Replication Timing Analysis of the Major Histocompatibility Complex on Human Chromosome 6

Peter Johonnett

Thesis submitted for the degree of Doctor of Philosophy at the University of London

Imperial Cancer Research Fund

April 2001
Corrections

Page 22, section 1.1.3, paragraph 2: deleted first sentence. For clarity first sentence should read as follows: "The absence of methyl groups is associated with the ability to be transcribed, while methylation results in gene inactivity."

Figures 4.1, 4.4, 4.6 and 5.4. The standard errors were calculated using the following formulae

The variance between doublet and singlet signals was calculated using

\[ \Theta = \frac{SD + 2DD}{2(SS + SD + DD)} = \frac{K_1 + 2K_2}{2N} \]

\[ \text{var} (\hat{\Theta}) = \frac{\text{var}(K_1)}{4N^2} + \frac{4\text{var}(K_2)}{4N^2} + \frac{4\text{covar}(K_1, K_2)}{4N^2} \]

\[ \text{var} (\hat{\Theta}) = \frac{1}{4N} \left\{ 4 \times \frac{SS}{N} \times \frac{DD}{N} + \frac{SD}{N} \times \left(1 - \frac{SD}{N}\right) \right\} \]

Where SS, SD and DD correspond to the number of nuclei with two singlets, a singlet and a doublet and two doublets, respectively. N equals the total number of nuclei. This formula assumes that the nuclei behave independently of each other. The standard error was calculated by finding the square root of the variance i.e. \( \{\text{var}(\hat{\Theta})\}^{1/2} \)

Page 83, section 4.3.2.2. paragraph 1, line 2: Should read as follows: "Lane 4 in Figure 4.2 and the analysis..."

Acknowledgements

Many thanks to my supervisor Dr Denise Sheer for giving me the opportunity to work in her laboratory and for all her expert guidance. I am particularly grateful for her unlimited support during the writing up period. Thanks also to my external supervisor Professor Sue Povey, who always provided me with balanced and helpful advice.

I would like to give a special thank you to Dr Peter Sasieni, with whom I enjoyed many stimulating and illuminating discussions. He provided some fascinating insights into my work, and showed patience beyond the call of duty.

Many thanks to all my friends in the Human Cytogenetics Laboratory: Pat Gorman, Rebecca Roylance, Edith Chevret, Emmanuela Volpi, Ceri Edwards, Tania Jones and Carol Shiels for helpful discussions. Thank you to Radost Vatcheva for her analysis of HLA-DRA mRNA. And a special thank you to my very good friend Jill "get yer gun" Williamson, without whose help I'd probably still be counting spots.

Thank you also to: John Trowsdale, Jethero Herberg, Stephan Beck, Roger Horton, Duncan Campbell, Jiannis Ragoussis, Rossen Donev and Douglas Higgs, who kindly shared ideas, materials and/or results. Thanks to everyone in the ICRF's Cell Production and FACS laboratories.

Thank you to the directors of Countrywide Porter Novelli for giving me the time I needed to finish this off.

Finally, thanks to Maggie for everything else.
for mum & dad.

this thesis is a product of their hard work.
Abstract

We employ fluorescence in situ hybridisation (FISH) to assay DNA replication in the human major histocompatibility complex (MHC). We examine whether features in the primary sequence and/or differences in transcription effect changes in replication timing. We perform these analyses on interphase chromosomes in the fixed S-phase nuclei of B-cells and fibroblasts.

The MHC shows later replication in the distal end of classical class II region, termed here the "HLA-DQ/DR" locus. The rate of change in mean replication time from the class III region to the HLA-DQ/DR locus is consistent with a replication fork moving within the accepted rate of between 0.3–6Kb/min. This is contrary to the suggestion that there is a "precise switch" from early to late replication across this region.

We show tissue-specific loci replicating earlier in B-cells than fibroblasts. However, the HLA-DQ/DR locus is later replicating regardless of whether or not it is expressed, providing the first evidence for an active gene cluster that is late replicating. We incubate fibroblasts with interferon-gamma in order to induce class II expression. Such cells adopt a replication timing profile similar to that in B-cells, including the late replication of the expressed HLA-DQ/DR locus. These data question the extent to which transcription influences replication time and suggest additional factors may be involved (at least in determining later replication).

In the class III region, the pattern of peaks and troughs in the data are remarkably similar in each replication profile, helping us to identify a replicon in the MHC. Two-colour FISH analysis confirm the existence of this replicon and provide the first demonstration that replicons separated by >1Mb of interphase chromatin can be simultaneously replicated. We suggest that in some cells, replicons in the class II and III regions may share the same replication machinery.
Table of Contents

Acknowledgements 1
Dedication 2
Abstract 3
Tables and Figures 10
Abbreviations 13

Chapter 1. Introduction 14

1.1. The Structural Organization of the Genome 14
  1.1.1. The banded Metaphase Chromosome 14
    1.1.1.1. Bands in other species 16
    1.1.1.2. Chromosome bands and the AT-queue 16
    1.1.1.3. The resolution of chromosome bands 17
  1.1.2. Isochores 19
    1.1.2.1. Isochores in other species 21
  1.1.3. CpG Islands 21
  1.1.4. Heterochromatin & Repetitive Sequences 22
    1.1.4.1. Tandemly repeated sequences 22
    1.1.4.2. Interspersed repeats 23
  1.1.5. Genes and Genetic Activity 24
  1.1.6. Nuclear Organization 25
    1.1.6.1. Chromosome territories 26

1.2. The Temporal Organization of the Genome 27
  1.2.1. Replication Banding 27
    1.2.1.1. Replicons and replication bands 28
    1.2.1.2. Replication and isochores 29
    1.2.1.3. Replication timing 30
  1.2.2. Eukaryotic DNA Replication 31
    1.2.2.1. Origin discovery 31
1.2.2. Origin usage 33
1.2.3. Replication and the Nucleus 33
1.3. The Human Major Histocompatibility Complex 34
  1.3.1. The Classical Class II Region and its Products 35
    1.3.1.1. Genes involved in the class I presentation pathway 36
    1.3.1.2. The extended class II region 37
  1.2. Control of HLA class II Gene Expression 37
    1.2.1. The role of IFN-γ in HLA class II expression 39
    1.2.2. IFN-γ and the class II transactivator (CIITA) 40
  1.3. The Class III Region 41
  1.4. The Class I Region and its Products 41
  1.5. The Isochore Structure of the MHC 42
  1.6. Repeat Elements in the MHC 44
1.4. Aims of this Thesis 45

Chapter 2. Materials and Methods

2.1. Materials 48
  2.1.1. Chemicals and Reagents 48
  2.1.2. Genomic Clones 49
2.2 Methods 51
  2.2.1. Cell culture 51
  2.2.2. Harvesting and slide making 53
  2.2.3. DNA extractions for FISH analysis (Midiprepes) 53
  2.2.4. Interferon induction of the class II region 54
    2.2.4.1. mRNA analysis of HLA-DRA 55
    2.2.4.2. Immunofluorescence analysis of class II expression 55
  2.2.5. Fluorescence in situ hybridization 56
Chapter 3. Using FISH to Assay Replication Time

3.1. Introduction 58
3.2. FISH Patterns in Interphase Nuclei 58
3.3. Development of the FISH Protocol 59
   3.3.1. The requirements for good quality FISH 57
   3.3.2. Scoring criteria 60
3.4. Replication Patterns in S-phase Nuclei 59
3.5. Pilot Study 63
   3.5.1. Doublets without replication 64
   3.5.2. Replication without doublets 65
3.6. Early and Late Replicating Controls 66
   3.6.1. Comparisons with established controls 67
3.7. Discussion 68
   3.7.1. Why FISH? 68
   3.7.2. Potential inconsistencies 70

