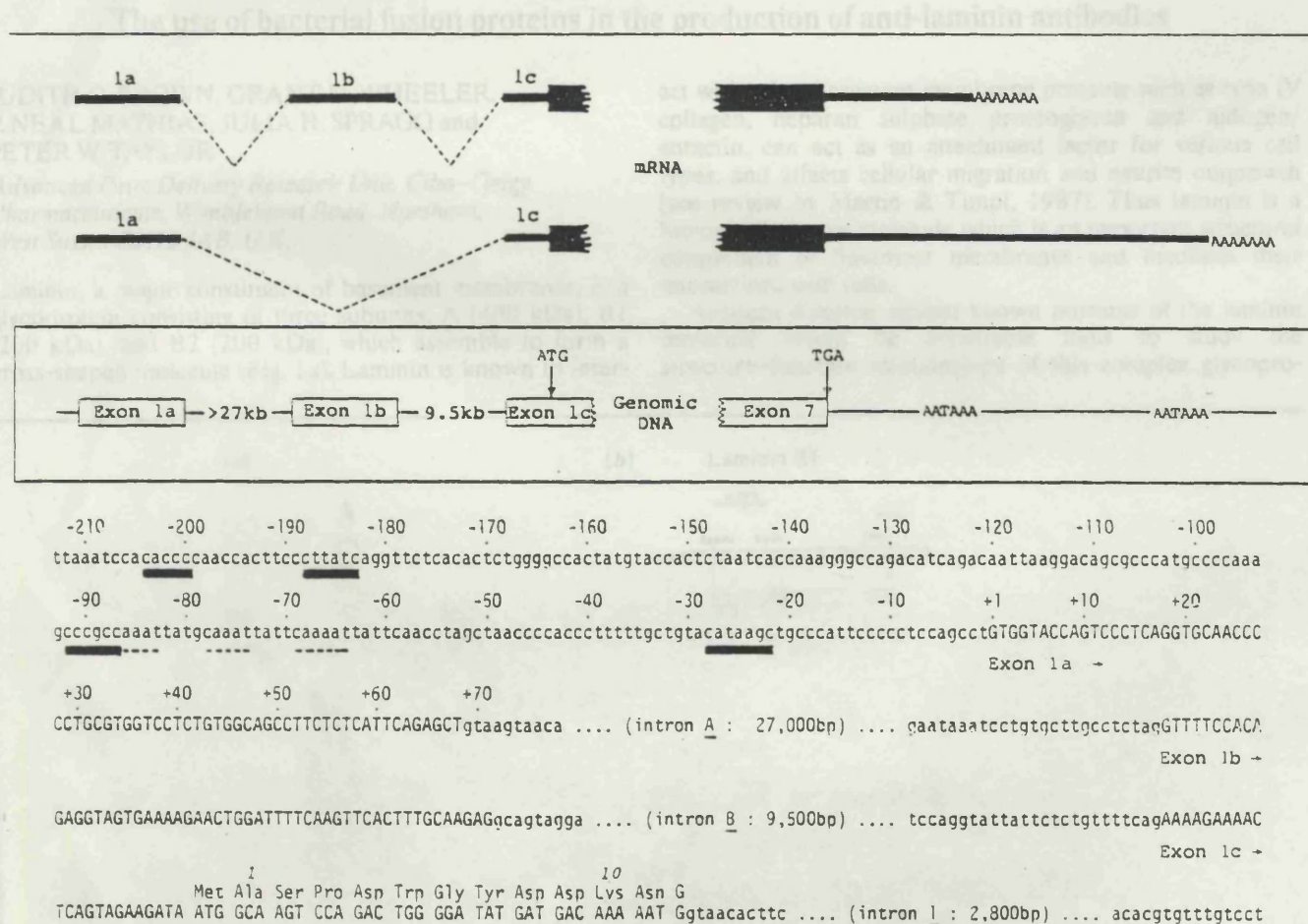


Fig. 1. Organization of the human CAI gene

Diagrams of the different 5'- and 3'-ends of CAI mRNAs derived from cDNA cloned sequences are shown at the top of the Figure: untranslated regions of the transcript are shown as narrow bold lines, protein-coding sequence as broad bold lines and sequences eliminated by splicing as dotted lines. A diagram of the organization of the CAI gene derived from the analysis of CAI-containing genomic clones is shown in the boxed part of the Figure. Sequence data are provided of regions flanking the transcription start site (+1), of the exon/intron junctions of the elements (1a, 1b and 1c, shown in capital letters) which make up the 5'-leader and of the beginning of the protein-coding sequence. Promoter motifs 5' to the transcription start site which are referred to in the text are underlined.

tion more than 9 kb upstream from the beginning of the protein-coding region (including fragment 1c), the recombinant containing the 5'-end of the transcription unit has not yet been placed on a genomic fragment contiguous with the remainder of the 5'-leader (see Fig. 1). In fact, the 5'-end is more than 27 kb upstream from the rest of the gene. Scrutiny of the sequence of the CAI 'optional' exon (1b) reveals that it has all the characteristics of an intron (see 5'- and 3'-splice site consensus sequences, polypyrimidine tract at the 3'-end and appropriately positioned potential 'A-residue' for a branch-point in the sequence shown in Fig. 1), but as an intron it is probably too short for efficient splicing (see Wieringa *et al.*, 1984).

