Analysis of the 16p telomere to examine the relationship between DNA replication, chromosome structure and gene expression

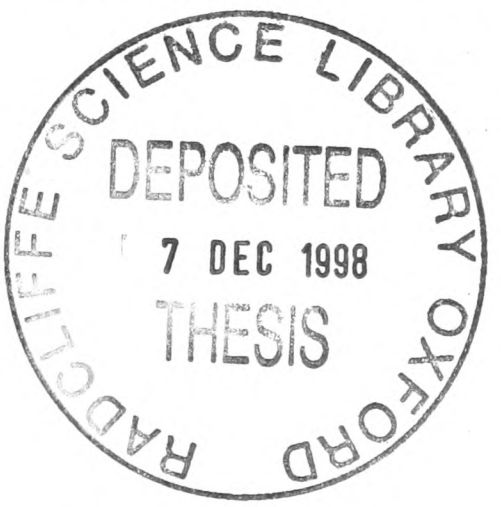


Zoë Elizabeth Smith
Wolfson College

A thesis submitted for the degree of Doctor of Philosophy of the University of Oxford

MRC Molecular Haematology Unit,
Institute of Molecular Medicine,
University of Oxford.

Trinity Term 1998



Analysis of the 16p telomere to examine the relationship between DNA replication, chromosome structure and gene expression

Abstract

Zoë Smith
Wolfson College

Doctor of Philosophy
Trinity Term 1998

The overall aim of the laboratory is to understand how the human α globin genes are expressed in a tissue-specific and developmental-stage specific manner from their natural chromosomal environment. This gene cluster has been studied in great detail from this point of view, and it now represents an important general model to address how eukaryotic genes are regulated. Ultimately, this information may be important for the further development of strategies for effective gene transfer, particularly for patients who suffer from inherited anaemias which result from abnormal globin gene expression (thalassemias).

The human α globin gene cluster is located close to the telomere, in the terminal band of the short arm of chromosome 16 (16p13.3). Many aspects of α gene regulation have been studied in detail, including the identification of the structural genes, their proximal regulatory elements (promoters) and distal regulatory element (HS-40). Erythroid-specific *trans*-acting factors which bind these elements *in vitro* and *in vivo* have been identified. In an attempt to define the regulatory domain containing the α genes, the surrounding chromosomal environment (~400kb) has also been extensively studied. The region is fully sequenced, and the chromatin structure, methylation patterns and scaffold attachment sites have been established. Over 100 natural mutants that perturb α globin expression have been characterised and numerous fragments from this region have been analysed in transgenic mice.

One potentially important aspect of α globin gene expression that has not been previously examined is DNA replication. Study of replication throughout this particular region provides an excellent opportunity to integrate this information with what is already known about the regulation of the α cluster. Initially, studies were undertaken to analyse the pattern and timing of DNA replication of the human α globin gene cluster and the surrounding chromosomal region in erythroid and non-erythroid cell-types. Replication timing of this region has also been compared in the individual paternal and maternal alleles of chromosome 16. Once the *in vivo* pattern of replication was established, the role of chromosomal position upon replication timing was assessed by analysing various constructs of the α cluster in transgenic mice. The role of the distal regulatory element HS-40 in influencing the pattern and timing of replication of the extended region has been examined. Naturally-occurring deletions within this region have facilitated the study of the importance of the telomere and subtelomeric region in determining the replication characteristics of this region. Replication has also been studied in another natural mutant, which contains a deletion which may perturb α gene expression by a novel mechanism.

Acknowledgements

This thesis would not have been written without the never-ending support of my supervisor, Professor Doug Higgs. He dragged me through the process of writing by my hair whilst remaining good humoured, and for that I am truly grateful. I was also lucky enough to have a supervisor whose door was always open for advice throughout my four years in Oxford. As a result, I learnt about scientific research from a man with a great deal experience and expertise.

I feel very privileged to have worked with the many others in lab 302, who taught me the witchcraft behind benchwork and offered me their friendship, especially Helena Ayyub, Rachael (ClarisDraw wizard) Daniels, Tarra McDowell, Delia O'Rourke and Cristina Tufarelli. My thanks to Wendy Bickmore and Judy Fantes for getting me started with FISH. I have also very much enjoyed working with Veronica Buckle, Lyndal Kearney and the cytogenetics lab at the IMM. Many others have offered me kindness and companionship along the way: chef extraordinaire Vicki Gamble; death row partner in crime Dombo Maloney; ice hockey hooligans Ann Benton and Jane Lee-Prudhoe; and party animals Sharon, Regina and Mike. Thanks to Peggy, Gail and Neil for clearing up after me, and Adam and Louis for frequently jumping to my rescue in times of IT difficulty.

Thanks to those who have provided me with technical resources from outside the MHU: Drs Bernie Morley, Howard Cedar, Itamar Simon, Miriam Meisler, and Ann Harris. Wolfson College and all the administrative staff have been fantastic in providing me with financial assistance for conferences, and facilities for work and play. I have fond memories of friends from there and rowing days: Schamski, Karen, Lizbuf, Cindy, Sarah, Philip and Leanne to name but a few. Friends from further afield who remain close and kept me sane include Sums, Stu, Sam, Sal, Nickers, Gilesy and Matt.

Thanks to the MRC and MRF who have funded the research and kept me financially afloat, along with subsidies from my Mother and Brother. Mum and Guy have offered me boundless support throughout my life, and I owe a great deal to them.

Last but not least, my thanks to Mark, my old faithful security blanket. 'They think it's all over. It is now'.

Abbreviations

ACS	ARS consensus sequence
APRT	adenine phosphoribosyl transferase
ARS	autonomously replicating sequence
dATP	2'-deoxyadenosine 5'-triphosphate
bp	base-pair
BCIP	5-bromo-4-chloro-3-indolyl phosphate
BdU	5-bromo-2'-deoxyuridine
BSA	bovine serum albumin
CCD	charge-coupled device
CF	cystic fibrosis
CO_2	carbon dioxide
cm	centimetres
conc	concentrated
dCTP	2'-deoxycytidine 5'-triphosphate
DAPI	4',6-diamidino-2-phenylindole
DIG-11-UTP	digoxigenin-11-2'-deoxyuridine 5'-triphosphate
DMSO	dimethylsulphoxide
DNA	deoxyribonucleic acid
DTT	dithiothreitol
EBV	Epstein-Barr virus
EDTA	ethylene N,N,N',N' diaminetetraacetic acid
EtBr	ethidium bromide
EtOH	ethanol
FCS	fetal calf serum
FISH	fluorescence *in situ* hybridisation
FITC	fluorescein isothiocyanate
dGTP	2'-deoxyguanosine 5'-triphosphate
HCl	hydrochloric acid
HS	hypersensitive site
IPTG	isopropyl ß-D-thiogalactopyranoside
IMM	Institute of Molecular Medicine
KCl	potassium chloride
kDa	kiloDaltons
KH_2PO_4	potassium dihydrogen orthophosphate
K_2HPO_4	di-potassium hydrogen orthophosphate
LCR	locus controlling region
M	molar
MAR	matrix attachment region
MEL	mouse erythroleukaemia
mg	milligram
$MgCl_2$	magnesium chloride
$MgSO_4$	magnesium sulphate
ml	millilitre
mM	millimolar
NaCl	sodium chloride
NaOH	sodium hydroxide
NBT	nitroblue tetrazolium
ng	nanograms
^oC	degrees Centigrade
o/n	overnight
ORC	origin recognition complex
PBS	phosphate-buffered saline
PEG	polyethylene glycol
pg	picograms
psi	pounds per square inch

PYGM	muscle glycogen phosphorylase gene
rpm	revolutions per minute
RPMI	Rosswell Park Memorial Institute
RT	room temperature
SAR	scaffold attachment region
SDS	sodium dodecyl sulphate
SSC	sodium chloride/sodium citrate
TAE	tris/acetate/EDTA
TE	Tris-EDTA
Tris	Tris[hydroxymethyl]aminomethane
dTTP	2'-deoxythymidine 5'-triphosphate
Tween 20	polyoxyethylene (20) sorbitan monolaurate
µg	micrograms
µM	micromolar
UV	ultraviolet
V	volts
v/v	volume per volume
w/v	weight per volume
X-Gal	5-Bromo-4-chloro-3-indolyl-ß-D-galactopyranoside

Table of contents

CHAPTER 1 INTRODUCTION 1

1.1 THE COMPLEXITY OF EUKARYOTIC GENE REGULATION 1
1.2 WHERE DOES DNA REPLICATION FIT INTO GENE REGULATION? 6
1.3 A REVIEW OF EUKARYOTIC DNA REPLICATION 10
1.3.1 Replication initiates from multiple sites 10
1.3.2 Replicon size 10
1.3.3 Temporal firing of replicons 12
1.3.4 Mammalian origins of DNA replication 13
1.3.5 Origin utilisation 16
1.3.6 Replication origins, nuclear localisation and scaffold attachment sites 18
1.3.7 Role of chromosomal position in replication timing 19
1.3.8 Summary 20
1.4 THE HUMAN GLOBIN COMPLEX AS A MODEL OF EUKARYOTIC GENE REGULATION 20
1.4.1 Structural and functional organisation of the human α globin cluster 21
1.4.2 Chromosomal environment surrounding the human α globin cluster 22
1.4.3 Comparison with the human β globin cluster 25
1.4.4 Developmental regulation of the α and β globin clusters 26
1.4.5 Alpha-thalassemia 27
1.4.6 Natural mutations of the α cluster 27
1.4.7 Regulation of α globin gene expression 28
1.4.8 Discovery of an X-linked protein critical for α globin gene expression 29
1.4.9 Expression of α globin gene constructs in stable MEL transfectants and transgenic mice 31
1.4.10 Summary 32
1.5 AIMS OF THIS THESIS 34

CHAPTER 2 MATERIALS AND METHODS 37

2.1 CHEMICALS AND REAGENTS 37
2.2 GENERAL LABORATORY TECHNIQUES 37
2.2.1 Bacterial culture 37
2.2.2 Preparation of nucleic acids 38
2.2.3 Manipulation of DNA 39
2.2.4 Tissue cell culture 40
2.3 FISH ANALYSIS TO MEASURE REPLICATION TIMING 41
2.3.1 Nick translation labelling of probes 41
2.3.2 Colourimetric assay to measure label incorporation 41
2.3.3 Preparation of interphase nuclei 42
2.3.4 Assay for percentage of nuclei in S phase 42
2.3.5 Preparation of slides 43
2.3.6 Fluorescence *in situ* hybridisation 43
2.3.7 Detection 44
2.3.8 Fluorescence microscopy and slide scoring 44

CHAPTER 3 DEVELOPMENT OF A FISH PROTOCOL TO MEASURE REPLICATION TIMING 46

3.1 INTRODUCTION 46
3.2 DEVELOPMENT OF THE FISH PROTOCOL 47
3.2.1 Technical refinement of standard FISH 47
3.2.2 Detection of S phase nuclei 48
3.3 DEVELOPMENT OF CRITERIA FOR SCORING 50
3.3.1 Selection of slides to score 50
3.3.2 Slide scoring 50
3.3.3 How many nuclei to score? 50
3.3.4 Independent scorers 51
3.4 REPRODUCIBILITY OF RESULTS 52
3.5 CONTROLS 53

3.5.1 Measuring hybridisation efficiency 53
3.5.2 Selection of early and late replicating controls 54
3.6 DISCUSSION 58

CHAPTER 4 ESTABLISHING THE NORMAL PATTERN OF REPLICATION THROUGHOUT THE TELOMERIC REGION OF 16P 62

4.1 INTRODUCTION 62
4.1.1 Rationale for research 62
4.1.2 Replication timing of the globin gene clusters 63
4.1.3 Research objectives 63
4.1.4 Selection of an erythroid cell-line 63
4.1.5 Individual allelic replication timing 64
4.2 METHODS 65
4.3 RESULTS 67
4.3.1 Replication of telomeric 16p in a normal non-erythroid cell-line. 67
4.3.2 Replication of telomeric 16p in a normal erythroid cell-line 69
4.3.3 Replication timing of the individual human 16p alleles 73
4.3.4 Measuring allelic asynchrony of the control probes 78
4.3.5 Measuring allelic asynchrony throughout the terminal region of 16p 79
4.4 DISCUSSION 82
4.4.1 The chromosomal band 16p13.3 is early replicating 82
4.4.2 The terminal region of 16p is generally early replicating in both erythroid and non-erythroid cell-lines 82
4.4.3 Does the 325kb region replicate synchronously? 83
4.4.4 Replication asynchrony and genetic imprinting 85
4.4.5 Replication timing and gene expression 85
4.4.6 Replication differences between the α and β globin clusters 87
4.4.7 Temporal order of replication throughout the 325kb region 88
4.4.8 Chromosomal organisation during interphase 90
4.4.9 Summary 91

CHAPTER 5 ANALYSIS OF THE REPLICATION TIMING OF α GLOBIN GENE CONSTRUCTS IN TRANSGENIC MICE 92

5.1 INTRODUCTION 92
5.2 METHODS 96
5.2.1 Culture of primary erythroid cells from transgenic mice 96
5.2 RESULTS 97
5.3 DISCUSSION 100

CHAPTER 6 THE ROLE OF THE α GLOBIN REGULATORY ELEMENT (HS-40) IN DNA REPLICATION 103

6.1 INTRODUCTION 103
6.2 METHODS 104
6.3 RESULTS 105
6.3 DISCUSSION 108
6.3.1 HS-40 does not influence replication timing 108
6.3.2 Does HS-40 influence where replication initiates? 108
6.3.3 Replication timing and chromosome structure 110
6.3.4 Do transcription factors play a direct role in replication? 111
6.3.5 Summary 113

CHAPTER 7 THE ZF DELETION: RELATIONSHIP BETWEEN DNA REPLICATION AND α GLOBIN GENE EXPRESSION 114

7.1 INTRODUCTION 114
7.1.1 Mutations giving rise to α-thalassemia 114
7.1.2 The ZF deletion 114
7.1.3 A heterochromatin-like region downstream of the α genes 117
7.1.4 Further examination of the replication characteristics of the heterochromatin-like region 119
7.1.5 Replication timing of the ZF fragment in transgenic mice 119
7.2 METHODS 120

7.2.1 Replication timing of the ZF mutant hybrid 120
7.2.2 Replication of the region spanned by cos12 120
7.2.3 Replication timing of the pZFBam10 construct in transgenic mice 121
7.3 RESULTS 123
7.3.1 Replication timing of the ZF mutant hybrid 123
7.3.2 Replication timing of the region spanned by cos12 124
7.3.3 Replication timing of the ZF fragment in transgenic cells 127
7.4 DISCUSSION 130
7.4.1 The ZF mutant chromosome 16 has a normal replication profile 130
7.4.2 A repressive chromatin structure can replicate early in S phase 130
7.4.3 GC content and replication timing 132
7.4.4 Mechanism of transcriptional silencing 132
A. The 'spreading' mechanism 133
B. The loss of a boundary/insulating element 133
C. Abnormal localisation within the interphase nucleus 134
D. Loss of a positive regulatory element 135
E. Models arising from studies of imprinted genes 136
F. Loss of a replication termination site 136
7.4.4. Summary 137

CHAPTER 8 REPLICATION TIMING OF THE HUMAN 16P SUBTELOMERIC REGION AND THE EFFECT OF TELOMERE TRUNCATION UPON DNA REPLICATION TIMING 139

8.1 INTRODUCTION 139
8.2 METHODS 143
8.2.1 Measuring the replication timing of the proterminal region of 16p type A allele 143
8.2.2 Replication timing of the mutant MEL-16 hybrid lines IC and TAT 144
8.3 RESULTS 146
8.3.1 Replication timing of the subtelomeric region in the type A allele 146
8.3.2 Replication timing of the truncated telomeres IC and TAT 147
8.4 DISCUSSION 150
8.4.1 Replication timing of the subtelomeric region of 16p 150
8.4.2 What determines late replication at telomeres? 151

CHAPTER 9 FINAL DISCUSSION 154

9.1 NATURE AND SPECIFICITY OF ORIGINS 154
9.2 FUTURE WORK TO ESTABLISH ORIGINS OF REPLICATION THROUGHOUT 16P 156
9.3 THE RELATIONSHIP BETWEEN DNA REPLICATION TIMING AND GENE EXPRESSION 157

REFERENCES 160

Chapter 1 Introduction

Although DNA replication has been extensively characterised in lower organisms, it has been particularly difficult to study in higher eukaryotes. At present, mammalian origins of replication are poorly characterised, replicons have not been identified and the influence of replication on other aspects of DNA metabolism (such as transcription, repair, recombination) are poorly understood. The development of a FISH-based assay by Selig *et al.* in 1992, allowed the pattern and timing of replication to be studied relatively easily in mammalian cells, and created the possibility of investigating the influence of replication on other aspects of cellular and molecular biology. To date, using this assay, replication has been analysed in relatively few well characterised regions of the human genome.

This thesis has addressed the role of replication at the human α globin locus. The structure of this telomeric region is well established and the functional genomics (including methylation, chromatin structure, presence of matrix attachment sites, analysis of natural mutants and constructs in transgenic mice) have also been extensively characterised. This wealth of information means it is an ideal region to evaluate the influence of replication on many different aspects of chromosome structure and function.

1.1 The complexity of eukaryotic gene regulation

A diagram to summarise the current general model of eukaryotic gene regulation is shown in fig 1.1, which will be referred to in the following text. Originally, efforts to understand eukaryotic gene regulation concentrated on analysing the structural genes (1) and identifying their local *cis*-acting sequences, including promoters (2), local enhancers (4) and silencers (3).

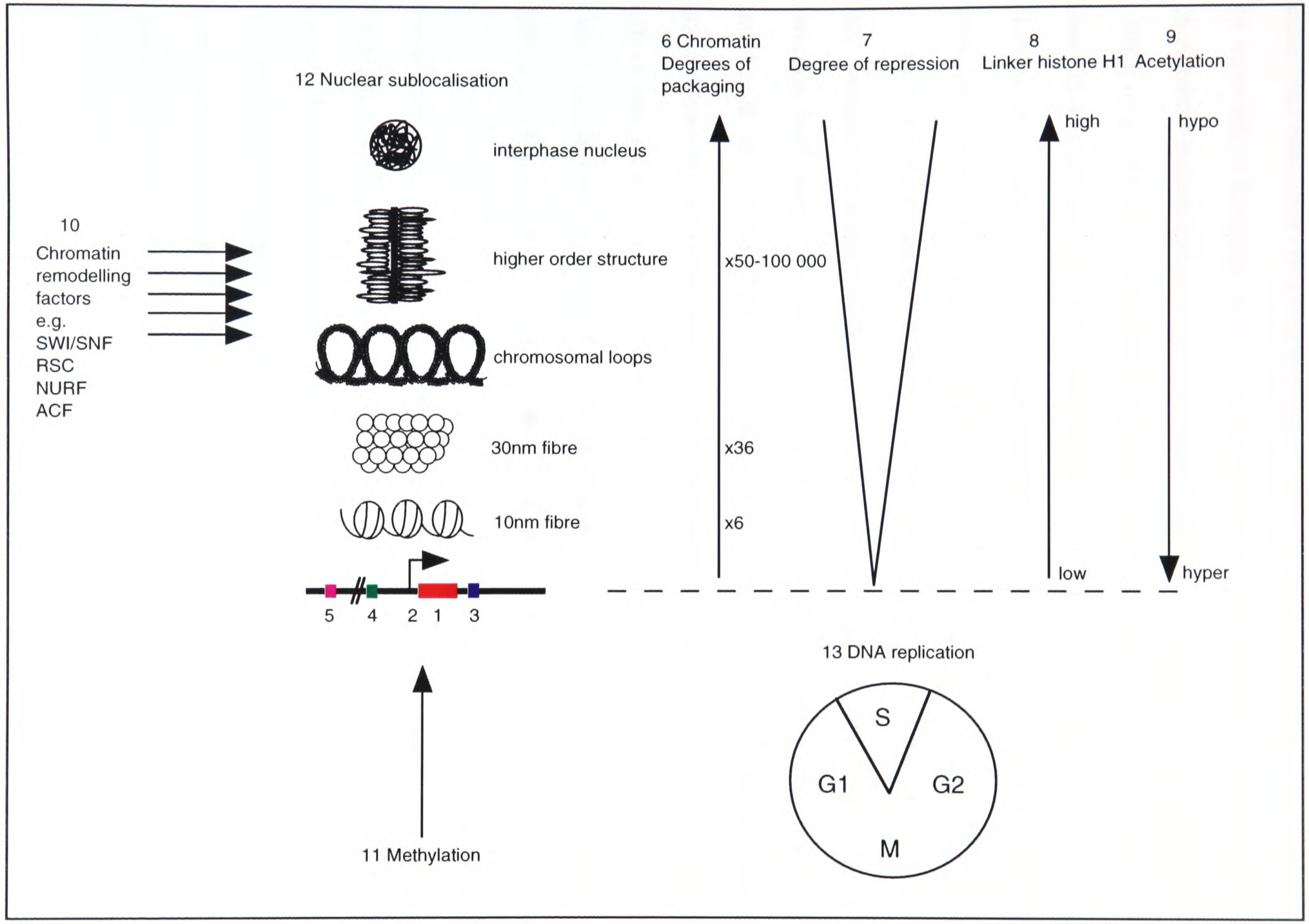


Fig 1.1 Generalised scheme of regulation of eukaryotic gene expression. Refer to text for further information.

As the human genome project extended and characterised cloned regions and sequences around individual genes, and as natural mutants of these genes were examined, it became clear that some *cis*-acting sequences regulating gene expression could be tens or even hundreds of kilobases from the structural genes they regulate (5). Examples of such elements include locus controlling regions such as the β globin LCR (Forrester *et al.* 1987, reviewed in Crossley and Orkin 1993) and long distance enhancers such as α globin HS-40 (Jarman *et al.* 1991). Each of these elements bind a variety of *trans*-acting factors *in vitro* and *in vivo* to form complex DNA/protein structures. Analysis of these factors has revealed the presence of DNA binding domains (such as homeodomains, zinc finger, leucine zipper and helix-loop-helix motifs) which target the proteins to *cis*-acting sequences within regulatory elements. Other domains are necessary for the interaction with other proteins. Activation domains (for example acidic domains), interact with components of the transcriptional machinery or other activator proteins to stimulate transcription (reviewed in Tjian and Maniatis 1994). Repressive domains (such as alanine rich or proline rich motifs), are thought to recruit other proteins to create a repressive chromatin structure or interact directly with the transcriptional machinery to inhibit transcription (reviewed in Hanna-Rose and Hansen 1996).

Over the last few years, the importance of chromatin (6) in the regulation of gene expression has emerged. Originally chromatin was thought to be merely a way of packaging DNA to fit within the nucleus, with little or no effect on gene expression or DNA metabolism in general. This view has now completely changed, and it is now becoming clear that chromatin plays a major role in DNA metabolism. In general, it is thought that higher order chromatin structure represses gene activity (7) with higher degrees of packaging causing increasing levels of repression (see Wolffe 1995).

At present, the precise chromosomal structure produced by chromosomal packing in the cell nucleus is unclear (6). It is known that DNA complexes with histones to form 10nm and 30nm chromatin fibres. Beyond this, much less is known about the higher levels of packaging, although several models exist. The most widely accepted of these

is the 'scaffold-loop' model proposed by Saitoh and Laemmli (1994), whereby chromatin forms loops which attach to a nuclear scaffold. Alternative models of chromatin compaction include successive helical coiling (reviewed in Earnshaw 1988).

Chromatin modification may be associated with different packaging states and may even be primarily involved in unfolding repressive chromatin structures. The most extensively studied of these is acetylation (9). In general, condensed transcriptionally inactive chromatin is associated with low levels of histone acetylation, whereas transcriptionally poised or active genes are associated with hyperacetylated histones (Davie 1998). The histone composition of chromatin also appears to vary with levels of compaction (8). The linker histone H1 is abundant in highly condensed chromatin, whereas in less condensed structures it is depleted (reviewed in Travers 1994).

How does this repressive chromatin structure become de-repressed? Since the cloning of the SWI/SNF complex in yeast, a new field has arisen, analysing a group of proteins that are thought to play an important role in de-repressing chromatin structure (10) and allowing classical transcription factors to gain access to their cognate binding sites within critical *cis*-acting sequences (such as promoters and enhancers discussed earlier). These proteins are known as chromatin remodelling factors (reviewed in Armstrong and Emerson 1998).

Once the DNA is accessible for transcription factor binding, other so-called epigenetic modifications may influence transcription factor accessibility at *cis*-acting sequences, thereby influencing transcriptional activation of genes. By far the most extensively studied of these is DNA methylation (11), which occurs at CpG dinucleotide sequences (reviewed in Kass *et al.* 1997). In general, methylated DNA templates are associated with inactive genes (for example, X inactivated genes, imprinted genes, and genes that become switched off in disease, such as genes involved in tumour suppression and fragile X syndrome). Conversely, *cis*-acting sequences of active genes are generally unmethylated. However, this form of modification is by no means a universal finding.

In fact, the importance of methylation in gene expression is not clear; for example, despite the complexity of gene regulation in *Drosophila Melanogaster,* there is no methylation in this multicellular eukaryotic organism.

As techniques have developed to visualise nuclear metabolism in higher eukaryotes, it has become of increasing interest to investigate the relationship between the position of a gene within the nucleus, and its transcriptional activity (12). Originally it was thought that the nucleus consisted of a random arrangement of chromosomes, with nuclear position playing no role in regulation. Opinion has now changed to a general consensus that chromatin is arranged to form a functional framework within the nucleus, which is important in the regulation of metabolic processes such as transcription, recombination and repair. Chromatin found near the nuclear periphery is generally thought to be metabolically inert, whereas chromatin within the internal nuclear compartment is thought to be more active. Individual chromosomes occupy discrete nuclear territories, and genes within those chromosomes, are thought to occupy a defined region within the territory (reviewed in Lamond and Earnshaw 1998).

Of the postulated mechanisms of gene regulation discussed in this section (chromatin structure, acetylation and methylation), most have been considered in unsynchronised cells, where one is measuring an average of what happens during the life cycle of each cell. However, we know that many of these parameters vary during the cell-cycle and this must be considered when constructing models of eukaryotic gene regulation. A further complexity concerns the cellular heterogeneity in gene expression. For example, if one considers erythropoiesis, developmental switching of embryonic to fetal to adult genes occurs, giving rise to different haemoglobin molecules (reviewed in Stamatoyannopoulos and Nienhuis 1994). A population of erythropoietic cells may consist of individual cells with different developmental programmes; analysis of erythropoiesis using the whole population would not necessarily reflect the situation in each individual cell.

1.2 Where does DNA replication fit into gene regulation?

An important stage of the cell-cycle when chromatin assembly and epigenetic modification can be dismantled and reprogrammed is during DNA replication (13). It is widely assumed that the process of replication during S phase causes a transient disruption of chromatin structure (fig 1.2), which may provide a window of opportunity for classical transcription factors to compete with histones to bind their cognate *cis*-acting sequences within promoter or enhancer elements. It also seems plausible that if DNA replication occurs at a time in the the cell-cycle when the local concentration of transcription factors is high, it would favour the assembly of a protein complex at promoters or enhancer elements of a gene to stimulate transcription. Conversely, if replication occurs when transcription factors are not available for binding at promoters or enhancers, this may lead to inactivation of a gene (reviewed in Wolffe 1995).

This model of replication-transcription coupling is supported by the fact that replication of a particular gene always occurs at the same time in S phase in differentiated cells (Epner *et al.* 1981, Furst *et al.* 1981, Braunstein *et al.* 1982). Initial replication timing studies also showed that active genes replicate early in S phase and inactive genes replicate later in S phase (Goldman *et al.* 1984, Holmquist 1987). As more genes have been examined, it seems this is not entirely the case, as some inactive genes replicate early (Hatton *et al.* 1988, reviewed in Eyre-Walker 1992). Genes or gene clusters which have a tissue-specific pattern of expression, appear to fall into two categories. The first category demonstrate an expression-linked pattern of replication, replicating early in S phase when expressed, and replicating late in S phase when unexpressed. Examples of these genes/gene clusters include the human β globin complex (Holmquist 1987, Epner *et al.* 1988), the human cystic fibrosis locus (Selig *et al.* 1992), the mouse albumin gene and the mouse T-cell receptor β-chain constant region gene cluster (Hatton *et al.* 1988). The second category are genes or gene clusters that replicate early in S phase irrespective of whether they are expressed or not. These include the mouse α globin cluster (Epner *et al.* 1981, Furst *et al.* 1981, Holmquist 1987, Hatton *et al.*

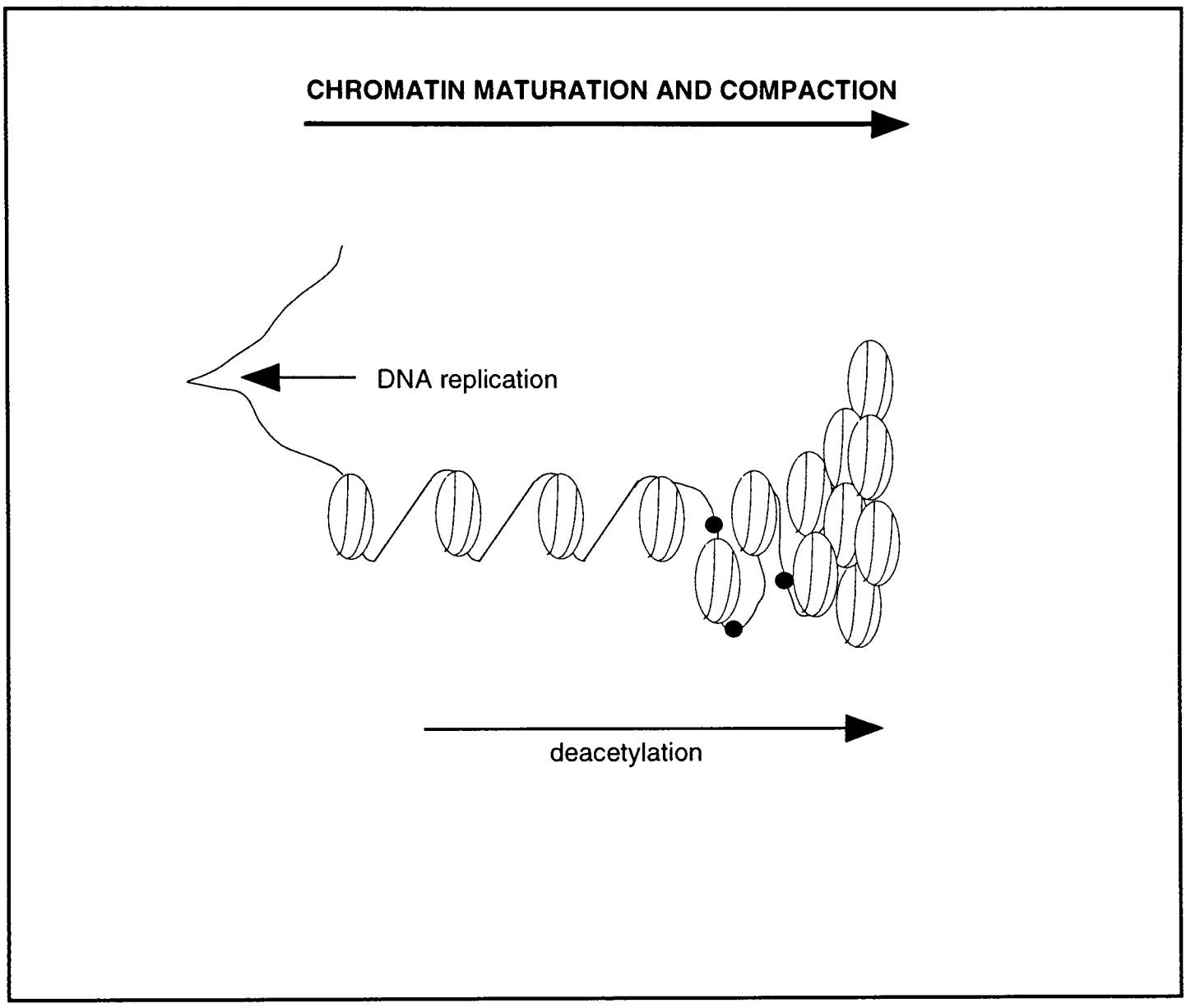


Figure 1.2 The role of DNA replication in transcription and chromatin assembly. During replication *cis*-acting regulatory elements become more accessible to *trans*-acting factors. Following replication, chromatin matures and compacts. Histone H1 is added (black filled circles) and histones are deacetylated. As compaction proceeds, chromatin becomes increasingly repressive against transcription (adapted from Wolffe 1995).

1988) and the mouse immunoglobulin Kappa constant chain gene cluster (Holmquist 1987, Hatton *et al.* 1988). Replication timing data suggest then, there is a correlation between early replication and gene expression. This has led to the proposal that early replication may be a requirement for gene activity.

Chromatin remodelling factors such as chromatin assembly factor 1 (CAF-1) may also take advantage of the disruption caused by replication, to redirect chromatin assembly (reviewed in Krude 1995). Therefore, in some instances, DNA replication could provide an opportunity to change the higher order chromatin state from repressive to active, and vice-versa. However, this is not a universal mechanism of chromatin remodelling factors, since the SWI/SNF complex does not require replication to exert its effect. Instead, it utilises ATP hydrolysis as an energy source to modulate chromatin structure (reviewed in Armstrong and Emerson 1998).

Replication may also be an opportunity for changing the acetylation status of chromatin. The CAF-1 protein complex possesses acetyltransferase activity, with the potential of modifying chromatin during the passage of the replication fork. This could be another mechanism of influencing the subsequent packaging of chromatin (reviewed in Grunstein 1997).

The process of DNA replication may also provide an instant in the cell-cycle for resetting methylation patterns at CpG dinucleotides, which occurs in X inactivation during normal development (Hansen *et al.* 1992) and in disease conditions such as tumourigenesis (Laird and Jaenisch 1996) and fragile X syndrome (Flannery *et al.* 1995). It has recently been shown by Chuang *et al.* (1997) that DNA methyltransferase is targeted to the replication machinery via a protein-protein interaction with proliferating cell nuclear antigen (PCNA). This interaction is thought to be controlled during the cell-cycle. It is possible that the time at which a gene replicates determines whether it becomes methylated and silenced, or unmethylated and active (Hansen *et al.* 1997).

Several proteins have been identified that are involved in DNA replication (reviewed in Dutta and Bell 1997). One of the key components is the multisubunit origin recognition complex (ORC), which was originally described in yeast (Bell and Stillman 1992, Diffley and Cocker 1992) and is now known to be highly conserved between species, including humans (Gavin *et al.* 1995, Gossen *et al.* 1995). In addition to functioning in DNA replication, this complex may also be involved in gene silencing, by recruiting other proteins which are involved in establishing a heterochromatic, condensed structure. In yeast, ORC has been shown to interact with the silencing protein SIR1 (Triolo and Sternglanz 1996) and in *Drosophila* and *Xenopus,* the complex interacts with another protein involved in heterochromatin formation, HP-1 (Pak *et al.* 1997).

It has been shown in this section, how the process of replication and the time at which it occurs in the cell-cycle, may be important in determining gene activity by several different mechanisms. It may allow direct access of transcription factors to bind their cognate sequences to activate transcription, or influence the higher order of packaging and level of repression of chromatin or finally, determine whether gene regulatory elements (such as promoters or enhancers) become methylated or not. Proteins which function in DNA replication may also be important in coupling the process to other functions such as transcriptional repression, by influencing chromatin structure.

From this viewpoint, it can be postulated that the inheritance of a replication timing programme from a parent to a daughter cell will be important for the maintenance of the same pattern of transcriptional activity, as seen in terminally differentiated, proliferating cell types. If the replication timing programme of a cell changes from a parent to daughter cell, it could be a mechanism to change the pattern of gene expression, as a part of a developmental process (such as erythropoiesis or X inactivation).

Here I have outlined the potential importance of DNA replication timing in regulating chromosome structure and gene expression. However, it is still unclear at present, whether early replication is the primary determinant of chromatin packaging and gene

expression. It could merely be a secondary consequence of a permissive chromatin environment where genes are active. The interplay that exists between different metabolic processes such as replication and transcription is complex and the chronological order of events such as replication, methylation, acetylation, chromatin configuration and nuclear position, which lead to gene activation is not fully understood.

The following section of the introduction reviews in more detail what is known about DNA replication in higher eukaryotic systems and how it relates to the other facets of gene regulation (nuclear sublocalisation, chromatin structure, methylation etc).

1.3. A review of eukaryotic DNA replication

1.3.1 Replication initiates from multiple sites

Evidence from electron microscopy and fibre autoradiography in the 1970s, indicated that initiation of DNA replication occurs at multiple sites in higher eukaryotes. Replication forks proceed bidirectionally from each initiation point until they meet with forks initiated from adjacent sites, whereupon the newly synthesised DNA fuses to form a long continuous strand. Each segment of DNA replicated from a single initiation site, or origin of replication, is known as a replicon (Hand *et al.* 1975, Hand *et al.* 1978 and references therein). This replicon model of replication is shown in fig 1.3. From the study of yeast, a general model for DNA replication initiation has been developed, which is analogous to the initiation of transcription. It involves the stepwise assembly of a core initiation protein complex at the origin DNA sequence, which may require auxiliary *trans*-acting factors for activation (Heintz 1992, Dutta and Bell 1997).

1.3.2 Replicon size

It has been discussed that timing of replication may be important in determining whether a gene is active or not. This is governed by two factors. The first is the point in time during S phase, when replication is initiated at an origin, or in other words, when a origin 'fires'. The second is how far a gene lies from an origin (the farther a gene lies,

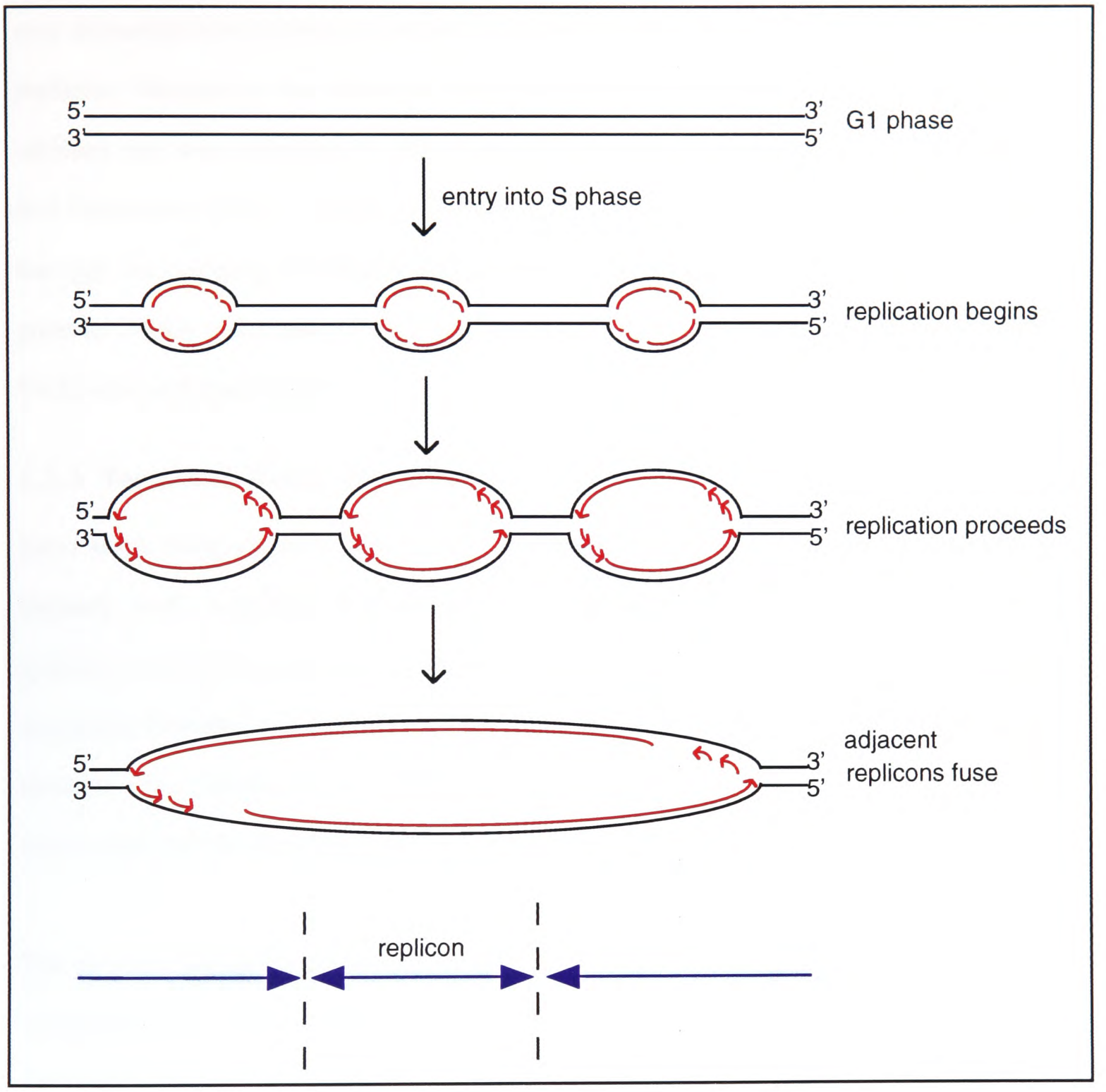


Fig 1.3. The replicon model of DNA replication as originally proposed by Jacob, Brenner and Cuzin (1963). Co-ordinate firing of adjacent origins occurs during S phase. Leading strand synthesis is shown by continuous arrows, and lagging strand synthesis as discontinuous arrows. As S phase proceeds, DNA replication continues, until newly synthesised DNA from adjacent origins fuse. The region of DNA replicated from a single origin is defined as a replicon (adapted from Adams 1991).

the longer it will take for the replication machinery to reach and replicate it). It is known that origins can fire at different times during S phase (Reynolds *et al.* 1989, Kitsberg *et al.* 1993b). As individual replicons have different replication timing programmes, this could influence the transcriptional activity of genes within them. In this sense, replicons may demarcate transcriptional domains and it becomes important to delimit the size of a replicon. Measuring the distance between initiation sites by electron microscopy, replicon size was estimated to vary between 50-300kb in mammalian cells (Edenberg and Huberman 1975). Another estimation by electron microscropy and more recently through the mapping of adjacent replication origins, replicon size is thought to be as great as 500kb in humans (Yurov and Liapunova 1977, Bickmore and Oghene 1996, Verbovaia and Razin 1997).

1.3.3 Temporal firing of replicons

Early work using electron microscopy, showed that adjacent replication bubbles had a uniform size, indicating that initiation of replication in adjacent replicons occurs synchronously (Blumenthal *et al.* 1974). Subsequent work using fibre autoradiography supported this idea, and also demonstrated that initiation occurs at preferred, rather than random sites (Hand *et al.* 1975). These preferred initation sites, or origins of replication, will be discussed later.

The idea that adjacent replicons share the same temporal firing of origins, is supported cytogenetically, where reproducible, distinct banding patterns of early and late replication are seen on metaphase chromosomes (Latt 1976, Drouin *et al.* 1994). It is thought that each replication band consists of a cluster of 2-100 synchronously firing replicons (Craig and Bickmore 1993). In addition, pulse labelling of synchronised cells using ^{3}H-thymidine reveals two peaks of DNA synthesis during S phase, corresponding to the first half (S_E) and the second half (S_L) of the phase. These peaks are reported to be separated by a drop in DNA synthesis called the 3C pause (Goldman *et al.* 1984, Holmquist 1987). From this result, it is proposed that DNA sequences

replicated in S_E are clustered together into early replicating chromosomal bands and sequences replicated during S_L are clustered into late replicating bands.

Although replication banding suggests that regions of the genome with similar replication timing are grouped together, using standard resolution, an average band at metaphase consists of 5-10 megabases of DNA (Bickmore and Sumner 1989). Therefore a banding pattern can only be a crude average of the whole region. There is substantial evidence that a gene and its flanking sequence within a band can switch replication timing throughout development (Hyrien *et al.* 1995) and according to pattern of expression (Goldman *et al.* 1984, Holmquist *et al.* 1987, Hatton *et al.* 1988, Epner *et al.* 1988, Michaelson *et al.* 1997). These changes are not accompanied by an observable change in the replication banding pattern (Craig and Bickmore 1993).

Early replicating bands seen during metaphase localise to the internal compartment of the nucleus during interphase, whereas late replicating bands assemble at the nuclear periphery (Ferreira *et al.* 1997). As early replicating bands are more gene-rich than late replicating bands (Holmquist 1992, Craig and Bickmore 1993), positioning within the nucleus may be important for chromosome function, such as gene transcription. What is not clear yet, is the sequence of events which result in gene activation. Does nuclear position determine the time at which at region is replicated, which in turn influences whether a gene is transcribed or not? Or does early replication result in the correct positioning of a region within the nucleus which allows transcription to occur?

1.3.4 Mammalian origins of DNA replication

The next stage in understanding DNA replication, is to pinpoint where in the genome origins are found. Characterisation of these regions and the proteins that bind them might shed some light on how DNA replication is regulated, and how it may be coupled to the other DNA metabolic processes, such as transcription.

Unlike the identification of origins in the simple eukaryote *Saccharomyces Cerevisiae*, identification of origins in higher eukaryotes has been very difficult. In *S.Cerevisiae*,

the functional test for origins relies on their ability to behave as autonomously replicating sequences (ARSs). The ARS element confers origin function to a plasmid, allowing it to replicate when it is transferred to cells or cell extracts containing the required replication proteins (reviewed in Diffley 1996). Analysis of ARS sequences has revealed they are 100-200bps in size and contain an essential 11bp ARS consensus sequence (ACS), as well as motifs which function as DNA unwinding elements (Lin and Kowalski 1997). Many ARS elements function as origins *in vivo* and are where DNA replication initiates (Bielinsky and Gerbi 1998). Nevertheless, other sequences that behave as ARS elements in experimental assays do not normally function as chromosomal origins.

When looking for origins it higher eukaryotes, it has been found that most large (>10kb) fragments will provide some ARS activity in mammalian cells (Masukata *et al.* 1993, Krysan *et al.* 1993). Subcloning of smaller fragments from these larger sequences however, does not yield functional origins (Krysan *et al.* 1993). To compound the problem, fragments do not show any strict DNA sequence conservation. Due to the lack of sequence specificity, and the lack of a robust assay to detect origins of replication in higher eukaryotes, other methods have been developed to map individual origins (reviewed in Falaschi *et al.* 1993). Of the thousands of mammalian origins that must exist, only around thirty have been identified. These are presented in table 1.1. From these mapping studies, it seems that DNA replication does initiate at specific sites. They do not have a strictly conserved DNA sequence like yeast, although a comparison of sequences of six mapped origin regions has led to the development of a loose consensus sequence (Dobbs *et al.* 1994). It appears that in somes areas of the genome, origins can be localised to within a few kilobases (such as the human β globin and ppv1 origins), whereas in other areas of the genome, there is a wider 'zone' where initiation can occur (the dihydrofolate reductase gene region). Replication can initiate anywhere within the zone, although a few sites are preferentially used (reviewed in Gilbert 1998). From table 1.1, it can be seen that almost half of the origins that have been identified are associated with the promoters of genes, where transcription

species	origin	Associated with a CpG island?	Associated with a promoter element?	Reference
human	hsp70 gene	yes	yes	Taira *et al.* 1994
human	ppv1 gene	yes	yes	Giacca *et al.* 1994
human	c-myc gene	yes	yes	Vassilev and Johnson 1990
human	cDNA 343			Wu *et al.* 1993
human	dystrophin gene 1	no	no (intronic between exon 7/8)	Verbovaia and Razin 1997
human	dystrophin gene 2	no	no (intronic between exon 28/29)	Verbovaia and Razin 1997
human	dystrophin gene 3	no	no (intronic between exon 43/44)	Verbovaia and Razin 1997
human	dystrophin gene 4	no	no (intronic between exon 46/48)	Verbovaia and Razin 1997
human	dystrophin gene 5	no	no (intronic between exon 64/68)	Verbovaia and Razin 1997
human	β globin	no	yes	Kitsberg *et al.* 1993b
human	rRNA genes			Gencheva *et al.* 1996 and refs therein
human	Igf2 gene	yes	yes	Delgado *et al.* 1998
chicken	α globin	(hypersensitive site)	yes	James and Leffak 1986; Verbovaia and Razin 1995
chicken	β globin	no	yes	Aladjem *et al.* 1995
chicken	histone H5 genes			Trempe *et al.* 1988
mouse	rRNA genes			Gencheva *et al.* 1996
mouse	Ig heavy chain			Ariizumi *et al.* 1993
mouse	ADA gene			Virta-Pearlman *et al.* 1993
mouse	ADA gene			Carroll *et al.* 1993
mouse	CAD gene			Kelly *et al.* 1995
hamster	rhodopsin gene	no		Gale *et al.* 1992
hamster	DHF (ori γ)			Handeli *et al.* 1989
hamster	ribosomal S14 gene			Tasheva and Roufa 1994
hamster	DHFR gene	no		Wang *et al.* 1998 and refs therein
hamster	APRT gene	yes	yes	Delgado *et al.* 1998
hamster	TK gene	yes	yes	Delgado *et al.* 1998
hamster	GADD gene	yes	yes	Delgado *et al.* 1998
hamster	EF2 gene	yes	yes	Delgado *et al.* 1998

Table 1.1. Known mammalian origins of replication. CpG islands have been found at origins by Delgado *et al.* 1998. The hypersensitive site at the chicken α globin origin was determined by Berberich and Leffak 1993.

complexes assemble. A recent paper by Delgado *et al.* (1998) has shown that CpG islands can function as replication origins, which are also associated with promoter elements. What is unclear at this stage is whether all CpG islands can function as origins, or only CpG islands that are part of a promoter sequence.

It has previously been argued that the process of DNA replication may allow access of classical transcription factors to bind DNA regulatory elements, through the disruption of chromatin structure. However, in viruses and prokaryotes, the argument is reversed: the presence of transcription factors or the act of transcription itself seems to stimulate the process of replication. Transcription may facilitate DNA replication by unwinding DNA or by remodelling chromatin at origins allowing replication factors to bind cognate DNA elements. Activation domains of transcription factors may also activate replication initiation by direct interaction with the replication machinery (DePamphilis 1988, Heintz 1992, DePamphilis 1993). If transcription factors function both in transcription and replication, it would allow the precise co-ordination of both processes during the cell-cycle, with the time at which an origin fires being dependent upon the presence of specific transcription factors. If the process of DNA replication is analogous to the process of transcription, initiation of replication at an origin may require distal regulatory elements, such as enhancers or locus controlling regions. This appears to be true for the activation of the human β globin origin (Aladjem *et al.* 1995).

1.3.5 Origin utilisation

It has been shown that rather than being fixed, sites that are utilised as origins can vary during development. In *Xenopus* early embryos, cell division occurs very rapidly (every 30 minutes), with the length of S phase around 15 minutes. This high rate of replication is achieved by having a vast number of origins, arranged into small, uniformly sized replicons of around 12kb (Hyrien and Mechali 1993). Twelve divisions later (the midblastula transition), when transcription begins, a transition occurs when replicon size increases and becomes more variable (Hyrien *et al.* 1995). The same phenomenon has been observed in *Drosophila,* where in early development, S phase is

9 minutes long and replicon size is uniformly sized at around 9 kb (Blumenthal *et al.* 1974, Zakian 1976, Shinomiya and Ina 1991). What triggers the developmental switch in origin utilisation is unclear. It could be that the structural features of the chromosome determine where initiation can begin, hence the regularity of initiation in the early embryo is imposed by a repeating chromatin structure (Hand *et al.* 1978, Coverley and Laskey 1994, Burhans and Huberman 1994). As chromatin templating occurs during development, initiation becomes restricted to a smaller number of sites.