Chapter 4. Cell Type-Specific Replication Timing Profiles of the Human Major Histocompatibility Complex

4.1. Introduction 72
   4.1.1. Hypotheses on the replication behaviour of the MHC 73
4.2. Methods 75
4.3. Results 75
   4.3.1. Replication of the MHC in B-cells 75
      4.3.1.1. The MHC is not entirely early replicating in B-cells 75
      4.3.1.2. B-cells have a temporal boundary in their MHC 77
      4.3.1.3. B-cells express class II molecules 79
4.3.2. Replication of the MHC in Fibroblasts

4.3.2.1. The classical class II is late replicating

4.3.2.2. Fibroblasts do not express MHC class II molecules

4.3.3. Replication Timing in B-cells Versus Fibroblasts

4.3.3.1. Comparing the classical class II region data

4.3.3.2. Comparing the MHC class III region data

4.4. Discussion

4.4.1. The Relationship Between Transcription and Replication

4.4.2. The Mouse H-2 Complex on Chromosome 17

4.4.3. The Temporal Change Across the Class II/III Boundary

4.4.3.1. Is there a sharp switch in replication timing?

4.4.3.2. Replication timing studies in other chromosomal regions

Chapter 5. The Effect of IFN-γ Induction on the Replication Time of the MHC Class II and Class III

5.1. Introduction

5.1.1 Hypotheses concerning IFN-γ induction in fibroblasts

5.2. Methods

5.3. Results

5.3.1. IFN-γ incubation times

5.3.2. Induced fibroblasts express MHC class II molecules

5.3.3. Chasing the IFN-γ induction
5.3.4. The MHC class II and III regions in induced fibroblasts

5.3.4.1. Induction fails to make the HLA-DQ/DR locus earlier replicating

5.3.4.2. Induction causes other MHC sequences to replicate earlier

5.3.4.3. The class III region in induced and non-induced fibroblasts

5.3.4.4. Induced fibroblasts adopt a B-cell replication profile

5.4. Discussion

5.4.1. Why is the HLA-DQ/DR locus unusual?

5.4.1.1. Linkage disequilibrium in the HLA-DQ/DR locus

5.4.1.2. MARs in the MHC

5.4.1.3. Are we measuring sister chromatid separation?

5.4.2. Patterns in the genome

5.4.2.1. Random initiation in two massive replicons?

5.4.2.2. Synchronously activated adjacent replicons

5.4.2.3. Asymmetry in the replicon structure of the MHC

Chapter 6. Evidence for at Least One Replicon in the MHC Class III Region

6.1. Introduction

6.1.1. Hypotheses on a putative replicon

6.2. Methods

6.3. Results

6.3.1. Rejecting the null hypothesis

6.3.2. KM2 and M7B are in the same replicon as G10s

6.3.3. G10s and co-probe doublets are on the same homologue
6.4. Discussion 124
   6.4.1. Identification of human MHC replicons 124
      6.4.1.1. The KM2-G10s-M7B replicons 124
      6.4.1.2. The classical class II region (probes O27 & A1) 125
   6.4.2. Levels of control in replicon activation 125
   6.4.3. The later replication time of the HLA-DQ/DR locus 127

Chapter 7. Final Discussion 129
7.1. Review 130
   7.1.1. Features in the genome 130
   7.1.2. Replication timing and gene expression 131
   7.1.3. Replication organization 132
7.2. Future Work 134
7.3. Has Theory Changed? 134

References 136

Appendix A. Using the FISH Assay to Estimate Mean Replication Time 153
Appendix B. Calculating the chromosome odds (Chapter 6) 158
Tables and Figures

Table 1.1  Chromosome banding techniques  15
Table 1.2  Structural and functional properties of G- and R-bands  18
Table 1.3  Isochores and chromosome bands  20
Table 3.1  Nuclei counts and estimates of the probability of observing a doublet signal for cosmid 11421 in B-cells  64
Table 3.2  The proportion of doublet signals observed in BrdU +ve and BrdU −ve nuclei using probe c1421  65
Table 3.3  Comparing % doublets between established controls and the cell lines used in this thesis  68
Table 5.1  The percentage of S-phase nuclei with SD signals for the probes surrounding the HLA-DQ/DR compared to the average of all the other probes  110
Table 6.1  The combined odds of observing G10s replication before the co-probe  120
Table 6.2  The ipsi- and contra-chromosome odds of observing a G10s doublet with a singlet for the co-probe  122
Table 6.3  The proportion of SD nuclei in which we observed both the G10s and co-probe doublets on the same homologue  123
Figure 1.1  Schematic of the conserved class II proximal promoter elements  38
Figure 1.2  The isochore structure of the MHC  45
Figure 2.1  Physical map of the human MHC  50
Figure 2.2  Variation in the organization of the DRB genes in the different DR-haplotypes  51
Figure 3.1  The FISH patterns seen in normal diploid nuclei  59
Figure 3.2  The five patterns of DNA replication identified in human MRC5 cells  62
Figure 4.1  The replication profile of the entire MHC in B-cells  76
Figure 4.2 Northern analysis of DRA mRNA in B-cells and fibroblasts

Figure 4.3 Immunofluorescence analysis of DRA expression on B-cells

Figure 4.4 The replication profile of the MHC class II and III regions in human fibroblasts

Figure 4.5 Immunofluorescence analysis of DRA expression on the surface of fibroblasts

Figure 4.6 Comparing the replication profiles in B-cells and fibroblasts

Figure 4.7 Adapted from Fig 4 in Tenzen et al. (1997)

Figure 5.1 Northern analysis of HLA-DRA mRNA in IFN-γ induced and non-induced MRC5 fibroblasts

Figure 5.2.a Immunofluorescence analysis of DRA expression on non-induced MRC5 cells

Figure 5.2.b DRA expression on IFN-γ induced MRC5 cells

Figure 5.3 Immunofluorescence analysis of MRC5 cells 48h after removal of the exogenous IFN-γ stimulus

Figure 5.4 Comparing the replication profiles of the class II and III regions in human fibroblasts incubated with IFN-γ for 24h and non-induced fibroblasts

Figure 5.5 The effect of induction on fibroblasts

Figure 6.1 The possible combinations of observable two-colour FISH signals

Figure 7.1 A comparison with other replication timing studies

Figure A.1 Theoretical distribution of replication time measurements

Figure A.2 Estimating the mean replication time

Figure A.3 Theoretical cumulative curve of replication timing

Figure A.4 Estimating the mean replication time from synchronized experiments
Figure A.5  Estimating the mean replication time from the cumulative curve 156