The transcription start site (+1 on Fig. 1) has been mapped by primer extension analysis using oligonucleotide primers made to the 1a and 1c sequence elements and has been confirmed by S₁-nuclease mapping. The genomic sequence of the region immediately upstream from the transcription start site of the CAI gene reveals some common promoter motifs. At -27 there is a globin-like 'TATA' sequence (-CATAAGC-), several potential 'CAAT' sequences in the region of -60 to -90 and a potential Spl-binding site (-CCCGCC-) at -86 (see Jones & Tjian, 1985). Highly significant is the presence of: (i) an appropri-

ately located sequence motif -GATAAG- (as the reverse complement -CTTATC- at -182) immediately adjacent to (ii) a -CCACCC- sequence (at -200), both of which have been identified as potential binding sites for erythroid-specific transcription factors (see Kemper *et al.*, 1987; Mantovani *et al.*, 1988, respectively).

This work was supported by a grant from the Wellcome Trust.

- Barlow, J. H., Lowe, N., Edwards, Y. H. & Butterworth, P. H. W. (1987) *Nucleic Acids Res.* 15, 2386
 Davis, M. B., West, L. F., Barlow, J. H., Butterworth, P. H. W., Lloyd, J. C. & Edwards, Y. H. (1987) *Somatic Cell Mol. Genet.* 13, 173-178
 Jones, K. A. & Tjian, R. (1985) *Nature (London)* 317, 179-182
 Kemper, B., Jackson, P. D. & Felsenfeld, G. (1987) *Mol. Cell Biol.* 7, 2059-2069
 Mantovani, R., Magaretti, N., Nicolis, S., Giglioni, B., Comi, P., Cappellini, N., Bertero, M. T., Caligaris-Cappio, F. & Ottolenghi, S. (1988) *Nucleic Acids Res.* 16, 4299-4313
 Tashian, R. E. & Hewett-Emmett, D. (eds.) (1984) *Ann. N.Y. Acad. Sci.* 429, 1-625
 Wieringa, B., Hofer, E. & Weissmann, C. (1984) *Cell* 37, 915-925

Received 20 June 1988

The use of bacterial fusion proteins in the production of anti-laminin antibodies

JUDITH C. BROWN, GRANT N. WHEELER,
 P. NEAL MATHIAS, JULIA H. SPRAGG and
 PETER W. TAYLOR

*Advanced Drug Delivery Research Unit, Ciba-Geigy
 Pharmaceuticals, Wimblehurst Road, Horsham,
 West Sussex RH12 4AB, U.K.*

Laminin, a major constituent of basement membranes, is a glycoprotein consisting of three subunits, A (400 kDa), B1 (200 kDa), and B2 (200 kDa), which assemble to form a cross-shaped molecule (Fig. 1a). Laminin is known to inter-

act with other basement membrane proteins such as type IV collagen, heparan sulphate proteoglycan and nidogen/entactin, can act as an attachment factor for various cell types, and affects cellular migration and neurite outgrowth (see review by Martin & Timpl, 1987). Thus laminin is a biologically active molecule which is an important structural component of basement membranes and mediates their interactions with cells.

Antisera directed against known portions of the laminin molecule would be invaluable tools to study the structure-function relationships of this complex glycopro-

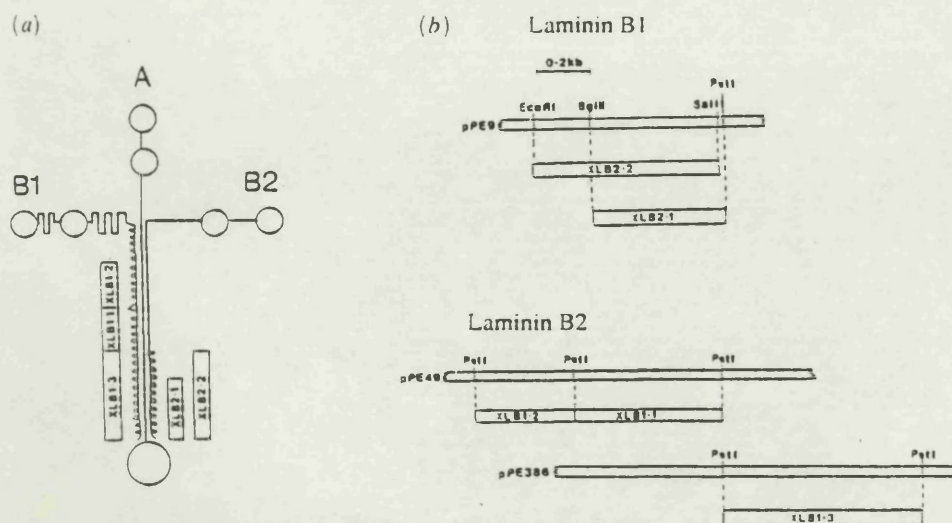


Fig. 1. Laminin sequences used to construct the fusion proteins

(a) Model of the cross-shaped structure of the laminin molecule as proposed by Sasaki *et al.* (1986). The positions of the sequences contained within the fusion proteins are indicated. (b) The laminin-encoding regions of plasmids pPE9, pPE49 and pPE386 (Barlow *et al.*, 1984) showing the restriction sites used in the construction of the fusion proteins. Plasmids pPE9, pPE49 and pPE386 were provided by Dr Brigid Hogan.