In mouse, the entire immunoglobulin heavy chain (Igh) locus is replicated from a single origin downstream of the cluster in non-B cells (Brown *et al.* 1987). As the locus is large (approximately 300kb) and appears to be part of a single replicon in non-expressing cells, it takes almost the whole length of S phase to complete its replication. Genes close to the origin are replicated in early S phase (constant chain genes), whereas the more distal genes (variable region genes) replicate much later (Brown *et al.* 1987, Hatton *et al.* 1988).When the cluster is expressed in B cells, replication initiates at multiple sites throughout the locus (Michaelson *et al.* 1997), so that all genes are replicated early in S phase. Whether this mechanism of utilising more origins in expressing cells to achieve early replication is common to other tissue-specific gene clusters, has yet to be determined. It is known that this is not the mechanism for early replication in the human β globin cluster. At this locus, a single origin is utilised to replicate the whole cluster in expressing and non-expressing cell-types (Kitsberg *et al.* 1993b), but the timing at which it fires is developmentally controlled to be early in expressing cell-types and late in others (Epner *et al.* 1981, Dhar *et al.* 1988). Deletion of this origin results in the region being replicated from forks extending from a neighbouring origin (Kitsberg *et al.* 1993b, Aladjem *et al.* 1995). It is not known whether replication is initiated by another 'bona fide' origin or whether an alternative site is augmented to function as an origin.

1.3.6 Replication origins, nuclear localisation and scaffold attachment sites

Visualisation of interphase nuclei suggests that replication occurs at distinct sites within the nuclear compartment (reviewed in Strouboulis and Wolffe 1996). Many groups have reported the fact that when cells are labelled with bromodeoxyuridine (BdU) or biotinylated deoxyribonucleotides (dNTPs) and stained with fluorescent antibodies, clusters of replication 'foci' are seen (Manders *et al.* 1992, Wansink *et al.* 1994, Bickmore and Carothers 1995, Jackson 1995, Ferreira *et al.* 1997). Each focus is thought to contain a cluster of replication origins, which are co-ordinately activated (Nakamura *et al.* 1986). In human somatic cells, the pattern of replication foci changes during the progression of S phase. Foci first appear in late G1 prior to replication (Jackson 1995) and when replication initiates at the start of S phase, foci are small, and appear evenly distributed throughout the interphase nucleus. As time progresses, foci become bigger and punctate. Targeting of a gene to the correct focus may be important in determining when during S phase replication initiates, which as previously discussed, may determine the transcriptional activity of the gene.

Evidence that the sites of DNA synthesis are fixed comes from electron microscopy of replicating nuclei, where most of the chromatin has been removed from the DNA by nuclease digestion (Hozak *et al.* 1993). Conservation of nuclear structure is obtained by carrying out digestion within agarose beads, which reveals incorporation of labelled nucleotides at replication 'factories'. The observation of nascent DNA being extruded from the factory suggests that that DNA is being spooled through a fixed architecture (Hozak *et al.* 1993). This model is somewhat contradicted by other data, which suggest that replication origins are fixed to the nuclear substructure via scaffold attachment sites (SARs), otherwise known as matrix attachment sites (MARs) throughout interphase and metaphase. Sequences that function as origins are fixed via a DNA/protein complex to the nuclear scaffold, whilst the replication machinery moves (Dijkwel and Hamlin 1988, Bickmore and Oghene 1996, Lagarkova *et al.* 1998). Perhaps scaffold attachment sites are necessary for directing genes to the replication machinery. This may

be a mechanism for certain genes, but we know that some large gene-rich regions do not contain scaffold attachment sites (Jarman and Higgs 1988).

1.3.7 Role of chromosomal position in replication timing

It has been discussed earlier in this section, how the temporal firing of origins within a eukaryotic cell is regulated during the cell-cycle. This programme may change during development and differentiation. Two levels of regulation appear to be in operation: the selection of a region to function as an origin and the timing of origin activation. From experiments performed in yeast, it appears that the temporal firing of an origin is not determined by the sequence itself, but by its chromosomal location. Origins that are located near a telomere fire late in S phase (McCarroll and Fangman 1988, Reynolds *et al.* 1989, Ferguson *et al.* 1991). Transferral of an early firing origin to a telomeric region causes the origin to become late firing. Conversely, removal of a telomeric origin to a circular plasmid causes a switch from late to early replication (Ferguson and Fangman 1992), implicating the telomere as the determinant of late replication. However, further analysis of two other telomeric origins found they maintained their late pattern of replication when moved from their normal chromosomal environment to a circular plasmid (Friedman *et al.* 1996).

How might chromosomal position affect DNA replication timing? Chromatin condensation involves the increasingly higher order packaging of the chromatin fibre (fig 1.1). Origin elements may not be accessible to replication *trans*-acting factors in highly condensed regions of the genome. Chromosomal position may also influence the position the region assumes in the interphase nucleus, and hence when replication occurs. Late replicating yeast telomeres are found at the nuclear periphery (Klein *et al.* 1992) and this compartmentalisation may play a role in regulating origin firing time. As discussed, origins may also require the presence of distal regulatory elements such as LCRs or enhancers, which may bind *trans*-acting factors required for origin initiation, or they may influence long range chromatin structure. Removal of an origin into an

ecotopic site in the genome away from these distal elements, could dramatically affect the firing time of an origin.

1.3.8 Summary

This section of the introduction has briefly reviewed what is currently known about DNA replication in eukaryotic cells. Initiation occurs at defined sites called origins. DNA that is replicated from a single origin is known as a replicon, and these are approximately 50-500kb in size. The time at which an origin fires in the cell-cycle is regulated. Early replication is associated with transcriptional activity, whereas late replication is associated with gene silencing. Replication is also spatially regulated within the interphase nucleus, occuring at specific sites which change during the cell-cycle. Origins of DNA replication are often associated with the promoters of genes, and it maybe that protein factors involved in gene transcription also function in DNA replication. Changing the position of an origin from one chromosomal location to another can influence the time during S phase when it fires, and it may alter the position the origin assumes within the interphase nucleus.

1.4 The human α globin complex as a model of eukaryotic gene regulation

So far, I have highlighted the complex interplay that exists between chromosome structure, DNA replication and nuclear sublocalisation, which in a concerted fashion, regulate gene expression. The next important stage in understanding the roles of these phenomena is to define the order and hierarchy in which they occur. It may be that the order of events which lead to gene expression differ from one region of chromatin (locus/domain) to another.

The human α globin cluster offers a particularly well characterised model of eukaryotic gene expression. Detailed replication analysis of the human α globin cluster has not been previously undertaken, although a provisional report (Goldman *et al.* 1984) suggested that it replicates early in the cell-cycle in non-erythroid cell-types. The aim of

this thesis is to describe the pattern of replication across the human α cluster in erythroid and non-erythroid cells. Since it has been suggested that chromosome context maybe important in regulating gene expression, these replication studies have been extended to include all of the surrounding well-characterised region, including the telomere. This has enabled me to provide the first detailed study of replication at a human telomere. Natural mutants of the human α globin cluster have been extensively studied and characterised, which has allowed me to examine the effect of distant regulatory elements on the pattern and timing of replication around the globin genes. In addition, other characterised mutants have allowed me to investigate the role of the telomere and subtelomeric region in controlling the pattern and timing of replication and the effect on gene expression in this telomeric region. Finally, I have investigated what happens to the pattern and timing of replication when various segments of DNA from this region are randomly integrated into the mouse genome, and how this relates to α globin gene expression. The following section of the introduction reviews what is already known about the human α globin complex and its surrounding chromosomal environment.

1.4.1 Structural and functional organisation of the human α globin cluster

The human α globin cluster consists of four functional genes: one embryonic gene (ζ2), two fetal-adult globin genes (α2, α1) and one gene of unknown function (θ1). It also includes three pseudogenes (ψζ1, ψα2, ψα1). The functional genes are arranged in the order in which they are expressed throughout development in erythroid cells: telomere-ζ2–α2–α1–θ1–centromere (reviewed in Higgs *et al.* 1989). The whole cluster is located in the terminal band of the short arm of chromosome 16 (16p13.3), at a short but variable distance of 170-430kb from the telomere (Buckle *et al.* 1988, Wilkie *et al.* 1991). The variation is due to polymorphism in the subtelomeric region, which gives rise to three common alleles, known as A, B and C (Wilkie *et al.* 1991). Type A occurs the most frequently, and sequences in this subtelomeric region are found on other non-

homologous chromosomes, including X, Y, 18, 10 and 9 (Wilkie *et al.* 1991, Kermouni *et al.* 1995, Flint *et al.* 1997a).

1.4.2 Chromosomal environment surrounding the human α globin cluster

A 300kb region of the type A allele, extending from the telomere, has recently been sequenced and analysed (Flint *et al.* 1997a, presented in fig 1.4). The region is extremely GC rich, averaging 53.9%, which is typical of a T band and an H3 isochore (Saccone *et al.* 1993, Bernardi *et al.* 1993, Flint *et al.* 1997a). Within the region, using a wide window of 20kb and steps of 10kb, variation in GC content can be visualised as a sinusoidal wave. Peaks occur every 80-90kb, with GC content varying from a maximum of 63% at a peak to an minimum of 46% at a trough. Using a smaller window of 100bp, and shifts of 10bp, fourteen sharp GC rich peaks can be seen. When these results are compared with CpG dincucleotide frequency, the peaks correspond to CpG islands, labelled A-N in fig 1.4. The 300kb region is also rich in partial and complete Alu-repeats, which comprise 25.7% of the total sequence (Flint *et al.* 1997a). CpG islands are often associated with promoter elements of genes and many of the CpG islands found in this region correspond to the 5' end of identified genes (Flint *et al.* 1997a). The remaining CpG islands may indicate where novel, as yet unidentified genes may be found. Every CpG island found in the region remains unmethylated in every cell-line examined, including the CpG islands associated with each of the α globin genes. Every CpG island is also associated with a hypersensitive site (HS). Many of these hypersensitive sites are constitutive, except the HSs which co-localise with the CpG islands at the α globin gene promoters, and the HS found at the HS-40 regulatory element, which are only present in erythroid cells (Vyas *et al.* 1992, Craddock *et al.* 1995). Additional HSs are found throughout the region that are not associated with CpG islands (Higgs *et al.* 1990, Vyas *et al.* 1992, Craddock *et al.* 1995, Flint *et al.* 1997a). These findings suggest the majority of the 300kb region where the α globin gene cluster is located, remains in a constitutively 'open' chromatin conformation, irrespective of α globin gene expression (Vyas *et al.* 1992, Craddock *et*

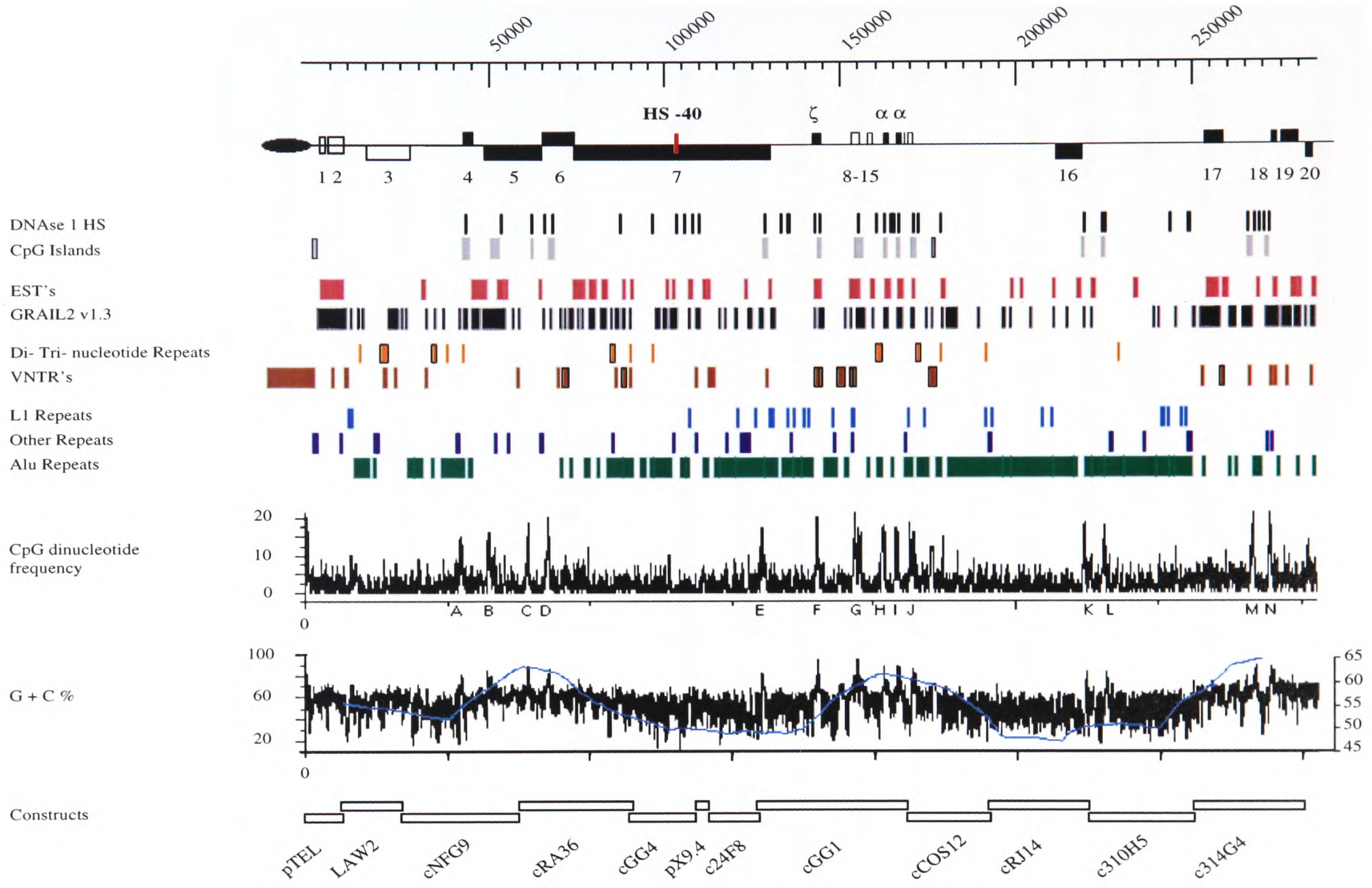


Fig 1.4 Structural and functional analysis of the terminal 300kb region of human chromosome 16p containing the human α globin cluster (from Flint *et al.* 1997a). In the top schematic, telomeric repeats (TTAGGG)n are represented as a black oval. Genes (filled boxes) and pseudogenes (open boxes) are shown above (orientated towards the centromere), below (orientated to the telomere) or on the line (orientation unknown) and are numbered. The α globin regulatory element (HS-40) is shown in red and the α–like genes are labelled. Below this, features are shown, some of which are discussed in the text . The CpG dinucleotide frequency plot shows positions of CpG islands, which are labelled A-N. In the G+C% plot, the blue line represents a window size of 20kb and a shift of 10kb (right scale). The graph in black measures GC content (%) using a window size of 100bp and shift of 10bp (left scale). Scale is in base pairs.

al. 1995). However, there are two regions of notable exception to the general observation of 'open' chromatin, where there are no HSs or CpG islands. These are the subtelomeric region and an approximately 38kb Alu-dense region, which lies 3' downstream of the α globin genes.

As can be seen in fig 1.4, the 300kb region is extremely gene dense (approximately 1 gene/20kb), which is a common feature of T bands (Saccone *et al.* 1992, 1996). Four ubiquitously expressed genes are found between the α globin structural genes and the telomere (genes 4-7); HS-40 is located in the intron of one of these genes (Vyas *et al.* 1992). A further five genes are found on the centromeric side of the cluster (genes 16-20, Vickers *et al.* 1993, Flint *et al.* 1997a). These genes show a diversity of function: in addition to the erythroid-specific, developmentally regulated α globin cluster, there is a membrane associated protein (gene 5), a DNA repair enzyme, MPG (gene 6), an inhibitor of GDP dissociation from the protein RHO (gene 18) and a protein disulphide isomerase (gene 19). The remaining genes have unknown function.

This comprehensive characterisation of the broader region where the α globin cluster is located, has shown that unlike the classical model of a tissue-specific gene cluster (Kitsberg *et al.* 1991), the α cluster is not organised into a discrete regulatory domain. It exists amongst other genes in an apparently unstructured isochore, where genes are arranged so that direction of transcription can change from one gene to the next and transcriptional units can physically overlap. These genes include housekeeping genes and other genes, which have a more restricted pattern of expression (Flint *et al.* 1997a). A similar arrangement has been observed previously in another GC rich isochore (Larsen *et al.* 1993), which may suggest this type of chromosomal organisational will be typical of GC rich regions.

1.4.3 Comparison with the human β globin cluster

The functional and structural organisation of the α globin cluster is in stark contrast to the β globin cluster. The β globin cluster is also located in a terminal band (11p15.5), but rather than the cluster being located very close to the telomere like the α cluster, it is located approximately 8 Mb from the telomere end, and is in fact close to the R/G band boundary of 11p15.5/11p15.4 (Quackenbush *et al.* 1995, chromosome 11 database http:\\shows.med.buffalo.edu\database.html). The β cluster fits the more classical definition of a tissue-specific cluster, as it organised into a discrete regulatory domain of 70kb, which solely contains the five β globin genes which are again, are arranged in the order of developmental expression: 5'-ε–Gγ–Aγ–ψβ–δ–β-3'. The major regulatory element consists of five different elements, which are collectively known as the β-locus controlling region (β-LCR); this element is located 4-20kb upstream of the ε gene (reviewed in Crossley and Orkin 1993). The domain is AT rich, in contrast to α cluster, which adopts a 'closed' chromatin environment (general DNase 1 insensitivity) in non-expressing cells, which undergoes decondensation to an 'open' chromatin conformation (DNase 1 hypersensitive) in expressing cells. This transition is thought to be mediated, at least in part, by the β-LCR (Forrester *et al.* 1990). The region displays an expression-linked pattern of replication, with replication taking place early in S phase when the locus is transcriptionally active, switching to late S phase when the cluster is transcriptionally inactive in non-erythroid cells (Epner *et al.* 1981, Furst *et al.* 1981). Replication is initiated at a single origin, which has been mapped to a 2kb region within the β globin gene promoter, which functions in both erythroid and non-erythroid cells (Kitsberg *et al.* 1993b). The contrasting features of the α and β globin clusters discussed here, and other differences are summarised in table 1.2 (from Higgs *et al.* 1998).

Feature	α globin cluster	β globin cluster
Location	16p13.3 telomeric	11p15.5 interstitial
GC content	53.9%	39.5%
CpG islands	Common	None
Gene density	High	Low
Alu family repeats	25%	approx 5%
LINE repeats	Rare	Present
Chromatin	Open	Closed→open
Matrix attachment sites	None detected	Common
Effect of hATRX mutations	Present	Absent
Regulatory element	Enhancer	LCR
Effect of deletion of regulatory element on long range chromatin structure	No	Yes
Predominant mutations	Deletions	Mutations
Expression in transient assays	Enhancer independent	Enhancer dependent
Expression in hybrids	Early	Late

Table 1.2. A summary of the differences in structure and function of the α and β globin gene clusters.

1.4.4 Developmental regulation of the α and β globin clusters

All human haemoglobins have a tetrameric structure, consisting of two identical α-like chains and two identical β-like chains. These structures change during development, to meet the differing oxygen transport requirement throughout fetal maturation. Both the α and β globin clusters are activated during the third week of embryogenesis to produce the embryonic haemoglobins Portland (ζ2ε2), Gower 1 (ζ2ε2) and Gower 2 (ζ2γ2), switching to fetal (α2γ2) and finally to adult major (α2β2) and minor (α2δ2) haemoglobins just prior to birth (reviewed in Stamatoyannopoulos and Nienhuis 1994). During normal development, both paternal and maternal alleles contribute to α and β globin gene expression. Considering the α globin cluster, where each allele has two copies of the fetal-adult genes (α2, α1), it has been shown that despite the protein products being identical, the two genes are not expressed equally. During fetal and adult life, the relative amounts of α2 mRNA and α1 mRNA vary by a ratio of approximately 3:1, which is thought to occur at the level of transcription (Orkin and Goff 1981, Molchanova *et al.* 1994). This expression differential creates a variance in severity of α-thalassemia seen in patients, according to which of the two genes is inactivated. A more severe phenotype is associated with loss of α2 expression. The disease phenotype which results from the perturbation of α globin gene expression is discussed in the following section.

1.4.5 Alpha-thalassemia

Mutations of the α globin cluster which down-regulate α globin gene expression, form the pathological basis of a disease called α-thalassemia (Online Mendelian Inheritance in Man entry number 141800). The lack of α globin chains results in a decrease in functional haemaglobin, and gives rise to an excess of γ chains in fetal life or β chains in adult life, which form non-functional, unstable haemoglobin tetramers Hb Bart's (γ4) or HbH (β4). According to the genotype, the spectrum of the clinical phenotype can vary from individuals who may be unaware of their condition, to those with severe anaemia who require regular blood transfusions, to a lethal condition called the Hb Bart's hydrops fetalis syndrome (Higgs 1993).

1.4.6 Natural mutations of the α cluster

Natural mutations have been extremely useful in identifying the critical *cis* and *trans* acting components which are necessary for the correct expression of the α globin genes. Deletions are more common than point mutations throughout the α globin cluster (reviewed in Higgs *et al.* 1989). The most frequent type of deletion causing α-thalassemia is removal of one or both of the fetal-adult structural genes (α2,α1) from one or both alleles of chromosome 16. This results in genotypes comprising 4, 3, 2, 1 or 0 copies of the α globin gene. Loss of one α globin gene (αα/-α) results in a mild α-thalassemia trait (otherwise known as α-thalassemia 2). Loss of two copies, either from the same allele (--/αα) or one from each allele (-α/-α) produces a severe form of α-thalassemia (α-thalassemia 1). Loss of three copies (--/-α) produces an even more severe phenotype called HbH disease; loss of all four copies, known as Hb Bart's hydrops fetalis (--/--) is lethal (reviewed in Higgs *et al.* 1989).

A second group of natural mutations were instrumental in the identification of the major positive regulatory element, HS-40. It is called HS-40 because it is located 40kb upstream of the start site of the ζ2 gene, and is associated with a DNase 1 HS site (Higgs *et al.* 1990). The first indication that this region might contain an important regulatory element came from natural mutants, where deletions of between 35-160kb,

upstream of the structural genes, resulted in the severe downregulation or complete loss of expression of both *cis*-linked α genes ($\alpha\alpha/(\alpha\alpha)^T$), despite the genes themselves being structurally normal (Higgs *et al.* 1990, Liebhaber *et al.* 1990, Hatton *et al.* 1990). These natural deletions helped to delimit the region of importance to a hypersensitive site 40 kb upstream of the ζ2 mRNA start site. To unequivocally demonstrate this was the single critical element in this upstream region, HS-40 was artificially deleted by homologous recombination. As expected, gene expression was severely downregulated (Bernet *et al.* 1995).

A novel type of deletion which results in α-thalassemia has recently been studied in the laboratory (Barbour *et al.* in preparation). This deletion was first identified by Indrak *et al.* (1993) and is described in full in chapter 7. Analysis has shown the deletion to be approximately 20 kb in size, which removes the α1 gene, and an 18kb region 3' downstream of the gene and silences the remaining, structurally normal α2 gene.

1.4.7 Regulation of α globin gene expression

Each globin gene contains three exons separated by two introns (reviewed in Higgs *et al.* 1989, Stamatoyannopoulos and Nienhuis 1994). Exons of the α2 and α1 genes are 100% homologous, and encode identical proteins, although these genes differ slightly within the second intron and the 3' non-coding sequence (Higgs *et al.* 1989). The α and ζ2 genes show a much lower amino acid homology (58%) between exons (Higgs *et al.* 1989).

Tissue specific and developmental stage-specific expression of each globin gene is achieved through the interaction of erythroid-specific and ubiquitous *trans*-acting factors with *cis*-acting regulatory elements. Known critical *cis*-acting regulatory elements within the α cluster are the promoters of each gene, and the more distal regulatory element, HS-40 (see above). HS-40 is approximately 350 bps in size, and binds two haemopoietic-restricted factors nuclear factor-erythroid 2 (NF-E2) and GATA-1, originally known as NF-E1 (Jarman *et al.* 1991, Strauss *et al.* 1992). GATA-1 is an X-

encoded DNA binding protein which contains a zinc finger motif that recognises the consensus sequence (T/A)GATA(A/G) (reviewed in Stamatoyannopoulos and Nienhuis 1994). NF-E2 is a heterodimer, which belongs to the B-zip family of transcription factors. It contains a basic DNA binding domain and a leucine zipper motif, which is involved in dimerisation of the 45kDa and 18kDa protein subunits. Expression of the 45kDa component is restricted to haemopoietic cells, whereas the 18kDa protein is thought to be ubiquitously expressed (reviewed in Stamatoyannopoulos and Nienhuis 1994). HS-40 also contains binding sites for ubiquitous transcription factors including CACC box binding factors (Jarman *et al.* 1991, Strauss *et al.* 1992, Higgs and Wood 1993).

The α globin gene promoters contain cognate binding sequences for ubiquitous transcription factors in addition to the minimal promoter region (TATA box) which binds the basal transcription machinery (RNA polymerase II, TFIID). A CCAAT box is found at the promoter of each globin gene, which binds the transcription factor CP1 (Chodosh *et al.* 1988). The α globin promoters also contain several consensus sequences for the transcription factor SP1, which contains a zinc finger DNA binding domain (Kadonaga *et al.* 1987). No erythroid-restricted transcription factor has been identified that binds within the α globin gene promoters, despite the formation of erythroid-specific hypersensitive sites at these elements.

1.4.8 Discovery of an X-linked protein critical for α globin gene expression

Expression of the α globin genes is also dependent on the production of a protein called ATRX (or XH2 or XHP), which is encoded by a gene found on the X chromosome, band Xq13.3 (Wilkie *et al.* 1990b, Gibbons *et al.* 1992, Gibbons *et al.* 1995a). Loss of function of this protein not only results in α-thalassemia (HbH disease), it has pleiotropic effects during development which result in other clinical manifestations, including mental retardation, urogenital abnormalities and a characteristic facial appearance. This clinical phenotype has been defined as ATR-X syndrome (Online Inheritance in Man entry number 300032, Weatherall *et al.* 1981, Wilkie *et al.* 1990,

Gibbons *et al.* 1995b). In this syndrome, although all of the α globin *cis*-acting regulatory elements are intact and remain unmethylated, and the structural genes are normal, expression is reduced to 30-65% (Wilkie *et al.* 1990). Interestingly, mutations within this gene do not appear to downregulate the β-globin genes (Gibbons *et al.* 1995a).

Characterisation of the ATRX gene has revealed that it is homologous to members of the SNF2 helicase superfamily (Gibbons *et al.* 1995a), which is a subgroup of the larger family of helicases and ATPases (Carlson and Laurent 1994, Ellis 1997). Many members of the subgroup are involved in regulation of transcription, by mediating chromatin remodelling. It is likely that this protein is a component of a large multiprotein complex and so far interactions have been shown between human ATRX protein and a chromatin associated protein EZH2 (Cardoso *et al.* 1998) and mouse ATRX and mouse HP-1 (Le Douarin *et al.* 1996). How the ATRX protein achieves specificity for the α globin gene cluster is unknown. It could bind DNA directly at *cis*-acting regulatory sequences or be targeted to these elements via protein-protein interactions with factors which specifically bind these sequences.

Analysis of the ATRX gene and identification of the mutations within it which result in ATR-X syndrome (Picketts *et al.* 1996, Gibbons *et al.* 1997), have highlighted several important conserved motifs. The first is the seven co-linear helicase motifs which defines ATRX as a member of the SNF2 family (Eisen *et al.* 1995). The second is a cysteine-rich motif, which resembles a particular type of zinc finger called a PHD finger (Aasland *et al.* 1995). This type of motif is involved in protein-protein interactions (Schmeichel *et al.* 1994) or it may be a DNA binding domain (Schwabe *et al.* 1994). Work is underway in the laboratory to identify protein or DNA targets which ATRX recognises, in order to understand how this protein regulates α globin gene expression.

1.4.9 Expression of α globin gene constructs in stable MEL transfectants and transgenic mice

Naturally occurring deletions helped to define essential *cis*-acting sequences that direct high-level, tissue-specific gene expression within the normal chromosomal environment. However, these elements may not be sufficient to drive high-level expression at an ectopic site within the human genome. Ideally for gene therapy, a fragment of DNA is required, which produces developmental-stage and tissue-specific expression, which is copy number dependent, and site of integration independent. A method to determine whether chromosomal position influences expression, is to randomly integrate the sequence of interest into another genome. Two systems commonly used to study expression of the human α globin genes are the stable transfection of MEL cells and transgenic mice.

Fragments of DNA containing the structural genes alone produce very low (<1%) or undetectable levels of α globin gene expression upon induction of stably transfected MEL cells (Charnay *et al.* 1984, Higgs *et al.* 1990). Linkage of the genes to HS-40 increases expression to 3-46% (Higgs *et al.* 1990, Sharpe *et al.* 1993). However, expression is still in general lower than the endogenous mouse α globin genes and it shows a large variation in levels, which may be dependent on the site of integration into the mouse genome. Moreover, expression is not dependent on copy number (Higgs *et al.* 1990, Sharpe *et al.* 1993). Inclusion of flanking DNA sequence in constructs in addition to the HS-40 element, does not improve expression levels or copy number dependence to any significant degree (reviewed in Higgs *et al.* 1998).

In parallel to the MEL stable transfectant data, constructs containing the α globin genes in the absence of HS-40, have been tested in transgenic mice; they do not express α globin (Palmiter and Brinster 1986, Hanscombe *et al.* 1989, Ryan *et al.* 1989, Higgs *et al.* 1990). The addition of HS-40 into constructs produces tissue specific expression, although as with the MEL stable transfectants, it does not confer copy number dependent expression and the levels are not developmentally stable (Higgs *et al.* 1990,

Sharpe *et al.* 1992, Sharpe *et al.* 1993). Increasing the size of constructs to include additional hypersensitive sites does not significantly improve results (Sharpe *et al.* 1993). Even a 70kb sequence which spans HS-40 and the structural genes and retains their normal genomic position does not produce fully regulated expression, despite the correct formation of CpG islands and HS sites (Gourdon *et al.* 1994). An even larger construct (approximately 150kb) has recently been tested. This PAC contains all the region in the 70kb construct and extends an additional 20kb in the 5' direction and 60kb in the 3' direction (Higgs *et al.* 1998). In general, levels of expression are higher in these transgenic lines, and they display a greater stability in expression throughout development (Higgs *et al.* 1998, Sharpe *et al.* in preparation). However, copy number dependency has not been achieved.

1.4.10 Summary

This part of the introduction has reviewed the current understanding of the regulation of the human α globin gene cluster. In the normal chromosomal environment, correct expression of the structural genes is dependent upon the promoter element of each gene and the distal regulatory element HS-40. These elements bind erythroid-specific and ubiquitous *trans*-acting factors. Each promoter element is associated with a CpG island, which are unmethylated in erythroid and non-erythroid cells. Analysis of the gene cluster and the surrounding chromosomal environment, has shown that the majority of the region is sensitive to endonucleases, in expressing and non-expressing cell-types, indicating that it adopts a constitutively 'open' structure. Erythroid-specific hypersensitive sites form at the promoters of each globin gene and at the HS-40 element, suggesting these regions are accessible for transcription factor binding in expressing cell-types. The hypersensitive sites at the gene promoters form in the absence of HS-40, suggesting HS-40 does not play a major role in influencing the local chromosomal structure throughout the α globin gene cluster or the wider region wherein the cluster is located. No scaffold attachment sites have been found in a 150kb region including and around the α cluster; nuclear position may be maintained by the

proximity of the α cluster to the 16p telomere, which could be attached to the nuclear matrix.

In an attempt to identify the minimum region of DNA that is required to produce the correct developmental-stage and tissue-specific expression of the α cluster, various DNA fragments have been tested in MEL cells and transgenic mice. In these systems, although HS-40 confers high level, tissue-specific expression of the α globin genes, expression is not developmentally stable and is not copy number dependent. As HS-40 does not appear to influence chromatin structure, expression of the constructs may be subject to chromatin-mediated repression effects, which is dependent upon the site of integration into the mouse genome. Prolonged expression throughout development seems to improve with increasing size of construct, suggesting that large fragments of DNA may be required in transgenes to re-create the permissive chromatin environment that normally contains the α locus *in vivo*.

A component of the work in this thesis is to determine whether the establishment of a chromatin environment which is permissive for α globin gene expression in transgenic mice, relies on the faithful reproduction of DNA replication timing of the α cluster seen at its natural chromosomal position. Although DNA methylation, and chromatin structure (as judged by DNase 1 hypersensitivity) has been examined in MEL stably transfected cells and transgenic lines containing different constructs, the pattern of replication has not been previously analysed. It is not known whether the vagaries of gene expression in transgenic mice are correlated with changes in the patterns of replication. Constructs may need to include origins of DNA replication to reproduce the pattern and timing of replication of the α locus at its normal chromosomal environment. The larger the construct, the better it may be at generating the normal pattern of DNA replication, which is reflected in increased levels of expression, which are more stable throughout development.

1.5 Aims of this thesis

This first part of the introduction gave a general review of the different components of gene regulation: essential *cis*-acting elements and *trans*-acting factors, higher order chromosome structure, acetylation of chromatin, methylation of DNA at CpG dinucleotides, nuclear sublocalisation and DNA replication. This thesis focuses on the importance of DNA replication in the regulation of gene expression, hence the second part of the introduction gave a more detailed review of replication in higher eukaryotes. The system being used as a model to examine the role of replication in gene expression is the extremely well-characterised human α globin gene cluster. The latter part of the introduction reviews the extensive work that has been carried out to elucidate how this gene cluster is regulated. This concluding part of the introduction summarises the work that was performed in this thesis, to understand the role of DNA replication in the regulation of α globin gene expression, and its relationship with the aforementioned mechanisms of gene regulation (methylation, chromosome structure, nuclear localisation).

Analysis of replication timing has recently become more accessible due to the development of a fluorescence *in situ* hybridisation (FISH) based assay, described in Selig *et al.* (1992). Using this technique, probes are hybridised to unsynchronised, interphase nuclei. A sequence that has not undergone DNA replication appears as a single signal and a replicated sequence as a double signal. Calculating the relative porportions of single to double signals determines at which stage of S phase the sequence of interest replicates. This technique has been further refined by selective labelling of actively replicating cells (those in S phase of the cell-cycle) within the asynchronous cell-culture prior to nuclei harvesting (Bickmore and Carothers 1995).

The first part of this thesis describes the development of this technique within the MRC Molecular Haematology Unit, and the testing that was performed prior to the commencement of novel research using this assay (chapter 3). Once the assay was established, the objectives of my research were firstly to obtain an accurate profile of

the pattern and timing of DNA replication throughout the human α globin gene cluster and its surrounding chromosomal environment at the tip of chromosome 16p, in expressing and non-expressing cell-types. The timing and pattern of replication of the individual paternal and maternal homologues were also compared (chapter 4).

Once the pattern of replication of the α locus and extended chromosomal region was known *in vivo,* I assessed whether this pattern of replication was maintained in transgenic constructs, which have been randomly integrated into the mouse genome. Replication timing analysis was performed on primary cells cultured from the spleens of transgenic mice, which contain various segments of DNA from the 16p region (chapter 5).

In chapter 6 I investigate the role of the distal regulatory element (HS-40) in determining the pattern and timing of replication of the α cluster and the extended chromosomal region, by analysing a mutant where the element has been artificially removed. Deletion of this element results in the severe downregulation of α globin gene expression, and work in this chapter determines whether this is associated with a perturbation in replication timing.

Chapter 7 describes the analysis of the replication characteristics of a novel deletion mutant, where α globin gene expression has been severely reduced. Gene silencing that occurs in this mutant appears to occur via a different mechanism from loss of HS-40, which is maintained in constructs introduced into transgenic mice. Replication timing of these constructs have also been examined.

In chapter 8, I examine the importance of the subtelomeric region and telomere in DNA replication by analysing the replication characteristics of two naturally occuring mutants, where the 16p chromosome has been truncated. These deletions also result in the loss of α globin gene expression as the deletion includes HS-40.

In summary, the human α globin gene cluster is an extremely well-characterised region, which provides an ideal opportunity to study DNA replication and its relationship to chromosome structure and gene expression. This is the first piece of detailed research that has been described, which investigates replication of a human telomere. This work is a valuable contribution to the understanding of the regulation of eukaryotic gene expression, with an important application in the development of strategies for gene therapy.

Chapter 2 Materials and Methods

2.1 Chemicals and reagents

All commonly used stock solutions and buffers were prepared as outlined in Current Protocols in Molecular Biology, Ausubel *et al.* eds (1994-1997) or Sambrook *et al.* eds (1989). All chemicals used were supplied by BDH, analytical grade or higher, unless otherwise stated. Addresses for all manufacturers and suppliers mentioned in this thesis can be found in Appendix 4 of the foremost publication above.

2.2 General laboratory techniques

2.2.1 Bacterial Culture

2.2.1.1 Growth media and conditions

Solid and liquid media were sterilised by autoclaving at 15psi, 121°C for 20 minutes. Heat labile solutions such as antibiotics were added to the culture media once it had cooled to 50°C or below. These solutions had been previously sterilised by filtration through a 0.45µm Acrodisc filter (Gelmann Sciences). *E.Coli* strains harbouring plasmids were grown at 37°C, with vigorous shaking (300rpm/rotary shaker), in LB media supplemented with the appropriate antibiotic. *E.Coli* strains harbouring cosmids were routinely grown in SB. 1.5% w/v bacto-agar (Difco) was added for solid medium. Where applicable, IPTG (0.5mM) (Gibco BRL) and X-Gal (40µg/ml) (Gibco BRL) were used for the selection of colonies containing recombinant vectors designed for blue/white selection.

Antibiotics were used at a concentration of:

Ampillicin (Amp)	40µg/ml (Gibco BRL)
Chloramphenicol (Cm)	50µg/ml (Boehringer Mannheim)
Kanamycin (Km)	50µg/ml (Boehringer Mannheim)

LB	1.0%	w/v bacto-tryptone (Difco)
	0.5%	w/v yeast extract (Difco)
	0.25%	w/v NaCl
SB	2.4%	w/v bacto-trytone (Difco)
	1.2%	w/v yeast extract (Difco)
	0.5%	v/v glycerol
	0.017M	KH_2PO_4
	0.072M	K_2HPO_4

2.2.1.2 Strain storage and revival

10% v/v glycerol was added to 800µl of mid-log culture. Mixture was vortexed and stored at -70°C. To revive, glycerol stock was thawed, mixed well, 50µl removed and streaked for single colonies onto a LB agar plate supplemented with the appropriate antibiotic.

2.2.2 Preparation of nucleic acids

2.2.2.1 Plasmid DNA isolation

A single colony was picked from a master plate and innoculated into 5ml of media (with the appropriate antibiotic added) which was grown o/n, as described in 2.2.1.1. Bacterial cells were harvested from 1.5ml of this culture at 14 000rpm, for 2 minutes at RT (Centrifuge 5417C, Eppendorf) and used to prepare plasmid DNA, which was isolated using a Hybaid Plasmid Mini Prep Kit, as per manufacturer's instructions.

2.2.2.2 Cosmid DNA isolation

Cosmid DNA was routinely isolated using the Qiagen Plasmid Maxi Prep Kit. This procedure is based on the optimised alkaline lysis method of Birnboim and Doly (1979). A single colony was picked from a master plate and used to innoculate 5ml of media (plus antibiotic), which was cultured during the day, as described in 2.2.1.1. 0.5ml of this starter culture was used to innoculate 150-200ml of SB, which was grown o/n. The following day, cells were harvested by spinning at 3000rpm for 15 minutes at 4°C (Beckman J-6B Centrifuge). Cosmid DNA was prepared from each cell

pellet using a Qiagen-tip 500 column, following the manufacturer's instructions. DNA was resuspended in 100-200µl of water or TE pH8.0 (10mM Tris-HCl, 1mM EDTA). DNA concentration was determined by measuring absorbance at 260nm using a spectrophotometer (SP6-500 model, Pye Unicam).

2.2.3. Manipulation of DNA

2.2.3.1 Restriction endonuclease digestion

Digestion of DNA with restriction endonucleases (Boehringer Mannheim) was carried out in reaction buffers recommended by the manufacturer. Digestion of cloned DNA was routinely performed in a 37°C waterbath, for 30 minutes, using 5-10U of enzyme per µg of DNA.

2.2.3.2 Agarose gel electrophoresis

DNA fragments of 100bp to 20kb were separated for analytical and purification purposes by horizontal agarose gel electrophoresis. 0.6-2% w/v agarose gels were made by dissolving agarose in 1X TAE buffer (40mM Tris-acetate, 2mM EDTA). Electrophoresis was performed in 1X TAE running buffer at 50V-120V, depending upon the percentage of the gel. To visualise DNA fragments, gels were made containing 0.5µg/ml EtBr and viewed under a UV transilluminator. Photographs of gels were taken using Polaroid cameras and instant film.

2.2.3.3 Isolation of DNA from agarose gels

DNA fragments were routinely recovered from agarose gels using either the Qiagen Gel Purification Kit, following the manufacturer's protocol, or for larger fragments using electroelution into dialysis tubing, as described in Sambrook *et al.* (1989). DNA was precipitated using 0.3M NaOAc and 2 volumes of ethanol, at -20°C, for 2-24 hours. Pellets were washed once in 70% ethanol and resuspended in 100-200µl water or TE pH8.0.

2.2.4 Tissue Cell Culture

2.2.4.1 EBV-transformed lymphoblastoid cell-lines

EBV-transformed lymphoblastoid cell-lines were grown in RPMI-1640 medium (Sigma) supplemented with 50U/ml penicillin (Gibco BRL), 50μg/ml streptamycin (Gibco BRL), 2mM L-Glutamine (Gibco BRL) and 15% v/v FCS (Sigma). Culture flasks were gassed with CO_2, sealed and incubated upright at 37°C. Cells were fed every 2/3 days to maintain asynchrony.

2.2.4.2 Somatic cell hybrid lines

Interspecific somatic cell hybrid lines containing a single copy human chromosome 16 in a MEL background were made by the method of Deisseroth and Hendrick (1978) as modified by Zeitlin and Weatherall (1983). Essentially, EBV transformed lymphocytes are fused with APRT negative MEL cells using 50% PEG. Cells are cultured in RPMI-1640 medium (Sigma) supplemented with 15% v/v FCS (Sigma), 50U/ml penicillin (Gibco BRL), 50μg/ml streptamycin (Gibco BRL) and 2mM L-glutamine (Gibco BRL). Selective pressure for hybrid cells is achieved by the addition of 1mM adenine (Sigma), 1mM methotrexate (Lederle Laboratories) and 30μM thymidine (Sigma). 0.5μM ouabain (Sigma) is also included in the first 14 days of culture to prevent the background growth of EBV-transformed lymphocytes in preference to mouse/human hybrid cells. Cultures are incubated, flask horizontal, at 37°C, gassed with 5% CO_2.

2.2.4.3 Cell storage and revival

For long term storage, 0.5 x 10^6 mid-log phase cells were harvested, resuspended in 0.6ml complete media as described in 2.2.4.1, an additional 0.3ml of FCS (40% total) and 0.1ml DMSO (BDH). Cell suspensions were frozen at -70°C overnight, then transferred to liquid nitrogen. To revive, frozen vials were thawed in a 37°C waterbath, and the culture transferred to a 25ml culture flask (Falcon), with 10ml of fresh media, prewarmed to 37°C. Cultures were gassed with 5% CO_2 and incubated at 37°C.

2.3 FISH analysis to measure replication timing

2.3.1 Nick translation labelling of probes

This nick-translation labelling reaction has been optimised to produce probes of <600bps by empirically varying the amount of DNase 1. For a single reaction, 2μg of cosmid or plasmid DNA were added to 4μl of 10X nick translation buffer (0.5M Tris pH7.5, 0.1M $MgSO_4$, 1mM DTT, 0.5mg/ml BSA fraction V), 4μl of each of the following: 2mM dATP, 2mM dGTP, 2mM dCTP, (all Pharmacia), 1mM DIG-11-UTP (Boehringer Mannheim), 2μl 0.5mM dTTP (Pharmacia), 1μl DNA polymerase (10U/μl, Gibco BRL) and 2μl DNase 1 (7.5U/μl, Pharmacia) diluted 1:30 with sterile distilled H_2O. This solution was made up to 40μl with sterile distilled H_2O, mixed and incubated at 16°C for 1.5 hours. Following incubation, 60μl TE pH7.5 was added and unincorporated nucleotides removed by centrifugation of the solution through a G50 Sephadex (Sigma) column, pre-equilibrated with TE pH7.5. Periodically, probe size would be checked by running 1/20th of the labelled probe on a 2% agarose gel alongside 0.5μg ØX174/*Hae III* marker.

2.3.2 Colourimetric assay to measure label incorporation

For each probe that was labelled, the efficiency of label incorporation was estimated in order to determine how much probe should be used to hybridise per slide. 1:100 and 1:1000 dilutions of labelled probe were made, of which 1μl and 2μl were spotted onto a Hybond N+ nylon filter (Amersham), which had been previously rinsed in distilled water, soaked in 20X SSC for 10 minutes, and dried. Standards of 10pg, 20pg, 100pg and 200pg of DIG-labelled pBR328 DNA (Boehringer Mannheim) were spotted alongside the tested probes. Following UV crosslinking (120mJ, UV Stratalinker 2400, Stratagene), the filter was equilibrated for 5 minutes at RT in buffer 1 (0.1M Tris pH7.5, 0.1M NaCl), blocked at 37°C for 30 minutes in buffer 1/3% BSA fraction V (Sigma), then incubated in buffer 1/1:5000 anti-BdU conjugated to alkaline phosphatase (Boehringer Mannheim) at RT for 30 minutes. The filter was then washed twice in buffer 1, 15 minutes each time, equilibrated for 5 minutes in buffer 2 (0.1M Tris

pH9.5) and placed in a sealed hybridisation bag containing 5ml of buffer 2 mixed with 2 drops each of BCIP, NBT and $MgCl_2$ from the BCIP/NBT alkaline phosphatase substrate kit IV (Vector). Colour was allowed to develop for half an hour in the dark, before rinsing the filter in distilled water and air drying.

2.3.3 Preparation of interphase nuclei

Approximately 2 x 10^7 cells were grown to mid-log phase. To identify S phase cells, BdU (Sigma) was added to the culture to give a final concentration of 0.1mM, 90 minutes prior to harvesting. Cells were harvested by centrifugation at 1200rpm, for 7 minutes, at RT in a benchtop centrifuge (Baird and Tatlock Mark IV). Cells were washed once with ice-cold PBS, collected by centrifugation as before, and swollen in hypotonic solution (0.07M KCl), containing 0.1μg/ml colcemid (Gibco BRL), which helps flatten the nuclei. Cultures were incubated in the dark at RT for either 10 minutes for MEL-hybrid lines/transgenic primary cells or 20 minutes for EBV-transformed lymphoblasts. Cells were spun again as before, and gently resuspended in freshly-made fixative (3:1 methanol: glacial acetic acid), and placed at -20°C for 30 minutes. Cells were spun down once more and resuspended in fresh fixative. This was repeated twice and the resultant fixed cell suspension used either straightaway to make slides, or stored at -20°C in the dark until needed.

2.3.4 Assay for percentage of nuclei in S phase

Prior to FISH, each fixed preparation was assayed to evaluate the percentage of nuclei in S phase (BdU positive nuclei). Only preparations with >30% S phase were used. Slides were denatured in 0.07M NaOH/EtOH (2 parts: 5 parts) for 90 seconds, then blocked in PBS/1% dried skimmed milk for 10 minutes. Anti-BdU-FITC (Boehringer Mannheim) was diluted 1:10 in 4X SSC/5% dried skimmed milk, and 30μl applied to each slide under a coverslip. Slides were incubated in a moist chamber at 37°C for 1 hour, then rinsed in 3 changes of 4X SSC/0.1% Tween 20, 2 minutes each, at 37°C. Slides were mounted in Vectashield antifade (Vector) containing 1μg/ml DAPI counterstain and coverslips sealed with rubber solution. Slides were viewed under

DAPI filter to assess total number of nuclei. The same field of view was then scanned through the FITC filter, to identify which of these were in S phase. Approximately 100 nuclei were scored per slide.

2.3.5 Preparation of slides

Slides were prepared when conditions in the laboratory were such that humidity was between 50-65%, and temperature between 20-26°C. Firstly, glass slides (Superfrost, Select Micro Slides, Chance Propper) were cleaned by soaking in 95% ethanol/1% conc HCl for at least one hour at RT. Slides were then removed, air dried, and polished with muslin. Fixed cell suspensions that had been in long term storage at -20°C, were centrifiuged at 1200rpm, for 7 minutes, at RT (Baird and Tatlock Mark IV), resuspended in fresh fixative and spun again to collect cells. Freshly fixed suspensions were spun once to collect cells. Each cell pellet was gently resuspended in a small volume of fixative (between 0.3-4ml). A clean slide was taken, breathed upon, and held at an approximate 45^0 angle to the bench. Using a pasteur pipette, a single drop of suspension was released onto the slide from 10cm above, and slides laid flat on the bench top to dry. Slides were examined using phase contrast microscopy (Olympus BH2, X10 objective, X12.5 eyepiece) to check for correct cell density, absence of cytoplasm, and large interphase nuclei, mid-grey in colour. Slides were aged by placing in a sealed, opaque slide box containing silica gel dessicant from between seven days to three months. Slides any older than three months were discarded and fresh ones made.

2.3.6 Fluorescence *in situ* hybridisation

For each slide, 20-50ng of labelled probe was co-precipitated with 0-5µg human Cot-1 DNA, 0-5µg mouse Cot-1 DNA (both Gibco BRL), 5µg salmon sperm DNA (Sigma) as carrier, with 2 volumes of EtOH. DNA was dried under vacuum (Savant DNA Speed Vac 120) and resuspended in 10µl of hybridisation buffer (50% formamide, 2X SSC, 1% Tween 20, 10% dextran sulphate) at 37°C for 10 minutes. DNA probe and competitor were denatured at 70°C for 5 minutes, and allowed to preanneal at 37°C for 15-30 minutes before applying to the slide. Slides were incubated with 100µl of

100µg/ml RNase A (Boehringer Mannheim) under a coverslip in a moist chamber at 37°C for 1 hour. Slides were then rinsed quickly in 2X SSC, dehydated through a 70%, 90%, 100% alcohol series, 2 minutes each, and air-dried. To denature the chromosomal DNA, slides were placed in 70% formamide/2XSSC at 70°C, for between 30 seconds and 2 minutes, then transferred quickly to ice-cold 70% EtOH for 2 minutes, then through 90% and 100% EtOH for 2 minutes each and air dried. Denatured probe was applied to each slide under coverslip, sealed with rubber solution, and slides incubated at 37°C overnight.

2.3.7 Detection

The following day, slides were washed in four changes of 50% formamide/2X SSC, followed by four more changes of 2X SSC, each change for 3 minutes at 45°C. Stringent washing was carried out using 3 changes of 0.1X SSC, 4 minutes each, at 60°C, before transferring to wash solution (4X SSC/0.1% Tween 20). Following a 10 minute incubation in blocking solution (4X SSC/5% dried skimmed milk), slides were treated with two antibody layers. For each layer, slides were incubated in a moist chamber at 37°C for 20 minutes. Antibodies were diluted in blocking solution and 30µl applied per slide, under a cover slip. Layer 1: 1:10 anti-BdU (Boehringer Mannheim)/1:50 sheep anti-DIG-FITC conjugate (Boehringer Mannheim). Layer 2: 1:100 rabbit anti-sheep-FITC conjugate (Vector)/ 1:100 anti-mouse-Texas Red (Vector). Between layers 1 and 2, and after layer 2, slides were washed in four changes of 4X SSC/0.1% Tween 20, 3 minutes each, at 37°C. Finally, slides were mounted in antifade Vectashield (Vector)/1µg/ml DAPI counterstain (Sigma) and coverslips sealed with rubber solution. Slides were stored at 4°C, in the dark until ready to view.