Figure B.1  Observable nuclei in two-colour FISH experiments 158
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARS</td>
<td>Autonomously replicating sequence</td>
</tr>
<tr>
<td>BrdU</td>
<td>Bromodeoxyuridine</td>
</tr>
<tr>
<td>CIITA</td>
<td>Class II transactivator</td>
</tr>
<tr>
<td>CCD</td>
<td>Cooled coupled devise</td>
</tr>
<tr>
<td>CEPH</td>
<td>Centre d'Etude Polymorphisme Humain</td>
</tr>
<tr>
<td>DAPI</td>
<td>4'6'-diamino-2-phenylindole</td>
</tr>
<tr>
<td>DMSO</td>
<td>Dimethyl sulphoxide</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>FISH</td>
<td>Fluorescence in situ hybridization</td>
</tr>
<tr>
<td>G-band</td>
<td>Giemsa-band</td>
</tr>
<tr>
<td>HLA</td>
<td>Human leukocyte antigen</td>
</tr>
<tr>
<td>ICRF</td>
<td>Imperial Cancer Research Fund</td>
</tr>
<tr>
<td>IFN-γ</td>
<td>Interferon-gamma</td>
</tr>
<tr>
<td>Kb</td>
<td>Kilobase (10^3)</td>
</tr>
<tr>
<td>LINE</td>
<td>Long interspersed nuclear element</td>
</tr>
<tr>
<td>MARs</td>
<td>Matrix attachment (or associated) regions</td>
</tr>
<tr>
<td>Mb</td>
<td>Megabase (10^6)</td>
</tr>
<tr>
<td>MHC</td>
<td>Major histocompatibility complex</td>
</tr>
<tr>
<td>MRS</td>
<td>MAR/SAR recognition sequence</td>
</tr>
<tr>
<td>PABL</td>
<td>Pseudoautosomal boundary-like sequence</td>
</tr>
<tr>
<td>PBS</td>
<td>Phosphate buffered saline</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>R-band</td>
<td>Reverse-band</td>
</tr>
<tr>
<td>SARs</td>
<td>Scaffold attachment (or associated) regions</td>
</tr>
<tr>
<td>SINE</td>
<td>Short interspersed nuclear element</td>
</tr>
<tr>
<td>SSC</td>
<td>Sodium citrate</td>
</tr>
</tbody>
</table>
Chapter 1. Introduction

The processes of DNA transcription, repair and replication are organized within the nucleus in both space and time. This organization has evolved in eukaryotes in order to meet the precise regulatory requirements needed for effective DNA metabolism, and is facilitated by specific interactions between the chromatin and the various structural and functional components of the nucleus e.g. (Spector, 1993; Jackson, 1995; Strouboulis & Wolffe, 1996; Sadoni et al., 1999; and Berezney & Wei, 1998). Do these DNA metabolic processes interact with each other? And to what extent are they a function of nuclear and genomic organization? One assay that could begin to address such questions was devised by Selig et al. (1992) – a variation on the theme of fluorescence in situ hybridization (FISH), it provides a relatively simple means of studying the timing of DNA replication in mammalian cells.

We have adopted this FISH approach and taken DNA replication as our functional reference point. We examine the human major histocompatibility complex (MHC) and test whether particular "genomic" features in its primary DNA sequence and/or differences in its transcriptional activity effect changes in its replication.

1.1. The Structural Organization of the Genome

1.1.1. The Banded Metaphase Chromosome

At mitosis the interphase nuclear chromatin condenses to form metaphase chromosomes. Various chemical treatments generate reproducible patterns of traverse bands along the length of the chromosome – the most familiar being the dye-based G-, or Giemsa and R-, or Reverse (to G) bands, see Table 1.1. Strongly staining G-bands are 'positive' and weakly or non-staining bands are 'negative'; negative G-bands are equivalent to positive R-bands and vice versa.
Banding Technique | Basis of Technique
---|---
G-banding | Giemsa staining after incubation with trypsin
R-banding | Giemsa staining after incubation with hot acidic saline
C-banding | Giemsa staining of heterochromatin after acid/alkali treatment and incubation in hot SSC
Q-banding | AT-specific fluorochromes: quinacrine, DAPI and Hoescht 33258
Negative Q-banding | GC-specific fluorochromes: chromomycin and mithramycin
Replication banding* | Incorporation of BrdU during either early or late S phase followed by Giemsa or antibody staining

*This is a function-based banding technique, see section 1.2.1.

Table 1.1. Chromosome Banding Techniques.

What do these banding patterns represent? Early studies with the highly AT-specific fluorochrome quinacrine showed positive G-bands to be relatively AT-rich (Caspersson et al., 1970). R-banding requires pre-treatment under conditions that denature AT-rich DNA, leaving the more resistant GC-rich bands stained by Giemsa. A subset of these R-bands, the T-bands, are the most resistant to denaturation and are mainly localized at telomeres, although a number are internal (Dutrillaux, 1973). In contrast to G-bands, R-bands are highlighted by fluorochromes with a preference for GC-rich DNA.

These early studies indicated that different chromosome bands have different molecular properties. These studies were important for stimulating initial interest in the genomic organization of DNA. However, it will become apparent in the following discussion that the reality is more complicated than the current cliché "GC-rich R-bands and GC-poor G-bands" would suggest (De Sario et al., 1997).
1.1.1.1. Bands in other species

The dye-based banding patterns seen on mammalian chromosomes are immutable during development and between cells from different tissues. This uniformity is thought to reflect a fundamental genomic organization. But how fundamental is banding, and therefore genomic organization, in other eukaryotes?

Bickmore and Craig (1997) examined banding from an evolutionary perspective. Summarizing the work of many authors, it appears that all eukaryotes with identifiable chromosomes contain a proportion of heterochromatin (see section 1.1.4 and Table 1.1) and so can be C-banded – this includes lower eukaryotes such as yeast. G- and R-banding appears in most mammalian and avian species, while in amphibia, fish, reptiles and plants, some species band and others do not e.g. (Maistro et al., 1999). Previously, all fish and plants had been thought refractory to banding, so perhaps those that still fail to band do so because of technical limitations rather than because they lack a band-like organization. Yeast chromosomes do not appear to have G- or R-bands, but they do share a fundamental banding pattern (and organization) with the higher eukaryotes: replication banding (see section 1.2.1).

1.1.1.2. Chromosome bands and the AT-queue

Currently, there is no consensus relating chromosome structure with the banding phenomenon. A model has been proposed by Saitoh and Laemmli based on their work with Indian muntjac chromosomes, (Saitoh & Laemmli, 1993 and 1994). They give a convincing account of banding in terms of higher-order chromatin structure, beginning with the accepted observation, originally by electron microscopy on histone-depleted metaphase chromosomes (Paulson & Laemmli, 1977) but since confirmed by gentler methods e.g. (Jackson, 1990), that the 30nm chromatin fibre is organized into loops (~100Kb) tethered to a proteinaceous scaffold (Laemmli et al., 1978; Marsden & Laemmli,
1979). These loops are anchored to the nuclear scaffold by short (~200bp) AT-rich sequences called scaffold-attachment (or associated-) regions (SARs) or matrix-attachment regions (MARs). As chromosomes condense the SARs line up to form what they term the AT-queue, and which can be usefully thought of as the backbone of the metaphase chromosome. They propose that the differential pathway and packaging of the AT-queue gives rise to the specific staining properties of G- and R-bands. In G-bands, the AT-queue is tightly packed, resembling a coiled spring. Here the DNA loops are short and parallel to the chromosome's length. Supporting this, a G-band like pattern was seen with immunofluorescence detection of topoisomerase II, a scaffold protein that co-operatively binds the AT-rich SARs (Adachi et al., 1989). In R-bands, the AT-queue is over-extended and the much larger DNA loops are at right angles to the chromosome's length. Saitoh and Laemmli use the image of a spring with periodic uncoiled gyres (R-bands) to illustrate the path of the core AT-queue.

This model is consistent with G-bands staining darkly with quinacrine or Giemsa because the AT-queue is densely packaged, and R-bands staining weakly because the AT-queue is stretched. It also provides a basis for banding without specifying the variation in base composition of the DNA, which is in any case small: G-bands being, on average, only 3.2% richer in AT than R-bands (Holmquist et al., 1982). Their model also accounts for some of the properties indicated in Table 1.2, such as DNase sensitivity and increased breakpoints resulting from the larger, more open chromatin loops in R-bands.

1.1.1.3. The resolution of chromosome bands

As chromosomes condense, bands coalesce and smaller bands become engulfed by larger ones. The resultant metaphase bands take on the appearance of the majority of their subbands. Evidence suggests that high-resolution R-bands fuse more often than G-bands (Drouin et al., 1991 and 1994). The human genome probably contains more than 2000
euchromatic chromosome 'bands' at the molecular level (averaging 1.3 Mb), but these are beyond the resolution of conventional cytogenetics. This makes the term 'chromosome band' problematic. The relatively low resolution of conventional cytogenetics (400 or 550 bands, averaging 4.7 to 6.5Mb) and the differing rates of compaction, means that in some cases assignment of a cloned DNA fragment to a light or dark band may be an oversimplification. Whilst cytogenetic studies on metaphase chromosomes suggest an underlying genomic organization, higher resolution studies are necessary to reveal the structural and functional minutiae.

<table>
<thead>
<tr>
<th>G-bands</th>
<th>R-bands</th>
</tr>
</thead>
<tbody>
<tr>
<td>GC-poor isochores (homogeneous)</td>
<td>GC-rich isochores (heterogeneous)</td>
</tr>
<tr>
<td>Low number of CpG islands</td>
<td>Higher number of CpG islands</td>
</tr>
<tr>
<td>Low gene density (tissue specific)</td>
<td>High gene density (house-keeping &amp; tissue specific)</td>
</tr>
<tr>
<td>Low level of histone acetylation</td>
<td>High level of histone acetylation</td>
</tr>
<tr>
<td>DNase I insensitive</td>
<td>DNase I sensitive</td>
</tr>
<tr>
<td>Few breakpoints/rearrangements</td>
<td>Higher number of breakpoints/rearrangements</td>
</tr>
<tr>
<td>LINE-rich</td>
<td>SINE-rich</td>
</tr>
<tr>
<td>Late replicating</td>
<td>Early replicating</td>
</tr>
</tbody>
</table>

**Table 1.2.** Structural and functional properties of G- and R-bands.

Important details uncovered by such analyses and successfully correlated with chromosome bands, are shown in Table 1.2, above. From this it is clear that chromosome bands represent distinct functional compartments of the nuclear chromatin. The distribution of particular sequences and, therefore, functions is not random along the
chromosome but is highly organized with respect to the bands. As these molecular properties are discussed their inter-relatedness will become clear.

1.1.2. Isochores

In the late 1970s, Bernardi and co-workers used density gradient centrifugation on nuclear DNA extracted from warm-blooded vertebrates, producing asymmetric profiles with four to five major fractions and several satellite fractions, each with its own modal density and base composition (Macaya et al., 1976). The major fractions were christened isochores (Cuny et al., 1981). In humans, the lightest (most buoyant) isochores are the GC-poor L1 (<38% GC) and L2 (~40%) families, representing about 63% of the genome. The heavy, GC-rich, isochores H1 (~45% GC), H2 (~50%) and H3 (~53%), represent 24.3, 7.5 and 4.7% of the genome, respectively. The remaining 3-4% is satellite and ribosomal DNA. Vertebrate genomes are mosaics of these compositionally homogenous isochores that extend for an average of >300Kb (Bernardi et al., 1985).

In humans, the GC-richest H3 isochore accounts for only 4.7% of the genome yet contains the highest gene concentration — 20-fold higher than the L isochores (Mouchiroud et al., 1991). Bernardi investigated the chromosomal distribution of the isochores by hybridizing them to metaphase (Saccone et al., 1993) and prometaphase chromosomes (Saccone et al., 1999). The hybridization patterns ranged from a relatively diffuse staining of every chromosomes through to an R-band like pattern. Specifically, the T-band subset hybridized the H3 isochore and part of the H2 and H1 families. R-bands, exclusive of T-bands, are formed to almost equal extents by the H1 isochore and GC-poor L1 and L2 isochores (with some contribution from the H2 and H3 families). G-bands consist of the GC-poor isochores L1 and L2 (with a minor H1 component).
Since gene concentration is low over GC-poor isochores and reaches a maximum in the GC-richest isochores (Zoubak et al., 1996), these results illustrate the distribution of genes along chromosomes. Supporting this distribution are earlier data showing that the GC levels of introns, exons and the third base pair of each codon show a positive correlation with the overall base composition of the isochrome in which they are found (Mouchiroud et al., 1991; Ikemura & Wada, 1991). This is summarized in Table 1.3.

<table>
<thead>
<tr>
<th></th>
<th>Isochrome</th>
<th>Gene Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>T-bands</td>
<td>H3 + H2 + H1 + L</td>
<td>++++</td>
</tr>
<tr>
<td>R-bands (exclusive of T-bands)</td>
<td>H3 + H2 + H1 + L</td>
<td>+++</td>
</tr>
<tr>
<td>G-bands</td>
<td>H1 + L</td>
<td>+</td>
</tr>
</tbody>
</table>

Table 1.3. Isochores and Chromosome Bands. L indicates both L1 and L2 isochores. Bold indicates the predominant isochrome family(ies), italics the minority family. Adapted from Bernardi (1995).

Thus, whilst it appears that G- and R-bands largely correspond to GC-poor and GC-rich DNA, this does not necessarily reflect the full complexity of sequence organization. The GC-richest isochores comprise approximately one third of human DNA, but R-bands make up almost half of the karyotype. Bernardi has repeatedly stressed that (i) the correlation between isochores and bands can only be considered an approximation; (ii) G-bands are homogeneous in their DNA composition because they house GC-poor isochores that differ little from each other; and (iii) R-bands, exclusive of T-bands, are heterogeneous because their constituent GC-rich isochores cover a wide GC range and (importantly) they contain some GC-poor isochores (Bernardi, 1993 and 1995).
1.1.2.1. Isochores in other species

The ancestral genomes of present-day mammals and birds underwent similar GC increases in their gene-rich regions relative to the gene-poor ones – effectively creating two subgenomes: the paleo- and the neogenome (Bernardi et al., 1993; D’Onofrio et al., 1999). This may have occurred in order to exploit the thermostability afforded to genes with greater GC versus AT. As a result of this non-uniform increase in GC, warm-blooded vertebrates show compositionally distinct isochore fractions. This includes birds, which is interesting since mammals and birds had independent origins of evolution. This implies that the GC increase was targeted to the gene-richest regions.

The paleogenomes of cold-blooded vertebrates, on the other hand, are characterized by narrow, symmetric profiles when prepared in CsCl gradients. That is to say, their genomes – which share a common ancestor with birds – show little compositional heterogeneity. Their GC-richest isochore is much lower in GC (~45%) than the GC-rich isochores of warm-blooded vertebrates (Bernardi, 1995). Despite this, their genes are also non-uniformly distributed: most are housed in their GC-richest fraction. This observation lends further weight to the theory that the compositional transition targeted the gene-richest parts of the paleogenome.

1.1.3. CpG Islands

The most striking epigenetic modification of mammalian DNA is the methylation of cytosines lying immediately 5' to guanines. Methylation occurs at between 60-90% of these CpG dinucleotides in the genome of a mammalian adult (Bird, 1986 and 1987). In the blastula, most of the DNA is unmethylated but as differentiation begins, a wave of methylation modifies most of the genome except for so-called CpG islands. These 0.2-2 Kb regions, found at the 5' end of mammalian genes, are dense clusters of unmethylated CpG dinucleotides (Cross & Bird, 1995). Around 60% of human genes (all house-
keeping and 40% of tissue-specific) possess most of the estimated 45,000 CpG islands in the genome (Antequera & Bird, 1993; Larsen, 1992).