2.3.8 Fluorescence Microscopy and Slide Scoring

Hybridised slides were examined using a fluorescence microscope (Olympus BX60) equipped with a Pinkel filter wheel containing DAPI, Texas Red, FITC, Dual (FITC and Texas Red) and Triple filters. Images were captured using a cooled CCD camera (Photometrics), and analysed using MacProbe version 3.3 software (Perceptive

Scientific Instruments). Scoring was carried out using the X100 oil eyepiece. Each slide was scored 'blind': the scorer being unaware of which cosmid the slide had been hybridised with. Slides were scanned under the dual filter to identify S phase nuclei (Texas Red) with cosmid signal (FITC). Cosmid signal was then further confirmed as either a singlet or doublet under the single FITC filter. A total of 200 chromosomes were scored per slide, with each slide taking approximately 1 hour to score. Nuclei with no signal or too many signals (due to either background or aneuploidy) were not scored. Nuclei with abnormal morphology were also not scored.

Chapter 3 Development of a FISH protocol to measure replication timing

3.1 Introduction

The use of FISH on interphase nuclei to measure replication timing was first introduced by Selig *et al.* (1992). Since then, it has been widely used in several laboratories (for example, see Kitsberg *et al.* 1993a, Boggs and Chinault 1994, Bickmore and Carothers 1995). In spite of this, the technique was still relatively new when this research was initialised in 1994. Thus, the feasibility of using it as an analytical method was first rigorously tested before proceeding with original experimentation.

To use FISH as a technique to measure replication timing, it is necessary to produce slides of the highest possible quality to allow successful scoring of nuclei. Slides need to be free of background signal within nuclei, achieved by pre-annealing probe with competitor DNA to suppress signal from repetitive sequences (Lichter *et al.* 1990) and the inclusion of stringent post hybridisation washing. They also need to be free from background signal all over the slide, which is due to non-specific binding of labelled probe or antibody to cytological material. This can be reduced by the introduction of a blocking step prior to antibody detection, thorough washing between antibody layers, and also by producing labelled probe of optimal fragment size (Buckle and Rack 1993). Whilst all of these precautions are carried out for most FISH applications, it is especially critical for this application, which relies on quantitative analysis of hybridisation signal. To this aim, time was initially spent optimising a FISH protocol to produce slides which could be confidently scored (section 3.2).

Section 3.3 describes the development of a scoring criteria for slides produced using the protocol designed in section 3.2. Sections 3.4 and 3.5 describe a series of experiments to assess whether data produced this way was reproducible. Reproducibility was assessed in two ways: it was first seen whether a result obtained with a particular probe could be reproduced when the experiment was repeated. Secondly, it was tested

whether results produced using this particular FISH protocol would produce results comparable with previously published FISH data. Probes that were successfully used to reproduce previously published data, were subsequently implemented as controls in future work. The replication timing of any probe used for the first time for FISH, was measured relative to these controls. The final FISH assay used throughout this thesis, was decided upon as a result of the work described in this chapter. This is summarised in Materials and Methods, section 2.3.

3.2 Development of the FISH protocol

3.2.1 Technical refinement of standard FISH

FISH was routinely being performed for analysis of metaphase chromosomes in the neighbouring cytogenetics laboratory at the IMM (Buckle and Rack 1993). However, it was found the protocol employed there had to be modified to improve signal strength and reduce backgound noise. A week was spent in the laboratory of Dr Wendy Bickmore, MRC Human Genetics Unit, Western General Hospital, Edinburgh, Scotland, where analysis of replication timing by FISH to interphase nuclei was already established (Bickmore and Carothers 1995). The resultant protocol outlined in this thesis is based upon the one in use there.

Technical refinement was achieved in several ways: firstly, by optimising the nick translation reaction used to produce labelled probes. Probes of optimal size (less than 1kb) with maximal label incorporation were produced by titrating the amount of DNase 1 and the amount of starting DNA, respectively. Probes were labelled with DIG and detected using FITC conjugates (green fluorescent signal). Secondly, the stringency of washing following the overnight hybridisation of the slide with labelled probe was increased by introducing washing steps at 45°C in the presence of 50% formamide. Finally, each probe was tested with various amounts of competitor DNA as another component towards acquisition of optimal signal : noise ratio. Fig 3.1 illustrates how slide quality determines whether a slide can be scored or not (see section 3.3.1). Fig

3.1A shows a slide produced using the original method (unscoreable) and fig 3.1B using the refined method (scoreable).

3.2.2. Detection of S phase nuclei

An additional modification of the original protocol was included to distinguish nuclei of S phase cells, when DNA replication is taking place, from nuclei in stages G1 and G2 of the cell cycle. Selective labelling of actively replicating cells within the asynchronous cell culture was carried out by adding 10^{-4}M bromodeoxyuridine (BdU), ninety minutes prior to harvesting. These cells were then used to produce interphase nuclei for FISH analysis. This procedure was first described in Kitsberg *et al.* (1993a).

Hybridised slides were counterstained with DAPI, so that all nuclei and metaphase spreads appear blue using the single DAPI filter. To identify S phase nuclei, BdU was detected using Texas Red which can be seen superimposed upon the blue under the triple filter set (DAPI, Texas Red and FITC), or more clearly, without blue, under the single (Texas Red) or dual filter (Texas Red and FITC). Fig 3.2 follows the replication of an autosomal locus within S phase nuclei prepared from a diploid cell-line (EBV transfomed lymphoblastoid line).

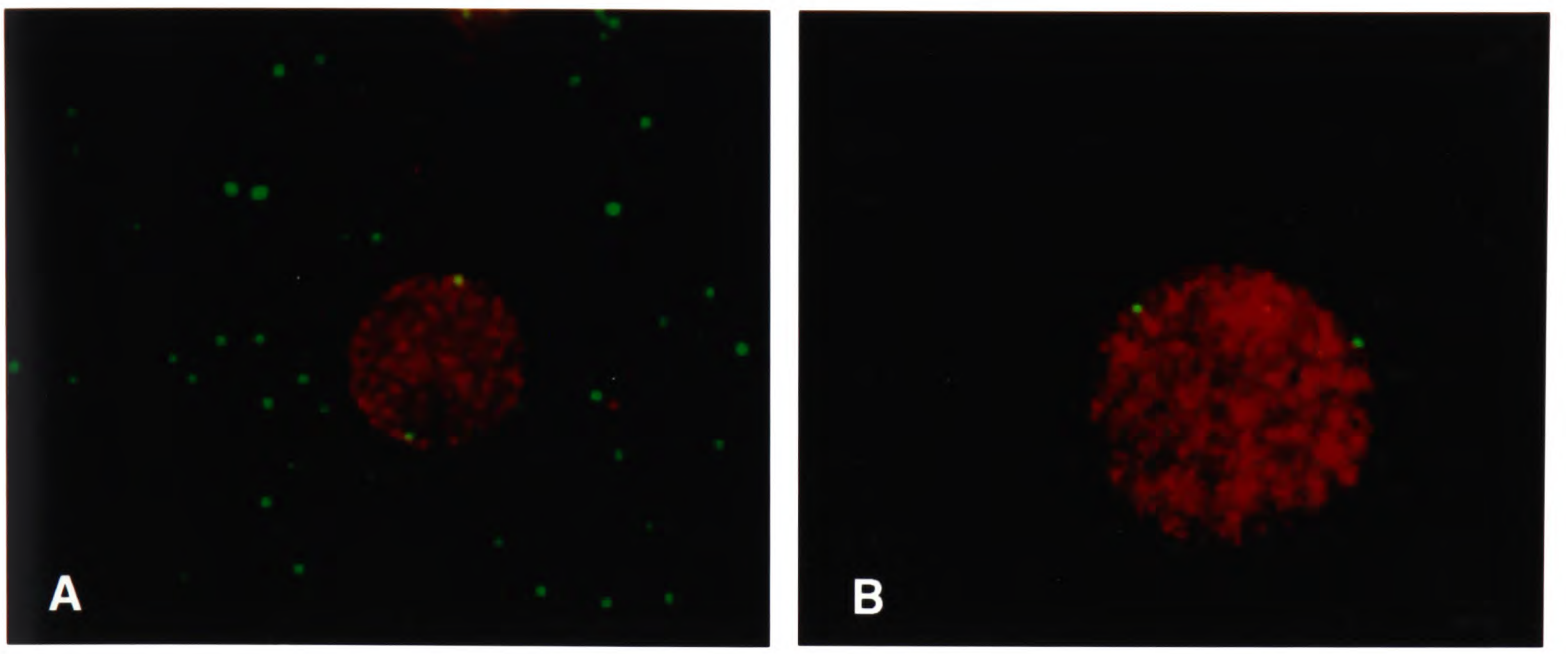


Fig 3.1 Development of a FISH protocol to measure replication timing.
A shows an example of a slide prepared using the original protocol. Non-specific background all over the slide is high (green speckle). Consequently, it is unknown whether the signal within the S phase nucleus (red) is specific or non-specific. For this reason, this slide is unscoreable.
B shows an example of a slide prepared using the refined FISH protocol. Non-specific binding has been eliminated through the introduction of stringent washing and blocking steps. This slide is scoreable.

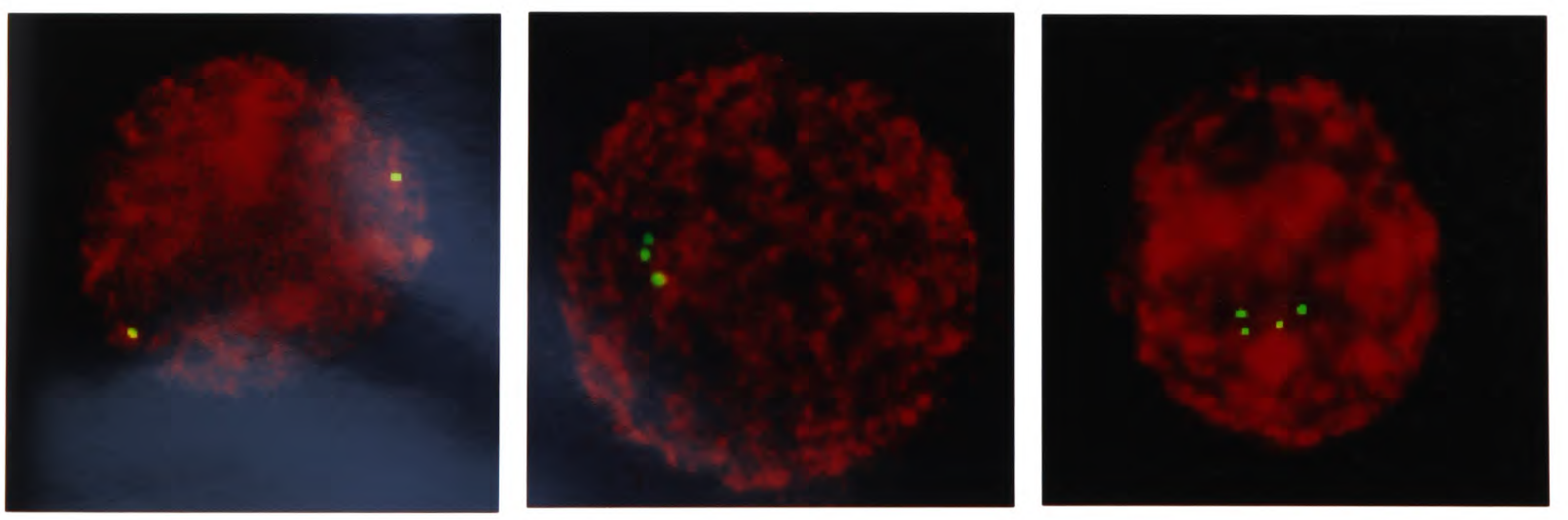


Fig 3.2 Replication patterns proceeding through S phase.
S phase nuclei (red) of a normal, diploid EBV-transformed lymphoblastoid cell-line (TA), hybridised with a chromosome 16p specific probe (green). Left hand nucleus: early stage of S phase, where both alleles are unreplicated (two single hybridisation signals). As S phase proceeds, replication begins. Centre nucleus : replication of one allele has occurred (top doublet signal), whilst the other has not (lower single hybridisation signal). Right hand nucleus: as S phase continues, both alleles become replicated (a pair of doublet signals).

3.3 Development of criteria for scoring

3.3.1 Selection of slides to score

Slides of overall poor quality were not scored. Poor quality was because of (i) weak hybridisation signal, (ii) high non-specific background within the nuclei which made the true signal difficult to identify, or (iii) high non-specific background over the whole of the slide. Slides were also discarded if identifying nuclei in S phase (BdU positive nuclei) was difficult due to a weak fluorescent signal. The primary antibody used to detect BdU recognises denatured DNA, therefore it was reasoned that a weak fluorescent signal was a result of under-denaturation of the chromosomal DNA upon the slide. Weak signal also occurred if slides were over-denatured, presumably as DNA is lost from the slide.

3.3.2 Slide scoring

To remove the possibility of scorer bias, each slide was scored 'blind' i.e. with the scorer being aware of which cell-line was being hybridised, but unaware of the probe being used. When scoring a slide, nuclei with a poor morphology were not scored. For diploid cells, when examining both copies of chromosome 16, only nuclei with a pair of signals were scored (i.e. either 1:1, 1:2 or 2:2). Nuclei displaying too few or too many signals were not scored.

3.3.3 How many nuclei to score?

Papers report the scoring of 100 nuclei to measure replication timing (Kitsberg 1993a); others score more: 100-200 (Selig *et al.* 1992, Torchia *et al.* 1994), >200 (Bickmore and Carothers 1995). Preliminary experiments were performed to address the question of how many nuclei to score. Three slides were taken and each hybridised with a different probe. A total of 400 *chromosomes* were scored per slide and the cumulative mean calculated every 50 chromosomes. Results are expressed as the percentage of doublets scored i.e. the percentage of loci that had replicated. Results are shown in fig 3.3.

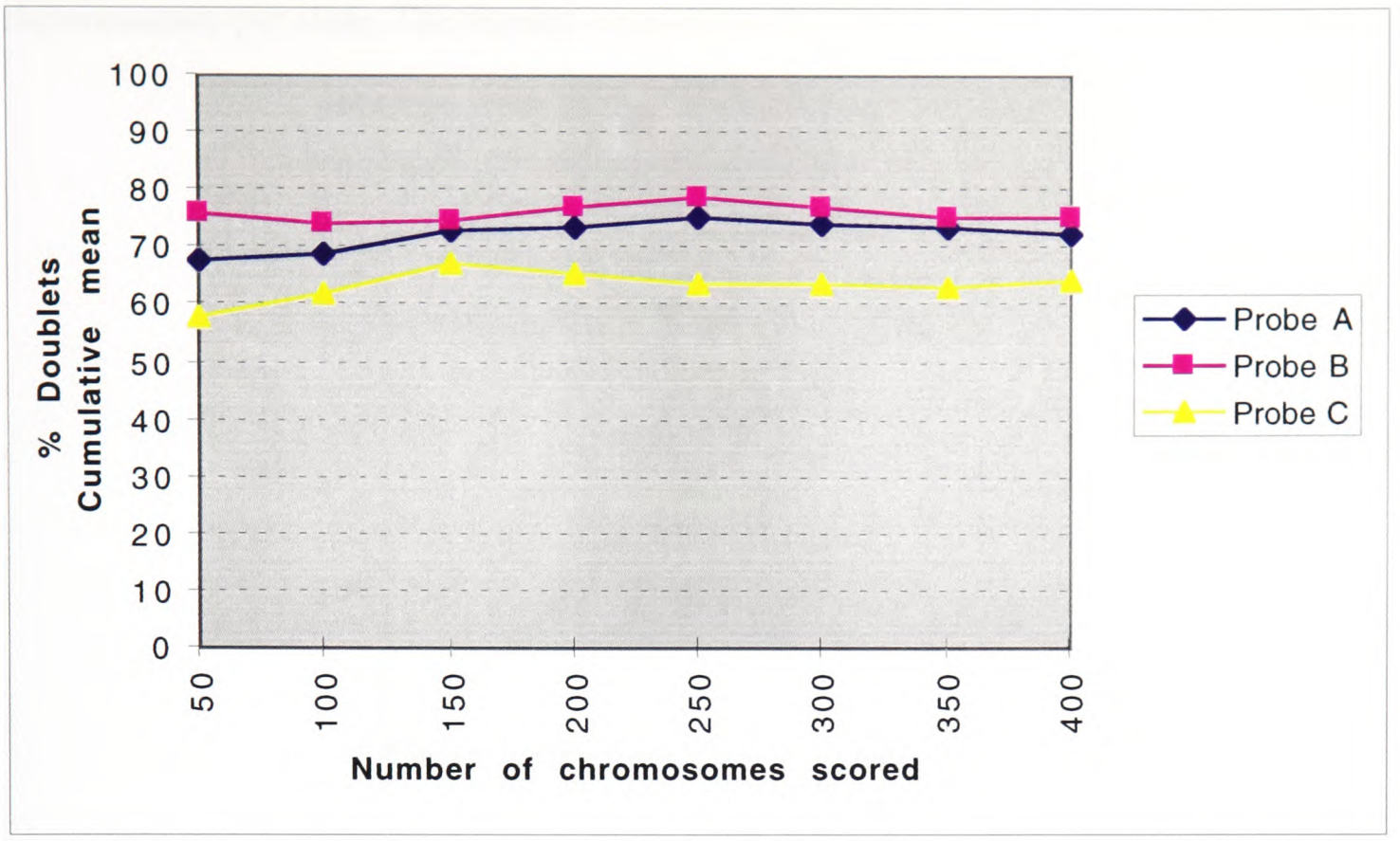


Fig 3.3 Graph to show how the % doublets scored changes with increasing sample size. Results are shown for 3 probes, denoted A, B, C.

Fig 3.3 illustrates that scoring 50 chromosomes is enough to ascertain whether a probe is early or late replicating by simple comparison with the early and late replicating controls (late controls typically give <25% doublets and early controls >55% doublets, shown in table 3.1). For all probes, very little variation is seen between cumulative mean values, although they do become slightly closer with increasing sample size. Therefore, it was reasoned to score 200 nuclei, as not only would this give an idea of replication timing relative to the controls, it may give a more accurate picture of each probe's replication timing relative to one another.

3.3.4 Independent scorers

An experiment was performed to ascertain whether values obtained were independent of the person scoring the slide. A batch of eight slides were made of the same cell-line. Each slide was hybridised with a different cosmid probe. Each slide was scored firstly by myself and then by my supervisor, Professor D.R. Higgs. Slides were scanned under dual filter (Texas Red and FITC) to identify S phase nuclei and each locus scored

either unreplicated (single signal) or replicated (double signal). Each person scored 100 chromosomes per slide. The results are given in fig 3.4.

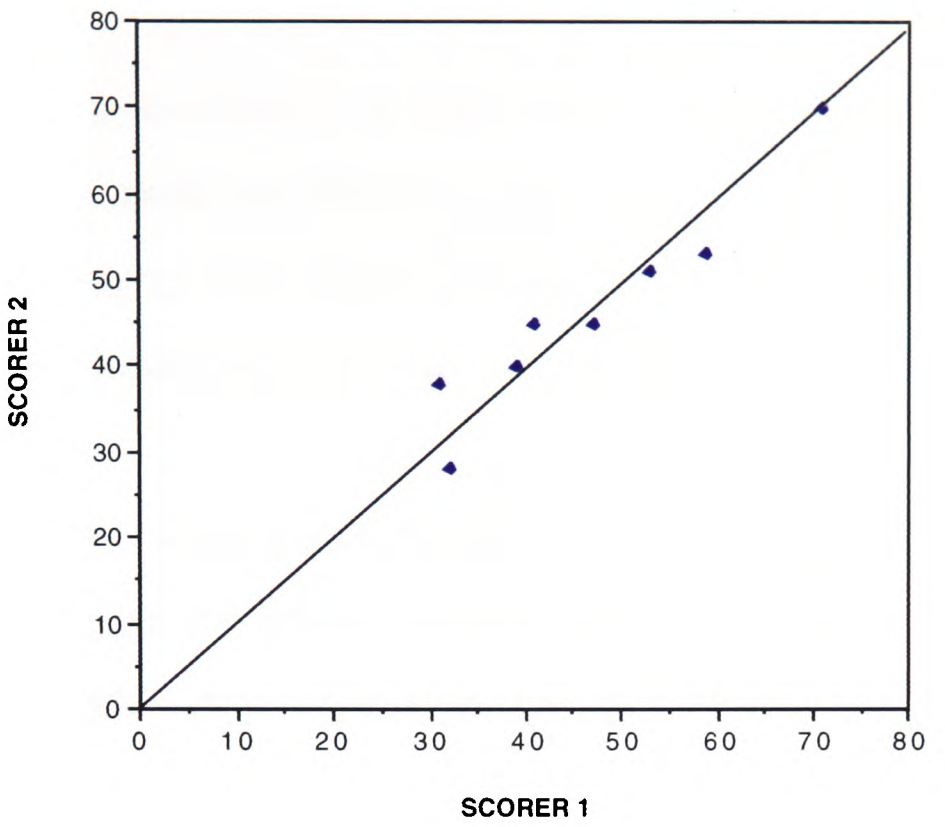


Fig 3.4 Comparison of results from independent scorers.This graph compares values of the % doublets scored by 2 different scorers, when analysing the same slide. Values are shown for 8 different slides, each slide hybridised with a different probe. Each value is obtained from scoring a total of 100 chromosomes.

The maximum difference recorded between scorers for any slide was 7%, but for most probes (6/8) differences were less than 5%, with neither scorer showing bias towards higher or lower scoring than the other. Slides showing the larger differences, were attributed to the probe, rather than inconsistencies in scoring *per se*. It was found that some probes produced easily resolvable signals, whilst others produced signals which were less clear to assign either a singlet or doublet. This is discussed in section 3.6.

3.4 Reproducibility of results

Once the basic protocol was established, a series of pilot experiments were carried out to ascertain how reproducible results were:

To confirm that two slides prepared from the same fixed nuclei preparation, **made at the same time** gave consistent results, two slides were hybridised with the same

probe and compared values obtained. The percentage doublets scored when scoring 200 chromosomes came within a few percent (72.5%, 74%).

To check variation between batches of slides made, the same nuclei preparation was taken and two **slides made on different days**, when the temperature and humidity conditions in the laboratory were slightly different. When these slides were hybridised with the same probe, values were also very consistent (60%, 62%).

To investigate whether variation could be detected from batch to batch of fixed nuclei preparation, one cell-line was taken, and BdU labelled S phase nuclei prepared on different occasions. When slides were made from the different nuclei preparations and hybridised with an identical probe, results were again very consistent (72.5%, 73%).

These experiments demonstrated that results were independent of conditions when the slides were made, and independent of the batch of fixed nuclei. Nonetheless, in order to reduce as many variables as possible, a single preparation of fixed nuclei was used to make a single batch of around 30 slides. These slides were then sequentially hybridised, each with an individual probe, in as short a time period as possible to reduce artefacts due to ageing of the slides.

3.5 Controls

3.5.1 Measuring hybridisation efficiency

As an internal control for each slide, hybridisation efficiency was assessed by scoring at least 10 metaphase spreads, as each chromosome should display two signals, one per sister chromatid. In diploid cell-lines, viewing non-S phase nuclei was also used as an indication of hybridisation efficiency as hybridisation patterns should show a lack of asynchrony: G1 nuclei a 1:1 pattern and G2 nuclei a 2:2 pattern.

3.5.2 Selection of early and late replicating controls

Timing of replication was measured relative to cosmids or plasmids which are specific to regions which are known to be early or late replicating. Hence, these recombinant clones acted as controls in each cell-line analysed. As far as possible, controls were selected whose replication timing had been determined using FISH, which corroborated earlier findings by the elutriation technique, described in Furst *et al.* (1981). For probes where elutriation and FISH data are available, techniques have been compared in at least one common cell-line (human lymphocyte, Manca).

When analysing human EBV-transformed lymphoblastoid cell-lines, as late replicating controls, two cosmids specific for the cystic fibrosis (CF) locus were used: CW44 and cJ21 (Rommens *et al.* 1989). Using FISH, these cosmids have been previously shown to be late replicating in cell-lines that do not express the CF protein. Elutriation demonstrated the region to be late replicating, using a plasmid probe approximately 50kb away (Selig *et al.* 1992). These cosmids were supplied by Dr A. Harris, Paediatric Molecular Genetics, Department of Paediatrics, Institute of Molecular Medicine, Oxford, UK. An additional late replicating control used for these particular cell-lines was a β globin specific cosmid, cosME2.5, described in Morley *et al.* (1991) and provided by Dr B. Morley, Rheumatology Unit, Royal Postgraduate Medical School, Hammersmith Hospital, London, UK. Although this particular probe has not been used for replication studies, the region it spans has been shown to be late replicating when β globin is not expressed, by FISH analysis (Selig *et al.* 1992) and by elutriation (Epner *et al.* 1988, Dhar *et al.* 1988). A cosmid specific for the muscle glycogen phosphorylase gene (PYGM), described by Lichter *et al.* (1990), was used as an early replicating control. This cosmid has been shown to be early replicating in the human lymphocyte Manca cell-line by FISH and the elutriation technique. FISH also confirmed it to be early replicating in several other cell-lines (Selig *et al.* 1992). This clone was supplied by Dr H. Cedar, Department of Cellular Biochemistry, Hebrew University, Jerusalem, Israel.

When analysing the timing of replication of probes within human chromosome 16 in an MEL background and smaller constructs of the terminal region of 16p in murine transgenic cells, cosmid cSamD4 was used as a late control. This cosmid is specific for the mouse salivary amylase gene (Amy-1) found on mouse chromosome 3 (Wiebauer *et al.* 1985) and was supplied by Dr M. Meisler, Department of Human Genetics, The University of Michigan Medical School, Michigan, USA. This region has been demonstrated to be late replicating in MEL cells by Forrester *et al.* (1990), Dhar *et al.* (1989) and Hatton *et al.* (1988), all by elutriation. For an early control, the cosmid P48 was used, which is specific for the mouse α globin gene cluster (Zhao *et al.* 1996). This region has been previously shown to be early replicating in MEL cells, again by the elutriation method (Furst *et al.* 1981, Epner *et al.* 1981). Metaphase spreads indicating the chromosomal location of the mouse and human controls are shown in fig 3.5 and fig 3.6 respectively. Table 3.1 is a comparison of the percentage doublet scores obtained with the control probes on three different lymphoblastoid cell-lines used in this thesis, with previously published FISH data. For the values given for CW44, cJ21, and PYGM, identical probes were used. In the case of β globin, previous work by Selig *et al.* (1992) used the probe K40 (Collins and Weissman, 1984) whereas work reported here was carried out using cosME2.5, described earlier. Figures quoted from Selig *et al.* (1992) are the percentage doublets scored in all nuclei, which were corrected to represent those in S phase only. Figures produced in this thesis are from direct scoring of S phase nuclei.

Cell-line	CW44	cJ21	PYGM	β globin	Reference
K562	21	26	68	61	Selig *et al.* 1992
Manca	18	13	77	4	Selig *et al.* 1992
Caco-2 *	62	66	64	15	Selig *et al.* 1992
Lymphoblastoid 1	23.5	22	68.5	29	This thesis
Lymphoblastoid 2	19.5	21.5	52.5	23	This thesis
Lymphoblastoid 3	20	19	56.5	18.5	This thesis

Table 3.1 Table to compare the values of % doublets of control probes. Shaded boxes represent early replication. In lymphoblastoid cell-lines, CW44, cJ21, and β globin are all are late replicating, whereas PYGM is early replicating. PYGM is ubiquitously expressed and is early replicating in every cell-line. The remaining probes also demonstrate expression-linked replication timing. Early replication is seen in cell-lines where the gene loci the probes are specific to, are being expressed. Late replication is seen in non-expressing cell-lines. * Epithelial-type cell-line, originally derived from a colon adenocarcinoma, that expresses the CF gene (Selig et al. 1992).

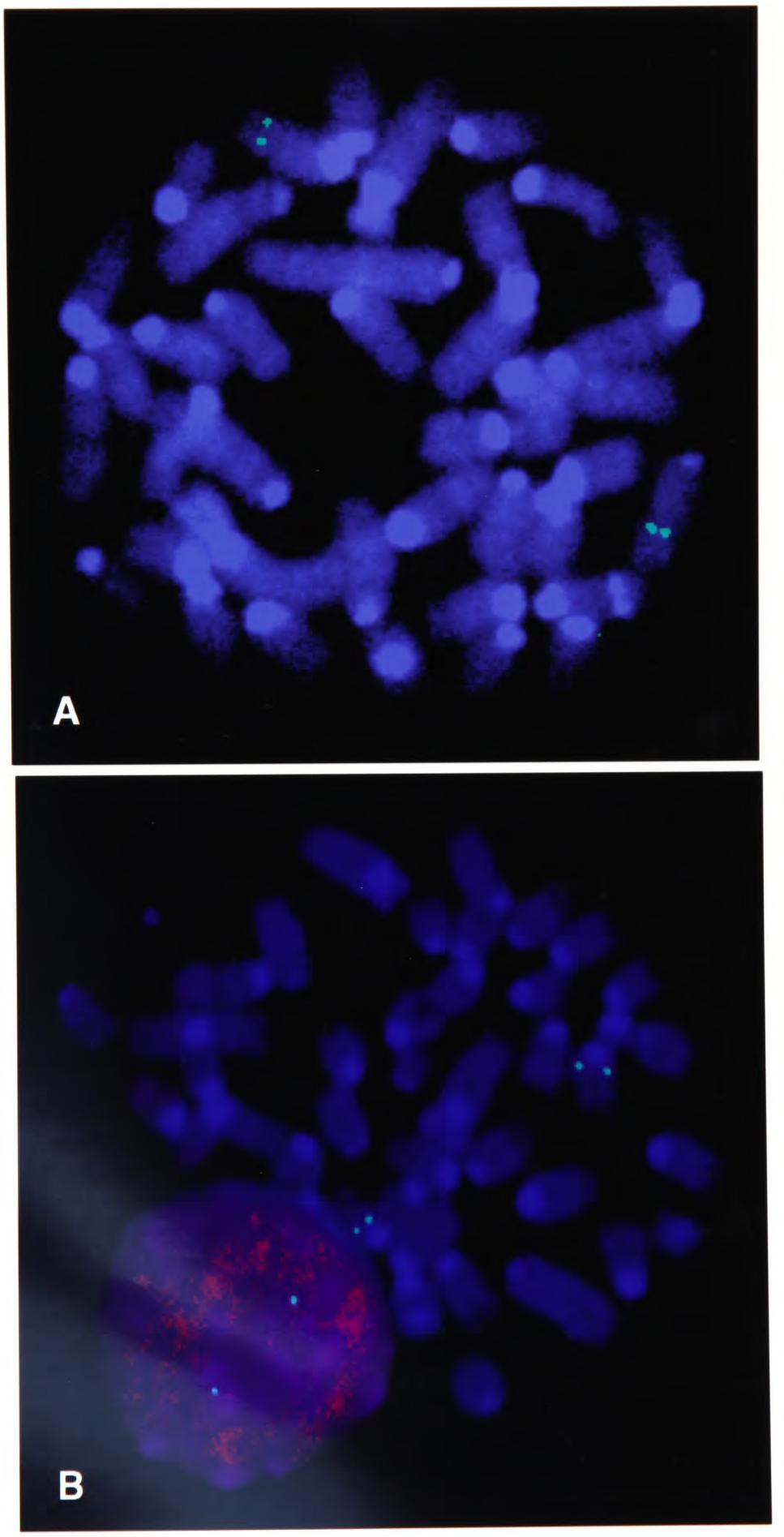


Fig 3.5 Chromosomal location of the mouse control probes. Metaphase spreads have been prepared from MEL-16 interspecific hybrid cell-lines. Probes are detected with FITC (green) and DNA is counterstained with DAPI (blue).

A. Early replicating control: cosmid P48, specific for the mouse α globin cluster (chromosome 11).

B. Late replicating control: cosmid cSamD4, specific for the mouse salivary amylase gene (chromosome 3).

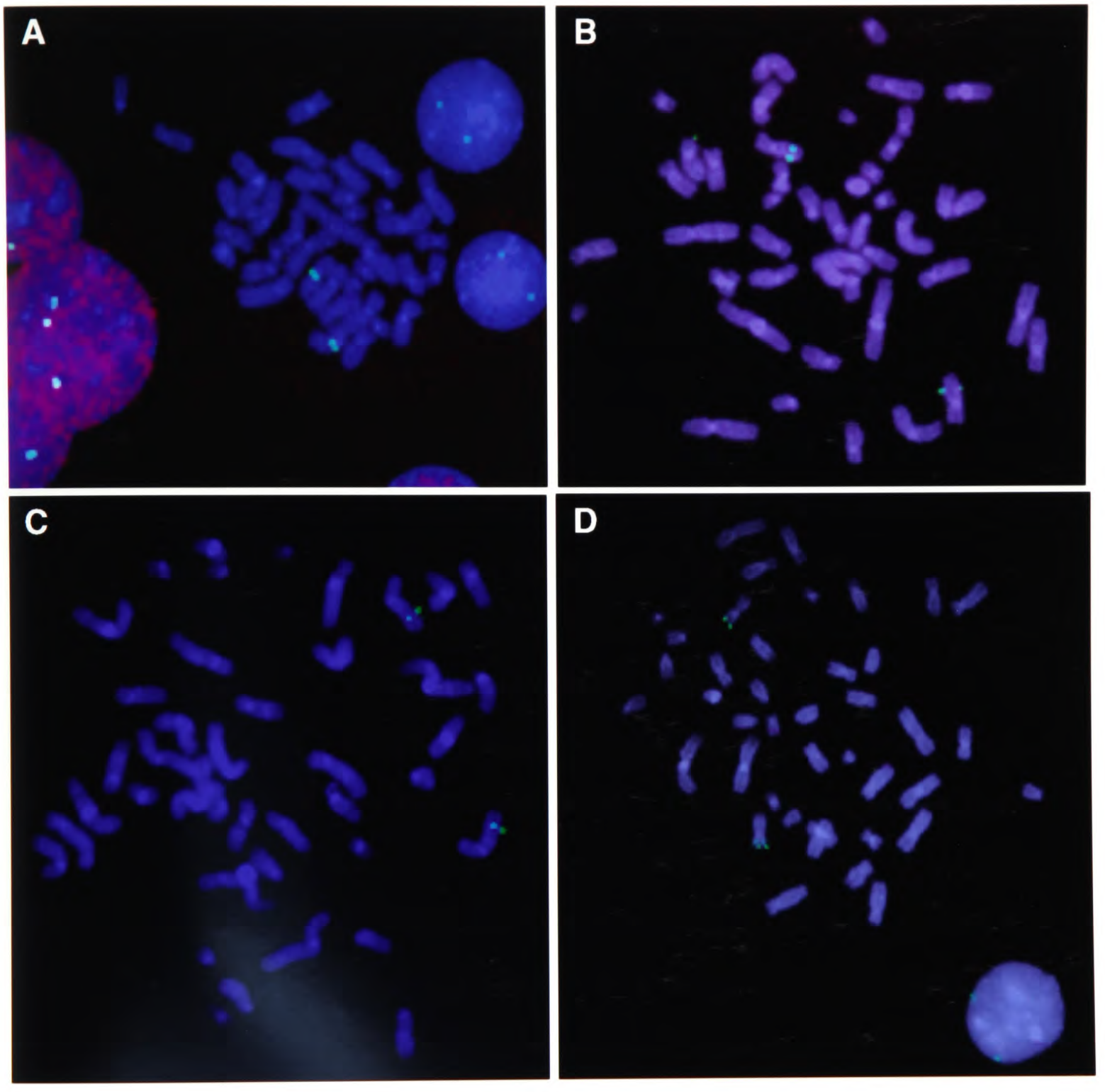


Fig 3.6 Chromosomal location of the human control probes. Metaphase spreads have been prepared from normal EBV-transformed lymphoblastoid cell-lines. Probes are labelled with DIG-11-UTP and detected with FITC (green). Slides are counterstained with DAPI (blue).

A. Early replicating control, cosmid PYGM, specific for the muscle glycogen phosphorylase gene (chr11q12-13).

B. Late replicating control, cosmid CW44, specific for the cystic fibrosis locus (chr7q31-32).

C. Late replicating control, cosmid cJ21, also specific for the CF locus.

D. Late replicating control, cosmid cosME2.5, specific for the β globin cluster (chr11p15.5).

3.6 Discussion

Using FISH as a technique to measure DNA replication timing has several advantages over other methods. Prior to its development, the methods most commonly employed were elutriation-based techniques, described by Furst *et al.* (1981) or similar, modified methods using flow cytometry with PCR as an method of quantitation (Hansen *et al.* 1993, Sinnett *et al.* 1993). All these methods are quite laborious and most require the synchronisation of cells, which may interfere with the cell cycle itself. The FISH technique uses asynchronous populations of cells, using a method is relatively quick and simple to perform. It also looks at replication timing in the context of individual nuclei, rather than the total cell population. Not only does this allow the analysis of individual autosomes, it also means results can be obtained from a relatively small number of cells, about ten fold less than elutriation. Moreover, there is no restriction on the type of cells to be examined, unlike elutriation, which works most efficiently with fast growing suspension cells.

The FISH technique was validated as a method to examine replication timing by Selig *et al.* (1992), by producing results in accordance with previously published elutriation data for β globin and other probes. This paper also extended the comparative data by analysing the CF locus, demonstrating that this technique was in broad agreement with the elutriation method. Since then, however, discrepancies have arisen, when using the various techniques available, to compare the timing of homologous regions on individual alleles of imprinted genes or genes subjected to X-inactivation.

Allelic differences in expression of X-linked genes are associated with differences in replication timing, in that the expressed allele generally replicates earlier than the silent allele, shown by conventional methods (Taylor 1960; Schmidt and Midgeon 1990; Hansen *et al.* 1993; Hansen *et al.* 1995) and by FISH (Boggs and Chinault 1994, Torchia *et al.* 1994). Many loci examined using the two techniques obey the general rule and produce concordant results of replication asynchrony. However, some loci

examined displayed asynchrony with one technique and synchrony with the other, or even reverse patterns of asynchrony. One example of this, where different techniques have produced contradictory findings is the *Xist* gene. This gene is expressed exclusively from the inactive X (X_i) and is essential for the initiation of X chromosome inactivation (Herzing *et al.* 1997). FISH studies have reported this gene to be earlier replicating on the X_i where it is transcriptionally active, than on the active X (X_a) where it is silenced (Torchia *et al.* 1994, Torchia and Midgeon 1995, Boggs and Chinault 1994). Conversely, Hansen *et al.* (1995) report a 'reverse' replication timing whereby the expressed allele of XIST replicates late in S phase and the silenced allele replicates early. This latter finding has been supported recently, using a novel technique, by Xiong *et al.* (1998).

Replication asynchrony has also been reported for imprinted regions found on autosomes, although the degree of asynchrony appears to be less pronounced than for X-linked genes (Kitsberg *et al.* 1993a, Knoll *et al.* 1994, Bickmore and Carothers 1995, Kawame *et al.* 1995, Greally *et al.* 1998). Once again, however, the extent of asynchrony remains controversial for some loci, such as the human imprinted region IGF2/H19, as FISH and conventional techniques report different findings (Kitsberg *et al.* 1993a, Kawame *et al.* 1995). Furthermore, controversy not only occurs between different techniques: disagreement over the magnitude of asynchrony of this region also occurs between different research laboratories using a similar FISH technique (Kitsberg *et al.* 1993a, Bickmore and Carothers 1995).

Comprehensive accounts detailing the reasons proposed for inconsistency are given in Bickmore and Carothers (1995) and Kawame *et al.* (1995). Some of the points raised regarding FISH will be discussed here.

One criticism of the FISH technique is the difficulty assigned to the scoring of spots. This may happen because of viewing the nucleus as a two dimensional image when it has a three dimensional shape. Hence, dots that may be separated in three dimensions

may appear coincident in two. This would give a negative bias to the results. However, it is assumed that this is a random process, so this artefact should effect all probes equally. Since, in this thesis, replication timing of probes are being measured relative to one other, rather than measuring their absolute replication timing, no artefactual bias should ensue.

Other reasons listed for artefactual FISH data include poor hybridisation efficiency and the use of large probes with central repetitive elements (these sequences become competed out during prehybridisation, leading to a false doublet result). A doublet signal could also appear in the absence of replication due to separation of the denatured DNA strands, or from chromosome decondensation (Houseal and Klinger 1994). However, provision has been made within the protocol developed in this chapter, for detection of any of these artefacts, by assessing the signal on metaphase spreads.

Structural chromatin differences have also been proposed to explain why asynchrony may appear using the FISH technique and be absent in others. Conformation of chromatin throughout and surrounding the sequence being examined, could influence the rate at which the two sister chromatids separate after replication. This directly influences how quickly replicated sequences appear as a doublet. We have observed this type of phenomenon, where replicated signals appear to remain closely opposed before resolving into two easily distinguished signals. This was mentioned in section 3.3.4 and will be discussed again later in the thesis. Whether the structure of chromatin is responsible for this phenomenon or not, it highlights the need to examine a region in detail, ideally with several different probes, and to interpret results with caution.

Using FISH to detect asynchrony within imprinted regions, can require the use of detailed statistical analysis, as the degree of asynchrony is quite subtle (Bickmore and Carothers 1995). In this thesis, for the most part, the technique is not being used to measure relative differences between homologous loci, but to measure relative timing of loci along a chromosome. However, chapter 4 does investigate the individual allelic

timing of replication throughout terminal region of human 16p. Nonetheless, it is not intended to analyse data using high resolution statistical methods. Replication timing data for different regions of the genome using FISH is still quite limited, and no extensive studies have been carried out to compare data obtained using the FISH method with data from elutriation studies. For these reasons, this thesis does not try to establish the actual time a DNA segment replicates within S phase. Replication timing is measured relative to regions which have have been shown to be early or late replicating by FISH and elutriation. Therefore, throughout this thesis, the terms 'early' and 'late' have been used to describe the replication timing of sequences with reference to these previously studied examples.

In summary, FISH is a very effective tool to measure DNA replication timing, with many advantages over other methods. Yet, it is acknowledged that previous work has highlighted the potential pitfalls that can arise when using FISH for this application. Artefacts intrinsic to the nature of the technique need to be eliminated as far as possible, by careful design of a protocol and scoring criteria. Regions analysed must be done so thoroughly, preferably using several probes. Caution must be taken when interpreting the data, especially when measuring small differences in timing of replication. The work carried out in this chapter is in response to these points. It therefore forms the foundation upon which the data in the following chapters has been collected, and interpreted.

Chapter 4 Establishing the normal pattern of replication throughout the telomeric region of 16p.

4.1 Introduction

4.1.1 Rationale for research

To date, there have been relatively few studies in higher eukaryotes to measure the temporal order of replication over an extensive region of chromatin. FISH and other methods have been used to examine replication timing of individual genes or gene clusters (for example, Goldman *et al.* 1984, Holmquist *et al.* 1987, Hatton *et al.* 1988, Selig *et al.* 1992, Hansen *et al.* 1993). Replication timing has also been measured across chromosomal band boundaries (Tenzen *et al.* 1997, Strehl *et al.* 1997) and preliminary work has been conducted to examine how replication timing changes traversing several light and dark Giemsa bands (Bickmore and Carothers 1995). However, little work has been carried out to analyse in detail a larger chromosomal region within a band (Brown *et al.* 1987, Dhar *et al.* 1988, Selig *et al.* 1992). The terminal region of the short arm of human chromosome 16, offers an exceptional model in which to study replication within a Giemsa light or T band, and relate the findings to chromosomal structure and gene expression. The region is extremely well characterised both structurally and functionally (reviewed in Flint *et al.* 1997a). It contains many housekeeping genes in addition to the tissue specific human α globin gene cluster. For the latter, a large amount of research has been carried out to elucidate how gene expression is regulated (reviewed in Higgs *et al.* 1998). Resources for this region which can be used for FISH analysis are also abundant. When the research began in 1994, overlapping cosmid clones were available for the terminal 325kb of the region. Physical mapping of the region has now extended much further towards the chromosome 16 centromere (R. Daniels, pers. comm.). In addition, this well characterised region provides an opportunity to examine replication at an individual human telomere. Whilst replication has been studied in some detail at yeast telomeres (McCarroll and

Fangman 1988, Ferguson *et al.* 1991, Friedman *et al.* 1996), very little is known about replication at human telomeres (Ten Hagen *et al.* 1990).

4.1.2 Replication timing of the globin gene clusters

Previous studies of the human and mouse β globin clusters, have demonstrated that the region adopts an expression linked pattern of replication, with replication taking place early in S phase when the locus is transcriptionally active, and late replicating in S when the cluster is transcriptionally inactive in non-erythroid cell-lines (Epner *et al.* 1981, Furst *et al.* 1981, Epner *et al.* 1988, Dhar *et al.* 1988, Dhar *et al.* 1989, Forrester *et al.* 1990). By contrast, experiments performed using mouse erythroleukaemia (MEL) cells and mouse B-cell plasmacytoma lines (S107 and MPC11) have shown that the mouse α globin cluster is early replicating in both expressing and non-expressing cell-types respectively (Epner *et al.* 1981, Furst *et al.* 1981, Holmquist *et al.* 1987). Hatton *et al.* (1988) also propose that the mouse α cluster is early replicating in non-expressing cells, and that it may become even earlier replicating in expressing cell-types. Studies using HeLa cells by Goldman *et al.* (1984), suggest that the human homologue is also early replicating in non-erythroid cells. No data are available for replication timing of the human α cluster in expressing cell-types.

4.1.3 Research objectives

The aims of the research described in this chapter were threefold. The first objective was to extend the previous findings that the human α globin gene cluster is early replicating in non-expressing cells. This would be acheived by characterising the replication kinetics throughout a larger 325kb region of the 16p telomere, where the gene cluster is located, using an EBV transformed lymphoblastoid cell-line. The second objective was to examine this extended region in an erythroid cell-line and compare the two datasets. The final objective was to compare maternally or paternally-derived alleles of chromosome 16.

4.1.4 Selection of an erythroid cell-line

Previous FISH-based replication studies of the human β globin cluster have been carried out using K562 cells as an erythroid cell-type (Selig *et al.* 1992). However, this cell-line is aneuploid, ranging from one to five copies of a given human chromosome per cell, most

frequently containing three. Not only does this introduce difficulties with the scoring of nuclei, it is unclear whether aneuploidy influences replication timing of individual chromosomes. Selig *et al.* (1992) noted that when K562 cells were used to analyse the replication timing of the human β globin gene (chromosome 11), asynchrony was found in >50% of S phase nuclei, compared to a highly synchronous pattern (>90% of nuclei) of the same region in the non-expressing diploid cell-lines HL60 and Manca. The reason for asynchrony is unclear and brings into question whether this cell-line accurately represents the *in vivo* situation. For this reason, a mouse-human interspecific hybrid was selected to measure replication timing of human 16p rather than the K562 cell-line. These hybrids contain a single copy of human chromosome 16 in an MEL background (MEL-16 hybrids). They are produced by fusing human non-erythroid cells with MEL cells. Resultant MEL-16 hybrid cells are selected for using the human APRT locus, which is also located on chromosome 16 (see chapter 2, section 2.2.4.2). MEL cells are mouse adult erythroid progenitors which upon induction, stimulate the expression of the human adult α globin genes from the human non-erythroid derived chromosome 16 (Deisseroth and Hendrick 1978, Zeitlin and Weatherall 1983, Charnay *et al.* 1984, Anagnou *et al.* 1985). Previous work has shown that these hybrids faithfully reproduce normal and mutant α globin gene expression and thereby constitute a good model system to study α globin gene regulation (Higgs *et al.* 1990, Sharpe *et al.* in preparation).

4.1.5 Individual allelic replication timing

As discussed in section 3.6, research suggests that in diploid cells, gene loci on homologous chromosomes display a synchronous pattern of replication timing (Selig *et al.* 1992), except in areas that are imprinted (Kitsberg *et al.* 1993a, Knoll *et al.* 1994, Bickmore and Carothers 1995). Using FISH to measure replication timing of 16p in a normal EBV line with a single 16p specific probe, the paternal and maternal alleles are indistinguishable. A percentage doublet value recorded for a given probe is the average value of replicated loci for both autosomes. However, the greater the number of individual nuclei displaying an asynchronous pattern of replication (double hybridisation signal: single hybridisation signal), the greater the apparent difference in temporal replication between the

two alleles (Kitsberg *et al.* 1993a). Yet, encountering nuclei with a 2:1 pattern, it is impossible to know which of the two alleles has replicated. Consequently, it is impossible to know whether asynchrony is due to one allele consistently replicating before the other (parent of origin dependent) or whether the 2:1 pattern is a mixture of either allele replicating first (mosaic). To circumvent this problem, the temporal pattern of replication of the individual alleles of human 16 has been examined using two cell-lines that are monosomic for the tip of chromosome 16p. One cell-line (JG) contains a single copy of 16p which is maternal in origin (Lamb *et al.* 1993) and the other (NB) has a single paternally inherited copy (not previously described).

4.2 Methods

Each cell-line was examined with twelve cosmid probes, which span a contiguous region of 325kb, including the human α globin cluster. Probes start at the unique sequence proximal to the subtelomeric region of chromosome 16p (type A allele). This is approximately 20kb from the 3' end of the tandem $(TTAGGG)_n$ repeats found at the telomere. From here, overlapping cosmid clones extend proximally towards the centromere, shown in fig 4.1. Each cosmid probe was hybridised individually to a slide. A total of 200 chromosomes were scored, either by myself, or jointly with D.R.H., as described in section 2.3.8. For the normal EBV cell-line (TA), which is diallelic for 16p, this is equivalent to 100 nuclei. For the mutant EBV lines and the MEL-16 hybrid lines which are monoallelic for 16p, 200 nuclei were scored. In addition to the 16p specific probes, each cell-line was hybridised with several control probes, to regions which were previously known to be early or late replicating (see section 3.5.2). Human EBV cell-lines were hybridised with late replicating probes CW44, cJ21 (both CF locus specific), cosME2.5 (β globin) and an early replicating probe, PYGM. For the MEL-16 hybrids, the control probes used were P48 (from the early replicating mouse α globin cluster) and cSamD4 (from the mouse salivary amylase gene (Amy-1), late replicating).