Not all CpG islands are unmethylated. Methylation functions as part of a genome-wide transcriptional repression mechanism and, therefore, the CpG islands of tissue-specific genes are unmethylated only in tissues where the genes are expressed (reviewed in Kass et al., 1997). However, because they are independent of the level and site of expression of their associated genes, CpG islands provide a readily isolated and quantifiable marker of a large sub-set of genes. Craig and Bickmore (Craig & Bickmore, 1994) used FISH to map CpG islands on human metaphase chromosomes. They showed that >80% of CpG islands are in the 45% of the genome that corresponds to R-bands and that marked variation exists within the karyotype. Most CpG islands appear to map to the T-bands — the same sites as the GC-rich, gene-rich H3 isochore. This is consistent with Aissani and Bernardi (Aissani & Bernardi, 1991), who previously showed that the frequency of CpG islands increases in concert with the GC increase in isochores. Some human chromosomes are generally CpG island-rich (e.g. chromosomes 17, 19 and 22), whereas others are particularly island-poor (e.g. chromosomes 18, 13 and 21). These differences parallel the differences in the numbers of genes assigned to these chromosomes.

1.1.4. Heterochromatin & Repetitive Sequences

1.1.4.1. Tandemly repeated sequences

Approximately 3-4% of the human genome is repetitive DNA organized in tandem arrays (Singer, 1982a). These so-called satellite repeats correspond to Bernardi’s satellite isochore fractions. There are a number of satellite repeat families; each characterized by the repetition of a long monomer over many megabases of DNA. The satellites often correspond to constitutive heterochromatin and nuclease resistant C-bands described in Table 1.1 (Yunis et al., 1971), and see below.
Chromatin exists in one of two states: euchromatin or heterochromatin. Euchromatin is that part of the genome that is de-condensed in the interphase of the daughter cell — it is potentially 'active', and includes both R- and G-band DNA. Heterochromatin, on the other hand, does not de-condense and can be seen as darkly staining material in interphase as well as mitosis. It has two forms: facultative and constitutive. Facultative heterochromatin is under developmental control — previously euchromatic (it contains coding sequences, not tandem repeats) it has since been inactivated and condensed according to the developmental programme. Facultative heterochromatin is found in G-bands and may serve to silence many tissue-specific genes. An example of facultative heterochromatin's gene-silencing effect is seen in the inactive X chromosome in female somatic cells, where it serves to maintain dosage compensation of X-linked genes between the sexes (reviewed in Heard et al., 1997; Riggs & Pfeifer, 1992).

Constitutive heterochromatin is present throughout the life of the organism (Brown, 1966) and contains large arrays of satellite sequences and no coding sequences. Most satellite families are found at the centromeres of the chromosomes, at the site where the kinetochore will form before it pulls the sister chromatids apart during anaphase.

1.1.4.2. Interspersed repeats
In addition to large arrays of satellite repeats, eukaryotes house a variety of interspersed repeats which self-replicate and retrotranspose between chromosomes. They account for at least 20% of the human genome, and include short and long interspersed nuclear elements (SINEs and LINEs) (Korenberg & Rykowsk, 1988; Singer, 1982b). Each repeat class has a dominant family member. The major human SINE is the relatively GC-rich (56% GC) Alu family, with a consensus of approximately 300bp reiterated between $10^5$-$10^6$ times in the genome (Hwu et al., 1986). The major LINE is the relatively GC-poor (42%
GC) L1 family (not to be confused with Bernardi’s L1 isochore). This has a 6.4Kb consensus sequence, many members having serial deletions of the 5' end or internal deletions and rearrangements, but with most sharing the 3' end. The frequency of the 5' end is \(2 \times 10^4\) increasing to about \(10^5\) for the 3' end (Grimaldi et al., 1984; Hwu et al., 1986).

Korenberg and co-workers (1988) examined the distribution of these interspersed repeats by hybridizing \(Alu\) and L1 fractions to human metaphase chromosomes. Again the inverse nature of G- and R-bands was broadly confirmed: \(Alu\) hybridized largely to R-bands (the brightest regions corresponding to T-bands) and L1 hybridized to G-bands (and a sub-set of R-bands. Note that at 42% GC, L1 is more GC-rich than the bulk of the human genome). As with the other sequence properties discussed in the above sections, this finding linked genomic organization with chromosome structure and function.

The significance of the symbiotic relationship between retroposing sequences and their host genomes remains unclear, but the presence of extant interspersed repeats suggests that rather than reducing its fitness, certain repeat elements may actually have contributed to the evolution of the human genome (Beck et al., 1996; Andersson et al., 1998).

### 1.1.5. Genes and Genetic Activity

The true functional significance of metaphase bands is evinced by the fact that G-bands house only \(~20\%\) of mapped genes (mostly tissue-specific) whereas virtually all the widely-expressed, house-keeping genes and many tissue-specific genes map to R-bands, particularly the T-band subset (Holmquist, 1992; Craig & Bickmore, 1993). In addition, 30 000 expressed sequences (ESTs) have been mapped, and their overall distribution shows a concentration in R-bands (Deloukas et al., 1998).
Naturally, gene-associated properties have a concomitant distribution, see Table 1.2. For example, the modification of core histones within nucleosomes, in particular the progressive acetylation of histone H4, has been associated with active chromatin from yeast to man (Braunstein et al., 1993; Jeppesen et al., 1992), reviewed in (Wolffe, 1997). Fluorescence in situ hybridisation studies, using nucleosome-derived DNA as the probe, reveal patterns of histone acetylation on human chromosomes: with DNA from low-acetylated chromatin labelling several pericentric regions, whereas DNA from highly acetylated chromatin results in a pattern similar to R-banding (Breneman et al., 1996). Although visualized on metaphase chromosomes, these signals represent a "snapshot" of the genetic activity of the interphase chromatin. Supporting this are data showing that the nucleosomes associated with active CpG islands have hyperacetylated core histones and are depleted in the chromatin-inactivating histone H1 (Tazi & Bird, 1990).

The above sections present the human metaphase chromosome as a highly differentiated structure, and can be summarized by answering the question posed near the beginning of this chapter: what do chromosome bands represent? They are the most obvious manifestation of organization in a large and complex genome. At higher resolutions, the distributions and inter-relatedness of the bands' molecular properties reflect the evolution, organization and function of the genome.

1.1.6. Nuclear Organization

This introduction has so far focused on the distribution of sequences (and functions) along metaphase chromosomes. However, this state of extreme condensation is only temporarily adopted by the chromatin in cycling cells. Recent studies on intact (and in some cases living) cells have sought corresponding structural and functional distributions in the 3-dimensions of interphase nuclei. The picture emerging from these studies is of a dynamic system in which an architectural matrix and a evanescent skeleton of functional protein
complexes, interact with the chromatin to provide the organization vital for effective DNA metabolism. This organization is not rigid. The specificity of the nuclear interactions varies in response to both differentiation and the changing functional needs of the cell (for reviews see, (Strouboulis and Wolffe, 1996; Singer and Green, 1997; Craig et al., 1997; Berezney & Wei, 1998; Lamond & Earnshaw, 1998 and Belmont et al., 1999).

1.1.6.1. Chromosome territories

Interphase chromosomes were first visualized in hybrid cell lines – with one human chromosome inserted into the nucleus of a mouse cell. In situ hybridization with total human DNA showed that this single chromosome occupied a discrete area of the nucleus (Schardin et al., 1985; Manuelidis, 1985). This observation was confirmed using the fluorescence in situ hybridization technique of "chromosome painting", in which chromosomes are specifically labelled in interphase nuclei, e.g. (Cremer et al., 1988; Lichter et al., 1988). The space occupied by the chromosome was termed a territory – with different territories thought to have active genes on their surfaces and be separated from each other by interchromosomomal domains or channels (Cremer et al., 1993; Zirbel et al., 1993). These interchromosomal channels allow access to the nuclear machinery and are a convenient means of transporting transcription products to the nuclear pores. While recent evidence supports the existence of a channel-like system between the chromosomes (Bridger et al., 1998), the initial prediction that "positioning of genes at the surface of the chromosome territories is essential for their transcription" has been re-evaluated in light of the observation that transcription takes place throughout territories, predominantly at the surfaces of compact chromatin domains (Verschure et al., 1999). In addition, Visser et al. (1998) established that replication foci are distributed throughout the entire chromosome territories, in early as well as late S-phase. These results demonstrate
that DNA metabolism takes place throughout the chromosome territory and is not confined to the surface as previously thought.

Emphasis has shifted towards understanding the possible significance of the organization of the structural and functional sub-domains found within the chromosome territories (Ferreira et al., 1997; Zink et al., 1999; Zink et al., 1998; Verschure et al., 1999). In addition, Berezney and colleagues have suggested that territories owe their structural integrity to the nuclear matrix, since its disruption corresponds to loss of the discrete chromosome territories. The acidic non-histone proteins released immediately prior to the disintegration of the territories may be involved, at some fundamental level, in maintaining this integrity (Ma et al., 1999).

1.2. The Temporal Organization of the Genome

1.2.1. Replication Banding

The clearest manifestation of temporal genomic organization is the replication banding produced by incubating cells with bromodeoxyuridine (BrdU) and visualizing its incorporation into the chromosomes with specific antibodies (Vogel et al., 1989). Various blocking agents can control in which half of S-phase the BrdU is incorporated, to produce either an early- or a late-banding pattern. Dutrillaux et al. (1976) showed that each band replicated in a discrete period of S-phase and classified them into one of 18 replication intervals; with all but one of the R-bands replicating in the nine early intervals, while G-and C-bands occupy the remaining late intervals. Despite these analyses, the exact boundaries between replication bands have not been identified, and little is known about the bands' structure-function relationships. Replication bands appear to relate to genomic organization since there is a general correlation of early-replicating bands with R-bands (and T-bands) and late-replicating bands with G-bands (Holmquist et al., 1982),
suggesting that both structural and temporal bands have approximately the same boundaries.

Whereas G- and R-bands are only seen in some higher eukaryotes, replication banding seems to exist in all organisms, including yeast (Bickmore & Craig, 1997). The ubiquity of this temporal organization underlines the fundamental importance of DNA replication. This realization of replication's arch-importance underlines our need to understand how it is organized – particularly at levels of resolution higher than the metaphase chromosome.

1.2.1.1. Replicons and replication bands

Electron microscopy and fibre autoradiography have shown replication initiating at multiple sites in eukaryotic DNA (Huberman & Riggs, 1966 and 1968). From an initiation point, or origin, replication forks proceed bidirectionally until they fuse with forks initiated by adjacent origins. Fork movement rates vary according to species, cell type and time within S-phase but generally fall within the range of 0.3-6 Kb/min (Edenberg, 1975).

This unit of DNA replication, initiated at an origin and bound by the junctions with its neighbours, is defined as a replicon (Hand, 1978). Replicons are generally considered to be between 50-300Kb in mammals (Edenberg & Huberman, 1975), though more recent estimates put them in the region of 500Kb (Bickmore & Oghene, 1996; Verbovaia & Razin, 1997); and there is even evidence to suggest that some individual replicons might be >1Mb (Yurov & Liapunova, 1977). Clusters of replicons are sequentially activated throughout S-phase, with the origins of the replicons in the same cluster 'firing' in synchrony (Blumenthal, 1974). This functional grouping of replicons is evinced by the replication banding studies described above in which the bands are thought to represent discrete clusters of replicons, with those in R-bands firing earlier than those in G-bands (Hand, 1978; Craig & Bickmore, 1993). When considering this idea, however, it is
important to remember that as with G- and R-bands, replication bands may also have a sub-band structure. Dutrillaux's assignment of G- and R-bands to replication intervals was performed on low-resolution metaphase chromosomes, and subsequent studies show that genes and their immediate flanking regions can have different replication times to the surrounding chromatin depending on the stage of development, or tissue-type (Goldman et al., 1984; Holmquist, 1987; Hatton et al., 1988; Epner et al., 1988; Selig et al., 1992; Hyrien et al., 1995; and Kitsberg et al., 1993a). Therefore, molecular level subtleties in replication time do not necessarily produce an observable change in the gross pattern of replication bands (Craig & Bickmore, 1993).

1.2.1.2. Replication and isochores

Is it possible that isochores (averaging 300Kb) and replicons (in the general region of 50-500Kb) are one and the same? It has been asserted that R- and G-band DNA replicates early and late, respectively (Holmquist, 1987). And Bernardi (1993) has shown R-bands to be mainly GC-rich isochores while G-bands are completely GC-poor. The corollary of these observations is that GC-rich isochores will replicate early in the S-phase and GC-poor isochores replicate late.

Federico et al. (1998) hybridized the GC-rich, H3 isochore to Dutrillaux's replication bands and found it localized mainly in those bands classified as the earliest replicating (congruent with T-bands), whereas bands with no H3 hybridization (some R- and all G-bands) were classified as late replicating. Despite this correlation between GC/AT content and replication time, they caution that replication bands occur in all vertebrates, including cold-blooded ones – which have poor isochore fractionation (Schmid & Guttenbach, 1988). They also remind us that replication time appears to be a function not of GC content, but of gene concentration and transcription (Bernardi, 1989).
In addition, Eyre-Walker (1992) had previously provided evidence for both GC-rich and GC-poor isochores replicating in both early and late S-phase (Eyre-Walker, 1992). Eyre-Walker revealed considerable heterogeneity in the GC content of isochores replicated throughout S-phase. This poses a problem for the theory that isochores are maintained in warm-blooded vertebrates because different (early versus late) replicons are replicated in compositionally distinct nucleotide pools, biasing the pattern of mutation and generating the isochores with different GC content (Wolfe et al., 1989). The evidence for different isochores replicating simultaneously argues against this. The most likely explanation for the variation across isochores is that natural selection acts in vertebrates with high body temperatures to stabilise DNA, RNA, and important proteins encoded by GC-rich coding sequences (D'Onofrio et al., 1999; Eyre-Walker, 1999). This theory, however, does not predict the causal relationship, if any, between replicons and isochores.

1.2.1.3. Replication timing

The idea that transcription and replication time are linked comes from studies which show active genes replicating earlier than inactive ones (Goldman et al., 1984; Holmquist, 1987). After examining many house-keeping and tissue-specific loci, Schildkraut and co-workers established that almost all house-keeping genes replicate in the first half of S-phase in many cell types, as do some inactive tissue-specific genes (Hatton et al., 1988). Tissue-specific loci appear to fall into one of two groups. The first replicate early in S-phase whether they are expressed or not e.g., the mouse α-globin cluster and the mouse Ig-constant chain region (Hatton et al., 1988). The second group have a developmentally regulated pattern of replication whereby they replicate early in expressing cells, and late in non-expressing cells. The most extensively studied of this 'switching' type is the human β–globin locus, where over 200Kb of DNA is early replicating in expressing
erytholeukemia cells but late replicating in lymphocytes and HeLa cells (Dhar et al., 1989). No-one has yet identified an active gene or gene cluster that is late replicating.

### 1.2.2. Eukaryotic DNA Replication

DNA replication has been extensively characterized in yeast. Much is known about their specific origin and *cis* regulatory sequences, and the various trans-acting proteins these sequences bind. As yet, a similarly comprehensive picture of the genetics of higher eukaryotic replication is currently unavailable. Considerable progress has been made in understanding the biochemistry - which is remarkably conserved in eukaryotes - and much is known about the events that determine when and how replication begins and how it is limited to once per cell cycle (for reviews see (Rowles & Blow, 1997; DePamphilis, 1998). But although many vertebrates have recognizable genetic regions of replication activity, an unequivocal consensus of what actually constitutes these origins is lacking.