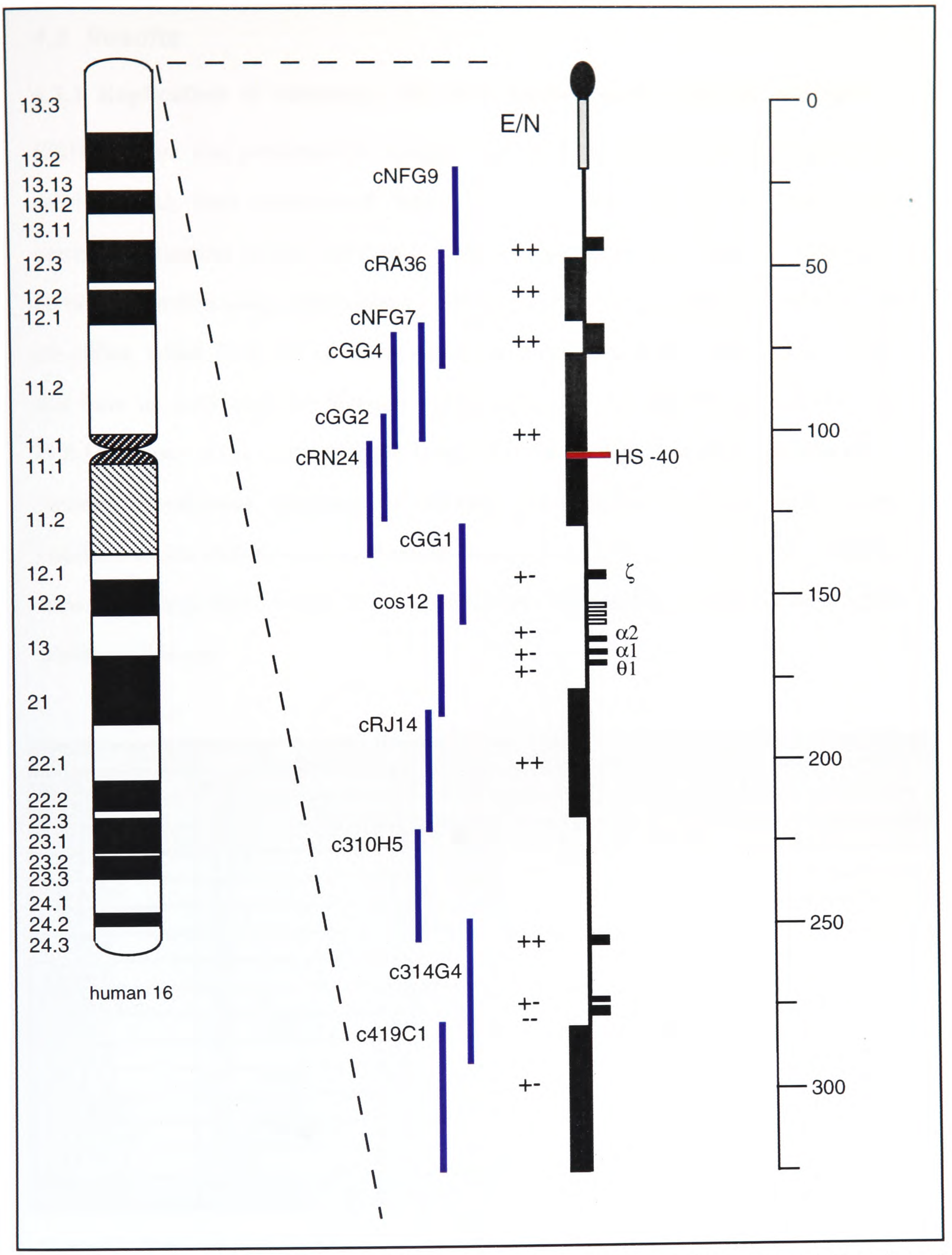


Fig 4.1 The left hand side of the figure shows a G band ideogram of human chromosome 16. Central panel shows a magnified schematic of the terminal 325kb of band 16p13.3 (type A allele), where the α globin cluster is located. Telomeric repeats (TTAGGG)n are depicted as a black oval and the subtelomeric region as a stippled box. Genes and pseudogenes are shown as filled and empty boxes respectively. Genes right of the line are transcribed towards the centromere. Genes left of the line are transcribed towards the telomere. The α globin regulatory element (HS-40) is shown in red, and the α globin genes are labelled. Overlapping cosmid clones used to measure replication timing are shown in blue. Expression of each gene in an erythroid (K562) and a non-erythroid (EBV-lymphoblastoid) cell-type is given (from Flint *et al.* 1997a). E= erythroid, N= non-erythroid. Scale is given in kb.

4.3 Results

4.3.1 Replication of telomeric 16p in a normal non-erythroid cell-line.

FISH analysis was performed in a single, normal EBV transformed lymphoblastoid cell-line (TA). Each cosmid was hybridised to at least two different slides. The percentage doublets scored for each cosmid, on each slide, is shown in table 4.1. Hybridisation efficiency (expressed as a percentage) has been calculated for over half of the slides, which is shown in square brackets alongside each % doublet score. Slides that have no hybridisation efficiency values were scored jointly between Z.E.S and D.R.H. These slides were analysed before the use of scoring metaphase spreads to measure hybridisation efficiency was formally implemented. However, both scorers considered each slide to be of good enough quality to score (see section 3.3.1). Fig 4.2 depicts a typical field of view for a chromosome 16p specific probe (c419C1) when analysing this line.

cosmid	*slide 1*	*slide 2*	*slide 3*	*slide 4*
CW44	16 [79]	31 [82]		
cJ21	23	21		
cosME2.5	30	28.5 [80]		
cNFG9	34.5	50 [93]		
cRA36	43	51.5 [85]		
cNFG7	39.5	46.5		
cGG4	56	59		
cGG2	49	55.5 [93]		
cRN24	46	59.5 [94]		
cGG1	54 [90]	66 [93]		
cos12	52	41		
cRJ14	67 [95]	65.5 [94]		
c310H5	64 [88]	58 [85]		
c314G4	69.5 [89]	61 [95]		
c419C1	70.5	69.5 [94]	80 [100]	
PYGM	67 [92]	62 [96]	57.5[93]	70 [86]

Table 4.1. Table to show % doublets scored for each cosmid in a normal EBV line. Where hybridisation efficiency (%) has been measured, it is shown in square brackets alongside each value. Shaded boxes show control probes. The remainder are human 16p specific, the most telomeric probe at the top of the table, moving proximally down the table. Each value represents the % doublets scored for at least 200 chromosomes on a single slide.

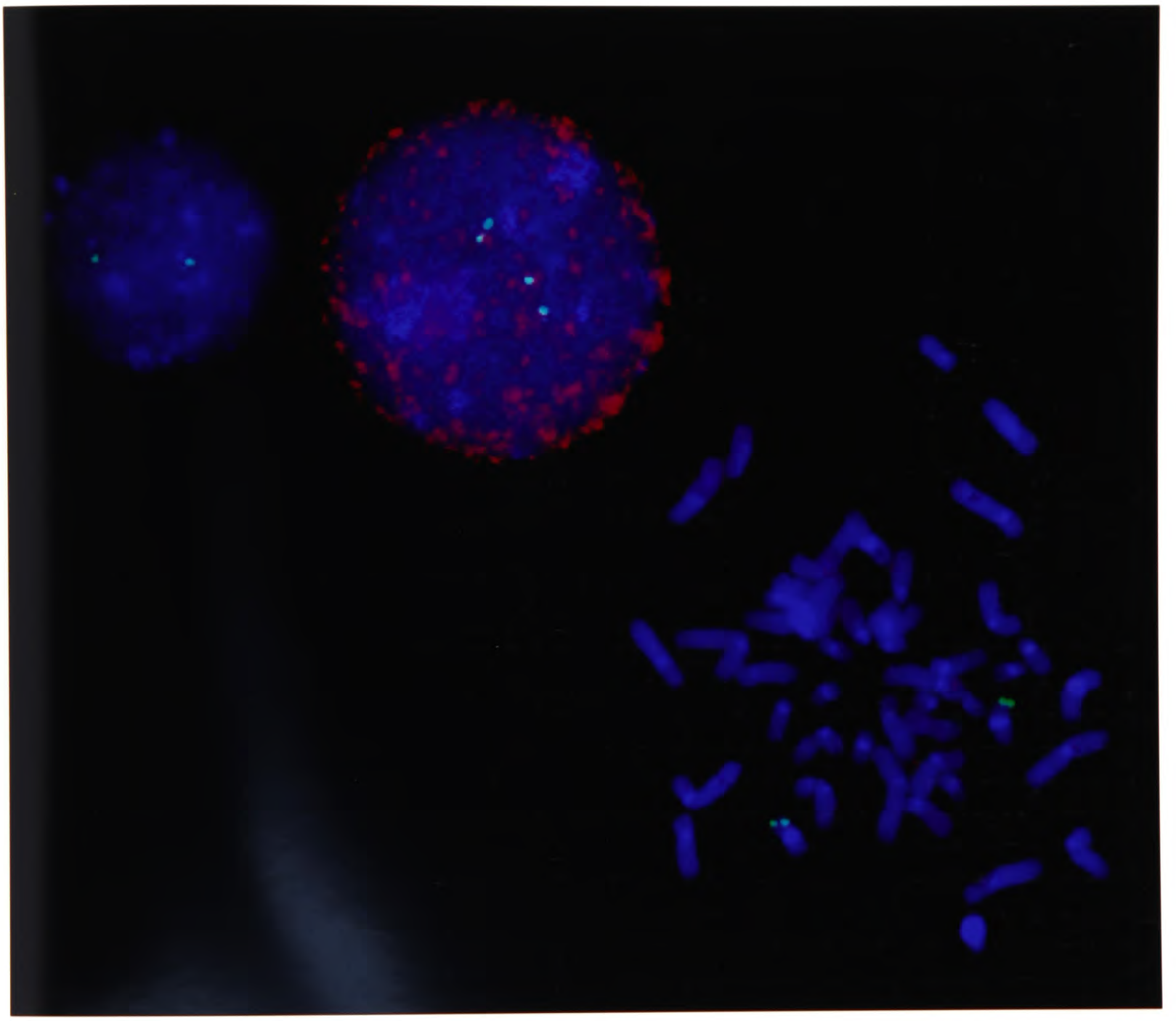


Fig 4.2 A typical field of view of nuclei prepared from an asynchronous culture of a normal EBV-transformed lymphoblastoid cell-line (TA). In this example, the slide has been hybridised with cosmid probe c419C1, specific for chromosome 16p. It has been labelled with DIG-11-UTP and detected with FITC (green). The centre of the figure contains an S phase nucleus (BdU positive, detected with texas red) where both alleles of 16p have replicated (a pair of doublet signals). On the left is a G1 nucleus, where both alleles are unreplicated (two single signals). Bottom right is a metaphase spread. DNA is counterstained with DAPI (blue).

4.3.2 Replication of telomeric 16p in a normal erythroid cell-line

Replication timing was measured in three MEL-16 hybrids, each containing a single, normal copy of human chromosome 16 (JY5-4, CW3-6, JG7). Results are shown in table 4.2. Hybridisation efficiency is shown in square brackets. The 'slide 2' column displays values for duplicate slides of either JY5-4 or JG7. Fig 4.3 shows a metaphase spread from an MEL-16 normal hybrid which has been analysed by FISH using one of the human 16p specific probes (c419C1).

cosmid	*slide 1 (JY5-4)*	*slide 2*	*slide 3 (JG7)*	*slide 4 (CW3-6)*
cSamD4	29.5 [81]		33.5 [98]	27.5 [98]
cNFG9	55.5 [88]		56 [92]	
cRA36	62 [100]	63 [100]	56 [93]	
cNFG7	66 [95]		68.5 [95]	
cGG4	74.5 [100]		78 [100]	
cGG2	66.5 [100]		70 [90]	
cRN24	66.5 [100]		73 [100]	
cGG1	76 [100]	72.5 [100]	67 [95]	72 [100]
cos12	60.5 [95]		63 [90]	77 [100]
cRJ14	78.5 [97]		81 [100]	
c310H5	68 [100]		74 [100]	
c314G4	73.5 [100]		81.5 [100]	
c419C1	72.5 [100]	74 [100]	75.5 [100]	75.5 [100]
P48	63 [100]		55 [100]	62.5 [90]

Table 4.2. % doublets scored in three normal erythroid cell lines (JY5-4, CW3-6, JG7). Results for the control cosmids are shown in the shaded boxes. Hybridisation efficiency (%) is shown in square brackets.

Results for both the erythroid and non-erythroid cell-lines have been expressed graphically in fig 4.4A to show how percentage doublets scored varies with distance along the 16p chromosome. The midpoint of each cosmid has been used as a coordinate. Mean percentage doublet scores and error bars of 1 standard deviation (sd) are shown for each. Fig 4.4B shows percentage doublets measured for the control probes. As seen in fig 4.4A, the percentage doublets scored for any given cosmid is higher in the erythroid line when compared to the non-erythroid line (except for the final cosmid c419C1, where values are identical). This suggests that replication occurs earlier in S phase in the erythroid line relative to the non-erythroid line, throughout the majority of the region examined. Nevertheless, the overall profile of replication timing across the region appears to be very similar in erythroid and non-erythroid cells. Hence,

although the absolute time in S phase in which the region replicates may differ between the cell-types, the temporal order of replication along the chromosome appears to be the same. From the graph it can also be seen that the most of the human α globin region is earlier replicating than the endogenous mouse α globin cluster in the MEL-16 hybrid lines. The majority of the region also appears to replicate earlier than the human early replicating control probe, which is specific for the ubiquitously expressed gene PYGM. However, the PYGM value has been obtained by scoring EBV lines as opposed to the MEL-16 lines, therefore it is not a direct comparison of replication timing in the same cell-line. In both cell-types, the most telomeric part of the 325kb region examined (approximately 80kb) becomes increasingly later replicating towards the telomere. In the MEL-16 lines, this 80kb region still replicates at an early part of S phase when compared to the controls. In the EBV line, this region appears to replicate midway between the early and late replicating controls. This probably equates to mid-S phase.

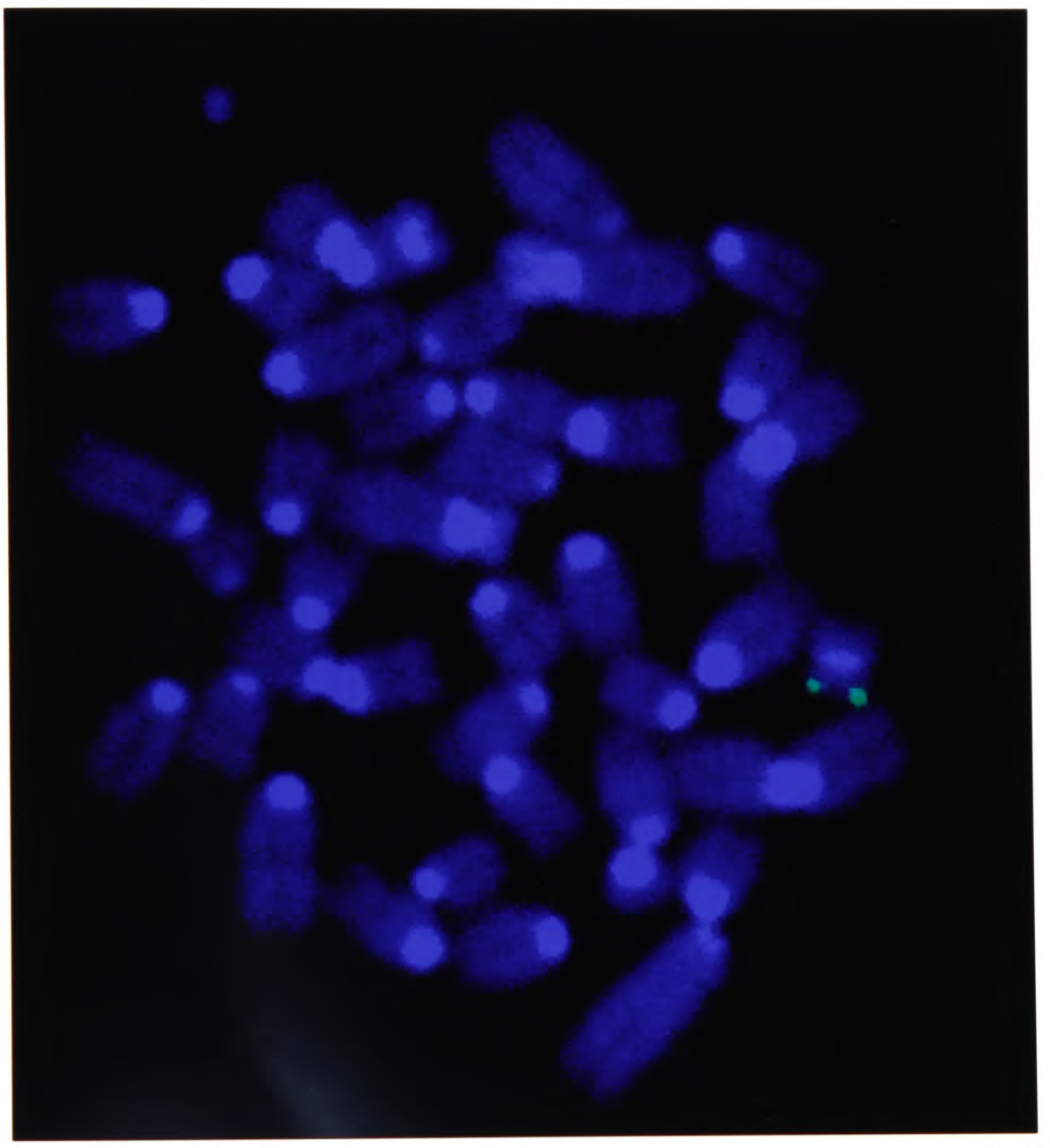


Fig 4.3 Metaphase spread of a normal MEL-16 interspecific hybrid line (JY5-4), probed with human 16p specific cosmid probe c419C1 (green). DNA is counterstained with DAPI (blue).

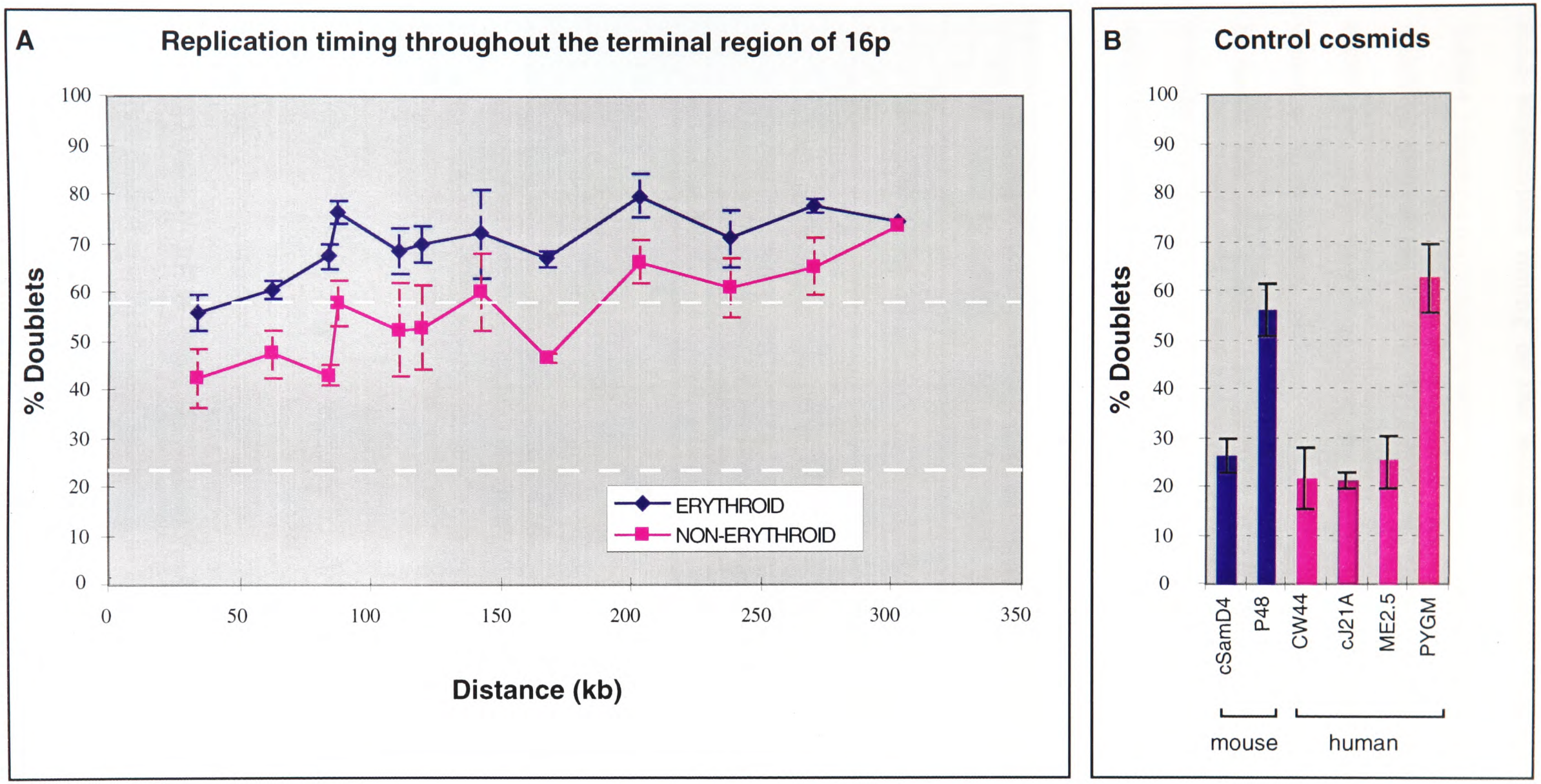


Fig 4.4A Replication timing throughout the terminal 325kb of human 16p. Each co-ordinate is the mean value recorded for each cosmid, plotted at its midpoint. Error bars represent 1 standard deviation. The greater the % doublet score, the earlier the region replicates. White dashed lines indicate the mean % doublet scores for the early (top line) and late (bottom line) control cosmids. Results for the erythroid line are shown in blue and the non-erythroid cell-line in pink.

Fig 4.4B Replication timing of the control cosmids. Mean values have been calculated from the data produced from all cell-lines examined in this thesis. Error bars represent one standard deviation. Control cosmids measured in the erythroid line are shown in blue and in the non-erythroid line are shown in pink. Cosmids P48 and PYGM are early replicating; the remaining cosmids are late replicating.

4.3.3 Replication timing of the individual human 16p alleles

Table 4.3 displays percentage doublets scored for each cosmid in the mutant EBV lines which are monosomic for 16p. Hybridisation efficiency could not been recorded for every slide, as the number of metaphase spreads on slides made from both cell-lines were low. Consequently, on some slides, 10 hybridised metaphase spreads could not be found. On most slides, however, there were at least 5 spreads, and from these hybridisation efficiency was found to be above 90%. These figures are not given in the table. FISH analysis of an EBV cell-line monosomic for 16p is shown in fig 4.5.

cosmid	*NB 1*	*NB 2*	*JG 1*	*JG 2*
CW44	20 [89]		19.5	
cJ21	19		21.5 [90]	
cosME2.5	18.5 [93]		23	
cNFG9	34.5		44 [100]	
cRA36	45		41.5 [100]	
cNFG7	38.5		47.5	
cGG4	44	48.5 [100]	50.5	
cGG2	34		50.5 [95]	
cRN24	40		48.5	
cGG1	46		55	
cos12	48.5		45.5	
cRJ14	62 [97]		69 [100]	
c310H5	44.5	56.5 [100]	67.5 [93]	
c314G4	53		61.5	
c419C1	67.5		62	63.5
PYGM	56.5	70.5 [100]	52.5 [96]	

Table 4.3 % doublets scored in two EBV lines which are monosomic for the terminal region of 16p. The region is either maternally (NB) or paternally inherited (JG). Where two values are given (e.g. NB1 and NB2) a cosmid has been hybridised and scored on two different slides. Hybridisation efficiency (%) is shown in square brackets.

There could be several reasons why only a small number of metaphase spreads were found. One explanation could be that there were few mitotic cells present when the cells were harvested. However, when the nuclei preparations were made, the percentage of nuclei in S phase was calculated to be greater than 30% in each case (for the method see section 2.3.4). This suggests that both cell-lines were in growth phase when the nuclei were prepared and thus a high mitotic index would be expected. Therefore, it is more likely that an error occurred during harvesting of the nuclei or during slide making.

During harvesting, if the hypotonic solution used to swell the cells is too hypo-osmotic (section 2.3.3), it may cause lysis of mitotic cells. Alternatively, a problem may have occurred during slide making (section 2.3.5), which is partly dependent upon laboratory conditions of temperature and humidity. If conditions are such that slides take a long time to dry, dispersal can occur, whereby chromosomes no longer appear as a discrete unit of total chromosomes from a single mitotic cell, but individual chromosomes spread out and finally settle some distance apart. Despite several attempts at making slides from the same preparation of fixed nuclei on different days, the number of metaphase spreads could not be increased. However, slides that did contain enough metaphases demonstrated that hybridisation efficiency was high. Therefore it was not thought necessary to prepare the nuclei again. Fig 4.6 displays the replication profiles of the individual alleles throughout 16p. These profiles are not the same, indicating that the pattern of replication along the individual 16p alleles may differ. However, differences recorded in absolute timing between the two alleles were never very large. The largest differences measured were for the regions spanned by cGG2 (16.5%) and c310H5 (17.5%). All other cosmids showed differences of less than 10%. Thus for the majority of the region examined, the results indicate that homologous regions on the different alleles replicate synchronously. It is interesting to note that when the values for the individual alleles are added back together, and the mean for each cosmid is re-calculated, the replication pattern produced is very similar to the pattern seen in the normal diploid EBV line (fig 4.7). It is important to know whether the different patterns observed from the mono-allelic cell lines is genuinely different from the diploid line, or if such differences could be generated by artefact. It could be argued that adding the two monoallelic values back together and taking the mean, essentially doubles the sample size and therefore reduces error. However, it has been demonstrated that increasing the number of nuclei scored does not significantly change the percentage doublets scored (section 3.3.3, figure 3.3). Therefore it seems that the pattern of replication across the maternally and paternally derived alleles may truly differ in a subtle way.

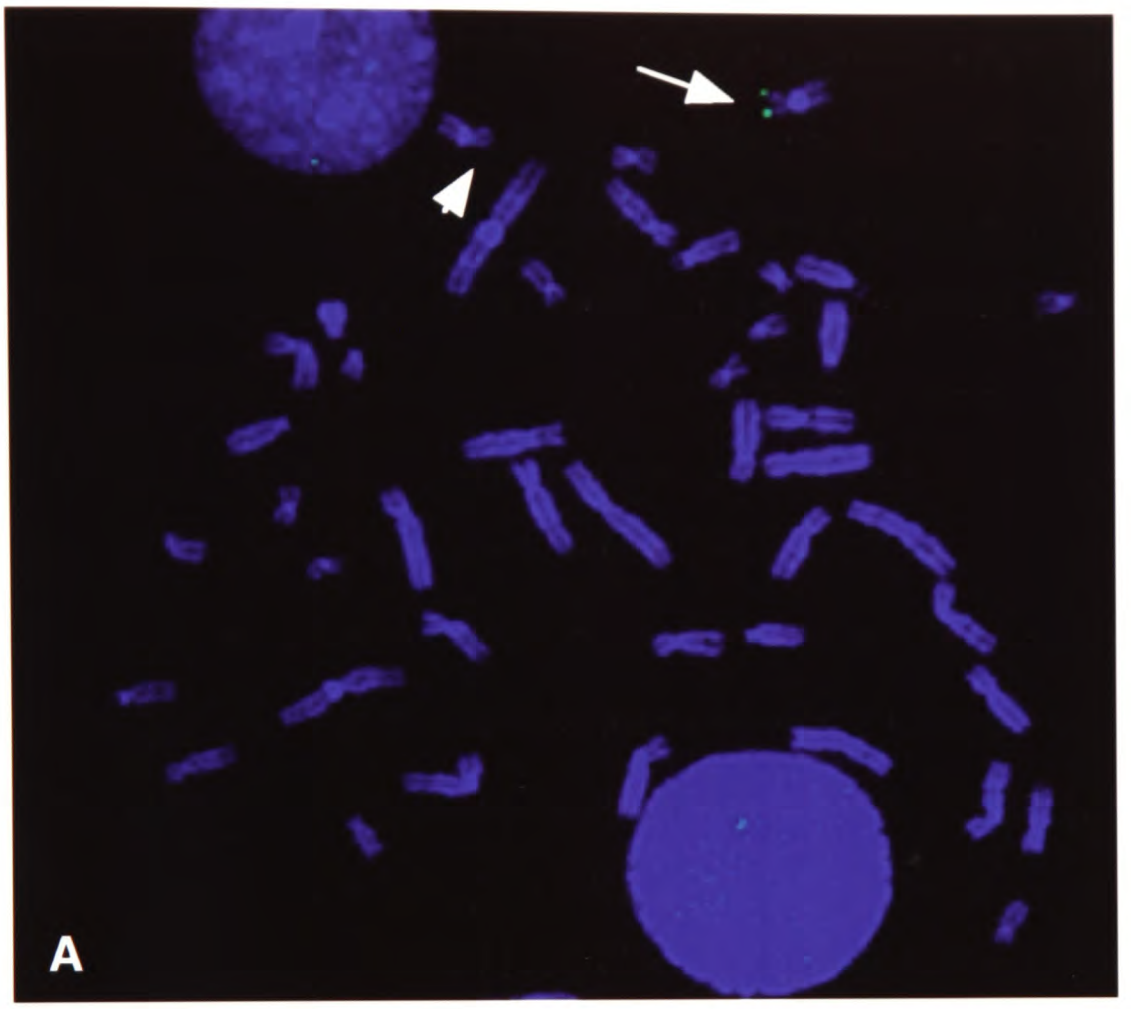


Fig 4.5 Measuring the replication timing of the individual alleles of chromosome 16p.
A. Metaphase spread of an EBV-transformed lymphoblastoid cell-line which is monosomic for a paternally inherited copy of 16p (NB). Hybridisation with a 16p specific cosmid probe (c314G4, green) demonstrates the lack of signal on one allele of 16p (arrowhead) and the presence of signal on the other (arrow).
B. FISH with another 16p specific probe (c310H5) of the EBV-line monosomic for the maternal copy of 16p (JG). Field of view contains two G1 nuclei (top left and top middle), one G2 nucleus (right) and two S phase nuclei (red), one where the maternal copy of 16p has not yet undergone replication (bottom left) and one where it has replicated (bottom middle). Both slides are counterstained with DAPI (blue).

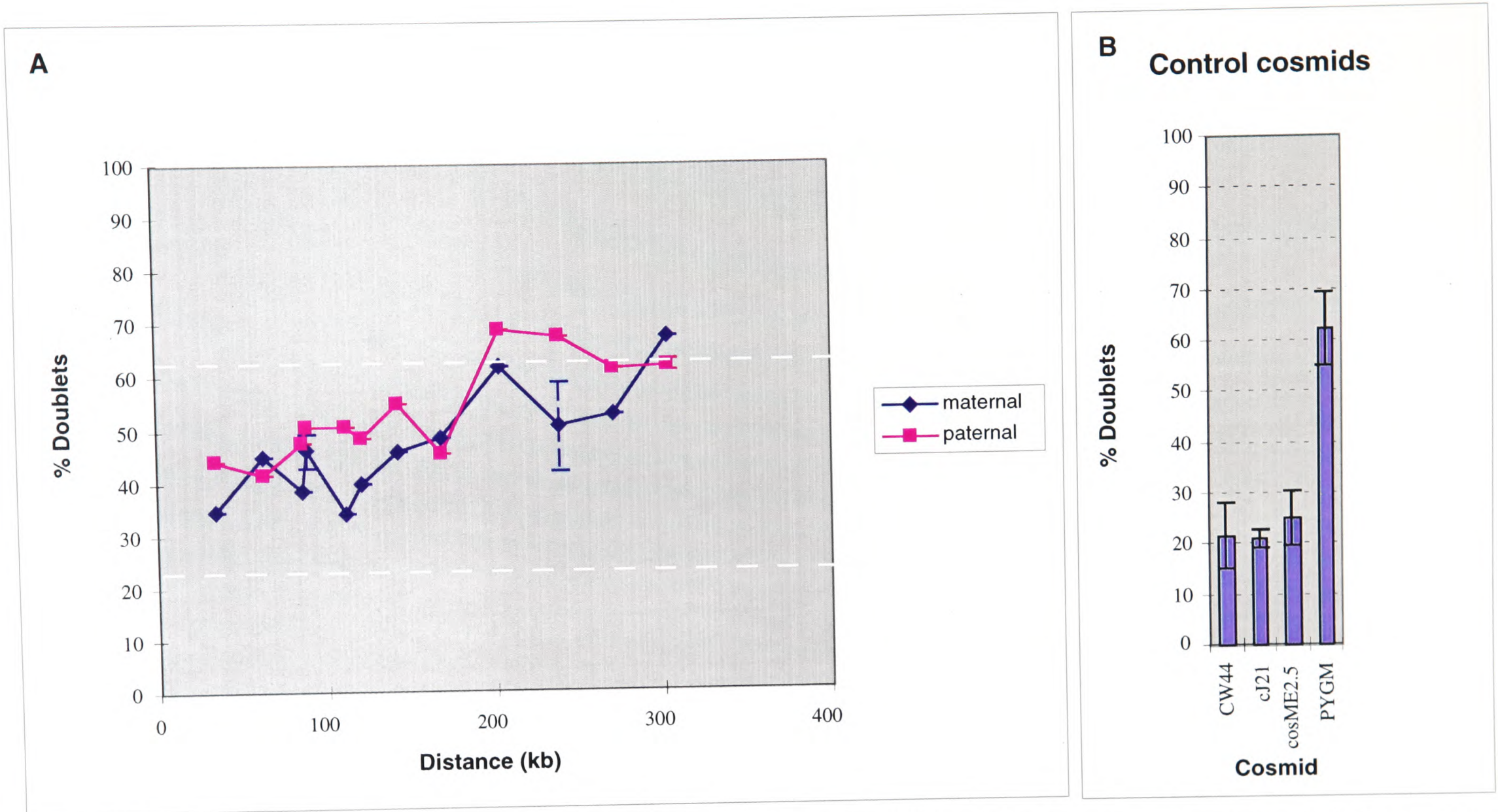


Fig 4.6A Replication characteristics of the individual alleles of 16p. Replication timing of the 325kb region has been measured in cell-lines derived from patients monosomic for this region. Results from the cell-line containing the paternally inherited copy (JG) is shown in pink and the maternally inherited copy (NB) is shown in blue. White dashed lines indicate the mean values of the early (top line) and late (lower line) replicating controls.
Fig 4.6B Replication timing of the control cosmids measured in the EBV lines. Mean values shown have been calculated by pooling scores obtained in all the EBV-lines. PYGM is early replicating; cJ21, CW44 (CF locus) and cosME2.5 (β globin) are late replicating.

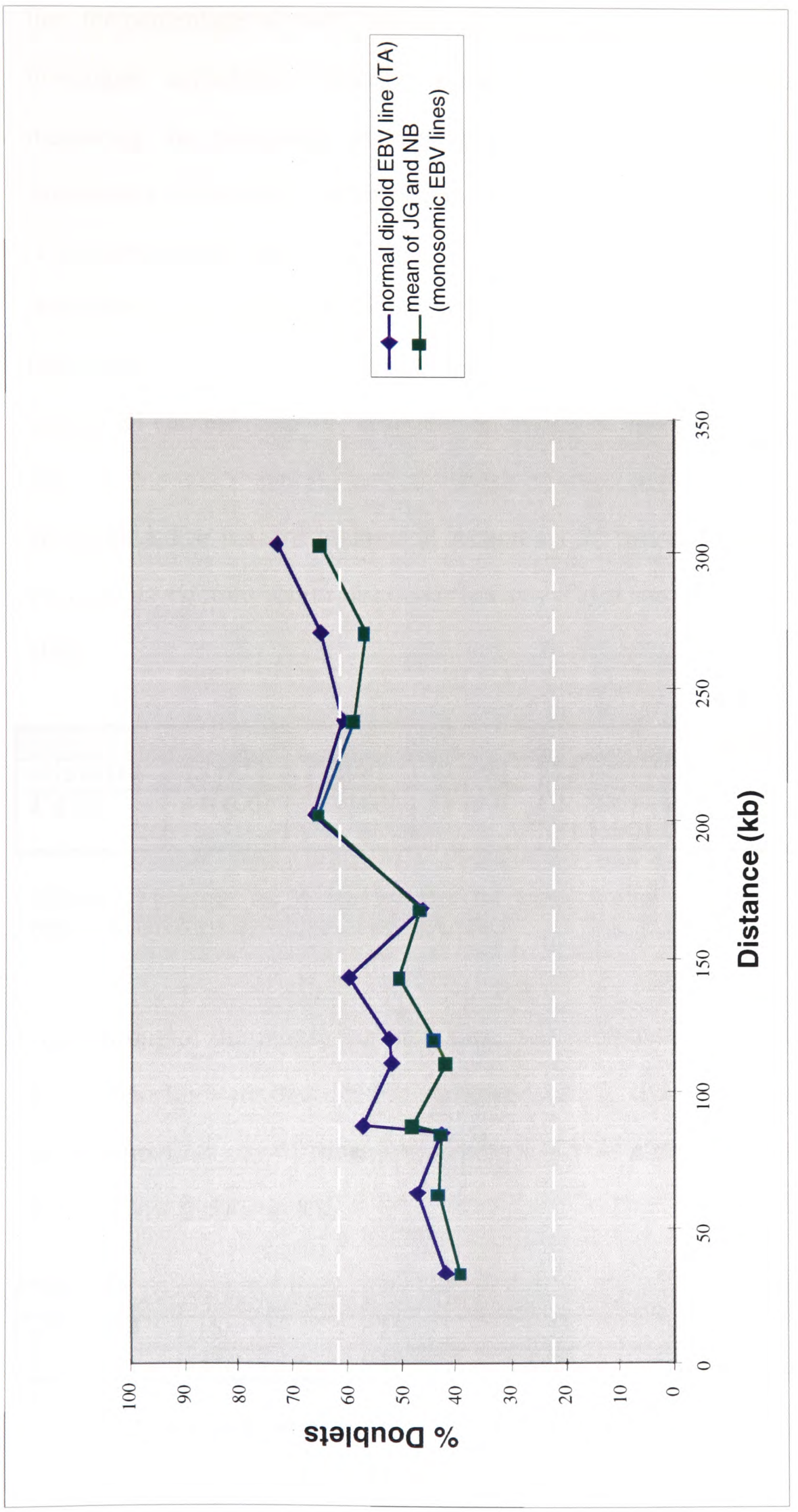


Fig 4.7 Comparison of the temporal order of replication and the absolute timing of the 325kb region in the normal diploid EBV line (TA), shown in blue and the mean of the two monosomic EBV lines (JG and NB). Absolute values of replication are within 10% for most cosmids. The pattern of the profiles is also very similar. Mean values of % doublets are shown for the early (top line) and late replicating (lower line) control cosmids.

4.3.4 Measuring allelic asynchrony of the control probes

In addition to scoring the percentage of replicated loci (% doublets) in the normal EBV line, the percentage of nuclei displaying a 2:1 signal was measured, as another means to investigate asynchrony between homologous loci on different alleles. Prior to measuring the percentage asynchrony of the human 16p region, the percentage asynchrony of the early and late replicating controls were measured.

(I) mouse control cosmids.

Asynchrony was measured for the mouse control cosmids in all of the MEL-16 hybrid lines analysed in this thesis, except one (table 4.4). Values for the hybrid containing the mutant IC chromosome 16, described in chapter 8, has been excluded. It was found that this line was aneuploid for both mouse chromosomes 3 and 11, where the controls are located. For reasons outlined in section 4.1.2, this cell-line has not been used to evaluate asynchrony. Hybridisation efficiency (%) is shown in square brackets for each slide.

cosmid	*JY5-4*	*JG7*	*CW3-6*	*C40*	*C10*	*FZ*	*TAT4A*
cSamD4	15 [81]	17 [98]	13 [98]	25 [95]	13	14 [100]	24 [95]
P48	34[100]	32 [100]	35 [90]	35 [96] 31 [100]	39	23 [93]	37 [99]

Table 4.4 Measuring the % asynchrony of the mouse control cosmids in MEL-16 hybrid lines. Hybridisation efficiency is given in square brackets.

Asynchrony of the mouse controls were also examined in the transgenic lines (table 4.5). These lines are described in chapters 5 and 7. Hybridisation efficiency could not be measured for any of these lines due to a lack of metaphase spreads, which was a result of low mitotic index.

cosmid	*α53 A*	*α53 B*	*α100*	*α107*	*α108*
cSamD4	15	25	13	22	17
P48	37	48	41	36	42

Table 4.5 Measurement of the % asynchrony seen with the mouse control cosmids in the transgenic lines.

(2) Human control cosmids

The same procedure was performed for the human control probes by scoring the percentage asynchrony in the normal EBV line (TA) and the mutant EBV lines (JG and NB). Results are shown in table 4.6.

cosmid	*TA1*	*TA2*	*TA3*	*TA4*	*NB1*	*NB2*	*JG1*
CW44	21 [79]	22 [82]			25 [89]		25
cJ21	21	22			18		11 [90]
cosME2.5	18	27 [80]			19 [93]		30
PYGM	36 [92]	30 [96]	27 [93]	30 [86]	31	19 [100]	34 [96]

Table 4.6 % asynchrony obtained for the human control probes in the normal EBV line (TA) and the mutant EBV lines which are monosomic for 16p (JG and NB). Hybridisation efficiency is given in square brackets.

4.3.5 Measuring allelic asynchrony throughout the terminal region of 16p

Table 4.7 displays the values of percentage asynchrony obtained with the human 16p specific probes when analysing the normal diploid EBV line (TA). Asynchrony has been measured for each cosmid probe on at least two slides. Fig 4.8 compares these values with the values obtained for the control probes.

cosmid	*TA 1*	*TA2*	*TA3*
cNFG9	28	42 [93]	
cRA36	48	41 [85]	
cNFG7	26	28 [93]	
cGG4	30	26	
cGG2	50	39 [78]	
cRN24	30	39 [94]	
cGG1	31 [90]	26 [93]	
cos12	36	30	
cRJ14	28 [95]	27 [94]	
c310H5	35 [88]	34 [85]	
c314G4	42 [89]	31 [95]	
c419C1	10	30 [94]	32 [100]

Table 4.7 Values of asynchrony obtained for human 16p in a normal EBV line (TA). Hybridisation efficiency is given in square brackets.

Mean values of percentage allelic asynchrony for each 16p specific cosmid, measured by the 2:1 pattern in the diploid cell-line, is consistently greater than percentage allelic asynchrony measured in the monoallelic cell-lines (shown in fig 4.8). The 'monoallelic'

method produces mean percentage asynchrony values ranging from 3-17%, compared to 24-45% using the '2:1' method. The reason for this increase could be due to artefacts when scoring the nuclei, or could genuinely indicate asynchrony. Whereas the 'monoallelic' method calculates the overall dominance of early replication of one particular allele over the other, the '2:1' method gives an overall picture of how replication timing may differ between the two alleles. This overall picture could indicate a parent of origin replication imprint, or it could indicate mosaic asynchrony, with either allele replicating first. Without performing a dual hybridisation of the diploid nuclei, whereby another probe is used to 'tag' the maternal or paternal allele, it is impossible to know which allele is replicating first to give the 2:1 pattern. However, as the '2:1' values were consistently higher than the 'monoallelic' method, it supports the idea of mosaic asynchrony occurring. It is interesting to note that one cosmid, cGG2, produces a high percentage asynchrony using both methods (16.5% 'monoallelic method', 45% '2:1' method). Conversely, cosmid c310H5 does not display a noticeably large 2:1 pattern (34.5%) despite a large 'monoallelic' difference (17%). Evidence to suggest that artefacts could be responsible for the large percentage asynchrony values by the '2:1' method comes from the controls. The late replicating controls (CW44, cJ21, cosME2.5, cSamD4) show the lowest percentage of 2:1 nuclei (<24%, fig 4.8). These cosmids also score the lowest percentage doublets. This could suggest that the more doublets scored, the larger the error in scoring, which appears as asynchrony. However, this correlation does not hold true for the 16p cosmids (fig 4.9).

Fig 4.8

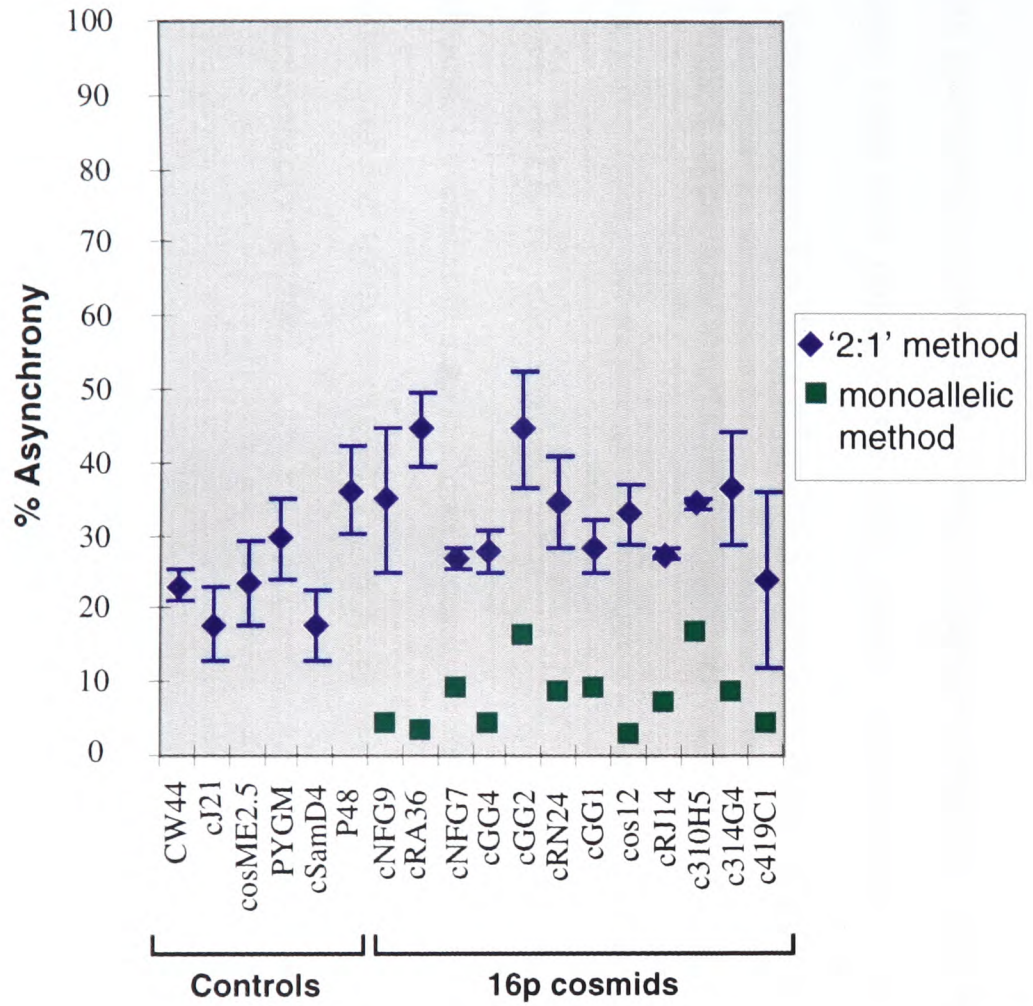


Fig 4.9

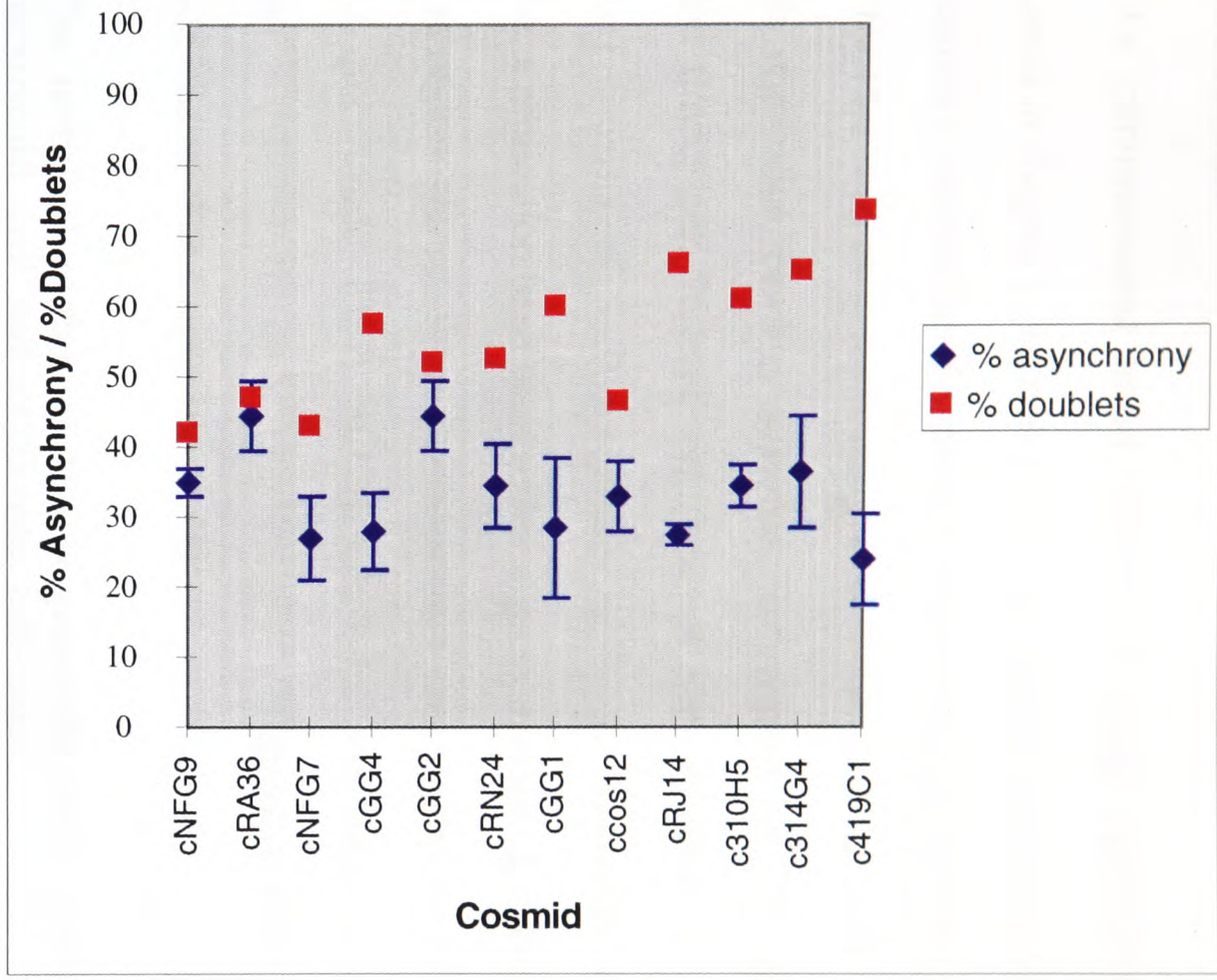


Fig 4.8 Measuring the percentage of asynchrony between the two alleles of 16p. The data series shown in blue is the mean % asynchrony recorded in the normal diploid EBV line, i.e. the % of nuclei displaying a '2:1' hybridisation pattern (error bars represent one standard deviation). The data series shown in green for the 16p cosmids, is the difference in % doublets recorded between the monosomic cell-lines. The 'monoallelic' method gives consistently lower values than the '2:1' method. Cosmid controls P48 and cSamD4 are mouse-specific, hence their '2:1' pattern was measured in the MEL-16 hybrid lines. The '2:1' pattern for the remaining controls was measured in the monosomic and diploid EBV-lines.

Fig 4.9 Is the % asynchrony recorded by the '2:1' method related to the % doublets scored? In the normal EBV line (TA), the % doublets and % asynchrony has been recorded for each cosmid. The mean % doublets scored for each cosmid is shown in red with the % asynchrony in blue (error bars of one standard deviation). The y axis is both % asynchrony and % doublets. A larger value of % doublets may result in a higher apparent asynchrony due to artefacts when scoring slides. This does not seem to be true for the 16p cosmids.

4.4 Discussion

4.4.1 The chromosomal band 16p13.3 is early replicating

As discussed in chapter 1 (section 1.3.3), in addition to morphological banding (R, G and T banding), mammalian chromosomes can be banded in terms of replication timing. Human 16p13.3 is an early replicating, T band (Drouin *et al.* 1994, Flint *et al.* 1997a). Although replication banding suggests that regions of the genome with similar replication timing are clustered together, an average band at metaphase (standard resolution) consists of 5-10 megabases of DNA (Bickmore and Sumner 1989). Therefore, a replication band must reflect a gross average of replication timing across a very large region. Hence, although informative, the cytogenetic resolution of replication timing is quite crude. FISH analysis has been used here as a method of higher resolution, to delineate the replication timing of a segment of DNA at least one order of magnitude smaller than a replication band.

4.4.2 The terminal region of 16p is generally early replicating in both erythroid and non-erythroid cell-lines

It was previously stated that, using elutriation analysis, the human α globin genes replicate early in S phase of the cell-cycle in non-erythroid cells, although no specific data were presented in this preliminary communication (Goldman *et al.* 1984). Using a FISH-based approach in this thesis, I have shown that the entire 325kb region of 16p13.3, including the whole of the human α globin cluster, is early replicating in both erythroid (MEL-16 interspecific hybrid) and non-erythroid (EBV-transformed B-lymphocyte) cell-lines. These findings are consistent with Hatton *et al.* (1988) who demonstrated that other multigene clusters (up to 250kb) replicate synchronously in the same portion of S phase.

It is interesting that most of the region containing the α cluster replicates earlier in S phase in the erythroid cell-line compared with the non-erythroid cell-line and it is worth considering this point in some detail. In this study I have used a FISH-based assay to determine the timing of replication across the entire 325kb region. It is acknowledged

that the reliability of such studies lies in comparing the replication of different regions relative to each other in the same cell-line. Since the non-erythroid cell-line (EBV) used is entirely human, whereas the erythroid line (MEL-16) is an interspecific hybrid, the replication timing was, of necessity, compared with different human and mouse controls. It is possible that the differences between the two cell-lines resulted from species differences rather than differences in tissue-type. However, this seems unlikely since the human early and late controls (used for the EBV experiments) gave similar results (% doublets) to the corresponding mouse controls (used in the MEL-16 line). Thus it appears that the 16p region genuinely replicates earlier in erythroid than non-erythroid cells. This is in agreement with previous elutriation studies, which demonstrated that the mouse α globin cluster similarly replicates earlier in erythroid (MEL) than non-erythroid (e.g. S107 and MPC11 plasmacytoma) cells. This pattern of early replication in non-expressing cells and even earlier replication in expressing cells, has also been observed for several murine immunoglobulin genes (Hatton *et al.* 1988). A similar long-range change in replication timing was noted by Strehl *et al.* (1997), who studied replication timing across a human R/G boundary at 13q14.3/q21.1. As for the 16p segment studied here, although the pattern of replication across the 4-5Mb 13q region did not change, the timing went from later (EBV transformed B-lymphocytes) to earlier (peripheral blood T-lymphocytes and fibroblasts). At present, although the α globin data suggest that this shift to earlier replication timing may be related to gene expression, there is insufficient information to know if this is generally the case.

4.4.3 Does the 325kb region replicate synchronously?

When replication timing of the individual alleles is analysed using the monosomic cell-lines, mean values obtained for most cosmids are within 10%. Similar FISH analysis of monosomic lines by other groups report <12% asynchrony for unimprinted regions, 25-40% asynchrony for imprinted regions found on autosomes, and up to 70% asynchrony between the active and inactive X chromosome (Kitsberg *et al.* 1993a, Torchia *et al.* 1994, Boggs and Chinault 1994, Knoll *et al.* 1994, Greally *et al.* 1998). By comparison, the whole 325kb region examined here replicates synchronously. Of

the small differences recorded, 9 out of 12 cosmids detect regions that appear to replicate earlier on the paternal allele. This is similar to findings reported by Kitsberg *et al.* (1993a), when examining imprinted regions in humans and mouse.

Using the '2:1' method to measure replication timing, percentage asynchrony values were considerably larger (24-45%). These values fall within the range found for imprinted genes. The results of percentage asynchrony for the control cosmids CW44, cJ21 and PYGM, in this thesis, can be compared with results of Kitsberg *et al.* (1993a), where these probes were used as synchronously replicating controls. They found percentage asynchrony in the human lymphocyte line (LAZ) to be: CW44 9%, β globin (probe HG-4) 9%, PYGM 9%. In this thesis, when measuring the percentage asynchrony in EBV-lines, the following values were obtained: CW44 23%, β globin (cosME2.5) 23%, PYGM 29%. This consistent scoring of higher percentage asynchrony described here, could be due to scoring criteria. As mentioned in section 3.3.4 and 3.6., difficulties can arise when scoring a doublet signal. With some probes, doublet signals can be scored very easily as the dots appear separated (suggesting that the replicated sequences pull apart quickly). Conversely, some probes produce doublet signals in many nuclei that are very close together, which are difficult to resolve as two distinct dots (suggesting that the replicated DNA pulls apart slowly). This kind of variation was seen within the control probes and within the cosmids that span the 16p region. In this thesis, unless hybridisation signals could be seen as two separated dots, the loci was scored as a singlet (unreplicated). This may explain the discordant results. However unlikely, it cannot be ruled out that mosaic asynchrony may be occurring throughout this region. Others have reported a very small percentage of 2:1 nuclei displaying a 'reverse' replication imprint to an overall parent of origin replication imprint (Bickmore and Carothers 1995). As mentioned in section 4.3.5, dual hybridisation FISH analysis would settle this query.

4.4.4 Replication asynchrony and genetic imprinting

Study of α globin gene expression has revealed that the transcriptional activity of the two adult α globin genes is different. Throughout development, the steady state level of α2 mRNA predominates over α1 mRNA by approximately 3:1 (Orkin and Goff 1981). Despite the unbalanced transcriptional activity between the two genes, both alleles are believed to contribute equally to α globin gene expression. A similar deletion of maternal or paternal 16p gives rise to a consistent α thalassemia phenotype and no differences in one form of epigenetic modification (methylation) are seen between the two autosomes, which also would suggest the α globin genes are not imprinted (D.R. Higgs, pers. comm). Patients that are monosomic for considerable stretches of the 325kb region examined in this thesis, also have no abnormal phenotype, which is evidence that the majority of the genes in this region are not imprinted (Wilkie *et al.* 1990a, Flint *et al.* 1994). There is one gene however, which could possibly be imprinted. This is a novel gene called Axin or Fused (Fu), which has recently been identified 3' to the α globin cluster, within cosmid c419C1 (Flint *et al.* 1997a, R.Daniels, pers. comm). The mouse homologue is located on chromosome 17 (Sussman *et al.* 1994) and it encodes a protein which functions as a negative regulator of embryonic axis formation (Zeng *et al.* 1997). Skewed genetic penetrance suggests this gene is imprinted in mouse, with the paternal allele being dominant (Ruvinsky *et al.* 1990). Curiously, this region shows one of the lowest values of asynchrony between the maternal and paternal alleles when measured using 16p monosomic cell-lines. The maternal allele apparently replicates slightly ahead of the paternal allele (4.25% difference between the two).

4.4.5 Replication timing and gene expression

As discussed in section 1.4.2 and can be seen in fig 4.1, the α globin gene cluster is located in a very gene dense region (approximately 1gene/20kb). Four housekeeping genes lie between between the α globin genes and the telomere. The α globin regulatory element (HS-40) is located in an intron of one of these genes (Vyas *et al.* 1992). A further five genes are found on the centromeric side of the α cluster. The pattern of

expression of these five genes has been examined in several cell-lines, including K562 cells (erythroid) and EBV-transformed lymphoblastoid lines (non-erythroid). As seen in fig 4.1, two of the five genes are expressed in both K562 and EBV lines; two others are expressed in K562 cells only; one gene is expressed in neither cell-line. This latter gene is a protein disulphide isomerase (PDI) and is expressed in liver only (fig 4.1, from Flint *et al.* 1997a). The majority of the 325kb region examined, including all of the region where these genes are located, is early replicating in a non-erythroid line (EBV) and an erythroid line (MEL-16 hybrid). This demonstrates that replication timing is not entirely linked to the pattern of gene expression. It is evident that unexpressed genes can be part of an early replicating region, including the tissue-specific α cluster. This is in contrast to the β globin cluster, where replication timing is tightly linked to expression (Epner *et al.* 1981, Furst *et al.* 1981, Epner *et al.* 1988, Dhar *et al.* 1988, Dhar *et al.* 1989, Forrester *et al.* 1990).

A compilation of mammalian replication timing studies by Eyre-Walker (1992), suggests that all genes expressed in a tissue are replicated early, whereas unexpressed genes may replicate at any time during S phase. Of the 44 mammalian genes considered, 17 of these were house-keeping genes, and all were found to replicate early in S phase. Of the 25 tissue specific genes, in non-expressing cells, approximately half were early replicating (13) and half late replicating (12). The expression status of the remaining two genes was unknown. The results shown here are entirely consistent with this report. The human α globin cluster, in addition to the other tissue specific genes in this region, are examples of genes that are early replicating even when transcriptionally silent. This is probably due to the presence of the four house-keeping genes. All the evidence to date suggests that house-keeping genes are constitutively early replicating, in accordance with their constitutive transcriptionally active pattern of expression (Eyre-Walker 1992 and references therein). Hence, although the 325kb region has only been examined in two cell-lines, the fact it contains four ubiquitously expressed genes, suggests at least part of the region will be constitutively early replicating.

Goldman *et al.* (1984) hypothesed that early replication is required, but not sufficient for gene transcription. For vertebrates, no data are available to refute this theory. Our results support this hypothesis, since unexpressed genes have been found to be early replicating, whereas no expressed genes have been found to be late replicating.

4.4.6 Replication differences between the α and β globin clusters

The fact that the β globin cluster has an expression linked pattern of replication and the α cluster does not, appears to be related to chromosomal environment. As seen here, the α globin cluster is located in the terminal band of the p arm of chromosome 16. This is a Giemsa light band and a T band (Saccone *et al.* 1992). The β globin cluster is also located in a terminal band (11p15.5) which is a Giemsa light band and a T band (Saccone *et al.* 1992, Quackenbush *et al.* 1995). Although superficially similar locations, at a higher resolution these environments are very different. Sequence analysis shows that the α cluster is located near the tip of the telomere (170kb away in the type A allele, shown in fig 4.1), at the very end of the terminal band. Less accurate physical mapping estimates that the β cluster lies approximately 8 megabases from the tip (Quackenbush *et al.* 1995) and is in fact, close to the R/G band boundary of 11p15.5 and 11p15.4 (chr 11 database, http:\\shows.med.buffalo.edu\database.html). Molecular analysis has also shown that the β cluster occupies a discrete regulatory chromosomal domain. This region is AT-rich, and contains no other genes apart from the β globin genes, which adopts a 'closed' chromatin environment (general DNase 1 insensitivity) in non-expressing cells, that undergoes decondensation to an open chromatin conformation (DNase1 hypersensitive) in expressing cells (Forrester *et al.* 1990). By contrast, the α globin cluster and flanking region is DNase 1 hypersensitive in erythroid and non-erythroid cells, suggesting that the region adopts a constitutively open chromatin conformation, irrespective of α globin expression (Vyas *et al.* 1992, Craddock *et al.* 1995). Replication timing therefore appears to mirror chromatin condensation, with late replication of condensed chromatin and early replication of less condensed chromatin.

4.4.7 Temporal order of replication throughout the 325kb region

FISH analysis has provided a broad picture of early replication throughout this 325kb region. Smaller differences in replication timing are seen within the region, which may indicate the temporal order in which the region is replicated. Both the erythroid and non-erythroid cell-lines display a similar pattern of replication throughout the region, suggesting they share the same temporal order of replication. Stretches of DNA displaying the earliest replication may contain a site where DNA replication initiates, as replication should be earliest at these initiation sites (origins). The farther a sequence lies from an origin of replication, the later it will replicate. If this technique is sensitive enough to detect direction of fork movement, the results may suggest that the erythroid and non-erythroid lines utilise the same origins of replication. This is true for the β globin cluster: the 140kb region contains a single origin, which is active in both erythroid and non-erythroid cell-types (Kitsberg *et al.* 1993b, Aladjem *et al.* 1995). Another origin, found on human chromosome 19, upstream of the *ppv1* gene (unknown function), was active in every proliferating cell-line that was examined. These lines were derived from three different cell-types: neural, epithelial, and connectival (Kumar *et al.* 1996).

Replication origins are reported to be 50-500kb apart (Edenburg and Huberman 1975, Bickmore and Oghene 1996, Verbovaia and Razin 1997). On this assumption, this region could contain between zero and six origins. The region also contains 11 sites which are very close (1 mismatch difference) to a proposed mammalian origin consensus sequence published by Dobbs *et al.* (1994). However, the replication patterns shown in fig 4.4 cannot be easily interpreted to identify any origins of replication or replicons (although Verbovaia and Razin (1997) provide evidence that mammalian replicons need not be symmetrical about the origin).

Considering the more general features of the profile, replication appears to become later approaching the telomere in both cell-lines, which could be explained by two mechanisms. Either the sequence lies a long distance away from an origin, so the

replicosome takes a long time to reach it and replicate it. Alternatively, the region lies near to an origin, but the origin is fired late in S phase. This suggests that either the 100kb telomeric third of this region contains origins which are fired later in S phase than origins in the more centromeric portion. This would require some sort of boundary to prevent replication forks generated from earlier firing origins from passively replicating the region. Alternatively, this region is passively replicated by forks generated from an origin further upstream. As differences recorded in replication timing over this region are small, and look rather gradual, it seems far more likely that replication is initiated in the more central portion of the 325kb region, and a replication fork is moving towards the telomere.

One of the latest replicating regions apart from the most telomeric region (spanned by cNFG9 and cRA36) is the region at and around 170kb from the telomere. The drop in replication timing over this region is more dramatic in non-erythroid line than in the erythroid line (fig 4.4A). The cosmid that spans this region (cos12) includes the two adult α globin genes and 20kb beyond, in the 3' direction. This region contains a gene of unknown function transcribed in the direction of the telomere (Flint *et al.* 1997a, C.Tufarelli, pers. comm). Despite the presence of a gene, the region has been shown to be Alu dense, and to adopt a less open chromatin structure compared to the rest of the 300kb region (shown in fig 1.4, from Flint *et al.* 1997a). Deletion of part of this heterochromatic region, is also strongly implicated in producing an α-thalassemia phenotype, described in chapter 7. From the dip in replication timing throughout this region, it could be further evidence to suggest that replication timing is related to chromatin structure.

One should be cautious of interpreting the data in this way. Whether the technique is sensitive enough to measure small differences in timing has not been rigorously tested. However, dual colour FISH has been successfully used for temporal ordering of replication along a chromosome, for probes that are approximately 50kb apart (Kitsberg *et al.* 1993b, Aladjem *et al.* 1995), and the replication of contiguous cosmids can be

temporally ordered on extracted metaphase chromosomes (Bickmore and Oghene 1996). Further analysis of the 325kb region could be performed using several other techniques to locate DNA initiation sites or direction of fork movement (reviewed in Falaschi *et al.* 1993).

4.4.8 Chromosomal organisation during interphase

During the scoring of S phase nuclei, qualitative differences in BdU appearance were seen between nuclei. Unreplicated loci were found within nuclei that had an even BdU patterning throughout. In many nuclei where loci had replicated, the BdU patterning was more mottled, with bright flourescent regions within a poorly fluorescent background (data not shown). These characteristic changes in BdU appearance during S phase progression have been observed by others (Bickmore and Carothers 1995, Jackson 1995, Ferreira *et al.* 1997). The bright fluorescent regions with the highest concentration of BdU have been termed 'replication foci'.

When studying replication timing of human 16p in the mouse-human hybrid lines, some cosmid probes that contain human-specific repetitive elements give a faint background signal (in addition to the true signal) all over the human chromosome but not on the mouse chromosomes. From this background, it is possible to see the area of the interphase nucleus the human chromosome occupies. Rather than chromosome 16 being dispersed throughout the whole of the interphase nucleus, it appears to occupy a compact, irregular shaped, 'territory' within the mouse nucleus (data not shown). This supports the findings of others, who have previously identified distinct chromosomal territories in interspecies somatic cell-hybrids using human total genomic DNA probes (Manuelidis 1985, Schardin *et al.* 1985). In the human lymphoblastoid cell-lines examined in this thesis, the individual alleles of 16p appear to be in different regions of the nucleus as signals seem far apart. This discovery that homologous chromosomes occupy distinct nuclear compartments, has also been reported by others (reviewed in Cremer *et al.* 1993).

Observations of interphase chromosomal organisation described in this thesis, are made from viewing a two dimensional image down the microscope, whereas the nucleus upon the slide is presumably three dimensional. Therefore, it is difficult to assign the spatial position of signals. In addition, it is not known to what extent *in vivo* nuclear topography is preserved after nuclei are fixed and FISH is performed. A more satisfactory technique would be confocal scanning microscopy of living cells (Cremer *et al.* 1993).

4.4.9 Summary

The replication timing of the terminal 325kb of the chromosomal band 16p13.3 has been determined. This region contains the human α globin cluster, in addition to many other genes. The whole region is generally early replicating in non-erythroid and erythroid cells. Allelic differences in replication timing have also been examined. The whole region is found to replicate synchronously, with homologous regions on the paternal and maternal allele replicating at the same time.

The α globin cluster is an example of a tissue-specific gene cluster, that replicates early in S phase even when the genes are not transcriptionally active. This is thought to be due to the close proximity of ubiquitously expressed genes, which may require early replication for expression. This is in contrast to the β globin cluster which has an expression-linked pattern of replication.

With increasing levels of resolution, replication timing appears to be related to chromosomal condensation. Replication timing shows a general correlation with structural bands (megabases of DNA). It also mirrors chromosome condensation within structural bands, over domains or isochores (100-300kb of DNA). It is also tentatively suggested that replication timing varies according to chromosomal structure within isochores (300kb of DNA).

Chapter 5 Analysis of the replication timing of α globin gene constructs in transgenic mice

5.1 Introduction

As a method to delineate the minumum region of DNA that is required for the correct developmental-stage and tissue-specific expression of the α cluster, various DNA fragments have been tested in transgenic mice. It has been found that although HS-40 confers high-level tissue-specific expression when linked to the α globin genes, expression is not developmentally stable, nor is it copy number dependent (Sharpe *et al.* 1992). When a larger 70kb fragment (GG1/GG2) has been tested, which includes additional sequence and maintains the normal genomic position of HS-40 in relation to the α globin genes, levels of expression are not significantly improved (Gourdon *et al.* 1994). As HS-40 does not appear to function primarily by influencing chromatin structure (Craddock *et al.* 1995, Bernet *et al.* 1995), expression of the constructs may be subject to chromatin-mediated effects, which is dependent on the site of integration into the genome (reviewed in Martin and Whitelaw 1996). An even larger 150kb construct (αPAC) which includes more flanking sequence surrounding the α cluster, appears to sustain a higher level of expression later in development (reviewed in Higgs *et al.* 1998) which may indicate this construct is more able to recreate the chromatin environment that normally contains the α locus *in vivo.* However, this construct still lacks position independence and expression is not linked to copy number (reviewed in Higgs *et al.* 1998).

If HS-40 is not responsible for establishing the permissive chromatin structure that allows α globin gene expression in its normal chromosomal environment, what is? As discussed more fully in the main introduction, the process of DNA replication and the time at which it occurs in the cell-cycle, may play a role in creating a chromatin structure that allows transcription to occur. It may be important to faithfully reproduce the timing of DNA replication seen at the normal chromosomal position in the fragments of DNA

introduced into transgenic mice. This may be achieved by including origins of DNA replication found within and around the α cluster and any *cis*-acting sequences which may be required for early activation of origins. To assess whether the failure to obtain complete regulation of the α cluster in transgenic mice is associated with a failure to reproduce the normal pattern of replication (which was determined in the previous chapter), I have analysed the replication timing of primary cells from two transgenic lines. One line contains the GG1/GG2 construct and the other contains the αPAC, shown in fig 5.1.

Transgenic lines with a low copy number were selected for analysis. This is because firstly, it makes the scoring of nuclei much easier: if transgenes are integrated in high copy number in tandem, it makes it impossible to examine the replication timing of a single copy. Secondly, it allows the analysis of the relationship between chromosomal position and replication timing: if a cell contains many copies of the same transgene integrated at different sites, it is impossible to distinguish signals from different chromosomes. If each transgene had a different pattern of replication according to its chromosomal position, it would be impossible to tell. Thirdly, the presence of high copy numbers inserted in tandem can lead to a 'repeat-induced silencing' effect. With increasing copy number, the level of transgene expression tends to fall, which is correlated with an increase in methylation and chromatin compaction at the transgene locus (Gourdon *et al.* 1994, Garrick *et al.* 1998). Therefore, to study the relationship of DNA replication with gene expression, without the added complications of multiple position effects and copy number effects, analysis was limited to low copy number mice.

Unfortunately, it is quite rare for transgenes to integrate singly or in few copies, so the number of lines available for analysis were very restricted. Line α53 is homozygous for a single copy of the 70kb GG1/GG2 construct (fig 5.1, described in Gourdon *et al.* 1994) and line αPAC100 contains two tandemly integrated copies of a 150kb PAC which spans the region (fig 5.1, reviewed in Higgs *et al.* 1998). The level of α globin

expression in these two lines has been measured in previous studies (Gourdon *et al.* 1994, Sharpe *et al.* in preparation), and is given in table 5.1. There were no single copy lines available to allow the analysis of the replication timing of a given construct at different chromosomal locations.

Line	Copy number	Number studied	Embryonic expression (range) (ζH + αH/ζM + αM)	Number studied	Adult expression (range) (αH/αM)
α53	1	5	13.4 (12-14)	14	7.5 (3-17)
αPAC100	2	6	6.4 (3-10)	6	1.5 (0-5.0)

Table 5.1 α globin mRNA expression levels in lines of transgenic mice. Embryos were analysed at 10.5d in line 53 and 12.5d in line αPAC100. Levels in the α53 line were measured in heterozygotic mice.

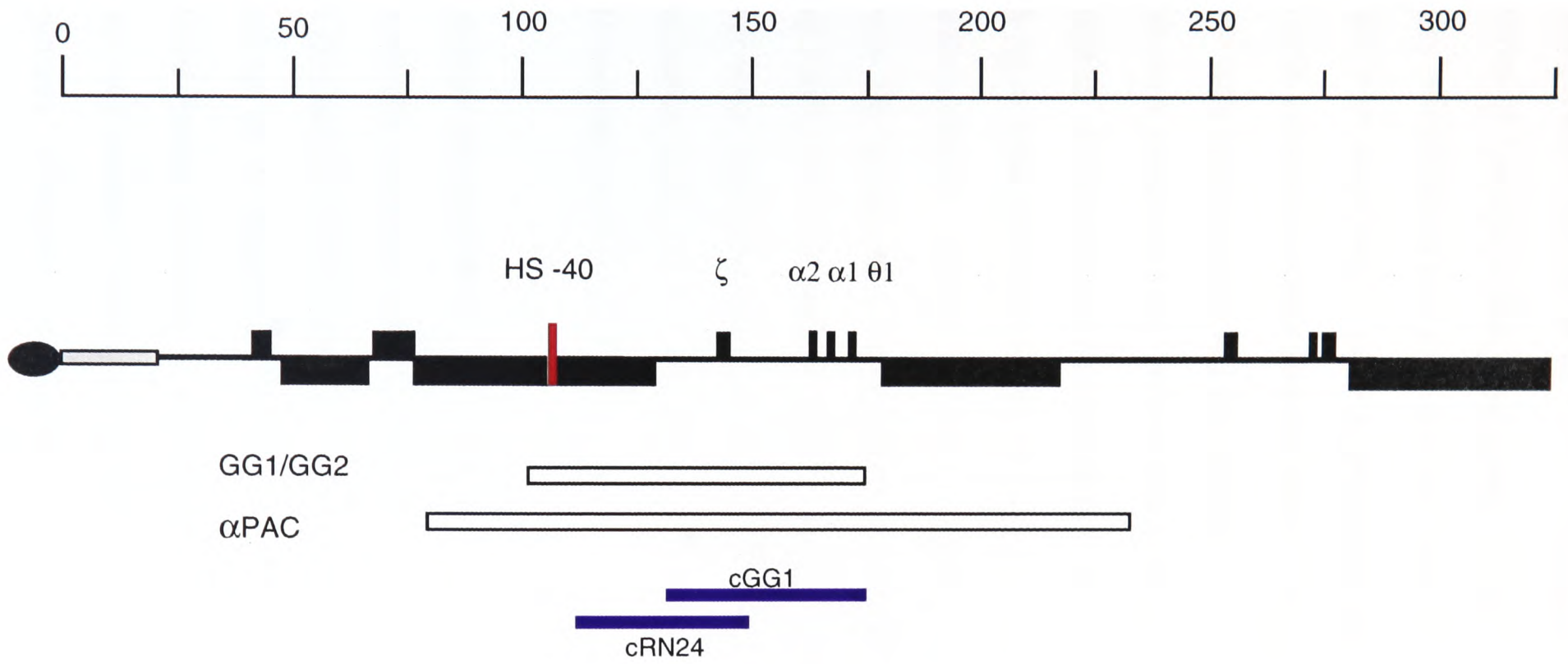


Fig 5.1 Two α globin constructs that have been analysed in transgenic mice. The top schematic shows the positions of the α globin genes and the HS-40 regulatory element in red. Flanking genes are shown as filled black boxes. Below, the black open boxes indicate the fragments of DNA analysed in transgenic mice. Probes used for FISH analysis are shown as blue filled boxes. Scale is in kilobases.

5.2 Methods

5.2.1 Culture of primary erythroid cells from transgenic mice

Erythropoiesis was stimulated in the spleen by inducing haemolytic anaemia using phenylhydrazine (Spivak *et al.* 1973). Transgenic mice received three intraperitoneal injections of 1-acetyl-2 phenylhydrazine (Sigma) at a dosage of 0.04mg/gram of body weight, administered at 5 p.m . on day 1 and at 9 a.m. and 5 p.m. on day 2. On day 5, mice were sacrificed by cervical dislocation, each spleen removed and placed into a petri dish containing approximately 20ml of complete media. Spleens were minced to release cells into the media and the solution transferred into a 50ml tube (Falcon) using a wide-bore pasteur pipette. After a few minutes to allow tissue debris to sediment to the bottom of the tube, the cell suspension was aspirated into a clean 50ml tube and made up to 30ml with more media. A few ml of cell suspension were taken at this stage to make cytospins, which were scored to estimate the percentage of erythroid cells (see below). The remaining cell suspension was cultured straight away in the presence of 0.1mM BdU to label S phase cells (see section 2.3.3). After ninety minutes, cells were treated with hypotonic solution and fixed in 3:1 methanol: glacial acetic acid, as described in section 2.3.3. Fixed nuclei suspensions were used to make slides for FISH analysis (see section 2.3.5).

To estimate the percentage of erythroblasts in the population of cells recovered from the mice spleens, cytospin slides were stained with Giemsa and viewed using an Olympus BH2 microscope, X100 objective (under oil). A minimum of 300 nucleated cells were scored by Dr A. Benton or Dr. W. Wood, and assigned to either the myeloid or erythroid lineage according to morphological criteria. Erythroblasts contain a large, central, oval nucleus, which contains chromatin which is very basophilic; nucleoli may be present. Cytoplasm surrounding the nucleus is also moderately basophilic. Myeloblasts also contain a large nucleus, but it is less uniform in shape and not so deeply staining. These cells have prominent nucleoli and tend to have a higher cytoplasmic/nuclear ratio; the cytoplasm is less basophilic than erythroblasts.

5.2 Results

The results from chapter 4 indicate that replication of the α cluster and surrounding chromosomal environment may be earlier in erythroid cells, when the α genes are expressed compared with non-erythroid cells, where the genes are transcriptionally inactive. Therefore, it was decided to measure the replication timing of the transgene in erythroblast cells, where the transgene is expressed and compare that to the replication timing of the endogenous mouse α cluster, which is also active and replicates early in S phase (chapter 4). A cosmid specific for the mouse salivary amylase gene was used as a late replicating control (described in chapter 4).

Transgenic mice were treated with phenylhydrazine (see methods), which causes the spleen to become a major site of erythropoiesis. Mice were then sacrificed, cells were extracted from the organ and analysed to assess the percentage of erythroblasts in the total cell population (given in table 5.2), which was subsequently used to prepare fixed nuclei for FISH analysis. Cells extracted from line α53 on the first occasion (α53 A) contained only 60% erythroblasts, so the procedure was repeated. The second time, the percentage of erythroblast cells was 82% (α53 B). Replication timing was measured in both fixed nuclei preparations (table 5.2). For most slides, the number of metaphase spreads was too infrequent to measure hybridisation efficiency effectively. However, analysis of the few metaphase spreads present, revealed the position of the transgene locus (fig 5.2). In line αPAC100, integration has occurred into mouse chromosome 4, within bands A3-A5. In line α53, position of integration is mouse chromosome 12, band E, shown in fig 5.2. Mapping of integration sites of transgenes in both lines has also been performed previously by V. Buckle (unpublished data). The tandem copy of the α PAC100 appears as a single signal both on metaphase spreads and in interphase nuclei. Hence, replication timing measured in this line is the average of the two copies.

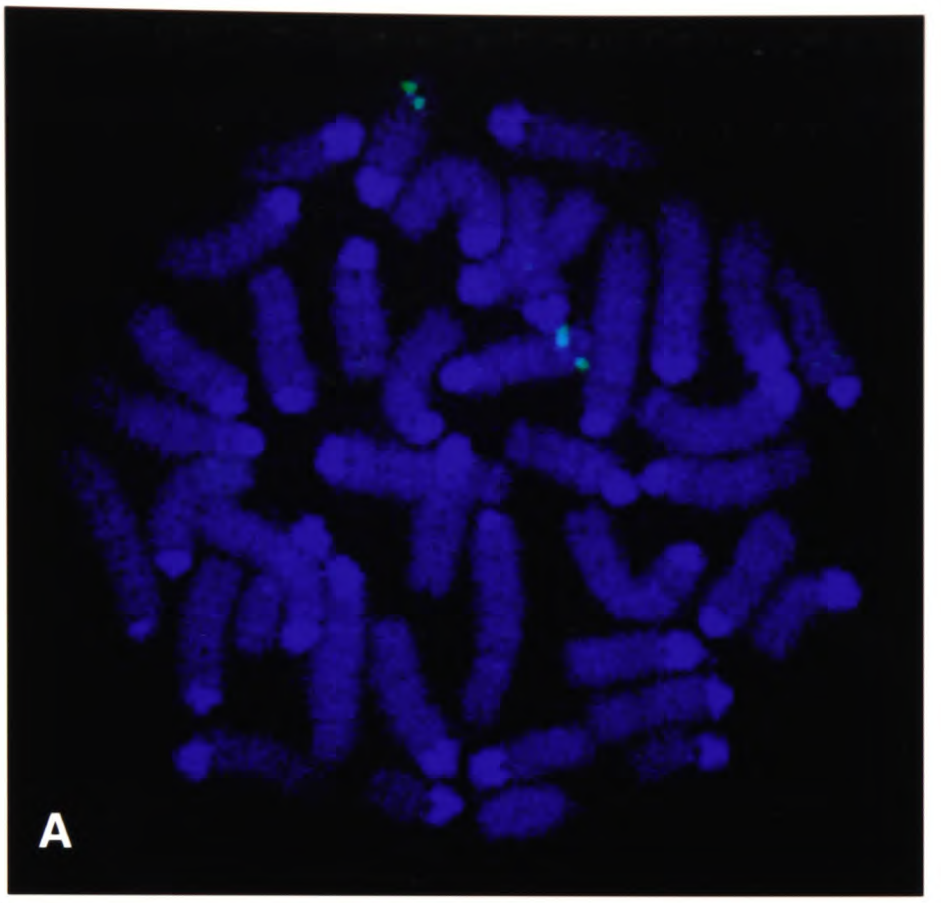


Fig 5.2 Mapping the position of integration of α globin constructs in transgenic mice using FISH. Metaphase spreads were prepared from primary cells extracted from the spleen of transgenic mice which had been treated with phenylhydrazine. Slides are hybridised with cosmid probe cGG1 (green), and counterstained with DAPI (blue).

A. Line α53 is homozygous for the GG1/GG2 construct, which maps to mouse chromosome 12, band E.

B. Line αPAC100, contains two copies of the αPAC construct, which have integrated in tandem, into mouse chromosome 4, within bands A3-A5.

	α53 A	α53 B	α PAC100
% erythroid	60	82	83
P48	56.5	53	43
cGG1	60 [100]	58	41 42.5
cRN24	N/A	N/A	41
cSamD4	23.5	28	19

Table 5.2 Replication timing of trangenes in primary cells extracted from the spleen of transgenic mice. P48 is the endogenous mouse α cluster probe (early replicating) and cSamD4 is the endogenous mouse late replicating control. Where hybridisation efficiency could be measured, it is given in square brackets alongside the percentage doublet score. Replication timing was measured twice (i.e on two separate slides) using cGG1 in line αPAC100. N/A = data not available.

The results in table 5.2 indicate that in both transgenic lines examined, replication timing of the transgene occurs early in S phase, at a similar time to the endogenous mouse α cluster. The two fixed nuclei preparations of the α53 line (A and B), which show a 22% difference in the percentage of erythroblasts of the total cell population, display a very similiar timing of replication. This may suggest that replication of the transgene occurs at the same time in myeloid and erythroid cell types, irrespective of whether the transgene is expressed or not.

Thus, the pattern of early replication seen at the normal chromosomal location appears to be maintained in both transgenic lines. The analysis, however, has only been performed in one line per construct. Further analysis of several other lines, where the position of integration is different, is needed to assess whether early replication of the transgene is position independent.

In the two lines that have been analysed here, it would seem both constructs contain the origin (or origins) of DNA replication and any distal *cis*-acting sequences which are required for early replication. Alternatively, it is possible that both constructs have integrated near a mouse early activating origin, which is responsible for the early replication of the human sequence. However, the αPAC is quite large at 150kb and is integrated in tandem as a 300kb piece. Replication timing has been measured using probes within the central portion of the αPAC, which will be present twice, 150kb apart. Signals from both copies co-localise in the interphase nuclei, to give a single

signal. The percentage doublets recorded for this signal is similar to the mouse early replicating control (table 5.2). If the replication fork moves at approximately 2kb/min (Adams 1991), it would take a replicosome generated from external mouse origin approximately two hours to replicate the double insert. It might be expected that this assay would detect the delay in replication, and therefore it seems more likely that replication is initiating within the αPAC. A comparison of replication times of the same construct integrated at different sites within the mouse genome would provide a more definitive answer.

5.3 Discussion

Chromatin analysis of the GG1/GG2 construct in erythroblasts from the α53 line, (performed previously by Gourdon *et al.* 1994), demonstrated that most of the DNase 1 hypersensitive sites present *in vivo* reform after integration into the mouse genome. These include the erythroid-specific hypersensitive sites associated with HS-40 and the α1 and α2 globin gene promoters, and all of the constitutive hypersensitive sites located throughout the construct. The hypersensitive site at the ζ gene promoter was detectable in embryonic cells, but not in adult erythroblasts, reconstituting the switching pattern seen in normal development (Gourdon *et al.* 1994). Erythroid hypersensitive sites that did not reform were those at 10kb and 8kb upstream of the ζ gene (Gourdon *et al.* 1994), although when these sites have been tested in previous constructs, they do not appear to influence α globin gene expression (Sharpe *et al.* 1993). Analysis of the CpG islands associated with the α globin gene promoters using restriction analysis, show they remain unmethylated (Gourdon *et al.* 1994). Despite the apparent reconstitution of the normal chromatin structure ascertained by DNase 1 sensitivity, the levels of α globin mRNA expression in lines containing the GG1/GG2 construct are still below the levels of the endogenous mouse α globin cluster, lack position independence, and decline throughout development. They are not significantly greater than those of smaller constructs, which contain HS-40 directly linked to the α genes (Sharpe *et al.* 1992). These data imply that although HS-40 has some activity in reconstituting the chromatin structure of the α globin cluster in transgenic mice, it is not able to reproduce the

chromatin configuration which permits α globin gene expression *in vivo*. In humans, the α globin cluster is located in a constitutively open chromatin environment in expressing and non-expressing cells. Natural deletions which remove HS-40, do not result in the loss of the erythroid hypersensitive sites at the α globin promoters, suggesting these sites form independently of HS-40. In addition, HS-40 appears to have no effect on the wider chromosomal region where the α cluster is located (Craddock *et al.* 1995, Bernet *et al.* 1995).

The focus of the work in this chapter was to assess whether the vagaries in α globin gene expression in transgenic mice were associated with a change in the normal pattern of replication in erythroid cells, which was established in the previous chapter. If the replication timing of the construct in transgenic mice differed from the normal timing of replication, it would support the mechanism that early replication is important in establishing a chromatin structure at the α globin locus, which allows the formation of protein/DNA complexes at the α globin promoters and HS-40, and provides an environment so these assemblies can interact. It has been shown that two different α globin constructs in two transgenic lines, reproduce the early pattern of replication seen in the normal chromosomal environment. This demonstrates that the lack of complete regulation of α globin expression is not associated with a disturbance in the pattern of replication. It can be inferred from these results, that early replication is not the sole, primary determinant of the permissive chromosomal structure at the α globin locus. However, as we do not know whether expression can occur when the locus is late replicating, early replication cannot be eliminated as a regulatory mechanism, which works in concert with other forms of regulation. Alternatively, early replication may be a consequence of the chromatin structure that is created at the transgene locus. Although this structure does not appear to permit fully regulated α globin gene expression, it may allow the assembly of replication factors at origins, which activate early in the cell-cycle. It is still not certain whether the presence of HS-40 contributes to early replication, like the β globin regulatory element, the β-LCR (Forrester *et al.* 1990), by either binding transcription factors which can activate replication aswell as transcription,

or by acting as an origin itself. From previous studies outlined above, it seems unlikely that HS-40 contributes to early replication by influencing chromatin structure.

The data also indicate the GG1/GG2 and αPAC constructs contain the necessary origins and *cis*-acting elements which are required for early replication. If correct nuclear targeting is required for early replication, the constructs also appear to include the relevant *cis* targeting sequences, which may be the origins themselves, rather than the telomere, as previously suggested (Higgs *et al.* 1998). As only two lines were examined, there is a possibility that replication timing of the constructs may be position dependent, and in some lines constructs may be late replicating. The mouse plasmacytoma MPC11 cell-line contains two copies of the Cγ2b gene, at different chromosomal locations. One copy, which is actively transcribed, replicates very early in S phase. The second copy, on a different chromosome, which is not transcribed, replicates late in S phase (Calza *et al.* 1984). Thus, two apparently identical genes and their flanking sequences, in the same cell, can replicate at different times during S phase. This may be dependent on their chromosomal location, but it may also be determined by their transcriptional state. Genes that are present in multiple copies at different chromosomal locations in mouse, such as the histone or rRNA genes, replicate throughout S phase (Hatton *et al.* 1988). This is consistent with the idea that a group of genes at a single chromosomal location replicate at different time from another group at a different position.

Chapter 6 The role of the α globin regulatory element (HS-40) in DNA replication.

6.1 Introduction

In chapter four it was established that the α globin cluster and surrounding chromatin, replicates early in the cell cycle in non-erythroid cells and maybe even earlier in erythroid cells. In chapter five, it was shown that the early pattern of replication seen *in vivo*, can be maintained in α globin constructs which include HS-40, when they are integrated into the mouse genome. Work described in this chapter determines whether HS-40 is required to produce the characteristic pattern of early replication.

Studies of the human β globin cluster have demonstrated the importance of the β-LCR in determining the replication timing of the locus, and where DNA replication initiates. In erythroid cells, absence of this regulatory element results in the loss of β-globin expression (Grosveld *et al.* 1987), shifts replication timing from early to late S phase (Forrester *et al.* 1990) and prevents initiation of replication from a site located 50kb away (Aladjem *et al.* 1995). The domain in which the cluster is located, also switches from DNase 1 sensitive to DNase 1 insensitive, indicating a transition from 'open' to 'closed' chromatin (Forrester *et al.* 1990).

Deletion of the α globin cluster regulatory element (HS-40) also results in the loss of α globin gene expression in erythroid cells (Higgs *et al.* 1990, Liebhaber *et al.* 1990, Hatton *et al.* 1990, Bernet *et al.* 1995). However, unlike the deletion of the β-LCR, DNase 1 sensitivity studies have shown that the long range chromatin structure of the isochore surrounding the α cluster, and local chromatin structure at the α globin promoters, remains 'open' (Craddock *et al.* 1995, Bernet *et al.* 1995).

How does the β-LCR influence the timing of DNA replication and where replication begins? It could either be through the initation of the chromatin decondensation, which allows replication *trans*-acting factors to bind origin sequences and initiate DNA synthesis. Alternatively, it may bind transcription factors which function in activating a basal replication initiation complex which is already formed at the origin. If the latter is true, it might be expected that HS-40 functions in a similar manner, as both elements bind a similar repertoire of erythroid-specific and ubiquitous factors (reviewed in Stamatoyannopoulos and Nienhuis 1994). To evaluate whether HS-40 plays a role in the replication timing of the isochore which contains the α cluster, a cell-line has been analysed (C40), from which HS-40 has been artificially removed. This is an MEL-16 hybrid line, which was constructed by removing a 1.1kb restriction fragment containing HS-40, and replacing it with a 2.7kb fragment containing a *neomycin* marker, flanked on both sides by FLP recombinase target (FRT). The *neomycin* gene was originally used as a selectable marker during the construction of the hybrid (Bernet *et al.* 1995); to ensure the presence of this marker does not influence replication timing, the replication characteristics of part of the isochore have been confirmed in a second hybrid cell-line called C10. This cell-line is a derivative of C40, where the *neomycin* gene has been removed by transient transfection of a construct containing the FLP recombinase gene (Bernet *et al.* 1995).

6.2 Methods

Fixed nuclei preparations were made from both the C40 and C10 lines as described in section 2.3.3. FISH analysis of the C40 line was performed using twelve cosmid probes which span a region of 325kb, as described in section 4.2. See also figure 4.1, which shows the physical location of these probes. As previously, each cosmid was hybridised to an individual slide. The percentage of replicated loci was determined by scoring 200 S phase nuclei. The replication timing of several cosmids was confirmed by hybridisation to a second C40 slide. To investigate the influence of the *neomycin* marker on replication timing, several cosmids were then analysed using FISH in the C10 line.

6.3 Results

Results are shown in table 6.1. Hybridisation efficiency has not been measured for many of the C10 slides, due to a lack of metaphase spreads. The reasons for the lack of metaphase spreads are the same as those outlined for the 16p monosomic EBV-lines in section 4.3.3. The data in table 6.1 have been presented graphically in fig 6.1, to show how replication timing varies with distance along the 16p chromosome. The normal MEL-16 profile has been included, to allow comparison between the normal and 'HS-40 deleted' profiles. Comparing the C40 profile with the normal profile, it can be seen that deletion of HS-40 does not have a long range effect on replication timing: the time during S phase when the region replicates appears the same. Comparison of the C40 profile to the early and late replicating controls, indicates that the whole region is still very early replicating. As for the temporal order of replication, there are no drastic differences between the C40 profile and the normal profile. The difference between the mean values of percentage doublets scored in the normal MEL-16 and the C40 line are <10% for ten out of the twelve cosmids. For one cosmid (cGG4), C40 records 11.75% fewer doublets than the normal line. This corresponds to the loss of a peak of early replication seen at around 90kb from the telomere. The other cosmid with a difference of over 10% from the normal line, is cos12, which also records a fall in doublet score (11.3%). This results in an accentuation of late replication at the 170kb region.

Early replication is also seen in the C10 line. This demonstrates that the presence of the *neomycin* gene in the C40 line does not influence replication timing. Of the four cosmids that have been examined in the C10 line, replication appears to be earlier than both the normal and C40 lines. The characteristic dip in replication timing at around 170kb is also lost. This is rather strange considering two cosmids (cSamD4 and c419C1), one of which is the late control cosmid (cSamD4), gave similar values to the normal and C40 lines. However, this line did produce slides that were difficult to score, due to difficulty in identifying S phase nuclei. As part of the FISH procedure, S phase nuclei are detected using an antibody sandwich. Firstly, slides are incubated with mouse IgG anti-BdU. This primary antibody recognises denatured DNA substituted

with BdU (S phase nuclei). This is followed by incubation with a secondary antibody, anti-mouse IgG conjugated with Texas Red. When scoring the C10 line, the Texas Red emission was very weak. Successful detection of S phase nuclei using Texas Red is dependent upon the slide drying regimen, and the length of time taken to denature the chromosomal DNA upon the slide. Despite many attempts to make slides using different conditions, and adjusting the time of denaturation, the problem of weak fluorescence still persisted. If time allowed, the C10 results would be repeated and extended to include cosmids throughout the whole of the region. However, the results obtained confirm that deletion of HS-40 does not result in a switch to late replication.

cosmid	*C40 1*	*C40 2*	*C10*
cSamD4	26.5 [95]		24
cNFG9	49 [100]		
cRA36	51 [100]		
cNFG7	58.5 [95]	60 [100]	
cGG4	64.5 [98]		
cGG2	64.5 [100]		
cRN24	66.5 [100]		79
cGG1	63.5 [95]	65.5 [100]	80
cos12	55.5 [85]		80.5
cRJ14	86 [100]		
c310H5	75 [86]		
c314G4	70.5 [90]		
c419C1	70.5 [100]	76.5 [100]	78
P48	53 [96]	52.5 [100]	64.5 [94]

Table 6.1 Each % doublet value has been obtained by scoring 200 chromosomes on an individual slide. Some cosmids have been scored on a second C40 slide (column 2). Where hybridisation efficiency has been measured, it is shown in square brackets alongside each value. Shaded boxes show control probes. The remainder are 16p specific. They have been listed so that moving down the table is equivalent to moving towards the centromere.

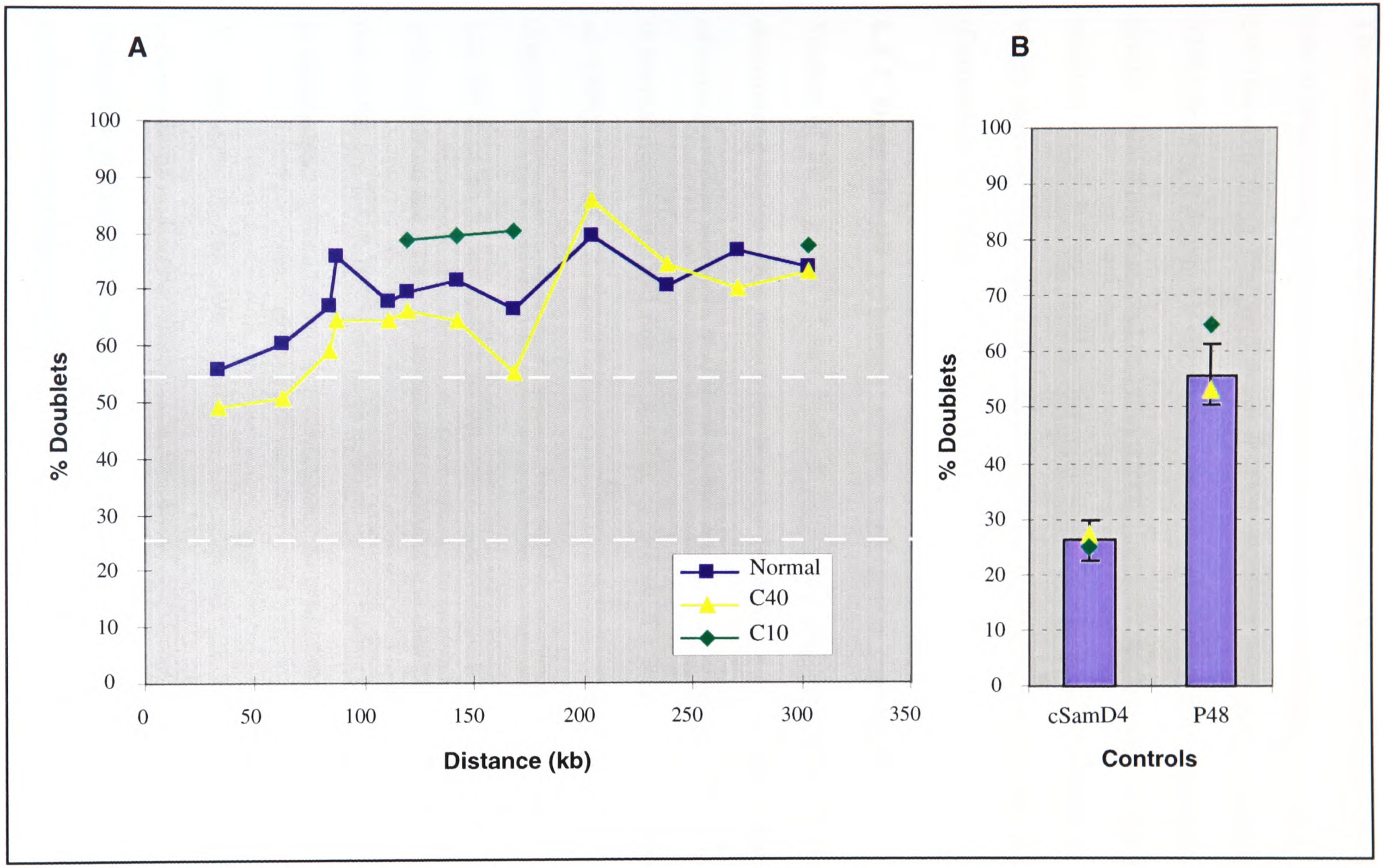


Fig 6.1A Replication profiles of the HS-40 deleted mutant lines, C40 and C10 (yellow and green respectively), compared to the normal profile (blue). The upper white dashed line indicates the mean value of the early control (P48). The lower white dashed line indicates the mean value of the late replicating control (cSamD4). On comparison with these controls, all profiles are early replicating. The C40 profile looks very similar to the normal profile. Only a few cosmids were tested in the C10 line, all of which are early replicating.

Fig 6.1B Accurate details of the mouse controls. The mean value (recorded from all MEL-16 hybrid lines) for the early (P48) and late (cSamD4) controls are given as filled columns. Measurements from the C40 and C10 lines are given individually (yellow and green respectively).

6.3 Discussion

6.3.1 HS-40 does not influence replication timing

The results described in this chapter demonstrate that HS-40 does not play an important role in determining replication timing of the isochore containing the α globin cluster. A cell-line containing a targeted deletion of HS-40 (C40), was used to examine the whole 325kb region. It displays a remarkably similar replication profile to the normal erythroid profile. Data from a second cell-line (C10), confirms that deletion of HS-40 does not result in a switch from early to late replication. This is in stark contrast to the β-LCR, which plays an important role in determining the replication timing of the β-globin locus (Forrester *et al.* 1990).

6.3.2 Does HS-40 influence where replication initiates?

Studies of the β globin cluster have demonstrated the importance of the β-LCR in determining where DNA replication initiates. Deletion of the β-LCR results in the loss of initiation from an origin 5' of the β globin gene (Aladjem *et al.* 1995), which is used in normal expressing and non-expressing cell-types (Kitsberg *et al.* 1993b, Aladjem *et al.* 1995). Instead, the region is replicated from a replication fork advancing from the 3' direction, from an unidentified initiation site beyond the cluster. However, an insertion into the β-LCR, which results in the loss of expression of β globin, does not prevent initiation from the origin 5' of the gene (Aladjem *et al.* 1995). This suggests that the role of the β-LCR in transcriptional regulation can be dissociated from the role it plays in replication.

To speculate on the role of HS-40 on replication initiation, the relatively small differences in replication timing between the normal and C40 profile could be due to changes in where DNA replication initiates. However, loss of HS-40 does not result in a dramatic, long range alteration in chromatin structure like that seen with deletion of the β-LCR: constitutive DNase 1 hypersensitive sites remain present (Bernet *et al.* 1995, Craddock *et al.* 1995). Furthermore, loss of HS-40 does not result in the loss of erythroid-specific DNase 1 hypersensitive sites, even those associated with the

promoters of the α globin-specific (ζ2, α2, α1) genes. The CpG islands found at the α globin gene promoters also remain unmethylated (Bernet *et al.* 1995, Craddock *et al.* 1995). If chromosome structure determines where initiation occurs (Coverley and Laskey 1994), it seems unlikely that deletion of HS-40 will have any effect on replication initiation.

In chapter four, it was shown that the 325kb region including the α globin cluster replicates earlier in S phase in erythroid cells (MEL-16 hybrids) compared with non-erythroid cells (EBV-transformed lymphoblasts). It was suggested that this shift to earlier replication may be correlated to expression of the α globin genes. However, in this chapter, it has been demonstrated that severe downregulation of α globin gene expression, due to the deletion of HS-40, is not accompanied by a shift to later replication. The discernible differences in chromatin structure that persist in the absence of HS-40 between erythroid and non-erythroid cells, may help to explain this phenomenon. It is conceivable that in erythroid cells (including the HS-40 deleted line), tissue-specific hypersensitive sites which include the α globin gene promoters, function as origins of replication. If more origins are utilised in erythroid versus non-erythroid cells, replicon size would decrease, which would result in the region replicating earlier in S phase. This phenomenon of utilising additional origins in expressing cell types may be occurring at the mouse immunoglobulin heavy chain locus (Brown *et al.* 1987, Michaelson *et al.* 1997). However, FISH analysis produces the same pattern of replication throughout the 325kb region in erythroid and non-erythroid cells (fig 4.4A), which supports the view that both cell-types utilise the same origins, which fire slightly earlier in erythroid cells.