#### 1.2.2.1. Origin discovery

Origin identification in higher eukaryotes has been frustrated by the lack of a functional genetic assay. Yeast geneticists have an assay that relies on putative origin sequences behaving as autonomously replicating sequences (ARSs) when inserted into a suitable plasmid. These ARSs allow the plasmid to replicate when transfected into cells or incubated in cell extracts (Brewer & Fangman, 1987). Yeast ARSs are 100-200 bp in length, with an essential 11bp ARS consensus sequence, and include other sequence motifs that have accessory functions (e.g. transcription factor binding sites).

This assay faces a problem with the more complex genomes of higher eukaryotes in that — if they are >10Kb — most DNA fragments appear to confer autonomous replication (Krysan et al., 1989; Masukata et al., 1993). To compound the problem, these fragments lack any specific sequence conservation and smaller fragments sub-cloned from them have
no origin function (Krysan et al., 1993). This indicates that, in higher eukaryotes at least, sequence alone is not enough to specify where replication should begin. An alternative interpretation is that mammalian chromosomes contain many sequence motifs that can act as origins although some sequences are preferred (DePamphilis, 1993). Some origins have been localized to within a few kilobases e.g. human β-globin locus (Aladjem et al., 1998), whereas others resemble broad 'zones' in which initiation has an increased probability of occurring at a preferred site that may itself be surrounded by a number of other potential sites e.g. the dihydrofolate reductase locus (Dijkwel et al., 1994), reviewed in (Gilbert, 1998 and DePamphilis, 1999). The apparent lack of genetic precision may be resolved using the concept of a statistically favourable sequence of events in the initiation of replication.

After a careful comparison of six previously mapped origins, Dobbs et al. (1994) elaborated on an earlier proposal in which replication initiation in eukaryotes is determined by statistical probabilities: with any single initiation event dependent on the combined probabilities of unwinding of the DNA duplex and initiation of strand synthesis within the unwound region (Dobbs et al., 1994; Du et al., 1995 and Benbow, 1992). Their computational analysis identified clusters of modular sequence elements at origin regions from very different species. Each stretch of sequence contained one or more 'potential initiation regions' consisting of a putative DNA unwinding element aligned with clusters of scaffold associated regions (SARs), autonomously replicating (ARS) consensus sequences, and pyrimidine tracts. With the first draft of the human genome completed (International Human Genome Sequencing Consortium, 2001; Venter et al., 2001), more powerful computer programmes should be developed in order to examine the sequence for further clusters of potentially important sequence elements.
1.2.2.2. Origin usage
Sites used as origins are not necessarily fixed. This is clearly illustrated during development – when the actual number of origins used decreases as development progresses. In early Xenopus embryos, S-phase takes 15 of the 30 minute cell cycle. This rapid rate of replication is achieved by enormous numbers of origins arranged in small (~12Kb), uniform replicons (Hyrien & Mechali, 1993). These origins persist until the mid-blastula transition, when transcription begins. At this point, replicon size increases (as does the length of S-phase) and becomes more variable (Hyrien et al., 1995). What signals the developmental switch in origin usage is not known. Various investigators have suggested that functional origins are specified by the architectural configuration of the chromatin in the local environment (Hand, 1978; Coverly & Laskey, 1994; Burhans & Huberman, 1994; DePamphilis, 1998), which may be a function of clusters of elements which are themselves organized within the primary sequence.

1.2.3. Replication and the Nucleus
The process of DNA replication is temporally and spatially ordered in the nuclei of both mammalian and plant cells (Jackson and Pombo, 1998; Ma et al., 1998; Sparvoli et al., 1994; Zink et al., 1999; Zink et al., 1998). In the prevailing model, the biochemical necessities of DNA replication take place at "replication foci" – where it is thought that between 10-100 adjacent replicons are simultaneously replicated (Jackson & Pombo, 1998; Ma et al., 1998; Gobbi et al., 1999). In a recent reinterpretation of this model, Berezney and colleagues maintain that replication foci are heterogeneous. Foci are composed not only of clusters of small replicons as previously thought, but of ones that differ in both size and number (with some foci being composed of a single large replicon). The DNA in each of these replication foci is thought to be replicated in less than one hour (Edenberg & Huberman, 1975; Hand, 1978).
In the nucleus, three (Nakayasu & Berezney, 1989) to five (O'Keefe et al., 1992) spatial distribution patterns of replication foci can be recognized during S-phase. Replication starts at thousands of small foci evenly distributed throughout the nucleus (Jackson and Pombo, 1998; Ma et al., 1998), as S-phase progresses foci get bigger and become concentrated on the chromatin located near the nuclear envelope and around the nucleolus. By late S-phase, there are a few very large foci replicating the last chromatin domains deep within the nuclear interior (Manders et al., 1992).

Evidence that replication foci are associated with an underlying nuclear structure (be it matrix, scaffold or skeleton (Craig et al., 1997)) first came from Hozak et al. (1993) who encapsulated early S-phase HeLa cells in agarose, and incubated the permeabilized cells with labelled-dUTP. They used electron microscopy to visualize a number of dense, morphologically discrete "replication factories fixed to a skeleton" (Hozak et al. 1993). Other studies have confirmed that replication foci are maintained on the nuclear matrix after extraction of cells grown on cover-slips. Moreover, the isolated nuclear matrix was capable of synthesizing DNA at replication sites that were indistinguishable from those visualized in intact cells (Nakayasu & Berezney, 1989; Jackson & Cook, 1995). These and other results illustrating the stability of replication foci, support the hypothesis that they provide fundamental units of nuclear or chromosome architecture (Sparvoli et al., 1994; Jackson & Pombo, 1998; Zink et al., 1998 and 1999; Ma et al., 1998).

1.3. The Human Major Histocompatibility Complex (MHC)

The MHC (also known as the HLA complex) plays an essential role in the immune response, encoding the genes that process and present foreign and self antigens to T-lymphocytes. Initial interest in the MHC was generated by tissue transplantation experiments, hence the reference to histocompatibility. Today the entire 3.6Mb has been
sequenced and more than 200 genes and pseudogenes have been identified, many with important functions both immune and non-immune (see Figure 2.1). The sequence includes constitutively expressed, tissue-specific and inducible genes. An estimated 40% of the expressed genes have immune-related functions (Beck et al., 1999) – reflecting the fact that this region is the strongest genetic determinant in many auto-immune diseases.

Though all vertebrates appear to have an MHC, it has been most studied in human and mouse, for review see (Trowsdale, 1995). In humans the MHC maps to band 6p21.31. Historically, it is divided into three classes based on the clusters of genes found in each: (centromere to telomere) class II, III and I, see Figures 1.2 & 2.1 (Trowsdale & Campbell, 1997). There is an unusually high density of genes in the region relative to other parts of the genome and one family of genes, the human leukocyte antigen (HLA) genes, are the most polymorphic ever discovered, with nearly 900 alleles at eight loci (Parham & Ohta, 1996). The MHC loci are closely linked and, for this reason, an individual inherits the HLA alleles as two sets, each set being referred to as a haplotype (see Figure 2.2). In an outbred population, individuals are generally heterozygous for the HLA haplotypes and co-dominantly express both the maternal and paternal alleles. This results in a tremendous diversity of MHC molecules within the population.