Some of the constitutive hypersensitive sites throughout the isochore containing the α cluster, remain unaccounted for in terms of function (Flint *et al.* 1997a). Hypersensitive sites are often found where proteins bind DNA, which can be ascertained by footprinting analysis. ORC footprinting patterns have been demonstrated in budding yeast (Diffley and Cocker 1992, Diffley *et al.* 1994). In humans, two hypersensitive

sites upstream of the *ppv1* gene, found at chromosomal band 19p13.3, colocalise to a region of DNA where replication initiates. Footprinting shows this region binds protein, which may include human ORC (Dimitrova *et al.* 1996). Moreover, a constitutive DNase 1 hypersensitive site is associated with the replication origin found in the chicken α globin cluster, upstream of the α^{π} gene (Berberich and Leffak 1993). It may be that one or more of the constitutive hypersensitive sites found in the isochore containing the α cluster, are where human ORC binds and DNA replication initiates. It would be interesting to know where the origins are located within the isochore, and if hypersensitive sites are where replication initiates. If they are, do all sites function as origins, or only a subset? Replicons are reported to be between 50-500kb in size, which suggests that only a fraction of the sites would be used. It could be envisaged that sites that are used as origins differ between individual cells. Techniques have been developed for mapping initiation sites, such as those used to identify the β-globin origin, namely leading strand analysis and nascent strand analysis (Kitsberg *et al.* 1993b, Aladjem *et al.* 1995).

6.3.3 Replication timing and chromosome structure

The β-LCR fits the classical definition of an LCR, by its ability to modulate chromatin structure (Dillon and Grosveld 1994, Kamakaka 1997). In non-erythroid cells, the β-globin cluster is found within a DNase 1 insensitive, closed chromatin domain. In erythroid cells, mediated by the β-LCR, the domain becomes DNase 1 sensitive (Forrester *et al.* 1990), suggesting that *cis*-acting regulatory elements may become more accessible for transcription factors to bind their cognate elements. Once bound, the protein/DNA complexes assembled at the β-LCR and the β-globin promoters can interact to upregulate β-globin expression. In contrast, HS-40 does not require the ability to open a repressive chromatin structure, as the α globin cluster is located in an isochore of constitutively open chromatin (Vyas *et al.* 1992). However, HS-40 does play a pivotal role in potentiating transcription (Higgs *et al.* 1990, Bernet *et al.* 1995). When human α globin gene expression was measured in two C40 hybrid clones and one C10 hybrid clone where HS-40 has been removed, basal levels of both α1 and α2 mRNA

expression were severely downregulated (<3% of levels found in the control lines). More importantly, these basal expression levels could not be raised by induction (Bernet *et al.* 1995). Hence, HS-40 is more likely to function by upregulating α globin gene expression by stimulating protein/DNA complexes which are already assembled at the α promoters. Therefore, HS-40 is more typical of a strong, inducible enhancer (Bernet *et al.* 1995, Kamakaka 1997, Higgs *et al.* 1998).

The differing roles of HS-40 and the β-LCR in influencing the degree of chromosome condensation shows a tight correlation with replication timing:

α globin: *normal*	*α globin:* *loss of HS-40*	*β globin:* *normal*	*β globin:* *loss of β-LCR*
early replicating open chromatin expression	early replicating open chromatin loss of expression	early replicating open chromatin expression	late replicating closed chromatin loss of expression

Early replication is found throughout a region of open chromatin structure. Likewise, late replication is found throughout a condensed chromatin structure. Whether early replication creates an open chromatin structure, or whether an open chromatin structure allows early replication, has yet to be determined.

6.3.4 Do transcription factors play a direct role in replication?

The β-LCR consists of five DNase 1 hypersensitive sites. Each site contains multiple haematopoietic-restricted binding motifs (GATA, NF-E2) aswell as motifs (CCACC) for ubiquitously expressed factors (reviewed in Stamatoyannopoulos and Nienhuis 1994). Binding of protein factors at the β-LCR may relieve chromatin-mediated repression at the β globin origin to allow replication to initiate there. As the same origin is utilised in expressing and non-expressing cells (Kitsberg *et al.* 1993b), origin specification cannot be determined through the binding of haemopoietic factors, and is more likely to be due to the binding of ubiquitous factors. Some of the hypersensitive

sites within the β-LCR are present in non-erythroid cells (Forrester *et al.* 1987, Forrester *et al.* 1989, Dhar *et al.* 1989); factors which bind at these sites may be necessary for origin activation. This is not clear yet, as although the β-LCR has been shown to be essential for determining origin utilisation in erythroid cells (Aladjem *et al.* 1995), it is not known whether the same is true in non-erythroid cells.

In addition to chromatin remodelling at origins, transcription factors which bind at the β-LCR may play a more direct role in replication. In viruses, factors with helicase activity stimulate local DNA unwinding at origins (reviewed in DePamphilis 1993). Transcription factors may have dual functionality, activating a basal replication initiation complex in addition to stimulating a basal transcription initiation complex. In several viruses, *cis*-acting mutations which abolish protein factor binding, can affect transcription and replication (reviewed in DePamphilis 1993). In budding yeast, transcription factors such as ABF-1 can stimulate replication and transcription; both processes may be mediated via the same activation domain (Marahens and Stillman 1992). Alternatively, transcription factors could act as docking proteins, delivering replication proteins to origins, thus facilitating the assembly of the initiation complex (Mohr *et al.* 1990). Many of the characterised mammalian origins are associated with transcriptional promoters (table 1.1 this thesis), also supporting the view that transcription factors play a role in replication. Further experimentation is required to localise the sequence(s) within the β-LCR which are necessary for origin activation, and to identify critical sequences within the origin itself. Characterisation of the proteins which bind both regions is also required to see how these proteins function and interact.

From the results shown in this chapter, HS-40 does not appear to specify where DNA replication initiates, or when origins fire, further highlighting the contrast in function of the two regulatory elements. Although HS-40 shares common binding motifs with the β-LCR, HS-40 is much smaller (one hypersensitive site as opposed to five of the β-LCR). Variations in the abundance of factors which bind the two elements may explain the different roles in replication. On the other hand, the difference in extents to which

origins of replication require additional protein factors may reflect the contrasting chromatin environments between the α and β clusters. The β globin origin may require auxiliary factors to remove a repressive chromatin structure to allow replication factors to bind. Although origins have not been identified in the α globin cluster, the majority of the region has a constitutive DNase 1 sensitive chromatin structure (Vyas *et al.* 1992). Origins in this region may be directly accessible to replication factors, obviating the need for additional factors to remove chromatin-mediated repression.

6.3.5 Summary

The α globin cluster regulatory element, HS-40, does not influence the local replication timing of the α globin genes, nor does it influence replication timing of the broader 325kb region where the α globin genes are located. Preliminary results described in this chapter also indicate that HS-40 does not influence where DNA replication initiates. These collective findings, further illustrate the differences in function between HS-40 and the β globin regulatory element, the β-LCR .

It has been shown that loss of gene expression need not be accompanied by a switch to late replication. Hence transcriptional activity and replication timing can be dissociated. However, the timing at which a region replicates still remains closely associated with chromosome structure. Early replication accompanies an 'open' chromatin structure and late replication is found in areas of repressive chromosome structure.

Chapter 7 The ZF deletion: relationship between DNA replication and α globin gene expression

7.1 Introduction

7.1.1 Mutations giving rise to α-thalassemia

Until recently, mutations that perturb α globin gene expression have typically fallen into two categories: deletions that remove one or both of the fetal-adult structural genes (α2, α1), or deletions that remove the α globin regulatory element (HS-40). Removal of one structural gene (-α) from a single allele of chromosome 16 results in a mild α-thalassemia phenotype, whereas removal of HS-40 or two structural genes (--), produces a more severe α-thalassemia phenotype (reviewed in Higgs *et al.* 1989).

7.1.2 The ZF deletion

In the previous chapter, it was shown that loss of expression of both α genes, due to the deletion of HS-40, is not accompanied by a change in replication timing. This chapter analyses the replication timing of an interspecific hybrid cell-line from a patient with a novel deletion (ZF), originally described by Indrak *et al.* (1993). This patient has severe α-thalassemia, although the deletion removes only one of the structural genes (α1). This discrepancy between genotype $(-\alpha)^{ZF}$ and phenotype (--) has been investigated further (Barbour *et al.* in preparation). The deletion is 18.3kb long, which removes the α1 gene and approximately 15kb of the region downstream (3') of the gene (fig 7.1) and results in the transcriptional silencing of the remaining α2 gene, to produce the phenotype normally associated with downregulation of both α genes.

DNA sequence analysis of genomic DNA from the patient demonstrated that the α2 promoter and gene are normal, indicating that lack of expression was not due to a point mutation or deletion. However, further experiments found that the α2 gene and its promoter are no longer accessible to DNase 1 and restriction endonucleases,

demonstrating a change in chromatin templating (Barbour *et al.* in preparation). In addition, the CpG island associated with the α2 gene has become fully methylated (Barbour *et al.* in preparation). The effect of the ZF deletion appears to be very localised, as the CpG island associated with the ζ gene remains unmethylated, and the corresponding DNase hypersensitive site is still present. This is also true of HS-40, which lies approximately 50kb upstream of the deletion (D.R. Higgs, pers. comm).

An EBV-transformed lymphoblastoid cell-line established from the patient with the ZF deletion and an interspecific hybrid (MEL-16) containing the ZF chromosome, both maintain the changes in methylation and DNase 1 hypersensitivity at the α2 gene and promoter, which suggests the gene silencing associated with these changes is stably inherited (Barbour *et al.* in preparation).

The mechanism of transcriptional silencing arising from this deletion, is clearly different from the mechanism due to deletion of HS-40, which does not prevent the formation of hypersensitive sites or result in the methylation of CpG islands at the promoters of the α globin genes (Bernet *et al.* 1995, Craddock *et al.* 1995). This deletion mutant therefore provides an opportunity to analyse the pattern and timing of replication of the silenced α2 gene and the surrounding chromatin, to see if a perturbation in replication is required to produce the other epigenetic changes which lead to transcriptional repression.

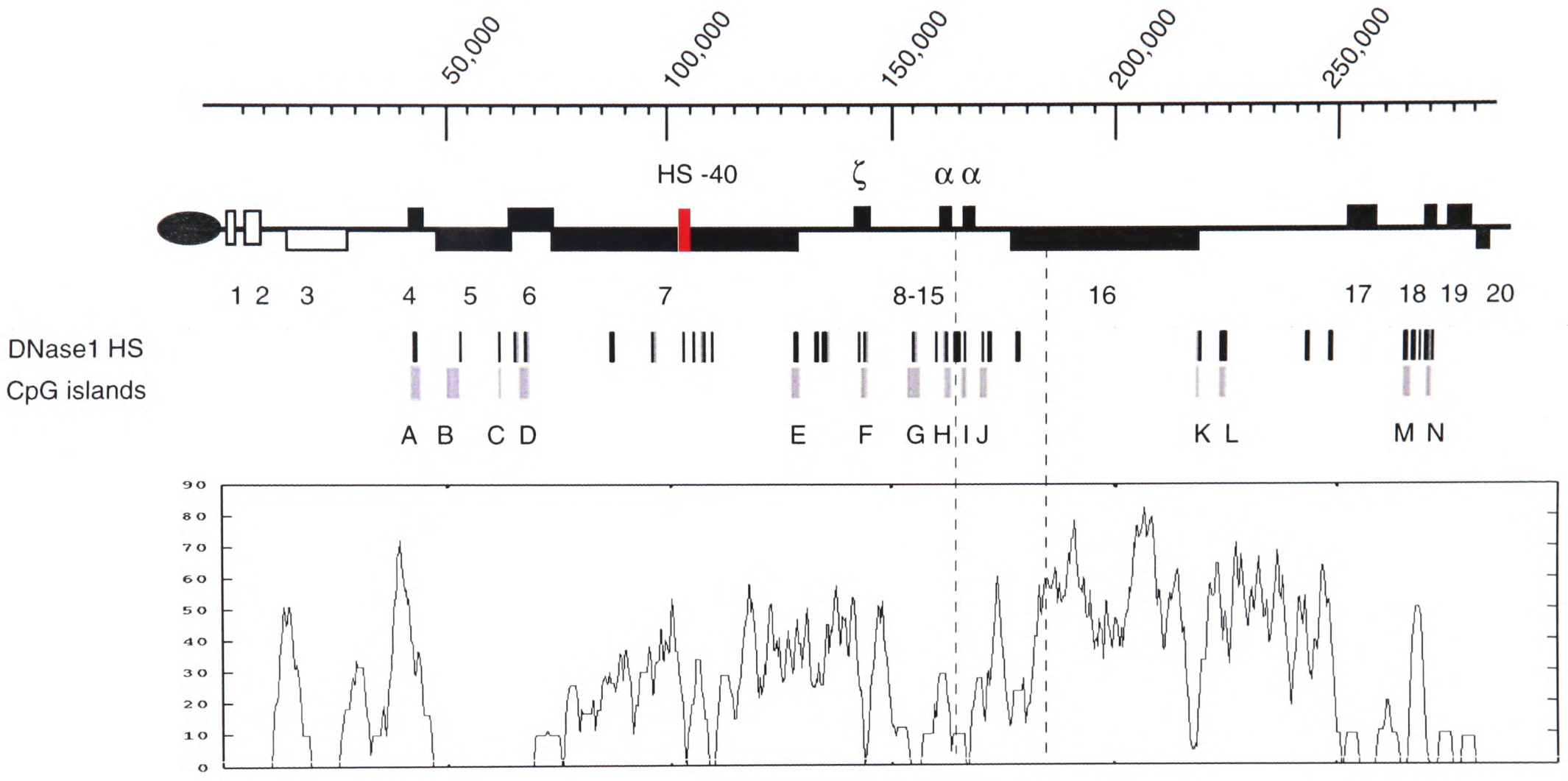


Fig 7.1 The ZF deletion. Scale, at the top of the diagram, is in base pairs. Below is a schematic of the terminal 325kb of human chromosome 16p. Telomeric repeats (TTAGGG)n are represented as an oval. Genes and pseudogenes are depicted as filled and hollow boxes respectively. These are numbered 1-20 (for clarity, not all genes found in the region have been included). Those above the line are transcribed towards the centromere; those below the line are transcribed towards the telomere; those on the line have no orientation. HS -40, the α globin regulatory element is shown in red, and the α structural genes are labelled. Positions of DNase 1 hypersensitive sites and CpG islands (A-N) are shown. The graph shows the percentage of Alu repeats within the region. The ZF deletion removes the segment between the two vertical dashed lines, which includes the second α gene (α1) and part of gene 16. This juxtaposes the remaining α gene (α2) against a region which is very Alu dense.

7.1.3 A heterochromatin-like region downstream of the α genes

Analysis of the 38kb region downstream of the α genes has shown that it contains a high density of Alu repeats, with no CpG islands or DNase 1 hypersensitive sites (fig 7.1 and 7.2, courtesy of D.R. Higgs). This is uncharacteristic of the 325kb region, which is typically euchromatic in structure (GC rich, high density of unmethylated CpG islands and genes, DNase 1 hypersensitive). Thus, this 38kb region resembles a heterochromatic structure located within a euchromatic environment.

Work has shown this Alu-dense region contains a gene, transcribed in the reverse direction to the α genes, shown in fig 7.1 (Flint *et al.* 1997a, C. Tufarelli, pers. comm.). This gene is approximately 42kb in size, containing at least 9 exons spread throughout this heterochromatic region (fig 7.2). It has been called the +80 gene, due to the presence of a DNase 1 hypersensitive site at its 5' end, which is 80kb downstream of the start of the ζ gene. The function of the gene remains unknown, although it is highly conserved throughout evolution (homologues are present in yeast, plant, worm, fly and mouse); it also contains a putative RNA binding domain (C.Tufarelli, pers. comm.). The ZF deletion, in addition to removing the α1 and θ1 genes, removes the final 3 exons of the +80 gene and creates a new junction between the remaining α2 gene and the putative heterochromatic region (fig 7.2).

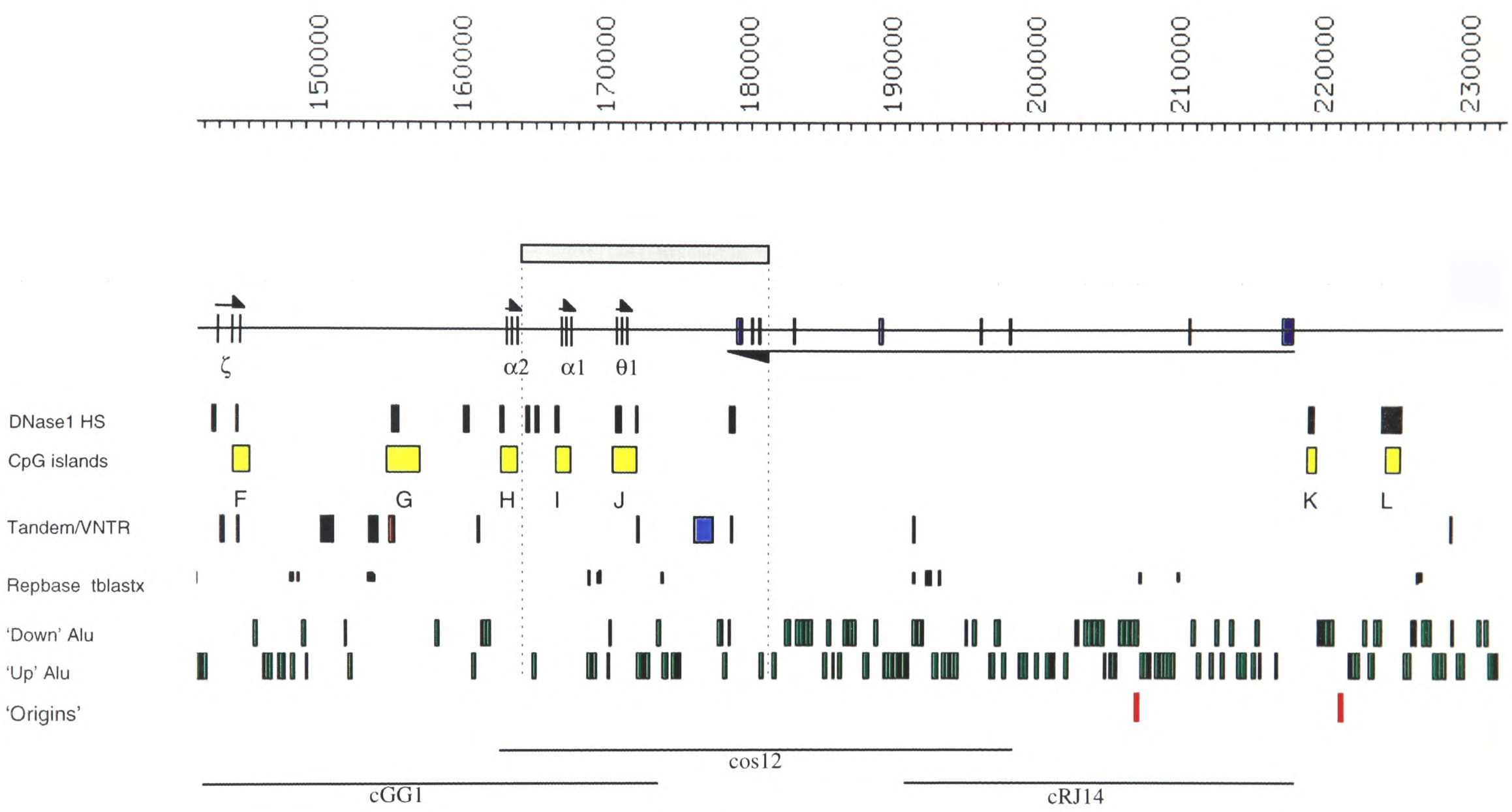


Fig 7.2 Magnification of the ZF deletion. Vertical lines indicate exons of genes, and arrows indicate the direction of transcription. Alpha globin genes are in black and are individually labelled. Gene 16, which is the +80 gene, is in blue. The stippled bar shows the extent of the ZF deletion, which removes two genes completely and the final three exons of the +80 gene. Tandem/ variable number tandem repeats (VNTRs) are shown, with the 3' hypervariable region (HVR) highlighted in blue. Alu repeats have been classified as 'down' (in the direction of the +80 gene) or 'up' (in the same orientation as the α genes). Exons of the +80 gene are scattered throughout a high density of Alu repeats. Two matches to the putative origin consensus sequence are found within the whole region. One is within the Alu dense region, in an intron of the +80 gene. The other is close to the CpG island (K) found at the promoter of the +80 gene. Recombinant clones used for FISH analysis are shown.

7.1.4 Further examination of the replication characteristics of the heterochromatin-like region

Two cosmids span the heterochromatin-like region 3' to the α genes: cos12 and cRJ14 (fig 7.2). The cosmid cos12 includes the genes α2, α1, θ1 and approximately 15kb of the Alu-dense region. The cosmid cRJ14 partially overlaps cos12, spanning 31kb of the total 38kb of repetitive region (fig 7.2). In chapter 4, it was shown in the normal MEL-16 lines, both cosmids are early replicating, relative to the early and late replicating controls (fig 4.4). However, comparing the replication timing of individual cosmids to each other along the 325kb region, cos 12 is one of the latest replicating cosmids (mean of 66% doublets). Curiously, cRJ14 is one of the earliest (mean of 80% doublets). During the scoring of slides, it was observed that when the cos12 region replicates, the hybridisation signals remain closely paired in many nuclei. This suggests more time is taken for the sister chromatids to separate than for other regions throughout the 325kb region. This observation is not seen with cRJ14; signals appear separated in the majority of nuclei. In this chapter, the replication timing of this putative heterochromatic region has been examined in closer detail. In order to identify which part of cos12 is the most late replicating, and to narrow down the region where signals remain closely paired, the cosmid has been divided into three parts (fig 7.3). Each part has been hybridised individually and results compared with that of the intact cosmid.

7.1.5 Replication timing of the ZF fragment in transgenic mice

If transcriptional silencing of the α2 gene is due to juxtaposition of the gene against the putative heterochromatic block, it should be possible to recreate the phenomenon in transgenic mice. A construct has been made which includes HS-40 ligated to the α2 gene and 10kb of the region downstream of the ZF breakpoint (pZFBam10). Two transgenic lines have been established, and expression of the α2 gene has been measured (Barbour *et al.* in preparation), shown in table 7.1. Compared to expression of other α constructs that have been tested which contain HS-40 linked to the α genes, levels in these lines are lower in embryos and significantly reduced in adults, indicating at least partial silencing of the α2 gene (Barbour *et al.* in preparation). Analysis of the chromatin stucture at the α2 promoter shows it is DNase 1 insensitive, recreating the

effect seen *in vivo*. However, methylation of the CpG island associated with the promoter is not reproduced. In this chapter, I have examined the replication timing of the ZF constructs in the two transgenic lines, and compared the results with the replication timing of the other α constructs in chapter 5.

Line (number studied)	Copy number of transgene	Number studied	Expression in embryos (range)	Number studied	Expression in adults (range)
α107	2	5	11.7 (9-14)	10	0.86 (0.06-1.9)
α108	2	3	10.5 (6.4-16.1)	7	0.98 (0.13-4.5)

Table 7.1 Levels of α2 mRNA expression per copy of construct in transgenic mice. Embryonic expression is calculated as the percentage of (ζH + αH/ζM + αM). Analysis was performed at 11.5d. Adult expression is calculated as the percentage of αH/αM .

7.2 Methods

7.2.1 Replication timing of the ZF mutant hybrid

Replication timing of the interspecific MEL-human hybrid, containing the mutant ZF chromosome was analysed using FISH as described in section 5.2.

7.2.2 Replication of the region spanned by cos12

To divide cos12 into three roughly equal parts, approximately 50μg of cosmid was digested with the restriction endonuclease *EcoR*1 (Sambrook *et al.* 1989). This enzyme cuts the cosmid four times, generating fragments of approximately 9kb; 8.6kb; 16kb and 2kb (fig 7.3). Fragments were resolved by 1% agarose gel electrophoresis. The three larger fragments were individually excised from the gel and purified using electro-elution (Sambrook *et al.* 1989). The 9kb and 8.6kb fragment ran very close to each other on the gel, so to confirm that the two fragments had been successfully purified and to demonstrate the position of each fragment within cos12, a 1μg aliquot of each DNA sample was digested with *Hpa1* (Sambrook *et al.* 1989). Agarose gel electrophoresis demonstrated that the larger 9kb *EcoR1* fragment was cut into two

smaller fragments, approximately 5 and 4kb. Digestion of the smaller 8.6kb *EcoR1* fragment with *Hpa1* produced a fragment of very similar size to the undigested *EcoR1* fragment (data not shown). This confirmed that the 9kb fragment was the 5' end and the slightly smaller 8.6kb fragment was the central portion of cos12 (fig 6.3). 2μg of each sample (16kb; 9kb; 8.6kb) was then used in a nick translation reaction to produce DIG labelled probes for FISH (method given in section 2.3.1). Four slides were made from the normal MEL-16 hybrid JY5-4. One slide was hybridised with labelled cos12; the other three slides were hybridised with the three parts of cos12. FISH was performed as described in section 2.3.6-2.3.8.

7.2.3 Replication timing of the pZFBam10 construct in transgenic mice

FISH analysis was performed on interphase nuclei prepared from primary cells extracted from the spleens of transgenic mice, as outlined in 5.2.1. The pZFBam10 DNA that was used to generate the transgenic mice, was labelled and used as a FISH probe. Slides were also probed with the early (P48) and late (cSamD4) replicating mouse controls.

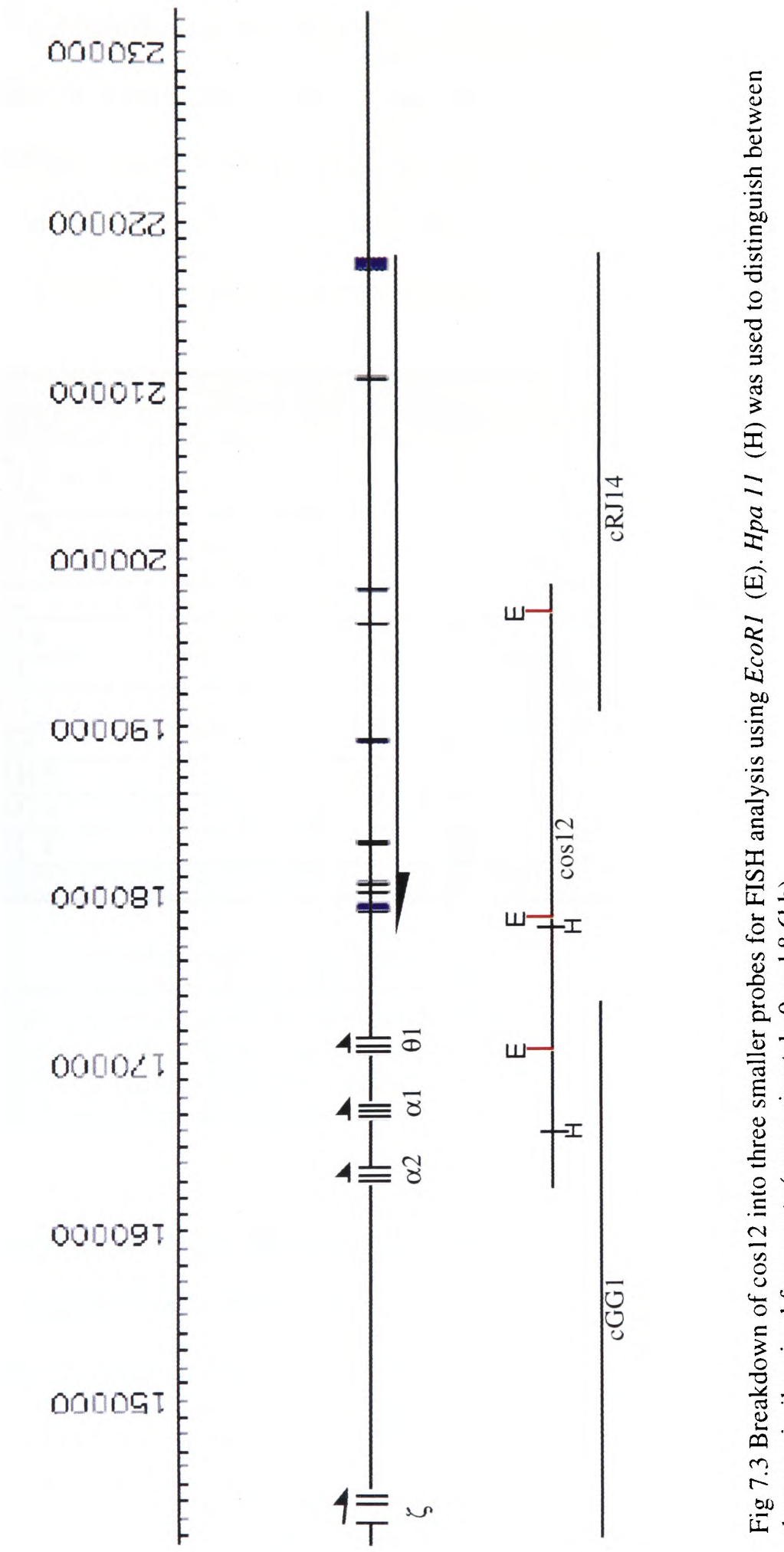


Fig 7.3 Breakdown of cos12 into three smaller probes for FISH analysis using *EcoR1* (E). *Hpa 11* (H) was used to distinguish between the two similar sized fragments (approximately 9 and 8.6kb).

7.3 Results

7.3.1 Replication timing of the ZF mutant hybrid

Results obtained for each cosmid are shown in table 7.1., and presented graphically in fig 7.4. Comparison of the ZF profile with the normal MEL-16 profile reveals no large changes in replication timing throughout the 325kb region, which remains early replicating compared to the early and late controls (fig 7.4). This was perhaps an expected result, considering the deletion only produces a localised effect of transcriptional silencing over a few kilobases.

cosmid	*ZF % doublets*
cSamD4	26.5 [100]
cNFG9	46.5 [95]
cRA36	61 [100]
cNFG7	62 [100]
cGG4	67 [100]
cGG2	67 [100]
cRN24	67.5 [95]
cGG1	64 [100]
cos12	spans the deleted region
cRJ14	85.5 [100]
c310H5	71 [90]
c314G4	73 [98]
c419C1	77 [100]
P48	58.5 [93]

Table 7.1 % doublets scored for each cosmid in the MEL-16 hybrid line containing the ZF mutant 16 chromosome. Shaded boxes show the mouse early and late replicating controls. The remaining cosmids are human 16p specific. They are listed in the order in which they map to the chromosome, the most telomeric at the top of the table. No data is given for cos12. This cosmid includes all of the region that is deleted in this mutant, therefore it was not used in FISH analysis. Hybridisation efficiency has been recorded for each slide, and is shown in square brackets alongside the % doublet score.

Considering the smaller differences in replication timing within the 325kb region, each cosmid scores within 10% of the mean value obtained in the normal line. This suggests that the ZF deletion causes no perturbation of replication timing throughout any of the region. The subsequence containing the silenced α2 gene is within cosmids cGG1 and cos12 (fig 7.2). Cosmid cos12 was not hybridised, as approximately 50% of this cosmid is not present in the ZF mutant due to the deletion (fig 7.2). This reduces the hybridisation signal (see section 7.3.2). However, cGG1 gives a percentage doublet figure of 64%, which is early replicating compared to the controls, and consistent with

the mean cGG1 value in the normal profile (fig 7.4). Therefore, despite transcriptional silencing, the α2 gene remains early replicating.

7.3.2 Replication timing of the region spanned by cos12

Table 7.2 shows the values obtained for the three parts of cos12 when hybridised to the normal MEL-16 hybrid line JY5-4.

probe	cos12 (whole)	5'	central	3'	mean of 3 parts
% doublets	53 [100]	66 [100]	51 [95]	50 [85]	55.7

Table 7.2 % doublets scored for the three different parts of cos12 in the normal MEL-16 hybrid line JY5-4. The first column gives the value obtained for the whole cosmid, the next three columns give the values for the individual parts. The final column gives the mean of the scores obtained with the three parts. Hybridisation efficiency is given in square brackets.

The percentage doublet value obtained for the slide hybridised with intact cos12 (53%) is slightly lower than the value obtained on a previous JY5-4 slide (60.5% from table 4.2); it is also lower than the mean value for this cosmid, derived from analysis of several normal MEL-16 lines (66% from table 4.2). However, all 4 slides used for this analysis were made, hybridised and detected at the same time. They have been subjected to identical treatment, therefore they should be accurate as possible in determining the replication timing of one probe relative to the other. Assuming each portion of cos12 contributes equally to the score obtained for the whole cosmid, then the mean of the value obtained for the three parts should be equal to cos 12 value. The values are not significantly different (2.7%, table 7.2).

The 5' probe includes the α1 and α2 genes; the central probe includes the θ1 gene and the 3' probe includes the final six exons of the +80 gene (fig 7.3). When scoring each slide, the 5' probe gave double hybridisation signals that were easy to score and not closely opposed, suggesting that this region pulls apart when it replicates. The middle and 3' probes gave double signals which were more difficult to score, as they remained very close. In addition, the 3' probe gave a poorer hybridisation efficiency (85%, table 7.2). This probe contains less unique sequence and more repetitive elements than the

other two probes. It was prehybridised with human and mouse Cot-1 (as are all probes) which reduced the background 'noise' sufficiently, but also reduced the signal. This resulted in what appears to be a poorer hybridisation efficiency. However, it is not poorer hybridisation efficiency, but weak signal, which is undetectable on some chromosomes.

These results show that the later replicating part of cos12 lies within its 3'end. When this region replicates, it also seems that the replicated chromatids remain in close contact for an extended period of time, compared with the other probes spanning the 325kb region. This may suggest that there is some sort of chromosome or nuclear structure here, which prevents the rapid dissociation of sister chromatids. Alternatively, it could indicate the replicosome temporarily stalls in this region. This is not evident for the putative heterochromatic region spanned by cRJ14.

The percentage doublet scores of the individual parts of cos12 also suggest that a replication fork is proceeding from the 5' to the 3' end of the cosmid. This is supported by the replication profile produced from the cosmids: replication of cGG1 (5' of cos12) occurs before cos12 (table 4.2). Cosmid cRJ14, which overlaps cos12 at its 3' end (fig 7.2) is also earlier replicating (table 4.2). One close match (1bp mismatch) to the putative origin consensus sequence (Dobbs *et al.* 1994) is found within the 3' half of cRJ14, within the Alu-dense region (fig 7.2). Another match lies close to the CpG island and hypersensitive site at the promoter of the +80 gene (denoted K in fig 7.2), just 3' to the proximal end of cRJ14 . If one or both of these sites are *bona fide* DNA replication origins, then a replication fork is travelling from the 3' end of CRJ14, in the opposite direction to the fork travelling through cos12. It is proposed that these two convergent replication forks fuse somewhere in the region spanned by cos12, that is not present in either cRJ14 or cGG1. This would explain why the region is later replicating than the sequences flanking it, and it would also explain why doublet signals do not separate easily. This region may be removed by the ZF deletion (fig 7.2).

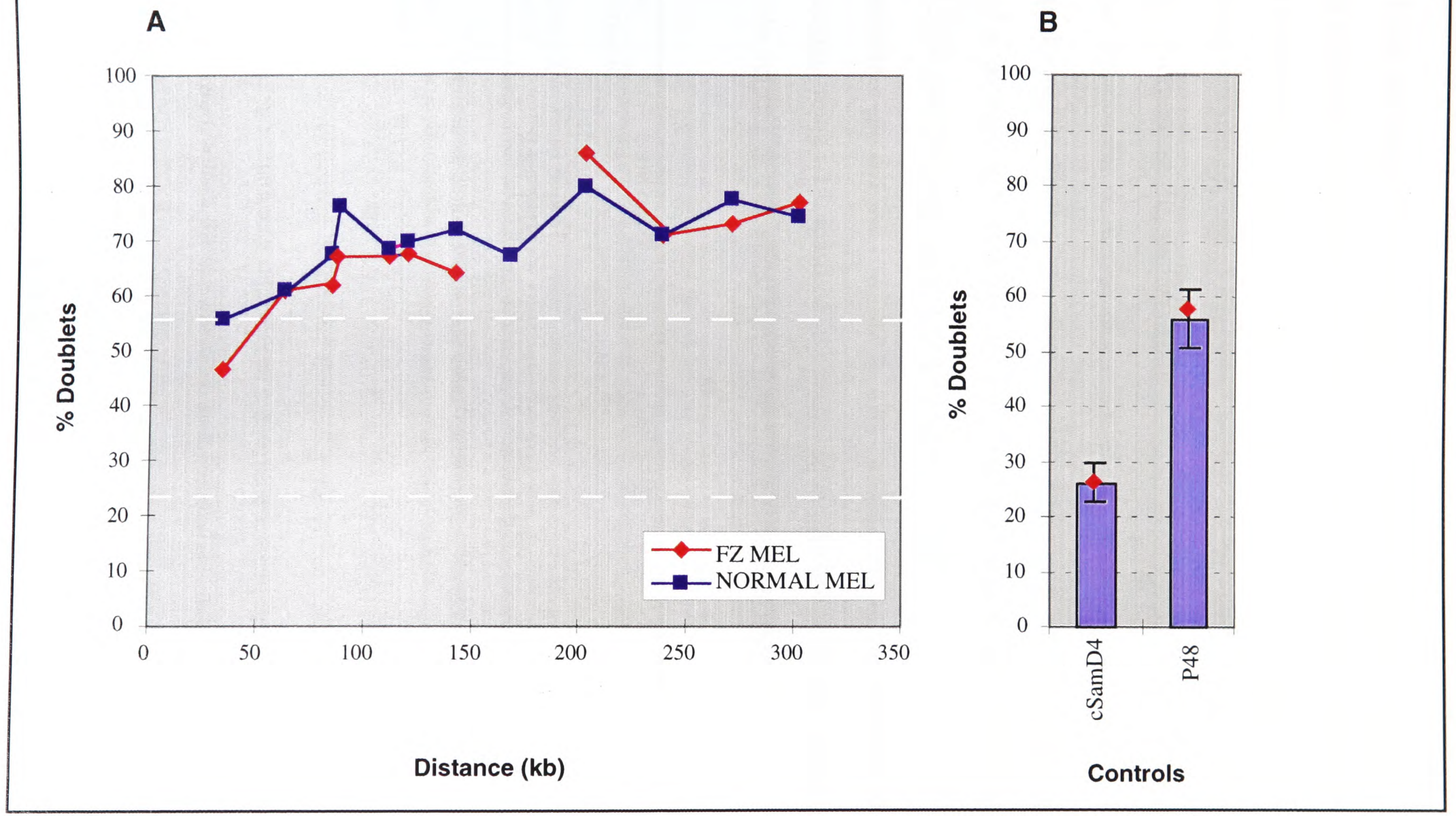


Fig 7.4A Replication pattern of the ZF MEL hybrid line (red) compared with the pattern produced in the normal MEL 16 hybrid line (blue). The cosmid value that is absent from the ZF MEL line (cos12) at approximately 170kb from the telomere corresponds to the position of the ZF deletion. White dashed lines indicate the mean % doublet scores for the early (top line) and late (bottom line) control cosmids.
Fig 7.4B Replication timing of the control cosmids. The mean value for each control is represented as a column. Values for the controls measured in the ZF MEL hybrid are shown in red .

7.3.3 Replication timing of the ZF fragment in transgenic cells

Line α107 and line α108 each contain 2 copies of the pZFBam10 construct (determined by Southern blot analysis, J. Sharpe, pers. comm). In both cases, copies have integrated in tandem, shown by restriction digestion analysis (J. Sharpe, pers. comm). FISH has previously been performed on metaphase spread preparations to provide a rough estimation of where in the mouse genome these transgenes have integrated (the level of resolution is such that multiple copies of the transgene appear as a single signal on metaphase spreads and in interphase nuclei). In line α107, the site of integration is mouse chromosome 10, within bands B1-4. In line α108, the site of integration is on mouse chromosome 8, within bands A1-A3 (V. Buckle, pers. comm). When FISH was performed in this thesis to measure replication timing and metaphase spreads were scored to measure hybridisation efficiency, the same results were produced (shown in fig 7.5). Hybridisation efficiency could not be calculated for each slide, as the number of metaphase spreads was low (due to a low mitotic index). As described in chapter 5, replication timing was measured in primary cells enriched with erythroblasts, extracted from the spleens of transgenic mice. Results are shown in table 7.3.

Probe	Line α107	Line α108
% erythroid	80	92
P48	53	54
pZFBam10	30.5 [95]	28.5
cSamD4	24.5 [100]	25.5

Table 7.3 Replication timing of the pZFBam10 construct in transgenic mice. Figures are percentage doublet scores. Hybridisation efficiency is given in square brackets. P48 and cSamD4 are the early and late replicating controls respectively.

In both lines, replication timing of the transgene is very similar to the late replicating control. This is remarkably different to the replication timing of the DNA fragment in its normal location within the ZF chromosome (table 7.1), when it replicates early within the cell-cycle (even earlier than the mouse endogenous early replicating control). These results also contrast to the results given in chapter 5, where other α constructs were found to replicate early in transgenic mice. The pZFBam10 construct also appears to replicate at approximately the same time in each line, suggesting that the time at which

the fragment replicates may be independent of the site of integration into the mouse genome.

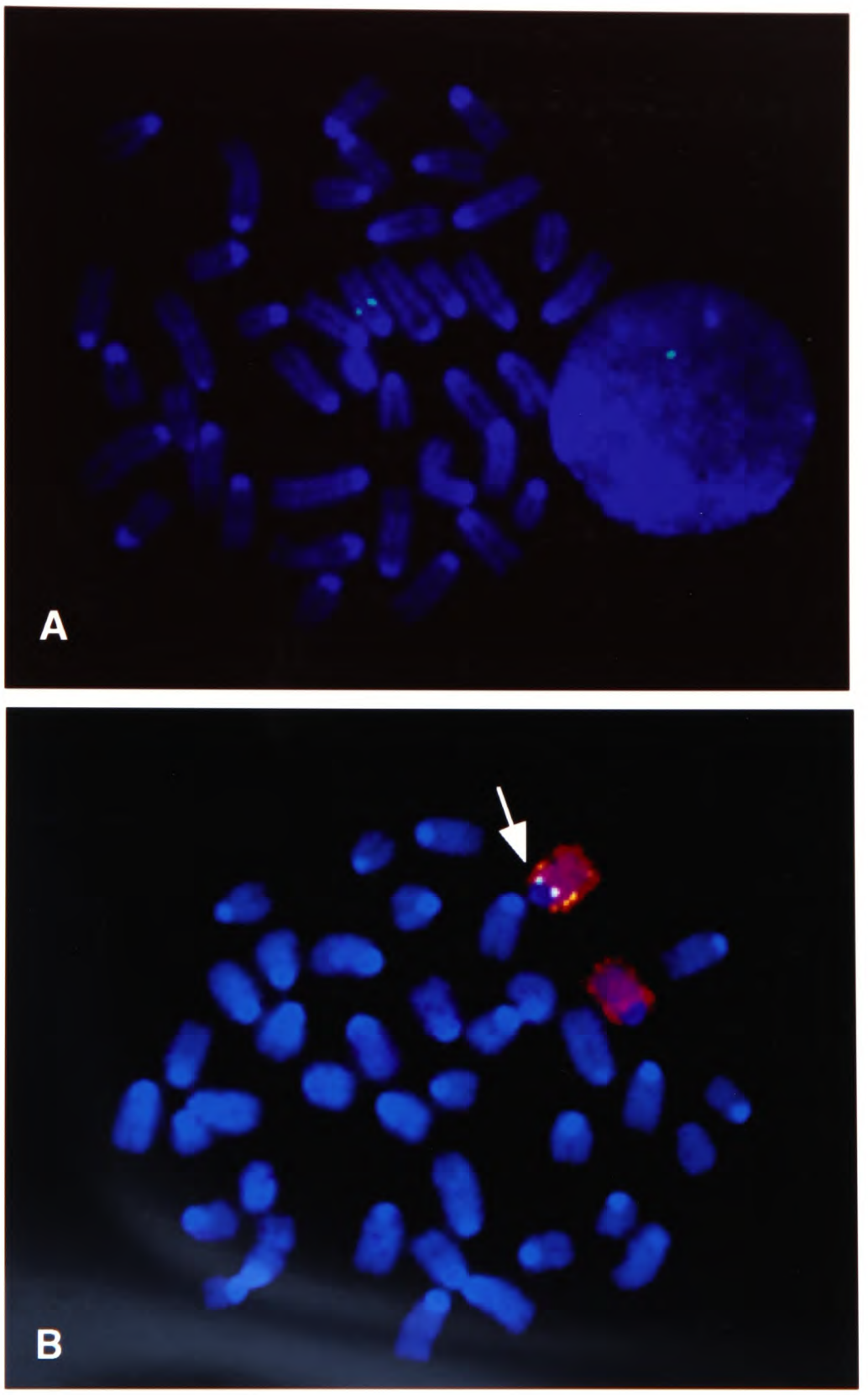


Fig 7.5 Mapping the position of integration of the pZFBam10 construct in transgenic mice using FISH. Metaphase spreads were prepared from primary cells extracted from the spleens of transgenic mice which had been treated with phenylhydrazine. The construct used to create the mice was labelled with DIG-11-UTP and used as a probe. Signal was detected using FITC (green). Slides are counterstained with DAPI (blue).
A: Line α107. Position of integration is chromosome 10, within bands B1-4.
B: Line α108. Position of integration is chromosome 8, within bands A1-3. This is confirmed using a mouse chromosome 8 paint labelled with biotin (Cambio) and detected using texas red.
Both lines are hemizygous for the transgene locus. Each line contains two copies of the transgene integrated in tandem, which appear as a single hybridisation signal.

7.4 Discussion

7.4.1 The ZF mutant chromosome 16 has a normal replication profile

No gross abnormality in replication timing throughout the 325kb region has been detected in this mutant. The whole region remains early replicating in S phase. Furthermore, no smaller differences in replication timing can be detected in the ZF mutant compared with the normal profile. This suggests that replication in the ZF mutant, initiates at the same sites and at the same time, as the wild type allele. It is not known whether these these sites are specific, or whether initiation occurs randomly throughout the region.

7.4.2 A repressive chromatin structure can replicate early in S phase

Heterochromatin has been classified cytologically into two distinct forms: dark staining constitutive or α-heterochromatin and diffuse, poorly staining facultative or β-heterochromatin. Constitutive heterochromatin consists of regions that are not expressed, and includes satellite DNA sequences. This is typically found at centromeres. Facultative heterochromatin consists of euchromatin that becomes heterochromatinised in a developmentally controlled way. The best example of this is X chromosome inactivation in the somatic cells of females (Lyon 1992). The term heterochromatin is now also used to describe regions of chromosomes, that if placed next to euchromatin, produce unstable gene repression (position effect variegation) or complete gene silencing, such as telomeres or the silent mating-type loci in yeast (Lohe and Hilliker 1995). One of the general characteristics of heterochromatin is late replication (McCarroll and Fangman 1988, Reynolds *et al.* 1989, Ferguson *et al.* 1991, Torchia *et al.* 1994, Boggs and Chinault 1994, Torchia and Midgeon 1995).

A small 38kb region located 3' downstream of the α genes on human chromosome 16p, resembles heterochromatin due its high density of Alu repeats, lack of CpG islands, DNase 1 insensitivity and its ability to silence α gene expression when adjacent to a structurally normal α gene. It is found within an isochore which is typically

euchromatic in structure. It has been demonstrated in chapter 4, that in the wild type allele, this putative heterochromatic region replicates early in S phase, although its timing of replication is later than the surrounding chromatin. It contains no DNase 1 hypersensitive sites or CpG islands, although it does contain one match to the putative higher eukaryotic origin consensus sequence (fig 7.2).

The +80 gene spans the whole of the heterochromatic region. Another close match to the proposed origin consensus sequence is found close to the DNase1 hypersensitive site and the CpG island (denoted K in fig 7.2) found at the promoter of this gene. The fact that this match is very close to a DNase 1 hypersensitive site and a CpG island, which are associated with a transcriptional promoter, upstream of a gene, all add credibility to this sequence being a *bona fide* origin (see chapter 1 for a fuller account of the features of mammalian origins). Moreover, the putative origin is just proximal to the cosmid cRJ14, which is one of the earliest replicating cosmids of the whole region (fig 4.4).

It is suggested here that early replication of the heterochromatic region is driven by this putative origin, which fires early in S phase. If this is true, a replication fork travels through the region in the direction the +80 gene is transcribed. Further evidence to support the idea that flanking DNA sequences are directing early replication of the heterochromatin is the replication timing of the ZF transgenic lines. When a construct containing a 10kb section of the 38kb region (approximately equivalent to the cos12 3' probe) attached to HS-40 and the α2 gene and promoter (pZFBam10), is introduced into transgenic mice, it is found to be late replicating in two transgenic lines, where the site of integration is different. Although it has been shown that CpG islands can function as replication origins (Delgado *et al.* 1998), it seems unlikely that the CpG island associated with the α2 promoter in this construct functions as such, due to the abnormal DNase 1 insensitive chromatin structure, which presumably precludes protein factor binding. As proposed here, this fragment of DNA has no replication origin and therefore cannot direct its own replication. Integration into a surrogate replicon in the

mouse genome, results in its timing of replication being determined by the distance from the mouse origin and the time at which this origin fires in S phase. It is interesting to note that late replication of this fragment does not result in methylation of the CpG island. Late replication of this segment of DNA in transgenic mice is in contrast to the results described in chapter 5, where two larger constructs containing the α globin genes, are found to replicate early.

7.4.3 GC content and replication timing

Through the observation that heterochromatin can replicate early, which is typically AT rich, the question of the relationship between GC content and replication timing arose. The GC content of the 325kb region examined here is very high, averaging 53.9%, (Flint *et al.* 1997a). As a whole, the 325kb region is early replicating in all the normal and mutant lines examined so far, which is typical of a band with a high GC content (Holmquist 1987, Drouin *et al.* 1994). At higher resolution within the region, the variation in GC content can be seen as a series of sinsoidal curves (fig 1.4, from Flint *et al.* 1997a). Peaks are spaced 80-90kb apart, with GC content varying from 63% at a peak to 46% at a trough. The heterochromatic region downstream of the α genes corresponds to one of the troughs, which I have shown in this thesis to be early replicating. In fact, throughout the whole of 325kb region, there seems to be no correlation between the sinusoidal varation in GC content and the pattern of replication timing. So although at the level of banding, GC rich regions replicate early and AT regions replicate late (Drouin *et al.* 1994), this relationship breaks down at higher resolution. This finding has been previously reported (Eyre-Walker 1992).

7.4.4 Mechanism of transcriptional silencing

The ZF deletion juxtaposes the α2 gene against a region of putative heterochromatin, which appears be involved in causing the transcriptional silencing. As the results in this chapter show, this silencing is not associated with a switch from early to late replication. Several theories can be proposed to explain the silencing phenomenon seen in the ZF mutant, based on the observations seen in mammals and other species:

A. The 'spreading' mechanism

The simplest mechanism which would explain the silencing, is that the ZF deletion allows the repressive chromatin structure of the putative heterochromatic region to propagate along a few kilobases further in the 5' region. This results in the α2 gene becoming part of a heterochromatic region, which would explain the DNase 1 insensitivity seen at the promoter. This 'spreading mechanism' has been used to explain X inactivation in females (Lyon 1992), the phenomenon of position effect variegation (PEV) in *Drosophila* (Weiler and Wakimoto 1995) and telomeric silencing in *S.Cerevisiae* (Grunstein 1997).

Heterochromatic regions of the human inactive X chromosome are found to be late replicating (Boggs and Chinault 1994, Torchia *et al.* 1994), as are yeast telomeres (McCarroll and Fangman 1988, Ferguson *et al.* 1991). Furthermore, in yeast, transferral of an early firing origin to a telomeric position, causes it to become late firing (Ferguson and Fangman 1992). However, the region of heterochromatin the α2 gene has become juxtaposed to, is early replicating in the wild type allele (fig 4.4) and remains early replicating in the ZF allele (fig 7.4). Therefore, it is maybe not so surprising that silencing in this case is not accompanied by late replication.

B. The loss of a boundary/insulating element

An extension of the 'spreading mechanism' could include the ZF deletion removing a boundary or insulator element (reviewed in Wolffe 1994). In the normal allele, this element would prevent the heterochromatic structure propagating along the chromosome as part of the normal regulation of α globin gene expression. Removal of the boundary element in the ZF mutant, would allow the spreading of the heterochromatin into the α2 gene, causing it to become silenced. This boundary element would have to be located somewhere proximal of the α genes, within the 15kb spanned by the deletion, upstream of the putative heterochromatic block. It is interesting to note that this is where the later replicating part of cos 12 is found and where replicated chromatids appear to take a long

time to dissociate. This latter observation may support the idea of a restraining chromosomal or nuclear structure.

It has been proposed that matrix or scaffold attachment regions (MARs or SARs) can function as boundary elements (reviewed in Geyer 1997). However, when this region has been examined, no such sites were found (Jarman and Higgs 1988). In addition to MARs or SARs functioning as boundary elements and demarcating a transcriptional domain, they are also reported to be close to replication origins. Origins are thought to be attached to the nuclear scaffold or matrix, during interphase and metaphase (for a fuller account see section 1.3.6), although there is no direct evidence to show that attachment is required for origin function. If a matrix attachment site is located in this region, it corresponds to a later replicating region, which resembles more of a termination site than an origin.

C. Abnormal localisation within the interphase nucleus

It is suggested that the human interphase nucleus is compartmentalised as a method of metabolic regulation (reviewed in Strouboulis and Wolffe 1996). This gives rise to an alternative model for silencing, which involves the aberrant localisation of the α2 gene within the interphase nucleus. This model envisages that juxtaposition of the α2 gene against a heterochromatic region, results in the gene being targeted to a 'silencing' compartment of the nucleus. In humans, there is evidence of the nuclear periphery, or nuclear envelope being the transcriptionally inactive region (Ferreira *et al.* 1997). In addition, the insertion of a piece of heterochromatin into the *brown* gene (bw^D) in *Drosophila* causes PEV and targeting to the apical periphery of embryonic nuclei (Dernburg *et al.* 1996). However, in mammals, silencing compartments can also be found within the internal nucleoplasmic space (Kurz *et al.* 1996, Brown *et al.* 1997). Irrespective of the precise location within the nuclei, association of the gene with centromeric heterochromatin appears to be critical for silencing (Dernburg *et al.* 1996, Brown *et al.* 1997).

When scoring slides to measure replication timing, no obvious differences were observed between the position of hybridisation signals within the ZF nuclei compared to the normal 16 nuclei. However, as mentioned previously, FISH analysis was carried out using fixed interphase nuclei, viewed as a two dimensional image. For a more accurate localisation of the α2 gene within the nucleus, scanning confocal microscopy or three dimensional fluorescence microscopy would need to be performed, preferably on unfixed cells. Direct association of the silenced α2 gene with centromeric heterochromatin could be investigated using the immuno-FISH procedure described by Brown *et al.* (1997).

Whereas PEV alleles are generally recessive, in the case of the *brown* gene, PEV is dominant: homologous pairing of a bw^D allele with a wild type allele (bw^+) in heterozygotic larva, results in the both alleles becoming silenced. Hence the mutant allele is called $brown^{Dominant}$ (bw^D). On examination, both alleles are associated with the centromeric heterochromatin (Dernburg *et al.* 1996). Although pairing of homologous chromosomes is reported in *Drosophila*, it is not routinely observed in human interphase cells (reviewed in Cremer *et al.* 1993). Certainly during the scoring of S phase nuclei to determine replication timing in the normal EBV-line, the individual alleles of chromosome 16 appear spatially separate. If any homologous pairing does occur during S phase, then it is for a very short period of time. Future experiments could examine the nuclear positioning of the wild type and ZF mutant allele in the ZF EBV-line. A far more striking piece of evidence favours the ZF allele not being dominant like the bw^D allele. If the ZF allele caused silencing of the normal allele (αα) in the ZF individual, there would be no expressing α globin genes, which would result in death. The ZF individual is still alive into his later years.

D. Loss of a positive regulatory element

Another explanation for the loss of expression of the α2 gene is that the ZF deletion removes a positive regulatory element. To test this hypothesis, a construct called DH2 (Barbour *et al.* in preparation, Higgs *et al.* 1998) which includes HS-40, the two fetal-

adult α genes and all of the heterochromatic region removed by the ZF deletion, was analysed in transgenic mice. It does not show a level of α gene expression higher than any of the similar constructs that have been tested, which do not contain the heterochromatic region.

E. Models arising from studies of imprinted genes

The ZF silenced α2 gene resembles a silenced, imprinted gene because of the methylation of the CpG island at the α2 promoter, and the fact the gene has a condensed chromatin appearance (Efstratiadis 1994). When germ-line DNA of this patient has been examined, methylation is not present, indicating that methylation is acquired during development (D.R. Higgs, pers. comm).

As discussed in section 3.6, imprinted regions on homologous alleles have been found to replicate asynchronously. However, the relationship between transcriptional activity and replication timing of such regions is not simple. In general, imprinted regions are quite large (hundreds of kilobases), and may contain genes that are expressed from opposite alleles, such as the Igf2/H19 locus (reviewed in Barlow 1997). Although the whole of the imprinted region displays replication asynchrony between the paternal and maternal alleles (Kitsberg *et al.* 1993a, Greally *et al.* 1998), it is not known whether the individual genes within the region switch replication timing according to their transcriptional activity. When other individual imprinted genes have been examined, replication always occurs earlier on the paternal allele, apparently independent of the pattern of expression (Kitsberg *et al.* 1993a).

F. Loss of a replication termination site

Despite the lack of direct evidence, it is proposed here, that the ZF deletion may remove a replication termination site. If a termination site has been removed, which may or may not be behave as a boundary element, which replication fork would be responsible for replication of the silenced α2 gene? In the model proposed, in the normal situation, replication of the gene occurs via a replication fork travelling in the direction of telomere to centromere. Under these conditions, in erythroid cells, the α2 gene remains

transcriptionally active, its promoter DNase 1 hypersensitive and the CpG island unmethylated. If, in the ZF deletion, the region became replicated by the other fork moving from the centromere through the heterochromatic region, this may contribute to silencing of the gene. It could be that ORC directs replication of the heterochromatic region (which the α2 gene and promoter have now become part of) and then contributes to establishing a condensed chromatin structure throughout this region. It would interesting to know what prevents further propagation of the repressive structure along the chromosome. This is a premature hypothesis and a good deal of experimental work would be required to prove it. Firstly, replication origins need to be confirmed using another method, such as those outlined in Falaschi *et al.* 1993. Secondly, experiments are required to characterise the nature of the heterochromatin-like region downstream of the α cluster, by investigating its pattern of acetylation, and identifying which proteins bind within the region.

7.4.4. Summary

It has been shown in this chapter, that silencing of the α2 globin gene seen in the ZF deletion mutant, occurs in the absence of a change in replication characteristics of the gene itself, or the surrounding chromosomal region. The mechanisms that bring about the changes in chromatin structure at the promoter of the gene and methylation of the CpG island, must occur independently of replication timing.

The ZF deletion places the α2 gene against a heterochromatic-like region of DNA, which appears to play a role in establishing silencing. When a construct containing HS-40 linked to the α2 globin gene and promoter and a portion of the heterochromatic region is introduced into transgenic mice, α2 mRNA expression is lower than other α constructs, and DNase 1 insensitivity at the α2 promoter is found, mimicing the effect seen *in vivo*. Methylation of the CpG island however, is not recreated. Replication of the construct occurs late in S phase, apparently irrespective of the position of integration into the mouse genome. This is in contrast to the other segments of DNA containing the

α cluster tested in transgenic mice, which display an early pattern of replication, matching the pattern seen within the normal chromosome.

Chapter 8 Replication timing of the human 16p subtelomeric region and the effect of telomere truncation upon DNA replication timing

8.1 Introduction

Although the replication timing of telomeres has been studied in some detail in budding yeast (for example McCarroll and Fangman 1988, Ferguson *et al.* 1991, Friedman *et al.* 1996), very little is known about the replication timing of human telomeres (Ten Hagen *et al.* 1990). The ends of all human telomeres contain 2-20kb of a simple, conserved, repeat sequence of $(TTAGGG)_n$ (Moyzis *et al.* 1988, de Lange *et al.* 1990, Kipling and Cooke 1990). As conventional DNA polymerases extend DNA in the 5' to 3' direction and require a pre-existing template, they are unable to synthesise the extreme 5' end of a linear chromosome. Replication of the terminal repeat sequence occurs by an alternative mechanism using a ribonucleoprotein polymerase, telomerase (Morin 1989). The long arrays of telomeric repeats bind two related proteins TTAGGG repeat binding factors 1 and 2 (TRF1 and TRF2) (Zhong *et al.* 1992, Chong *et al.* 1995, Bilaud *et al.* 1997, Broccoli *et al.* 1997), which are thought to protect the chromosome from exonucleolytic degradation, prevent chromosome fusion, regulate telomere length and attach telomeres to the nuclear matrix (de Lange 1992, reviewed in Zakian 1995, Broccoli *et al.* 1996, Ludérus *et al.* 1996, van Steensel and de Lange 1997). When the replication timing of telomeric repeats have been collectively examined in a population of human cells, they are found to replicate throughout S phase (Ten Hagen *et al.* 1990). It is suggested that individual telomere ends replicate at different times, which may be determined by sequences proximal to the repeats.

DNA sequence adjacent to the $(TTAGGG)_n$ repeats, known as the subtelomeric region, contains a variety of repetitive elements, some of which are restricted to a single subtelomeric region, whilst others are shared amongst a subset of chromosomes

(Brown *et al.* 1990, Flint *et al.* 1997b). Thus, the lack of unique sequence makes replication timing analysis of an individual subtelomeric region very difficult. However, rodent and human subtelomeric regions share little DNA sequence homology, so this problem can be circumvented by studying the replication timing of an individual human telomere in an interspecific mouse-human hybrid cell-line.

The subtelomeric region of the short arm of human chromosome 16 (16p) has been studied in some detail (Wilkie *et al.* 1991, Flint *et al.* 1997a, Flint *et al.* 1997b). It is polymorphic with three common alleles (A, B and C), which differ in size and the variety of repeats they contain (Wilkie *et al.* 1991). In the previous chapters I have shown that the unique sequence adjoining the polymorphic subtelomeric region is in general early replicating, becoming earlier moving away from the telomere. To determine whether the characteristic of early replication extends further towards the telomere, in this chapter I have analysed the replication timing of the subtelomeric region of the most common allele of 16p, the type A allele, in several MEL-human interspecific hybrids which contain a single copy of human chromosome 16. This is the first work described, which analyses the replication timing of an individual human telomeric region.

The second part of the work described in this chapter analyses the importance of the telomere and subtelomeric region in determining the replication pattern and timing of the extended chromosomal region found at the telomere. A series of naturally occurring truncations have been characterised, which involve the loss of the terminal region of human chromosome 16p (16p13.3). The extent of the deletions, which range from 135-155kb, are shown in fig 8.1. Analysis of these mutants has shown that all chromosomes have been stabilised by the direct addition of telomeric repeats $(TTAGGG)_n$ to the site of breakage. Attention was drawn to these mutants because each truncation removes the α globin gene regulatory element HS-40, which abolishes expression of the two *cis*-linked α globin genes, resulting in the phenotype of α-thalassemia. Whereas the α globin genes are normally found 170-430kb from the

telomere (Wilkie *et al.* 1991), in these truncated chromosomes, the genes are now less than 30kb from the new telomere end. I have analysed the replication characteristics of two of the truncated mutants $(\alpha\alpha)^{IC}$ and $(\alpha\alpha)^{TAT}$, to assess whether loss of α globin gene expression is associated with a change in replication timing and how this may relate to chromosome structure and nuclear positioning in the interphase nucleus.

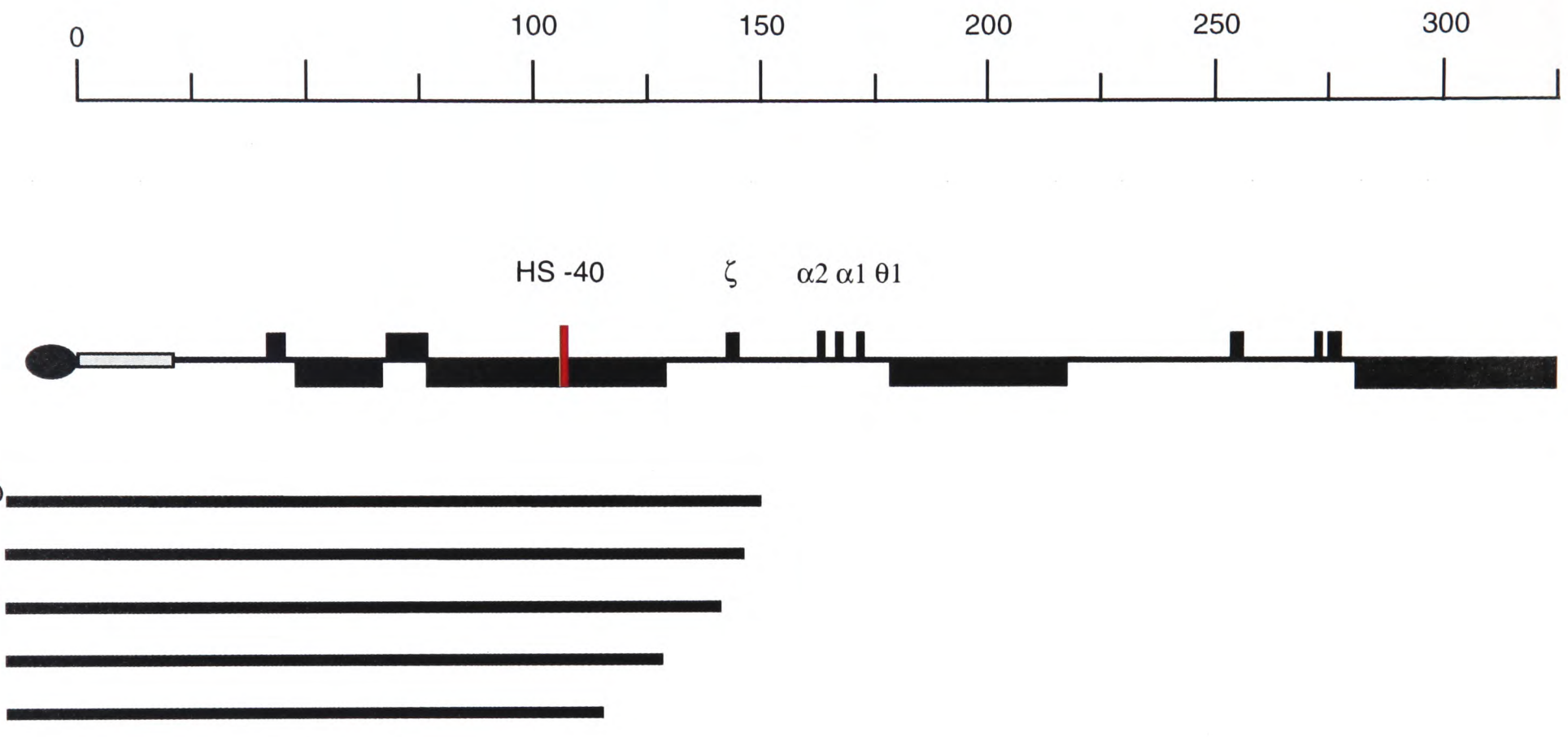


Figure 8.1 Naturally occurring truncations of telomere of 16p13.3. The telomere is represented as a black oval and the subtelomeric region as a stippled box. Positions of genes are shown. Those above the line are transcribed towards the centromere. Those below are transcribed towards the telomere. The α globin genes are labelled and the α globin regulatory element (HS-40) is shown in red. The extent of each truncation is shown as a thick black bar. (αα)TI was originally described in Wilkie *et al.* (1990a); provisional mapping of (αα) IdF was reported by Romao *et al.* (1992). More detailed molecular analysis of this chromosome, and the remaining chromosomes is reported in Flint *et al.* 1994. Scale is in kilobases.

8.2 Methods

8.2.1 Measuring the replication timing of the proterminal region of 16p type A allele

To identify which interspecific MEL-human hybrid lines contained a human chromosome 16 type A allele, DNA was isolated from various lines, digested with *Bgl2*, and used in southern blot analysis (Old and Higgs 1983) with probe pNFG400 (described in Wilkie *et al.* 1991). This probe recognises sequence which is present in all 16p alleles; each type of allele can be distinguished according to the size of fragment identified. Two lines containing a normal copy of human 16 (JY5-4, JG7) and one line containing the abnormal ZF chromosome were positive for the type A allele and selected for analysis (data not shown). Fixed nuclei preparations were already made for each of these lines, as they were used for FISH analysis as in previous chapters.

Probes within the subtelomeric region that gave a strong signal on human 16p and did not crosshybridise with rodent telomeres when tested in MEL-16 hybrid lines, were selected for FISH analysis and are shown in fig 8.2. The pBam series of recombinants contain various *BamH1* fragments cloned into pUC13 or pUC12. As these probes are small (<7kb), they needed to be pooled together to give a signal which was strong enough to be detected in interphase nuclei and on metaphase chromosomes. 2µg of each of the following recombinant plasmids pBam1, pBam2, and pBam 4 were labelled with DIG in separate nick translation reactions, as described in section 2.3.1. 0.5µl of each labelled probe was pooled into one eppendorf with 1.5µg of mouse Cot-1, 1.5µg of human Cot-1, 5µg salmon sperm DNA and precipitated with two volumes of ethanol. DNA was dried under vacuum, resuspended in 10µl of hybridisation buffer and used to hybridise a single slide (described fully in section 2.3). The same approach was taken with probes pTel Bam0.7 (a 0.7 kb *Pst*1 fragment sublcloned from pTel Bam3.4 (described in Brown *et al.* 1990) into pUC18), pTel Bam4.8 (a 4.8kb *BamH1* fragment cloned into pUC18) and pTel Bam8.5F. pTel Bam8.5F is derived from pTel Bam8.5, which could not used directly, as this cloned sequence contains an internal series of 11

$(TTAGGG)_n$ repeats, which if used as a FISH probe, would cross hybridise to the terminal repeats present at the human and mouse telomeres. To remove this sequence, the 8.5kb insert was released from the vector (pUC18) using *BamH*1 and cut with *EcoR*1 which generates two fragments, approximately 5.3kb and 3.2kb in size. The 5.3kb fragment which does not contain the repeats was gel purified (section 2.2.3.3) and 2μg used in the standard nick-translation reaction (section 2.3.1). For each series of slides that were hybridised, a control probe was included on a separate slide. The probe used was c419C1, whose replication timing has been established in the previous chapters.

8.2.2 Replication timing of the mutant MEL-16 hybrid lines IC and TAT

Fixed nuclei were prepared from two interspecific MEL-human hybrid lines containing a truncated chromosome 16 (IC and TAT). Slides were made and hybridised as described in section 2.3, using 6 out of 12 of the original recombinants detailed in chapter 4, which have not been removed by the truncation. Slides were scored as described in section 2.3.

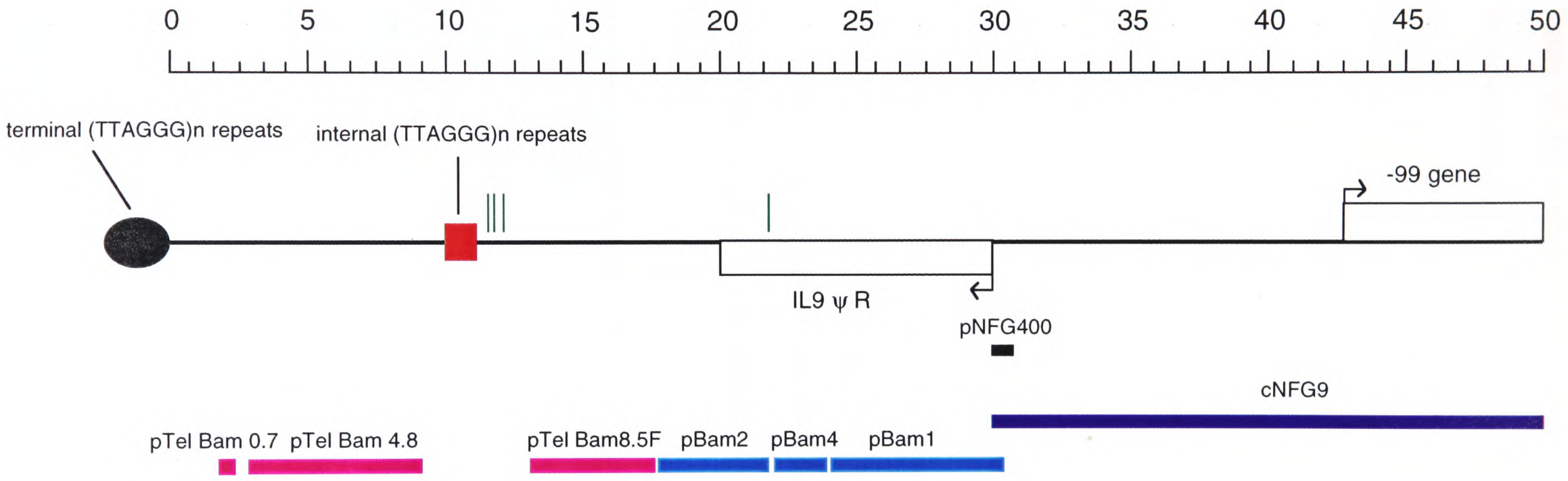


Fig 8.2 The subtelomeric region of human chromosome 16p type A allele. The terminal (TTAGGG)n repeats are depicted as a black oval and the position of the internal (TTAGGG)n repeats is shown, along with the IL9R pseudogene and -99 gene. Close matches (1 basepair mismatch) to the putative mammalian origin consensus sequence are shown as vertical green lines. Probes used for FISH analysis to measure replication timing are indicated. Probes of the same colour were pooled and hybridised together. Cosmid cNFG9 was used in previous chapters to measure replication timing. Probe pNFG400 was used in Southern blot analysis to identify MEL-16 hybrid lines containing type A alleles. Scale is in kilobases.

8.3 Results

8.3.1 Replication timing of the subtelomeric region in the type A allele

Replication timing of the subtelomeric region of the type A allele of chromosome 16p is shown in table 8.1. The table includes results for the cosmid cNFG9 from previous chapters, which is the first cosmid which spans the unique sequence adjacent to the subtelomeric sequence. It also includes results of the early and late replicating controls (P48 and cSamD4 respectively) from previous chapters. These results show that replication becomes later, moving through the subtelomeric region towards the terminal repeats. The pBam probes give a percentage doublet score which is not vastly different from cNFG9, which suggests this sequence may be part of the same replicon. The drop in percentage doublet score is more striking in the pTel probes, which may indicate this sequence is part of a different replicon. Two of the three probes that were pooled together to give this score (pTel Bam0.7, pTel Bam 4.8), lie on the distal side of the internal series of $(TTAGGG)_n$ repeats, closest to the telomere end. Late replication therefore, only seems to extend up to 20kb from the terminal repeats. Moving proximally from here, replication timing gradually increases to early, where it remains for the rest of the 325kb region.

	JY5-4 1 (normal)	*JY5-4 2 (normal)*	*JG7 (normal)*	*FZ (abnormal)*
P48	63 [100]		55 [100]	58.5 [93]
c419C1	73 [96]		79.5 [96]	75.5 [100]
cNFG9	55.5 [88]		56 [92]	46.5 [95]
pBam probes	43 [95]	49 [86]	47 [100]	47 [89]
pTel probes	20 [86]		29 [80]	N/A
cSamD4	29.5 [81]		33.5 [98]	26.5 [100]

Table 8.1 % doublets scored using subtelomeric probes (pBam probes = pooled pBam1, 2, and 4; pTel probes = pooled pTel Bam 0.7, 4.8, 8.5F). c419C1 which is 16p specific (approximately 250kb away from the subtelomeric region) was used as an early replicating control. Results from previous chapters are also included for comparison. cNFG9 is the first cosmid to span unique sequence directly adjacent to the subtelomeric region. cSamD4 and P48 are the mouse controls (late and early replicating respectively). Hybridisation efficiency measured for each slide is shown in square brackets alongside each score. N/A indicates data not available.

Three close matches (1 mismatch difference) to the putative mammalian origin consensus sequence are found 3' to the internal repeat sequence. Another one is found

within the IL9 receptor pseudogene (fig 8.2). The replication timing data (table 8.1) infer that if any of these matches function as origins of replication, they must fire later in S phase than upstream origins within the unique sequence. Alternatively, this region contains no functional origins, and is passively replicated from a fork generated from a origin that is found within the unique sequence 3' to the subtelomeric region.

8.3.2 Replication timing of the truncated telomeres IC and TAT

Table 8.2 shows the percentage doublet values obtained in two MEL-16 hybrid cell-lines, each containing a truncated copy of chromosome 16 (IC and TAT). These results are shown graphically in figure 8.3. In both lines, the most distal region (approximately 50kb), where the α globin genes are now located, replicates significantly later than the corresponding sequence in the normal allele (fig 8.3), although it is not as late replicating as the subtelomeric sequence which is normally located next to the terminal repeats (table 8.1). Replication timing of the remaining sequence in both chromosomes appears to be early. This is similar to the replication characteristics of the normal type A allele. Only sequences within less than 50kb of the terminal repeats seem to show a depressed percentage doublet score.

The pattern of replication in both mutant lines is different to the normal. No characteristic peaks of early replication are seen, and the four most proximal cosmids in IC and five in TAT, appear to have a very similar replication timing. This could indicate that preferred sites of replication initiation are lost in these truncated chromosomes, and replication initiates at random sites throughout the 200kb region. This may occur if chromatin structure throughout this region has been disrupted due to the deletion. However, analysis of the TAT allele in the MEL-16 hybrid shows that the erythroid-specific DNase 1 hypersensitive sites form normally at the α gene promoters (Higgs *et al.* 1995), which are less than 20kb from the newly formed telomere end. This suggests that the deletion does not drastically influence chromatin structure surrounding the α genes, despite their close proximity to the telomere, and the loss of HS-40.

cosmid	IC slide 1	IC slide 2	TAT
cSamD4	23.5		30 [95]
cGG1	40.5 [90]		52 [90]
cos12	48.5		63.5 [93]
cRJ14	63 [100]	56.5 [100]	66.5 [100]
c310H5	61 [95]		68.5 [96]
c314G4	56.5 [100]	62.5 [95]	70 [95]
c419C1	60 [96]		71 [93]
P48	56.5		59.5 [99]

Table 8.2 % doublets scored for each cosmid in the interspecific MEL-16 hybrid lines containing a truncated chromosome 16 (IC and TAT). Some cosmids have been scored twice in the IC line (column 3). Hybridisation efficiency has been measured for each slide, and is shown in square brackets. Shaded boxes show control probes. The remainder are 16p specific, and have been listed in the order in which they map to the chromosome, the most telomeric at the top of the table.

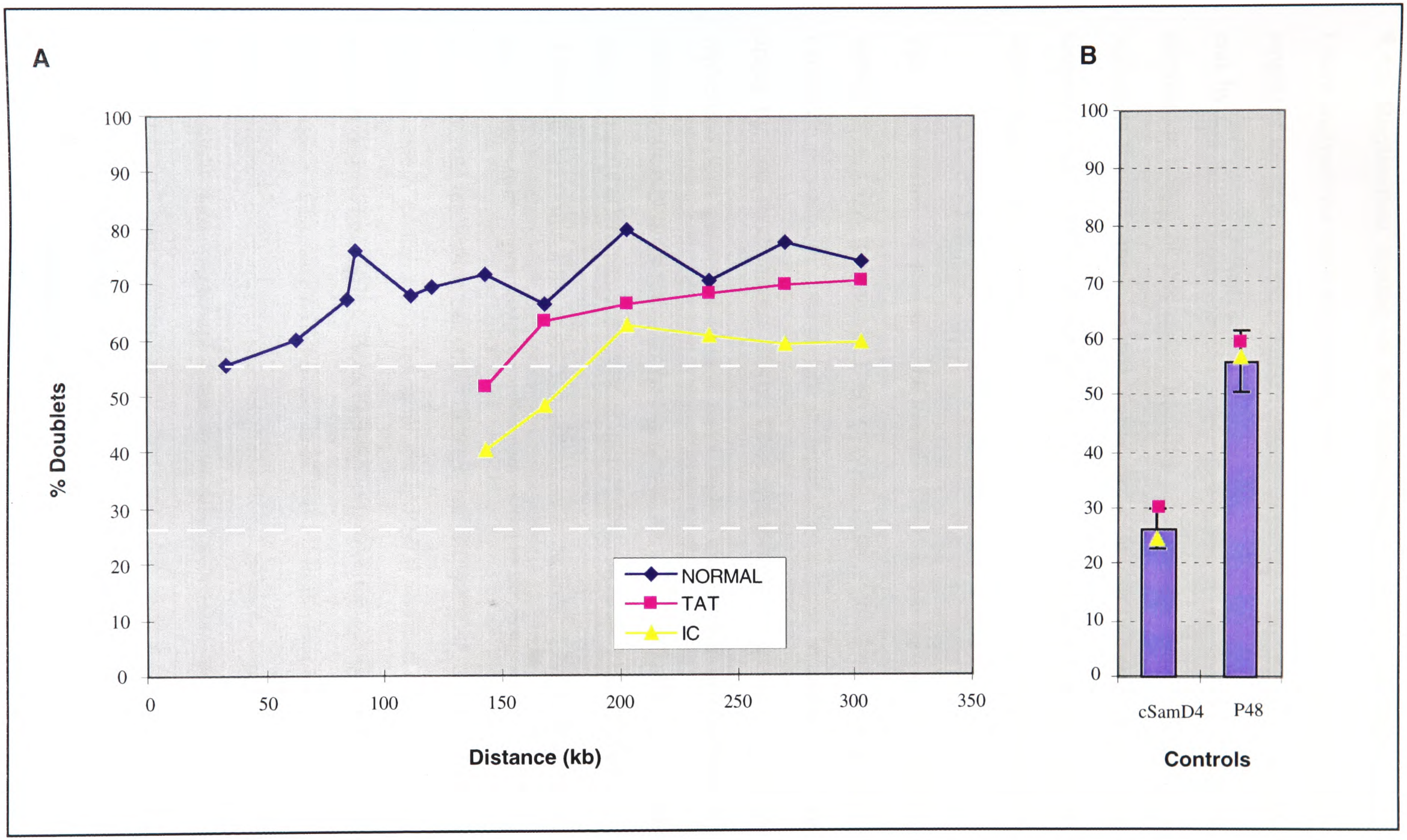


Fig 8.3A Replication profiles of the telomere truncated hybrid lines TAT and IC (yellow and pink respectively) compared with the normal hybrid line (shown in blue). White dashed lines indicate the mean values of the early (top line) and late (lower line) replicating controls taken from fig 8.3B.
Fig 8.3B Replication timing of the mouse controls. The mean value (recorded from all the MEL-16 hybrid lines) for the early (P48) and the late (cSamD4) controls are given as filled columns. Measurements from the TAT and IC lines are given individually (yellow and pink respectively).

8.4 Discussion

8.4.1 Replication timing of the subtelomeric region of 16p

I have analysed the replication timing of the subtelomeric region (approximately 40kb in length) of the type A allele of human chromosome 16p. This region is flanked at the 5' end by the terminal $(TTAGGG)_n$ repeats and on the 3' side by unique sequence. Replication timing of this region becomes progressively later towards the telomere. Although I have only examined one telomere, the 16p subtelomeric region shares sequence and structural similarities with several other human telomeres (Flint 1997b), which suggests this replication timing pattern may be common to other chromosomes.

The organisation of human subtelomeric DNA is similar to the organisation of subtelomeric DNA in budding yeast (Brown *et al,* 1990, Flint *et al.* 1997b), which exhibits a reproducible late pattern of replication at most telomeres. Late replication arises by two mechanisms: firstly, DNA sequences are passively replicated by a replication fork generated from a distal early firing origin. Replication of the left telomere of chromosome III occurs using an an origin (ARS305) which lies approximately 40kb from the telomere end (Huberman *et al.* 1988, Dubey *et al.* 1991). Although this origin fires in early S phase, slow replication fork progression through the subtelomeric region delays replication until late S phase (Reynolds *et al.* 1989). Secondly, late replication can occur through late activation of origins. The right end of chromosome V contains an efficient origin (ARS501) located 31kb from the telomere end (Ferguson *et al.* 1991). Here, late replication of the subtelomeric region is due to the late firing of this origin (Ferguson *et al.* 1991). In the human subtelomeric region I have examined, there are four close matches to the putative origin consensus sequence. It is possible these sequences function as origins *in vivo,* in which case they must activate late in S phase. Alternatively, they may not be used at all. In yeast, the left arm of chromosome III contains three ARS elements within 15kb of the telomere, none of which are active as origins (Dubey *et al.* 1991). Also in yeast, experimental analysis has shown that replication of the terminal repeats by telomerase occurs after the replication

fork reaches the end of the molecule (Wellinger *et al.* 1993), suggesting it may be the last event in S phase. In this thesis, although the replication timing of the telomeric repeats has not been measured directly, the region adjacent to the telomeric repeats is the latest to replicate, suggesting this phenomenon may also be true in humans.

8.4.2 What determines late replication at telomeres?

If the subtelomeric region does contain origins which activate in late S phase, what determines their late activation? One explanation is that the delay in replication timing is a secondary effect, caused by the repressive chromatin structure that is propagated from the telomere. An alternative but not mutually exclusive possibility is that nuclear sublocalisation is the primary cause of late replication. However, the causal relationships between these observations is not clear. It may be that replication timing is the primary event which leads to heterochromatinisation at telomeres. Replication of the telomere may also precede nuclear positioning. Mutations in the yeast chromatin remodelling factor CAF-1, which is replication dependent, results in the loss of targeting to the nuclear periphery and the loss of a repressive chromatin structure (Enomoto *et al.* 1997, Kaufman *et al.* 1997).

Direct analysis of chromosomal position by three dimensional microscopy suggests that telomeres are preferentially positioned near the nuclear periphery in budding yeast (Klein *et al.* 1992, Gotta *et al.* 1996), fission yeast (Funabiki *et al.* 1993) and flies (Mathog *et al.* 1984). Furthermore, telomeres are reported to form tight clusters in yeast (Funabiki *et al.* 1993, Gilson *et al.* 1993). This aspect of telomere biology does not appear to be conserved in mammals. Although telomeres are thought to be attached to the nuclear matrix (de Lange 1992, Ludérus *et al.* 1996), they are not found to predominantly associate with the nuclear periphery and do not appear to cluster during interphase (Vourc'h *et al.* 1993, Ludérus *et al.* 1996).

In yeast, transferral of an early replicating origin (ARS1) from its normal chromosomal position (chromosome IV) to the right arm of chromosome V, 17kb away from the

telomere, causes the origin to activate late. Conversely, removal of a telomeric, late activating origin (ARS501) to a circular plasmid, causes the origin to switch to early replication (Ferguson and Fangman 1992). Both of these results suggest that late activation is caused by proximity to the telomere, and indicate that late replication is a consequence of the association. To ascertain whether the same is true in humans, I have analysed two truncated chromosomes where an early replicating, gene-rich region has been 'healed' by the addition of telomeric repeats. Sequences less than 50kb from the new telomere end show a slight depression in replication timing, although the replication of sequences farther from the telomere appear to be unaffected. These results suggest that it is the presence of the telomeric repeats rather than the subtelomeric region, which are important in establishing a later pattern of replication. This is consistent with the observation that it is the (TTAGGG)n repeats which target the telomere to the nuclear matrix, which may be mediated through TRF1 or 2. Subtelomeric DNA is not thought to play a role in tethering, even though this sequence may contain an internal series of (TTAGGG)n repeats (de Lange 1992, Louis 1995, Ludérus *et al.* 1996). If the loss of the subtelomeric region and acquisition of telomeric repeats adjacent to the gene-rich region results in the genes and flanking sequence being placed in a new nuclear location, this does not seem to prevent early replication.

In yeast, proximity to the telomere has a profound effect on transcription. It appears that the closer a gene is placed to a telomere, the more efficiently it is silenced (Gottschling *et al.* 1990, Renauld *et al.* 1993, Maillet *et al.* 1996). Moreover, repressed genes can be maintained and clonally inherited (reviewed in Laurenson and Rine 1992). The mechanism of gene silencing is not fully understood, although it is thought to involve the same mechanisms that cause late replication, i.e. the formation of a repressive chromatin structure and/or localisation to silencing compartments of the nucleus. The DNA replication protein ORC is known to play a role in establishing gene silencing at the mating-type loci (for example, Micklem *et al.* 1993, Fox *et al.* 1995). One method of involvement is through the recruitment of SIR-1, a protein involved in formation of the repressive structure (Triolo and Sternglanz 1996, Fox *et al.* 1997). Less is known

about its role in silencing at telomeres, which does not require the SIR-1 protein (Aparicio *et al.* 1991, Fox *et al.* 1997), although subtelomeric regions contain ARS elements which bind ORC and mutations in ORC decrease telomeric silencing (Fox *et al.* 1997). However, interactions between ORC and other silencing factors have not yet been demonstrated. Furthermore, ORC binds at all active origins in yeast, but only a few origins repress expression of adjacent genes. How selective binding of silencing proteins occurs at certain origins is unknown.

It is known that the truncations of human 16p shown in fig 8.1, result in the loss of α globin gene expression, which is thought to be due to the removal of the critical regulatory element HS-40, rather than the increased proximity to the telomere, although this may play a part in gene silencing. The effect of the truncation upon expression of the other genes in this region is unknown. The fact the region containing these genes maintains an early pattern of replication, suggests they may remain transcriptionally active. In the normal type A allele, the pattern of replication mirrors the pattern of transcription. Early replication is seen in the unique sequence, which contains many expressed genes. Replication becomes progressively later throughout the subtelomeric region, which is transcriptionally quiescent.

Chapter 9 Final discussion

The first part of this thesis described the establishment of a protocol within the Molecular Haematology Unit, based on a technique originally developed by Selig *et al.* (1992), which uses FISH to measure the timing of DNA replication. Using this protocol, it has been shown that the human α globin cluster and its surrounding chromosomal region (approximately 325kb) replicates early within S phase of the cell cycle in both erythroid and non-erythroid cells. The results also suggest that the whole region examined becomes slightly earlier replicating in erythroid cells. The 325kb region has been examined in the individual alleles of chromosome 16: the maternal and paternal homologues replicate at a very similar time in S phase.

These findings are consistent with the fact this region is found at the tip of a Giemsa light band (16p13.3), which are typically early replicating (Drouin *et al.* 1994). Within the Giemsa light band, the α cluster is part of a GC rich H3 isochore, with a high density of CpG islands. Work by Delgado *et al.* (1998) has shown that unmethylated CpG islands can function as origins of replication, which fire synchronously, early in S phase. The work presented here is compatible with this report, and suggests that early replication throughout the region studied in this thesis, may be due to the synchronous firing of origins at CpG islands, early in the cell cycle. Similarly, high densities of CpG islands have been found within all Giemsa light bands (Craig and Bickmore 1994), suggesting that their well established pattern of early replication may be explained by CpG islands acting as early firing origins of replication.

9.1 Nature and specificity of origins of replication

Several questions worthy of further investigation have been raised by these findings. Does every CpG island function as an origin? The 325kb region examined here contains fourteen CpG islands, thirteen of which are constitutively unmethylated, with many of these being associated with a constitutive HS (CpG islands at the α globin gene

promoters are only associated with a HS in erythroid cells). The presence of a HS suggests these sequences are available for replication proteins to bind. If every CpG island in the 16p13.3 region functioned as an origin, replicons in this area would have an average size of ~25kb. Replicons are reported to be between 50-500kb in size, which might argue against the utilisation of every CpG island within a single cell, and imply that only a subset are used. What is unclear at present, is whether each individual cell utilises the same CpG islands, or whether initiation occurs randomly, with the number and selection of CpG islands varying from cell to cell. This model of replication initiation, where the number of potential origins outweighs the number actually used in each cell, fits with the previously proposed Jesuit model of replication origins, where 'many are called but few are chosen' (Burhans and Huberman 1994). Analysing a whole population of cells using FISH in this thesis demonstrates a reproducible pattern of replication, suggesting that replication may initiate more frequently at certain CpG islands than at others. Parallels can be drawn between this result and the result described at the hamster DHFR locus, where although replication can initiate anywhere within a 'broad zone' of 55kb, the region contains two sites which are preferentially used (reviewed in Burhans and Huberman 1994).

In humans, does origin utilisation change during development and differentiation? It seems feasible that the CpG islands at the α globin promoters only have the potential to function as origins in erythroid cells, when they are associated with a HS. If these CpG islands are used in addition to the other CpG islands with a constitutive pattern of hypersensitivity in erythroid cells, it would explain why replication occurs somewhat earlier in these cells compared with non-erythroid cell-types. The CpG island at the ζ2 gene promoter shows a developmental pattern of hypersensitivity, linked to its pattern of expression. This CpG island may only function as an origin in embryogenesis when the HS is present.

Is replication initiation within the 325kb region only restricted to CpG islands? The region contains other HSs which are not associated with CpG islands. As mammalian origins show little DNA sequence specificity, there seems no reason why these sequences could not function as origins.

There is a good deal of speculation over whether replication origins are attached to the nuclear matrix, and whether attachment is important for origin function, perhaps by localising origins to sites within the interphase nucleus which contain a high concentration of protein factors required for replication. For example, individual matrix attachment sites have been mapped near, or in the close vicinity of origins, such as the β globin origin, which is not associated with a CpG island (Jarman and Higgs 1988) and the origin upstream of the *ppv1* gene, which is associated with a CpG island (Lagarova *et al.* 1998). If CpG islands function as origins in the 16p region, there is no evidence to suggest these sequences are attached to the nuclear matrix (Jarman and Higgs 1988). The techniques used to detect MARs and SARs in the past are somewhat controversial and may be prone to artefacts (reviewed in Hozak 1996 and references therein). Interactions *in vivo* have not yet been demonstrated, either between origin *cis*-acting sequences or replication *trans*-acting factors with proteins of the matrix such as lamins. Further evidence of this kind is needed to understand the relationship between DNA replication and nuclear localisation.

9.2 Future work to establish origins of replication throughout 16p

To be able to answer any of the questions outlined above, it is necessary to map the positions of origins of replication throughout the 16p region. During the course of the DPhil, a proportion of time has been spent developing a technique to identify origins, known as the leading strand assay, developed by Handeli *et al.* (1989). Emetine, a general protein synthesis inhibitor, is used to inhibit lagging strand synthesis (Burhans *et al.* 1991), whilst leading strand DNA synthesis is allowed to proceed in the presence of BdU. BdU-substituted DNA is collected by alkaline isopycnic density centrifugation

and slot-blotted onto membranes. The direction of replication within a genomic region can be determined by hybridising the DNA with strand specific probes. If sequential probing along a genomic region reveals a switch in strand specificity, this indicates the position of an origin. Unique sequences to be used as probes throughout the 16p region have been cloned to allow the production of single stranded RNA probes, and hybridisation conditions have been optimised for each probe. Considerable time has been devoted to refining the method to prepare leading strand DNA. Completion of this work will complement and extend the replication timing data presented in this thesis.

An alternative approach to mapping origins of replication is now possible due to the cloning of human replication proteins and the establishment of antibodies directed against them (Gavin *et al.* 1995, Quintana *et al.* 1997). Identification of where these proteins bind can be performed using an immunoprecipitation procedure based on that outlined in Strutt *et al.* (1997). In brief, chromatin associated proteins (such as the proteins involved in DNA replication), are cross-linked to cognate DNA binding sites *in vivo* using formaldehyde. The modified chromatin is subsequently purified and sheared, and DNA sequences associated with replication initation proteins isolated using immunoprecipitation. Purified DNA from the immunoprecipitated fraction can then be assayed for the enrichment of target sequences using quantitative PCR. It would be possible to screen the whole 325kb region of 16p for origins using PCR, since the entire sequence is available to facilitate the design of PCR primers.

9.3 The relationship between DNA replication timing and gene expression

The primary objective of this thesis was to investigate the relationship between replication timing and gene expression. Is gene expression regulated by replication timing (Holmquist 1987, Wolffe 1995), or do changes in replication timing occur as a secondary consequence of transcriptional activity and a pre-formed chromatin structure (Coverley and Laskey 1994)? Once the normal pattern of replication of the α globin cluster and the surrounding chromosomal region was established, this question could

be addressed. If downregulation of α globin expression is associated with a perturbation in replication timing, it would indicate that reproduction of the normal pattern of replication is important, and might imply that replication timing is a determinant of gene expression. Analysis of two α globin constructs in transgenic mice, show they maintain the normal pattern of early replication, despite the lack of fully regulated α globin gene expression. This indicates that early replication may be required, but it is not sufficient for correct expression to occur. What determines early replication of these constructs? Deletion of the distal regulatory element (HS-40), which is critical for α globin gene expression, does not perturb replication timing of the α globin cluster or the larger chromosomal region where the cluster is located. Therefore, HS-40 does not function by establishing or maintaining early replication of the α cluster and again, the result indicates that early replication is not sufficient for gene expression to occur. A natural mutant of the region, genotype $(-\alpha)^{ZF}$, where silencing of the α2 gene is associated with a repressive chromatin structure and methylation of the CpG island at the gene promoter, maintains an early pattern of replication. Thus, a perturbation in replication timing is not a prerequisite for these epigenetic modifications and gene silencing. In two other naturally occuring mutants, where deletions remove the telomere and subtelomeric region of 16p, and place the α globin genes proximal to an newly formed telomere, the timing of replication still remains generally early, although sequences juxtaposed to the new telomere appear to replicate later than in their normal interstitial chromosomal position. It cannot be said that early replication is not required for gene expression, as it has not been demonstrated that expression can occur when the cluster is late replicating. However, as early replication occurs during the suboptimal expression of the α genes, it cannot be the sole determinant of expression. It would have to work in conjunction with other mechanisms. Collectively, these results indicate that although the pattern of early replication timing may be a component of the general permissive chromosomal environment in which α globin gene expression may occur, DNA replication does not play a primary role in the regulation of their expression.

Rather than being a determinant of gene expression, early DNA replication may be a consequence of transcriptional acitivity. The chromosomal region containing the α globin cluster also contains many other genes, several of which are constitutively expressed. Transcription factors stimulating expression of these genes, may maintain the open chromatin environment of this region. This environment may allow replication factors to bind origin sequences. It is possible that transcription factors involved in the expression of the housekeeping genes, function in assembling or activating a replication initiation complex. Certainly the timing at which genes replicate seems tightly linked to chromatin structure. The constitutively 'open' chromatin structure found throughout this region displays a constitutively early replication pattern.

In summary, the work presented in this thesis provides a strong foundation for future experiments, to fully elucidate the relationship between gene expression, chromosome structure and DNA replication throughout the human 16p region. The identification of DNA sequences that function as origins of replication and the proteins that bind these sequences, will take us one step closer to understanding the process of replication, how origin firing time is controlled, and whether DNA replication is an important player in the regulation of gene expression.

References

Aasland, R., Gibson, T.J. & Stewart, A.F. (1995) The PHD finger: implications for chromatin-mediated transcriptional regulation. *Trends Biochem Sci*, **20**, 56-9.

Adams, R.L.P. (1991) *DNA replication.*
In focus series. IRL press at Oxford University Press, Oxford.

Anagnou, N.P., Yuan, T.Y., Lim, E., Helder, J., Wieder, S., Glaister, D., Marks, B., Wang, A., Colbert, D. & Deisseroth, A. (1985) Regulatory factors specific for adult and embryonic globin genes may govern their expression in erythroleukemia cells. *Blood,* **65**, 705-12.

Aladjem, M.I., Groudine, M., Brody, L.L., Dieken, E.S., Fournier, R.E., Wahl, G.M. & Epner, E.M. (1995) Participation of the human beta-globin locus control region in initiation of DNA replication. *Science*, **270**, 815-9.

Aparicio, O.M., Billington, B.L. & Gottschling, D.E. (1991) Modifiers of position effect are shared between telomeric and silent mating-type loci in S. cerevisiae. *Cell*, **66**, 1279-87.

Ariizumi, K., Wang, Z. & Tucker, P.W. (1993) Immunoglobulin heavy chain enhancer is located near or in an initiation zone of chromosomal DNA replication. *Proc Natl Acad Sci U S A*, **90**, 3695-9.

Armstrong, J., A. & Emerson, B.M. (1998) Transcription of chromatin: these are complex times. *Curr Opin Genet Dev,* **8**, 165-172.

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. & Struhl, K., (ed) (1994-1997) Current Protocols in Molecular Biology. Wiley Interscience, New York.

Barlow, D.P. (1997) Competition--a common motif for the imprinting mechanism? *EMBO J*, **16**, 6899-905.

Bell, S.P. & Stillman, B. (1992) ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. *Nature*, **357**, 128-34.

Berberich, S. & Leffak, M. (1993) DNase-sensitive chromatin structure near a chromosomal origin of bidirectional replication of the avian alpha-globin locus. *DNA Cell Biol*, **12**, 703-14.

Bernardi, G. (1989) The isochore organization of the human genome. *Annu Rev Genet*, **23**, 637-61.

Bernardi, G. (1993) The isochore organization of the human genome and its evolutionary history--a review. *Gene*, **135**, 57-66.

Bernet, A., Sabatier, S., Picketts, D.J., Ouazana, R., Morle, F., Higgs, D.R. & Godet, J. (1995) Targeted inactivation of the major positive regulatory element (HS-40) of the human alpha-globin gene locus. *Blood*, **86**, 1202-11.

Bickmore, W.A. & Carothers, A.D. (1995) Factors affecting the timing and imprinting of replication on a mammalian chromosome. *J Cell Sci*, **108**, 2801-9.

Bickmore, W.A. & Oghene, K. (1996) Visualizing the spatial relationships between defined DNA sequences and the axial region of extracted metaphase chromosomes. *Cell*, **84**, 95-104.

Bickmore, W.A. & Sumner, A.T. (1989) Mammalian chromosome banding--an expression of genome organization. *Trends Genet*, **5**, 144-8.

Bielinsky, A.-K. & Gerbi, S.A. (1998) Discrete start sites for DNA synthesis in the yeast ARS-1 origin. *Science*, **279**, 95-98.

Bilaud, T., Brun, C., Ancelin, K., Koering, C.E., Laroche, T. & Gilson, E. (1997) Telomeric localization of TRF2, a novel human telobox protein. *Nat Genet*, **17**, 236-9.

Birnboim, H.C. & Doly, J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res*, **7**, 1513-23.

Blumenthal, A.B., Kriegstein, H.J. & Hogness, D.S. (1974) The units of DNA replication in Drosophila melanogaster chromosomes. *Cold Spring Harb Symp Quant Biol*, **38**, 205-23.

Boggs, B.A. & Chinault, A.C. (1994) Analysis of replication timing properties of human X-chromosomal loci by fluorescence in situ hybridization. *Proc Natl Acad Sci U S A*, **91**, 6083-7.

Braunstein, J.D., Schulze, D., Del Giudice, T., Furst, A. & Schildkraut, C.L. (1982) The temporal order of replication of murine immunoglobulin heavy chain constant region sequences corresponds to their linear order in the genome. *Nucleic Acids Res*, **10**, 6887-902.

Broccoli, D., Godley, L.A., Donehower, L.A., Varmus, H.E. & de Lange, T. (1996) Telomerase activation in mouse mammary tumors: lack of detectable telomere shortening and evidence for regulation of telomerase RNA with cell proliferation. *Mol Cell Biol*, **16**, 3765-72.

Broccoli, D., Smogorzewska, A., Chong, L. & de Lange, T. (1997) Human telomeres contain two distinct Myb-related proteins, TRF1 and TRF2. *Nat Genet*, **17**, 231-5.

Brown, E.H., Iqbal, M.A., Stuart, S., Hatton, K.S., Valinsky, J. & Schildkraut, C.L. (1987) Rate of replication of the murine immunoglobulin heavy-chain locus: evidence that the region is part of a single replicon. *Mol Cell Biol*, **7**, 450-7.

Brown, K.E., Guest, S.S., Smale, S.T., Hahm, K., Merkenschlager, M. & Fisher, A.G. (1997) Association of transcriptionally silent genes with Ikaros complexes at centromeric heterochromatin. *Cell*, **91**, 845-54.

Brown, W.R., MacKinnon, P.J., Villasante, A., Spurr, N., Buckle, V.J. & Dobson, M.J. (1990) Structure and polymorphism of human telomere-associated DNA. *Cell*, **63**, 119-32.

Buckle, V.J., Higgs, D.R., Wilkie, A.O., Super, M. & Weatherall, D.J. (1988) Localisation of human alpha globin to 16p13.3pter. *J Med Genet*, **25**, 847-9.

Buckle, V.J. & Rack, K.A. (1993) Fluorescent in situ hybridisation. *Human Genetic Disease Analysis* (ed. by K. E. Davies) p. 59-82. IRL Press at Oxford University Press, Oxford.

Burhans, W.C. & Huberman, J.A. (1994) DNA replication origins in animal cells: a question of context? *Science*, **263**, 639-40.

Burhans, W.C., Vassilev, L.T., Wu, J., Sogo, J.M., Nallaseth, F.S. & De, P.M. (1991) Emetine allows identification of origins of mammalian DNA replication by imbalanced DNA synthesis, not through conservative nucleosome segregation. *EMBO J*, **10**, 4351-60.

Calza, R.E., Eckhardt, L.A., Del, G.T. & Schildkraut, C.L. (1984) Changes in gene position are accompanied by a change in time of replication. *Cell*, **36**, 689-96.

Cardoso, C., Timsit, S., Villard, L., Khrestchatisky, M., Fontes, M. & Colleaux, L. (1998) Specific interaction between the XNP/ATR-X gene product and the SET domain of the human EZH2 protein. *Hum Mol Genet*, **7**, 679-684.

Carlson, M. & Laurent, B.C. (1994) The SNF/SWI family of global transcriptional activators. *Curr Opin Cell Biol*, **6**, 396-402.

Carroll, S.M., De Rose, M.L., Kolman, J.L., Nonet, G.H., Kelly, R.E. & Wahl, G.M. (1993) Localization of a bidirectional DNA replication origin in the native locus and in episomally amplified murine adenosine deaminase loci. *Mol Cell Biol*, **13**, 2971-81.

Charnay, P., Treisman, R., Mellon, P., Chao, M., Axel, R. & Maniatis, T. (1984) Differences in human alpha- and beta-globin gene expression in mouse erythroleukemia cells: the role of intragenic sequences. *Cell*, **38**, 251-63.

Chodosh, L.A., Baldwin, A.S., Carthew, R.W. & Sharp, P.A. (1988) Human CCAAT-binding proteins have heterologous subunits. *Cell*, **53**, 11-24.

Chong, L., van Steensel, B., Broccoli, D., Erdjument-Bromage, H., Hanish, J., Tempst, P. & de Lange, T. (1995) A human telomeric protein. *Science*, **270**, 1663-7.

Chuang, L.S., Ian, H.I., Koh, T.W., Ng, H.H., Xu, G. & Li, B.F. (1997) Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1. *Science*, **277**, 1996-2000.

Collins, F.S. & Weissman, S.M. (1984) The molecular genetics of human hemoglobin. *Prog Nucleic Acid Res Mol Biol*, **31**, 315-462.

Coverley, D. & Laskey, R.A. (1994) Regulation of eukaryotic DNA replication. *Annu Rev Biochem*, **63**, 745-76.

Craddock, C.F., Vyas, P., Sharpe, J.A., Ayyub, H., Wood, W.G. & Higgs, D.R. (1995) Contrasting effects of alpha and beta globin regulatory elements on chromatin structure may be related to their different chromosomal environments. *EMBO J*, **14**, 1718-26.

Craig, J.M. & Bickmore, W.A. (1993) Chromosome bands--flavours to savour. *Bioessays*, **15**, 349-54.

Craig, J.M. & Bickmore, W.A. (1994) The distribution of CpG islands in mammalian chromosomes. *Nat Genet*, **7**, 376-82.

Cremer, T., Kurz, A., Zirbel, R., Dietzel, S., Rinke, B., Schrock, E., Speicher, M.R., Mathieu, U., Jauch, A., Emmerich, P. (1993) Role of chromosome territories in the functional compartmentalization of the cell nucleus. *Cold Spring Harb Symp Quant Biol*, **58**, 777-92.

Crossley, M. & Orkin, S.H. (1993) Regulation of the beta-globin locus. *Curr Opin Genet Dev*, **3**, 232-7.

Davie, J.R. (1998) Covalent modifications of histones: expression from chromatin templates. *Curr Opin Genet Dev*, **8**, 173-178.

de Lange, T. (1992) Human telomeres are attached to the nuclear matrix. *EMBO J*, **11**, 717-24.

de Lange, T., Shiue, L., Myers, R.M., Cox, D.R., Naylor, S.L., Killery, A.M. & Varmus, H.E. (1990) Structure and variability of human chromosome ends. *Mol Cell Biol*, **10**, 518-27.

Deisseroth, A. & Hendrick, D. (1978) Human alpha-globin gene expression following chromosomal dependent gene transfer into mouse erythroleukemia cells. *Cell*, **15**, 55-63.

Delgado, S., Gomez, M., Bird, A. & Antequera, F. (1998) Initiation of DNA replication at CpG islands in mammalian chromosomes. *EMBO J*, **17**, 2426-2435.

DePamphilis, M.L. (1988) Transcriptional elements as components of eukaryotic origins of DNA replication. *Cell*, **52**, 635-8.

DePamphilis, M.L. (1993) How transcription factors regulate origins of DNA replication in eukaryotic cells. *Trends Cell Biol*, **3**, 161-167.

Dernburg, A.F., Broman, K.W., Fung, J.C., Marshall, W.F., Philips, J., Agard, D.A. & Sedat, J.W. (1996) Perturbation of nuclear architecture by long-distance chromosome interactions. *Cell*, **85**, 745-59.

Dhar, V., Mager, D., Iqbal, A. & Schildkraut, C.L. (1988) The coordinate replication of the human beta-globin gene domain reflects its transcriptional activity and nuclease hypersensitivity. *Mol Cell Biol*, **8**, 4958-65.

Dhar, V., Skoultchi, A.I. & Schildkraut, C.L. (1989) Activation and repression of a beta-globin gene in cell hybrids is accompanied by a shift in its temporal replication. *Mol Cell Biol*, **9**, 3524-32.

Diffley, J.F. (1996) Once and only once upon a time: specifying and regulating origins of DNA replication in eukaryotic cells. *Genes Dev*, **10**, 2819-30.

Diffley, J.F. & Cocker, J.H. (1992) Protein-DNA interactions at a yeast replication origin. *Nature*, **357**, 169-72.

Diffley, J.F., Cocker, J.H., Dowell, S.J. & Rowley, A. (1994) Two steps in the assembly of complexes at yeast replication origins in vivo. *Cell*, **78**, 303-16.

Dijkwel, P.A. & Hamlin, J.L. (1988) Matrix attachment regions are positioned near replication initiation sites, genes, and an interamplicon junction in the amplified dihydrofolate reductase domain of Chinese hamster ovary cells. *Mol Cell Biol*, **8**, 5398-409.

Dillon, N. & Grosveld, F. (1994) Chromatin domains as potential units of eukaryotic gene function. *Curr Opin Genet Dev*, **4**, 260-4.

Dimitrova, D.S., Giacca, M., Demarchi, F., Biamonti, G., Riva, S. & Falaschi, A. (1996) In vivo protein-DNA interactions at human DNA replication origin. *Proc Natl Acad Sci U S A*, **93**, 1498-503.

Dobbs, D.L., Shaiu, W.L. & Benbow, R.M. (1994) Modular sequence elements associated with origin regions in eukaryotic chromosomal DNA. *Nucleic Acids Res*, **22**, 2479-89.

Drouin, R., Holmquist, G.P. & Richer, C.L. (1994) High-resolution replication bands compared with morphologic G- and R-bands. *Adv Hum Genet*, **22**, 47-115.

Dubey, D.D., Davis, L.R., Greenfeder, S.A., Ong, L.Y., Zhu, J.G., Broach, J.R., Newlon, C.S. & Huberman, J.A. (1991) Evidence suggesting that the ARS elements associated with silencers of the yeast mating-type locus HML do not function as chromosomal DNA replication origins. *Mol Cell Biol*, **11**, 5346-55.

Dutta, A. & Bell, S.P. (1997) Initiation of DNA replication in eukaryotic cells. *Ann Rev Cell Dev Biol*, **13**, 293-332.

Earnshaw, W.C. (1988) Mitotic chromosome structure. *Bioessays*, **9**, 147-50.

Edenberg, H.J. & Huberman, J.A. (1975) Eukaryotic chromosome replication. *Annu Rev Genet*, **9**, 245-84.

Efstratiadis, A. (1994) Parental imprinting of autosomal mammalian genes. *Curr Opin Genet Dev*, **4**, 265-80.

Eisen, J.A., Sweder, K.S. & Hanawalt, P.C. (1995) Evolution of the SNF2 family of proteins: subfamilies with distinct sequences and functions. *Nucleic Acids Res*, **23**, 2715-23.

Ellis, N.A. (1997) DNA helicases in inherited human disorders. *Curr Opin Genet Dev*, **7**, 354-63.

Enomoto, S., McCune-Zierath, P.D., Gerami-Nejad, M., Sanders, M.A. & Berman, J. (1997) RLF2, a subunit of yeast chromatin assembly factor-I, is required for telomeric chromatin function in vivo. *Genes Dev*, **11**, 358-70.

Epner, E., Forrester, W.C. & Groudine, M. (1988) Asynchronous DNA replication within the human beta-globin gene locus. *Proc Natl Acad Sci U S A*, **85**, 8081-5.

Epner, E., Rifkind, R.A. & Marks, P.A. (1981) Replication of alpha and beta globin DNA sequences occurs during early S phase in murine erythroleukemia cells. *Proc Natl Acad Sci U S A*, **78**, 3058-62.

Eyre-Walker, A. (1992) Evidence that both G + C rich and G + C poor isochores are replicated early and late in the cell cycle. *Nucleic Acids Res*, **20**, 1497-501.

Falaschi, A., Giacca, M., Zentilin, L., Norio, P., Diviacco, S., Dimitrova, D., Kumar, S., Tuteja, R., Biamonti, G., Perini, G. (1993) Searching for replication origins in mammalian DNA. *Gene*, **135**, 125-35.

Felsenfeld, G. (1996) Chromatin unfolds. *Cell*, **86**, 13-9.

Ferguson, B.M., Brewer, B.J., Reynolds, A.E. & Fangman, W.L. (1991) A yeast origin of replication is activated late in S phase. *Cell*, **65**, 507-15.

Ferguson, B.M. & Fangman, W.L. (1992) A position effect on the time of replication origin activation in yeast. *Cell*, **68**, 333-9.

Ferreira, J., Paolella, G., Ramos, C. & Lamond, A.I. (1997) Spatial organization of large-scale chromatin domains in the nucleus: a magnified view of single chromosome territories. *J Cell Biol*, **139**, 1597-610.

Flannery, A.V., Hirst, M.C., Knight, S.J., Ritchie, R.J. & Davies, K.E. (1995) The fragile X syndrome. *Biochim Biophys Acta*, **1271**, 293-303.

Flint, J., Bates, G.P., Clark, K., Dorman, A., Willingham, D., Roe, B.A., Micklem, G., Higgs, D.R. & Louis, E.J. (1997b) Sequence comparison of human and yeast telomeres identifies structurally distinct subtelomeric domains. *Hum Mol Genet*, **6**, 1305-13.

Flint, J., Craddock, C.F., Villegas, A., Bentley, D.P., Williams, H.J., Galanello, R., Cao, A., Wood, W.G., Ayyub, H. & Higgs, D.R. (1994) Healing of broken human chromosomes by the addition of telomeric repeats. *Am J Hum Genet*, **55**, 505-12.

Flint, J., Thomas, K., Micklem, G., Raynham, H., Clark, K., Doggett, N.A., King, A. & Higgs, D.R. (1997a) The relationship between chromosome structure and function at a human telomeric region. *Nat Genet*, **15**, 252-7.

Forrester, W.C., Epner, E., Driscoll, M.C., Enver, T., Brice, M., Papayannopoulou, T. & Groudine, M. (1990) A deletion of the human beta-globin locus activation region causes a major alteration in chromatin structure and replication across the entire beta-globin locus. *Genes Dev*, **4**, 1637-49.

Forrester, W.C., Novak, U., Gelinas, R. & Groudine, M. (1989) Molecular analysis of the human beta-globin locus activation region. *Proc Natl Acad Sci U S A*, **86**, 5439-43.

Forrester, W.C., Takegawa, S., Papayannopoulou, T., Stamatoyannopoulos, G. & Groudine, M. (1987) Evidence for a locus activation region: the formation of developmentally stable hypersensitive sites in globin-expressing hybrids. *Nucleic Acids Res*, **15**, 10159-77.

Fox, C.A., Ehrenhofer-Murray, A.E., Loo, S. & Rine, J. (1997) The origin recognition complex, SIR1, and the S phase requirement for silencing. *Science*, **276**, 1547-51.

Fox, C.A., Loo, S., Dillin, A. & Rine, J. (1995) The origin recognition complex has essential functions in transcriptional silencing and chromosomal replication. *Genes Dev*, **9**, 911-24.

Friedman, K.L., Diller, J.D., Ferguson, B.M., Nyland, S.V., Brewer, B.J. & Fangman, W.L. (1996) Multiple determinants controlling activation of yeast replication origins late in S phase. *Genes Dev*, **10**, 1595-607.

Funabiki, H., Hagan, I., Uzawa, S. & Yanagida, M. (1993) Cell cycle-dependent specific positioning and clustering of centromeres and telomeres in fission yeast. *J Cell Biol*, **121**, 961-76.

Furst, A., Brown, E.H., Braunstein, J.D. & Schildkraut, C.L. (1981) Alpha-globin sequences are located in a region of early-replicating DNA in murine erythroleukemia cells. *Proc Natl Acad Sci U S A*, **78**, 1023-7.

Gale, J.M., Tobey, R.A. & D'Anna, J.A. (1992) Localization and DNA sequence of a replication origin in the rhodopsin gene locus of Chinese hamster cells. *J Mol Biol*, **224**, 343-58.

Garrick, D., Fiering, S., Martin, D.I. & Whitelaw, E. (1998) Repeat-induced gene silencing in mammals. *Nat Genet*, **18**, 56-9.

Gavin, K.A., Hidaka, M. & Stillman, B. (1995) Conserved initiator proteins in eukaryotes. *Science*, **270**, 1667-71.

Gencheva, M., Anachkova, B. & Russev, G. (1996) Mapping the sites of initiation of DNA replication in rat and human rRNA genes. *J Biol Chem*, **271**, 2608-14.

Geyer, P.K. (1997) The role of insulator elements in defining domains of gene expression. *Curr Opin Genet Dev*, **7**, 242-8.

Giacca, M., Zentilin, L., Norio, P., Diviacco, S., Dimitrova, D., Contreas, G., Biamonti, G., Perini, G., Weighardt, F., Riva, S. (1994) Fine mapping of a replication origin of human DNA. *Proc Natl Acad Sci U S A*, **91**, 7119-23.

Gibbons, R.J., Bachoo, S., Picketts, D.J., Aftimos, S., Asenbauer, B., Bergoffen, J., Berry, S.A., Dahl, N., Fryer, A., Keppler, K. (1997) Mutations in transcriptional regulator ATRX establish the functional significance of a PHD-like domain. *Nat Genet*, **17**, 146-8.

Gibbons, R.J., Picketts, D.J. & Higgs, D.R. (1995b) Syndromal mental retardation due to mutations in a regulator of gene expression. *Hum Mol Genet*, **Spec No 4**, 1705-09.

Gibbons, R.J., Picketts, D.J., Villard, L. & Higgs, D.R. (1995a) Mutations in a putative global transcriptional regulator cause X-linked mental retardation with alpha-thalassemia (ATR-X syndrome). *Cell*, **80**, 837-45.

Gibbons, R.J., Suthers, G.K., Wilkie, A.O., Buckle, V.J. & Higgs, D.R. (1992) X-linked alpha-thalassemia/mental retardation (ATR-X) syndrome: localization to Xq12-q21.31 by X inactivation and linkage analysis. *Am J Hum Genet*, **51**, 1136-49.

Gilbert, D.M. (1998) Replication origins in yeast versus metazoa: separation of the haves and the have nots. *Curr Opin Genet Dev*, **8**, 194-199.

Gilson, E., Laroche, T. & Gasser, S.M. (1993) Telomeres and the functional architecture of the nucleus. *Trends Cell Biol*, **3**, 128-134.

Goldman, M.A., Holmquist, G.P., Gray, M.C., Caston, L.A. & Nag, A. (1984) Replication timing of genes and middle repetitive sequences. *Science*, **224**, 686-92.

Gossen, M., Pak, D.T., Hansen, S.K., Acharya, J.K. & Botchan, M.R. (1995) A Drosophila homolog of the yeast origin recognition complex. *Science*, **270**, 1674-7.

Gotta, M., Laroche, T., Formenton, A., Maillet, L., Scherthan, H. & Gasser, S.M. (1996) The clustering of telomeres and colocalization with Rap1, Sir3, and Sir4 proteins in wild-type Saccharomyces cerevisiae. *J Cell Biol*, **134**, 1349-63.

Gottschling, D.E., Aparicio, O.M., Billington, B.L. & Zakian, V.A. (1990) Position effect at S. cerevisiae telomeres: reversible repression of Pol II transcription. *Cell*, **63**, 751-62.

Gourdon, G., Sharpe, J.A., Wells, D., Wood, W.G. & Higgs, D.R. (1994) Analysis of a 70 kb segment of DNA containing the human zeta and alpha-globin genes linked to their regulatory element (HS-40) in transgenic mice. *Nucleic Acids Res*, **22**, 4139-47.

Greally, J.M., Starr, D.J., Hwang, S., Song, L., Jaarola, M. & Zemel, S. (1998) The mouse H19 locus mediates a transition between imprinted and non-imprinted DNA replication patterns. *Hum Mol Genet*, **7**, 91-5.

Grosveld, F., van Assendelft, G.B., Greaves, D.R. & Kollias, G. (1987) Position-independent, high-level expression of the human beta-globin gene in transgenic mice. *Cell*, **51**, 975-85.

Grunstein, M. (1997) Histone acetylation in chromatin structure and transcription. *Nature*, **389**, 349-52.

Hand, R. (1975) Regulation of DNA replication on subchromosomal units of mammalian cells. *J Cell Biol*, **64**, 89-97.

Hand, R. (1978) Eucaryotic DNA: organization of the genome for replication. *Cell*, **15**, 317-25.

Handeli, S., Klar, A., Meuth, M. & Cedar, H. (1989) Mapping replication units in animal cells. *Cell*, **57**, 909-20.

Hanna-Rose, W. & Hansen, U. (1996) Active repression mechanisms of eukaryotic transcription repressors. *Trends Genet*, **12**, 229-34.

Hanscombe, O., Vidal, M., Kaeda, J., Luzzatto, L., Greaves, D.R. & Grosveld, F. (1989) High-level, erythroid-specific expression of the human alpha-globin gene in transgenic mice and the production of human hemoglobin in murine erythrocytes. *Genes Dev*, **3**, 1572-81.

Hansen, R.S., Canfield, T.K., Fjeld, A.D., Mumm, S., Laird, C.D. & Gartler, S.M. (1997) A variable domain of delayed replication in FRAXA fragile X chromosomes: X inactivation-like spread of late replication. *Proc Natl Acad Sci U S A*, **94**, 4587-92.

Hansen, R.S., Canfield, T.K. & Gartler, S.M. (1995) Reverse replication timing for the XIST gene in human fibroblasts. *Hum Mol Genet*, **4**, 813-20.

Hansen, R.S., Canfield, T.K., Lamb, M.M., Gartler, S.M. & Laird, C.D. (1993) Association of fragile X syndrome with delayed replication of the FMR1 gene. *Cell*, **73**, 1403-9.

Hansen, R.S., Gartler, S.M., Scott, C.R., Chen, S.H. & Laird, C.D. (1992) Methylation analysis of CGG sites in the CpG island of the human FMR1 gene. *Hum Mol Genet*, **1**, 571-8.

Hatton, C.S., Wilkie, A.O., Drysdale, H.C., Wood, W.G., Vickers, M.A., Sharpe, J., Ayyub, H., Pretorius, I.M., Buckle, V.J. & Higgs, D.R. (1990) Alpha-thalassemia caused by a large (62 kb) deletion upstream of the human alpha globin gene cluster. *Blood*, **76**, 221-7.

Hatton, K.S., Dhar, V., Brown, E.H., Iqbal, M.A., Stuart, S., Didamo, V.T. & Schildkraut, C.L. (1988) Replication program of active and inactive multigene families in mammalian cells. *Mol Cell Biol*, **8**, 2149-58.

Heintz, N.H. (1992) Transcription factors and the control of DNA replication. *Curr Opin Cell Biol*, **4**, 459-67.

Herzing, L.B., Romer, J.T., Horn, J.M. & Ashworth, A. (1997) Xist has properties of the X-chromosome inactivation centre. *Nature*, **386**, 272-5.

Higgs, D.R. (1993) Alpha-thalassaemia. *Bailliere's Clinical Haematology International Practice and Research: The Haemoglobinopathies* (ed. by D. R. Higgs, D. J. Weatherall) p. 117-150. Bailliere Tindall, London.

Higgs, D.R., Sharpe, J.A., Gourdon, G., Craddock, C.F., Vyas, P., Picketts, D.J., Barbour, V., Ayyub, H. & Wood, W.G. (1995) Analysis of the human α-globin cluster in transgenic mice. *Molecular Biology of Hemoglobin Switching* (ed. by G. Stamatoyannopoulos) in press. Intercept, Andover, Hants, UK.

Higgs, D.R., Sharpe, J.A. & Wood, W.G. (1998) Understanding alpha globin gene expression: a step towards effective gene therapy. *Seminars in Hematology*, **35**, 1-14.

Higgs, D.R., Vickers, M.A., Wilkie, A.O., Pretorius, I.M., Jarman, A.P. & Weatherall, D.J. (1989) A review of the molecular genetics of the human alpha-globin gene cluster. *Blood*, **73**, 1081-104.

Higgs, D.R. & Wood, W.G. (1993) Understanding erythroid differentiation. *Curr Biol*, **3**, 548-550.

Higgs, D.R., Wood, W.G., Jarman, A.P., Sharpe, J., Lida, J., Pretorius, I.M. & Ayyub, H. (1990) A major positive regulatory region located far upstream of the human alpha-globin gene locus. *Genes Dev*, **4**, 1588-601.

Holmquist, G.P. (1987) Role of replication time in the control of tissue-specific gene expression. *Am J Hum Genet*, **40**, 151-73.

Holmquist, G.P. (1992) Chromosome bands, their chromatin flavors, and their functional features. *Am J Hum Genet*, **51**, 17-37.

Houseal, T.W. & Klinger, K.W. (1994) What's in a spot? *Hum Mol Genet*, **3**, 1215-6.

Hozak, P., Hassan, A.B., Jackson, D.A. & Cook, P.R. (1993) Visualization of replication factories attached to nucleoskeleton. *Cell*, **73**, 361-73.

Hozak, P., Jackson, D.A. & Cook, P.R. (1996) Role of nuclear structure in DNA replication. *Eukaryotic DNA replication* (ed. by B.D. Hames and D.M. Glover) p. 124-38. IRL Press at Oxford University Press, Oxford.

Huberman, J.A., Zhu, J.G., Davis, L.R. & Newlon, C.S. (1988) Close association of a DNA replication origin and an ARS element on chromosome III of the yeast, Saccharomyces cerevisiae. *Nucleic Acids Res*, **16**, 6373-84.

Hyrien, O., Maric, C. & Mechali, M. (1995) Transition in specification of embryonic metazoan DNA replication origins. *Science*, **270**, 994-7.

Hyrien, O. & Mechali, M. (1993) Chromosomal replication initiates and terminates at random sequences but at regular intervals in the ribosomal DNA of Xenopus early embryos. *EMBO J*, **12**, 4511-20.

Indrak, K., Gu, Y.C., Novotny, J. & Huisman, T.H. (1993) A new alpha-thalassemia-2 deletion resulting in microcytosis and hypochromia and in vitro chain imbalance in the heterozygote. *Am J Hematol*, **43**, 144-5.

Jackson, D.A. (1995) Nuclear organization: uniting replication foci, chromatin domains and chromosome structure. *Bioessays*, **17**, 587-91.

Jacob, F., Brenner, S. & Cuzin, F. (1963) On the regulation of DNA replication in bacteria. *Cold Spring Harbor Symp. Quant. Biol.*, **28**, 329-348.

James, C.D. & Leffak, M. (1986) Polarity of DNA replication through the avian alpha-globin locus. *Mol Cell Biol*, **6**, 976-84.

Jarman, A.P. & Higgs, D.R. (1988) Nuclear scaffold attachment sites in the human globin gene complexes. *EMBO J*, **7**, 3337-44.

Jarman, A.P., Wood, W.G., Sharpe, J.A., Gourdon, G., Ayyub, H. & Higgs, D.R. (1991) Characterization of the major regulatory element upstream of the human alpha-globin gene cluster. *Mol Cell Biol*, **11**, 4679-89.

Kadonaga, J.T., Carner, K.R., Masiarz, F.R. & Tjian, R. (1987) Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain. *Cell*, **51**, 1079-90.

Kamakaka, R.T. (1997) Silencers and locus control regions: opposite sides of the same coin. *Trends Biochem Sci*, **22**, 124-8.

Kass, S.U., Pruss, D. & Wolffe, A.P. (1997) How does DNA methylation repress transcription? *Trends Genet*, **13**, 444-9.

Kaufman, P.D., Kobayashi, R. & Stillman, B. (1997) Ultraviolet radiation sensitivity and reduction of telomeric silencing in Saccharomyces cerevisiae cells lacking chromatin assembly factor-I. *Genes Dev*, **11**, 345-57.

Kawame, H., Gartler, S.M. & Hansen, R.S. (1995) Allele-specific replication timing in imprinted domains: absence of asynchrony at several loci. *Hum Mol Genet*, **4**, 2287-93.

Kelly, R.E., De Rose, M.L., Draper, B.W. & Wahl, G.M. (1995) Identification of an origin of bidirectional DNA replication in the ubiquitously expressed mammalian CAD gene. *Mol Cell Biol*, **15**, 4136-48.

Kermouni, A., Van Roost, E., Arden, K.C., Vermeesch, J.R., Weiss, S., Godelaine, D., Flint, J., Lurquin, C., Szikora, J.P., Higgs, D.R. (1995) The IL-9 receptor gene (IL9R): genomic structure, chromosomal localization in the pseudoautosomal region of the long arm of the sex chromosomes, and identification of IL9R pseudogenes at 9qter, 10pter, 16pter, and 18pter. *Genomics*, **29**, 371-82.

Kipling, D. & Cooke, H.J. (1990) Hypervariable ultra-long telomeres in mice. *Nature*, **347**, 400-2.

Kitsberg, D., Selig, S., Brandeis, M., Simon, I., Keshet, I., Driscoll, D.J., Nicholls, R.D. & Cedar, H. (1993a) Allele-specific replication timing of imprinted gene regions. *Nature*, **364**, 459-63.

Kitsberg, D., Selig, S., Cedar, H. (1991) Chromosome structure and and eukaryotic gene organisation. *Curr Opin Genet Dev,* **1**, 534-7.

Kitsberg, D., Selig, S., Keshet, I. & Cedar, H. (1993b) Replication structure of the human beta-globin gene domain. *Nature*, **366**, 588-590.

Klein, F., Laroche, T., Cardenas, M.E., Hofmann, J.F., Schweizer, D. & Gasser, S.M. (1992) Localization of RAP1 and topoisomerase II in nuclei and meiotic chromosomes of yeast. *J Cell Biol*, **117**, 935-48.

Knoll, J.H., Cheng, S.D. & Lalande, M. (1994) Allele specificity of DNA replication timing in the Angelman/Prader-Willi syndrome imprinted chromosomal region. *Nat Genet*, **6**, 41-6.

Krude, T. (1995) Chromatin. Nucleosome assembly during DNA replication. *Curr Biol*, **5**, 1232-4.

Krysan, P.J., Smith, J.G. & Calos, M.P. (1993) Autonomous replication in human cells of multimers of specific human and bacterial DNA sequences. *Mol Cell Biol*, **13**, 2688-96.

Kumar, S., Giacca, M., Norio, P., Biamonti, G., Riva, S. & Falaschi, A. (1996) Utilisation of the same DNA replication origin by human cells of different derivation. *Nucleic Acids Res,* **24**, 3289-94.

Kurz, A., Lampel, S., Nickolenko, J.E., Bradl, J., Benner, A., Zirbel, R.M., Cremer, T. & Lichter, P. (1996) Active and inactive genes localize preferentially in the periphery of chromosome territories. *J Cell Biol*, **135**, 1195-205.

Lagarkova, M.A., Svetlova, E., Giacca, M., Falaschi, A. & Razin, S.V. (1998) DNA loop anchorage region co-localises with the replication origin located downstream to the human gene encoding lamin B2. *J Cell Bioch,* **69**, 13-18.

Laird, P.W. & Jaenisch, R. (1996) The role of DNA methylation in cancer genetic and epigenetics. *Annu Rev Genet*, **30**, 441-64.

Lamb, J., Harris, P.C., Wilkie, A.O., Wood, W.G., Dauwerse, J.G. & Higgs, D.R. (1993) De novo truncation of chromosome 16p and healing with (TTAGGG)n in the alpha-thalassemia/mental retardation syndrome (ATR-16). *Am J Hum Genet*, **52**, 668-76.

Lamond, A.I. & Earnshaw, W.C. (1998) Structure and function in the nucleus. *Science*, **280**, 547-553.

Larsen, F., Solheim, J., Kristensen, T., Kolsto, A.B. & Prydz, H. (1993) A tight cluster of five unrelated human genes on chromosome 16q22.1. *Hum Mol Genet*, **2**, 1589-95.

Latt, S.A. (1976) Optical studies of metaphase chromosome organization. *Annu Rev Biophys Bioeng*, **5**, 1-37.

Laurenson, P. & Rine, J. (1992) Silencers, silencing, and heritable transcriptional states. *Microbiol Rev*, **56**, 543-60.

Le Douarin, B., Nielsen, A.L., Garnier, J.M., Ichinose, H., Jeanmougin, F., Losson, R. & Chambon, P. (1996) A possible involvement of TIF1 alpha and TIF1 beta in the epigenetic control of transcription by nuclear receptors. *EMBO J*, **15**, 6701-15.

Lichter, P., Tang, C.J., Call, K., Hermanson, G., Evans, G.A., Housman, D. & Ward, D.C. (1990) High-resolution mapping of human chromosome 11 by in situ hybridization with cosmid clones. *Science*, **247**, 64-9.

Liebhaber, S.A., Griese, E.U., Weiss, I., Cash, F.E., Ayyub, H., Higgs, D.R. & Horst, J. (1990) Inactivation of human alpha-globin gene expression by a de novo deletion located upstream of the alpha-globin gene cluster. *Proc Natl Acad Sci U S A*, **87**, 9431-5.

Lin, S. & Kowalski, D. (1997) Functional equivalency and diversity of cis-acting elements among yeast replication origins. *Mol Cell Biol*, **17**, 5473-84.

Lohe, A.R. & Hilliker, A.J. (1995) Return of the H-word (heterochromatin). *Curr Opin Genet Dev*, **5**, 746-55.

Louis, E.J. (1995) The chromosome ends of Saccharomyces cerevisiae. *Yeast*, **11**, 1553-73.

Ludérus, M.E., van Steensel, B., Chong, L., Sibon, O.C., Cremers, F.F. & de Lange, T. (1996) Structure, subnuclear distribution, and nuclear matrix association of the mammalian telomeric complex. *J Cell Biol*, **135**, 867-81.

Lyon, M.F. (1992) Some milestones in the history of X-chromosome inactivation. *Annu Rev Genet*, **26**, 16-28.

Maillet, L., Boscheron, C., Gotta, M., Marcand, S., Gilson, E. & Gasser, S.M. (1996) Evidence for silencing compartments within the yeast nucleus: a role for telomere proximity and Sir protein concentration in silencer-mediated repression. *Genes Dev*, **10**, 1796-811.

Manders, E.M., Stap, J., Brakenhoff, G.J., van, D.R. & Aten, J.A. (1992) Dynamics of three-dimensional replication patterns during the S-phase, analysed by double labelling of DNA and confocal microscopy. *J Cell Sci*, **103**, 857-862.

Manuelidis, L. (1985) Individual interphase chromosome domains revealed by in situ hybridization. *Hum Genet*, **71**, 288-93.

Marahrens, Y. & Stillman, B. (1992) A yeast chromosomal origin of DNA replication defined by multiple functional elements. *Science*, **255**, 817-23.

Martin, D.I. & Whitelaw, E. (1996) The vagaries of variegating transgenes. *Bioessays*, **18**, 919-23.

Masukata, H., Satoh, H., Obuse, C. & Okazaki, T. (1993) Autonomous replication of human chromosomal DNA fragments in human cells. *Mol Biol Cell*, **4**, 1121-32.

Mathog, D., Hochstrasser, M., Gruenbaum, Y., Saumweber, H. & Sedat, J. (1984) Characteristic folding pattern of polytene chromosomes in Drosophila salivary gland nuclei. *Nature*, **308**, 414-21.

McCarroll, R.M. & Fangman, W.L. (1988) Time of replication of yeast centromeres and telomeres. *Cell*, **54**, 505-13.

Michaelson, J.S., Ermakova, O., Birshtein, B.K., Ashouian, N., Chevillard, C., Riblet, R. & Schildkraut, C.L. (1997) Regulation of the replication of the murine immunoglobulin heavy chain gene locus: evaluation of the role of the 3' regulatory region. *Mol Cell Biol*, **17**, 6167-74.

Micklem, G., Rowley, A., Harwood, J., Nasmyth, K. & Diffley, J.F. (1993) Yeast origin recognition complex is involved in DNA replication and transcriptional silencing. *Nature*, **366**, 87-9.

Mohr, I.J., Clark, R., Sun, S., Androphy, E.J., MacPherson, P. & Botchan, M.R. (1990) Targeting the E1 replication protein to the papillomavirus origin of replication by complex formation with the E2 transactivator. *Science*, **250**, 1694-9.

Molchanova, T.P., Pobedimskaya, D.D. & Huisman, T.H. (1994) The differences in quantities of alpha 2- and alpha 1-globin gene variants in heterozygotes. *Br J Haematol*, **88**, 300-6.

Morin, G.B. (1989) The human telomere terminal transferase enzyme is a ribonucleoprotein that synthesizes TTAGGG repeats. *Cell*, **59**, 521-9.

Morley, B.J., Abbott, C.A. & Wood, W.G. (1991) Regulation of human fetal and adult globin genes in mouse erythroleukemia cells. *Blood*, **78**, 1355-63.

Moyzis, R.K., Buckingham, J.M., Cram, L.S., Dani, M., Deaven, L.L., Jones, M.D., Meyne, J., Ratliff, R.L. & Wu, J.R. (1988) A highly conserved repetitive DNA sequence, (TTAGGG)n, present at the telomeres of human chromosomes. *Proc Natl Acad Sci U S A*, **85**, 6622-6.

Nakamura, H., Morita, T. & Sato, C. (1986) Structural organizations of replicon domains during DNA synthetic phase in the mammalian nucleus. *Exp Cell Res*, **165**, 291-7.

Old, J.M. & Higgs, D.R. (1983) Gene Analysis. *The thalassemias: methods in hematology* (ed. by D. J. Weatherall) p. 74-102. Churchill Livingstone, Edinburgh.

Orkin, S.H. & Goff, S.C. (1981) The duplicated human alpha-globin genes: their relative expression as measured by RNA analysis. *Cell*, **24**, 345-51.

Pak, D.T., Pflumm, M., Chesnokov, I., Huang, D.W., Kellum, R., Marr, J., Romanowski, P. & Botchan, M.R. (1997) Association of the origin recognition complex with heterochromatin and HP1 in higher eukaryotes. *Cell*, **91**, 311-23.

Palmiter, R.D. & Brinster, R.L. (1986) Germ-line transformation of mice. *Annu Rev Genet*, **20**, 465-99.

Picketts, D.J., Higgs, D.R., Bachoo, S., Blake, D.J., Quarrell, O.W. & Gibbons, R.J. (1996) ATRX encodes a novel member of the SNF2 family of proteins: mutations point to a common mechanism underlying the ATR-X syndrome. *Hum Mol Genet*, **5**, 1899-907.

Quackenbush, J., Davies, C., Bailis, J.M., Khristich, J.V., Diggle, K., Marchuck, Y., Tobin, J., Clark, S.P., Rodkins, A., Marcano, S. (1995) An STS content map of human chromosome 11: localization of 910 YAC clones and 109 islands. *Genomics*, **29**, 512-25.

Quintana, D.G., Hou, Z., Thome, K.C., Hendricks, M., Saha, P. & Dutta, A. (1997) Identification of HsORC4, a member of the human origin of replication recognition complex. *J Biol Chem,* **272**, 28247-51.

Renauld, H., Aparicio, O.M., Zierath, P.D., Billington, B.L., Chhablani, S.K. & Gottschling, D.E. (1993) Silent domains are assembled continuously from the telomere and are defined by promoter distance and strength, and by SIR3 dosage. *Genes Dev*, **7**, 1133-45.

Reynolds, A.E., McCarroll, R.M., Newlon, C.S. & Fangman, W.L. (1989) Time of replication of ARS elements along yeast chromosome III. *Mol Cell Biol*, **9**, 4488-94.

Romao, L., Cash, F., Weiss, I., Liebhaber, S., Pirastu, M., Galanello, R., Loi, A., Paglietti, E., Ioannou, P. & Cao, A. (1992) Human alpha-globin gene expression is silenced by terminal truncation of chromosome 16p beginning immediately 3' of the zeta-globin gene. *Hum Genet*, **89**, 323-8.

Rommens, J.M., Iannuzzi, M.C., Kerem, B., Drumm, M.L., Melmer, G., Dean, M., Rozmahel, R., Cole, J.L., Kennedy, D., Hidaka, N. (1989) Identification of the cystic fibrosis gene: chromosome walking and jumping. *Science*, **245**, 1059-65.

Ruvinsky, A.O. & Agulnik, A.I. (1990) Gametic imprinting and the manifestation of the fused gene in the house mouse. *Dev Genet*, **11**, 263-9.

Ryan, T.M., Behringer, R.R., Townes, T.M., Palmiter, R.D. & Brinster, R.L. (1989) High-level erythroid expression of human alpha-globin genes in transgenic mice. *Proc Natl Acad Sci U S A*, **86**, 37-41.

Saccone, S., Caccio, S., Kusuda, J., Andreozzi, L. & Bernardi, G. (1996) Identification of the gene-richest bands in human chromosomes. *Gene*, **174**, 85-94.

Saccone, S., De, S.A., Della, V.G. & Bernardi, G. (1992) The highest gene concentrations in the human genome are in telomeric bands of metaphase chromosomes. *Proc Natl Acad Sci U S A*, **89**, 4913-7.

Saccone, S., De, S.A., Wiegant, J., Raap, A.K., Della, V.G. & Bernardi, G. (1993) Correlations between isochores and chromosomal bands in the human genome. *Proc Natl Acad Sci U S A*, **90**, 11929-33.

Saitoh, Y. & Laemmli, U.K. (1994) Metaphase chromosome structure: bands arise from a differential folding path of the highly AT-rich scaffold. *Cell*, **76**, 609-22.

Sambrook, J., Fritsch, E.F. & Maniatis, T., (ed) (1989) *Molecular Cloning: a laboratory manual.* 2nd ed. Cold Spring Harbour Laboratory Press, Cold Spring Harbour.

Schardin, M., Cremer, T., Hager, H.D. & Lang, M. (1985) Specific staining of human chromosomes in Chinese hamster x man hybrid cell lines demonstrates interphase chromosome territories. *Hum Genet*, **71**, 281-7.

Schmeichel, K.L. & Beckerle, M.C. (1994) The LIM domain is a modular protein-binding interface. *Cell*, **79**, 211-9.

Schmidt, M. & Migeon, B.R. (1990) Asynchronous replication of homologous loci on human active and inactive X chromosomes. *Proc Natl Acad Sci U S A*, **87**, 3685-9.

Schwabe, J.W. & Klug, A. (1994) Zinc mining for protein domains. *Nat Struct Biol*, **1**, 345-9.

Selig, S., Okumura, K., Ward, D.C. & Cedar, H. (1992) Delineation of DNA replication time zones by fluorescence in situ hybridization. *EMBO J*, **11**, 1217-25.

Sharpe, J.A., Chan, T.P., Lida, J., Ayyub, H., Wood, W.G. & Higgs, D.R. (1992) Analysis of the human alpha globin upstream regulatory element (HS-40) in transgenic mice. *EMBO J*, **11**, 4565-72.

Sharpe, J.A., Summerhill, R.J., Vyas, P., Gourdon, G., Higgs, D.R. & Wood, W.G. (1993) Role of upstream DNase I hypersensitive sites in the regulation of human alpha globin gene expression. *Blood*, **82**, 1666-71.

Shinomiya, T. & Ina, S. (1991) Analysis of chromosomal replicons in early embryos of Drosophila melanogaster by two-dimensional gel electrophoresis. *Nucleic Acids Res*, **19**, 3935-41.

Sinnett, D., Flint, A. & Lalande, M. (1993) Determination of DNA replication kinetics in synchronized human cells using a PCR-based assay. *Nucleic Acids Res*, **21**, 3227-32.

Spivak, J.L., Toretti, D. & Dickerman, H.W. (1973) Effect of phenylhydrazine-induced hemolytic anemia on nuclear RNA polymerase activity of the mouse spleen. *Blood*, **42**, 257-66.

Stamatoyannopoulos, G. & Nienhuis, A.W. (1994) Hemoglobin switching. *The Molecular Basis of Blood Diseases* (ed. by G. Stamatoyannopoulos, A. W. Nienhuis, P. W. Majerus, H. Varmus) p. 107-155. W.B. Saunders Company.

Strauss, E.C., Andrews, N.C., Higgs, D.R. & Orkin, S.H. (1992) In vivo footprinting of the human alpha-globin locus upstream regulatory element by guanine and adenine ligation-mediated polymerase chain reaction. *Mol Cell Biol*, **12**, 2135-42.

Strehl, S., La, S.J. & Lalande, M. (1997) High-resolution analysis of DNA replication domain organization across an R/G-band boundary. *Mol Cell Biol*, **17**, 6157-66.

Strouboulis, J. & Wolffe, A.P. (1996) Functional compartmentalization of the nucleus. *J Cell Sci*, **109**, 1991-2000.

Strutt, H., Cavalli, G. & Paro, R. (1997) Co-localisation of polycomb protein and GAGA factor on regulatory elements responsible for the maintenance of homeotic gene expression. *EMBO J*, **16**, 3621-32.

Sussman, D.J., Klingensmith, J., Salinas, P., Adams, P.S., Nusse, R. & Perrimon, N. (1994) Isolation and characterisation of a mouse homologue of the Drosophila segment polarity gene dishevelled. *Dev Biol*, **166**, 73-86.

Taira, T., Iguchi, A.S. & Ariga, H. (1994) A novel DNA replication origin identified in the human heat shock protein 70 gene promoter. *Mol Cell Biol*, **14**, 6386-97.

Tasheva, E.S. & Roufa, D.J. (1994) A mammalian origin of bidirectional DNA replication within the Chinese hamster RPS14 locus. *Mol Cell Biol*, **14**, 5628-35.

Taylor, J.H. (1960) Asynchronous duplication of chromosomes in cultured cells of Chinese hamster. *J. Biophysic. Biochem. Cytol.*, **7**, 455-464.

Ten Hagen, K.G., Gilbert, D.M., Willard, H.F. & Cohen, S.N. (1990) Replication timing of DNA sequences associated with human centromeres and telomeres. *Mol Cell Biol*, **10**, 6348-55.

Tenzen, T., Yamagata, T., Fukagawa, T., Sugaya, K., Ando, A., Inoko, H., Gojobori, T., Fujiyama, A., Okumura, K. & Ikemura, T. (1997) Precise switching of DNA replication timing in the GC content transition area in the human major histocompatibility complex. *Mol Cell Biol*, **17**, 4043-50.

Tjian, R. & Maniatis, T. (1994) Transcriptional activation: a complex puzzle with few easy pieces. *Cell*, **77**, 5-8.

Torchia, B.S., Call, L.M. & Migeon, B.R. (1994) DNA replication analysis of FMR1, XIST, and factor 8C loci by FISH shows nontranscribed X-linked genes replicate late. *Am J Hum Genet*, **55**, 96-104.

Torchia, B.S. & Migeon, B.R. (1995) The XIST locus replicates late on the active X, and earlier on the inactive X based on FISH DNA replication analysis of somatic cell hybrids. *Somat Cell Mol Genet*, **21**, 327-33.

Travers, A.A. (1994) Chromatin structure and dynamics. *Bioessays*, **16**, 657-62.

Trempe, J.P., Lindstrom, Y.I. & Leffak, M. (1988) Opposite replication polarities of transcribed and nontranscribed histone H5 genes. *Mol Cell Biol*, **8**, 1657-63.

Triolo, T. & Sternglanz, R. (1996) Role of interactions between the origin recognition complex and SIR1 in transcriptional silencing. *Nature*, **381**, 251-3.

van Steensel, B. & de Lange, T. (1997) Control of telomere length by the human telomeric protein TRF1. *Nature*, **385**, 740-3.

Vassilev, L. & Johnson, E.M. (1990) An initiation zone of chromosomal DNA replication located upstream of the c-myc gene in proliferating HeLa cells. *Mol Cell Biol*, **10**, 4899-904.

Verbovaia, L. & Razin, S.V. (1995) Analysis of the replication direction through the domain of alpha globin-encoding chicken genes. *Gene*, **166**, 255-9.

Verbovaia, L.V. & Razin, S.V. (1997) Mapping of replication origins and termination sites in the Duchenne muscular dystrophy gene. *Genomics*, **45**, 24-30.

Vickers, M.A., Vyas, P., Harris, P.C., Simmons, D.L. & Higgs, D.R. (1993) Structure of the human 3-methyladenine DNA glycosylase gene and localization close to the 16p telomere. *Proc Natl Acad Sci U S A*, **90**, 3437-41.

Virta-Pearlman V.J., Gunaratne, P.H. & Chinault, A.C. (1993) Analysis of a replication initiation sequence from the adenosine deaminase region of the mouse genome. *Mol Cell Biol*, **13**, 5931-42.

Vourc'h, C., Taruscio, D., Boyle, A.L. & Ward, D.C. (1993) Cell cycle-dependent distribution of telomeres, centromeres, and chromosome-specific subsatellite domains in the interphase nucleus of mouse lymphocytes. *Exp Cell Res*, **205**, 142-51.

Vyas, P., Vickers, M.A., Simmons, D.L., Ayyub, H., Craddock, C.F. & Higgs, D.R. (1992) Cis-acting sequences regulating expression of the human alpha-globin cluster lie within constitutively open chromatin. *Cell*, **69**, 781-93.

Wang, S., Dijkwel, P.A. & Hamlin, J.L. (1998) Lagging-strand, early-labelling, and two-dimensional gel assays suggest multiple potential initiation sites in the Chinese hamster dihydrofolate reductase origin. *Mol Cell Biol*, **18**, 39-50.

Wansink, D.G., Manders, E.E., van, d.K.I., Aten, J.A., van, D.R. & de, J.L. (1994) RNA polymerase II transcription is concentrated outside replication domains throughout S-phase. *J Cell Sci*, **107**, 1447-1456.

Weatherall, D.J., Higgs, D.R., Bunch, C., Old, J.M., Hunt, D.M., Pressley, L., Clegg, J.B., Bethlenfalvay, N.C., Sjolin, S., Koler, R.D. (1981) Hemoglobin H disease and mental retardation: a new syndrome or a remarkable coincidence? *N Engl J Med*, **305**, 607-12.

Weiler, K.S. & Wakimoto, B.T. (1995) Heterochromatin and gene expression in Drosophila. *Annu Rev Genet*, **29**, 577-605.

Wellinger, R.J., Wolf, A.J. & Zakian, V.A. (1993) Origin activation and formation of single-strand TG1-3 tails occur sequentially in late S phase on a yeast linear plasmid. *Mol Cell Biol*, **13**, 4057-65.

Wiebauer, K., Gumucio, D.L., Jones, J.M., Caldwell, R.M., Hartle, H.T. & Meisler, M.H. (1985) A 78-kilobase region of mouse chromosome 3 contains salivary and pancreatic amylase genes and a pseudogene. *Proc Natl Acad Sci U S A*, **82**, 5446-9.

Wilkie, A.O., Higgs, D.R., Rack, K.A., Buckle, V.J., Spurr, N.K., Fischel, G.N., Ceccherini, I., Brown, W.R. & Harris, P.C. (1991) Stable length polymorphism of up to 260 kb at the tip of the short arm of human chromosome 16. *Cell*, **64**, 595-606.

Wilkie, A.O., Lamb, J., Harris, P.C., Finney, R.D. & Higgs, D.R. (1990a) A truncated human chromosome 16 associated with alpha thalassaemia is stabilized by addition of telomeric repeat (TTAGGG)n. *Nature*, **346**, 868-71.

Wilkie, A.O., Zeitlin, H.C., Lindenbaum, R.H., Buckle, V.J., Fischel, G.N., Chui, D.H., Gardner, M.D., MacGillivray, M.H., Weatherall, D.J. & Higgs, D.R. (1990b) Clinical features and molecular analysis of the alpha thalassemia/mental retardation syndromes. II. Cases without detectable abnormality of the alpha globin complex. *Am J Hum Genet*, **46**, 1127-40.

Wolffe, A.P. (1995) *Chromatin: structure and function.* (2nd ed.)
Academic Press Inc., San Diego, California.

Wolffe, A.P. (1994) Gene Regulation. Insulating chromatin. *Curr Biol*, **4**, 85-87.

Wu, C., Zannis-Hadjopoulos, M. & Price, G.B. (1993) In vivo activity for initiation of DNA replication resides in a transcribed region of the human genome. *Biochim Biophys Acta*, **1174**, 258-66.

Xiong, Z., Tsark, W., Singer-Sam, J. & Riggs, A.D. (1998) Differential replication timing of X-linked genes measured by a novel method using single-nucleotide primer extension. *Nucleic Acids Res,* **26**, 684-686.

Yurov, Y.B. & Liapunova, N.A. (1977) The units of DNA replication in the mammalian chromosomes: evidence for a large size of replication units. *Chromosoma*, **60**, 253-67.

Zakian, V.A. (1976) Electron microscopic analysis of DNA replication in main band and satellite DNAs of Drosophila virilis. *J Mol Biol*, **108**, 305-31.

Zakian, V.A. (1995) Telomeres: beginning to understand the end. *Science*, **270**, 1601-7.

Zeitlin, H.C. & Weatherall, D.J. (1983) Selective expression within the human alpha globin gene complex following chromosome-dependent transfer into diploid mouse erythroleukaemia cells. *Mol Biol Med*, **1**, 489-500.

Zeng, L., Fagotto, F., Zhang, T., Hsu, W., Vasicek, T.J., Perry, W.L., Lee, J.J., Tilghman, S.M., Gumbiner, B.M. & Costantini, F. (1997) The mouse Fused locus encodes Axin, an inhibitor of the Wnt signalling pathway that regulates embryonic axis formation. *Cell,* **90**, 181-92.

Zhao, Q.Z., Liang, X.L., Mitra, S., Gourdon, G. & Alter, B.P. (1996) Cloning and characterization of the mouse alpha globin cluster and a new hypervariable marker. *Mamm Genome*, **7**, 749-53.

Zhong, Z., Shiue, L., Kaplan, S. & de Lange, T. (1992) A mammalian factor that binds telomeric TTAGGG repeats in vitro. *Mol Cell Biol*, **12**, 4834-43.