1.3.1. The Classical Class II Region and its Products

The classical class II region (HLA-DP to HLA-DRA) covers ~800Kb. Its precise length differs depending on the number of DRB alleles in the DR-haplotype. It has the lowest gene density of the three classes with a one gene every 40Kb, and encodes the classic HLA-DR, -DP and -DQ class II antigen presenting molecules. It also contains the class II-related sequences, HLA-DMA, -DMB, -DO and -DN, the products of which serve as accessory proteins in antigen presentation by class II molecules (Beck & Trowsdale,
1999). Their relationship to the classic class II genes suggests that they have arisen via
gene duplication and subsequent divergence of function (Kasahara, 1999).

The class II molecules are cell surface glycopolypeptides made up of two different
different polypeptide
class II genes, except for DO, are arranged in pairs of α and β
genes. Both genes have a 5' leader exon encoding a short signal peptide followed by two
exons encoding the extracellular domains. There follows an exon encoding the
transmembrane region, and finally one or two terminal exons encoding the cytoplasmic
domains and 3' UTR. The α and β chains associate by non-covalent interactions and
expression of the resulting αβ-heterodimer is restricted to the surface of professional
antigen-presenting cells, such as macrophages and B-cells. However, cytokines such as γ-
interferon can induce non-lymphoid cells to express class II molecules. The α1 and β1
subdomains form a groove that binds peptides derived from exogenous proteins degraded
by the endocytotic processing pathway, and presents them to CD4+ T-helper cells.

1.3.1.1. Genes involved in the class I presentation pathway

In the midst of the class II region is a cluster of genes involved in antigen presentation
through HLA class I molecules: LMP2/LMP7 and TAP1/TAP2. The LMP genes encode
components of the proteasome responsible for degrading cytosolic self-proteins; and TAP
encodes a transporter associated with antigen processing, which delivers the resulting self-
peptides into the lumen of the endoplasmic reticulum, see (Momburg & Hammerling,
1998). In the lumen, the peptides co-operate with newly synthesized β2-microglobulin to
complex with class I molecules, forming the class I/peptide heterodimer present on
virtually every nucleated cell.
1.3.1.2. The extended class II region
Though convention divides the MHC into the classes I, II and III, sequence analysis shows that the classical class II region extends further centromeric than previously thought (Stephens et al., 1999). The ~300 Kb from HSET to HLA-DP is increasingly referred to as the 'extended class II region', and will be referred to throughout this thesis as such. Immunologically important genes in the extended class II include Tapasin – a molecule important for association of class I molecules with the TAP transporter; and RXRB – a molecule that binds to the MHC class I enhancer region and up-regulates expression. It also encodes DAXX, a intercellular signalling molecule that binds to the cytoplasmic domain of Fas, eventually leading to apoptosis (Herberg 1998a and 1998b).

1.3.2. Control of HLA Class II Gene Expression
Even in immunologic cells, HLA class II expression is not static and varies depending on the developmental state of the cells, as well as in response to various external stimuli (Glickman et al., 1997; Latron et al., 1988). Regulation of HLA class II gene expression occurs primarily at the level of transcription and, under most circumstances, the different α and β genes are co-ordinately regulated. A number of highly conserved cis-acting sequences are required for both basal and inducible HLA class II expression (Ting & Baldwin, 1993). These sequence elements are collectively termed the proximal promoter, and lie 160bp upstream of the transcription start site. Numerous DNA-binding proteins have been identified that interact with these various sequences, as outlined in Figure 1.1. In addition to their DNA-binding activities, these trans-acting proteins can regulate transcription through their interactions with each other.
Figure 1.1. Adapted from (Boss, 1997). Schematic of the conserved class II proximal promoter elements and the transcription factors they bind, shown for the HLA-DRA gene. The regions shown to be responsible for B-cell and IFN-γ induced expression are shown. The dotted portion of the IFN-γ indicates the elements required for maximal induction, see section 1.3.2.1.

The highly conserved promoter is sufficient to confer both constitutive and inducible HLA class II expression (Glimcher & Kara, 1992, Ting & Baldwin, 1993; Benoist & Mathis, 1990). It contains four sub-elements termed the W/Z/S, X1, X2 and Y boxes, whose sequences, orientations, relative positions and spacing are conserved in all of the α and β chain genes examined (Glimcher & Kara, 1992: Ting & Baldwin, 1993 and references therein). This promoter is also found 5' of the HLA-DM gene (Kearns et al., 1996). Such conservation of the proximal promoter allows the HLA class II genes to be co-ordinately regulated. Irrespective of the system analyzed, the class II proximal promoter behaves as a single functional unit in which all the sub-elements contribute to optimal promoter activity (Mach et al., 1996).
1.3.2.1. The role of IFN-γ in HLA class II expression

Interferon-γ (IFN-γ) induces class II expression on macrophages, keratinocytes, endothelial cells, fibroblasts, eosinophils, kidney, lung and liver cells (Glickman et al., 1997). MHC class II mRNA is first detected by Northern analysis 8 to 12h after IFN-γ stimulation, it peaks at 24 to 48h and diminishes after 72h (Bottger et al., 1988; Vidovic et al., 1990). Additionally, Bottger and colleagues (1988) showed that induction of class II expression requires new protein synthesis, since incubation with cyclohexamide prevented IFN-γ-induced expression of class II mRNA in a macrophage cell line.

Multiple signalling pathways are involved in IFN-γ induction; including tyrosine kinase, protein kinase C, the Na+/H+ antiporter and Ca++/calmodulin (Ting & Baldwin, 1993), but there are cell type-specific differences in which pathway is used. Tyrosine kinase activity during IFN-γ stimulation has been carefully dissected and well documented (Darnell et al., 1994). The tyrosine kinase pathway begins with the binding of IFN-γ to its receptor, causing the phosphorylation of a transcription factor called STAT-1α (Lee & Benveniste, 1996) which then binds the γ-activation sequence of a number of IFN-γ responsive genes. However, the IFN-γ receptor lacks intrinsic kinase activity, and instead two other kinases (the Janus kinases, JAK1 and JAK2) are phosphorylated on IFN-γ binding which then mediate tyrosine phosphorylation of STAT-1α. This produces the primary transcriptional responses of the cell to IFN-γ. A number of IFN-γ responses are 'secondary' because they are not manifest until long after the initiation of IFN-γ treatment.
(and not before the necessary protein synthesis). IFN-γ induction of the HLA class II genes is one such 'secondary' response.

1.3.2.2. IFN-γ and the class II transactivator (CIITA)

It is now accepted that IFN-γ causes the expression of the class II transactivator (CIITA) – a protein which does not bind DNA, but instead interacts with the proximal promoter-bound RFX complex, see Figure 1.1 (Steimle & Mach, 1995). CIITA expression is usually restricted to constitutively class II-positive cells (Steimle et al., 1993), but it is also obligatory for IFN-γ induced class II expression in a variety of cell types (Steimle et al., 1994; Chang et al., 1994). Depending on the cell type, CIITA mRNA appears at approximately 1 to 2h after IFN-γ stimulation (Lee & Benveniste, 1996; Steimle et al., 1994 and Chang et al., 1994) – this accounts for the delay in the appearance of class II mRNA.

CIITA was originally identified as the defective gene responsible for the lack of class II molecules on the cells of patients with a particular form of Bare Lymphocyte Syndrome (BLS) – a genetically heterogeneous disease in which patients have normal numbers of B- and T-cells, none of which exhibit the class II isotypes, either constitutively or after IFN-γ treatment. Mutations in the CIITA gene were shown to be responsible for the class II-negative status of patients in BLS complementation group A (Steimle et al., 1993).

CIITA has no DNA binding sequence - agreeing with the observation that it does not bind the class II proximal promoter and that patients in BLS complementation group A have normal RFX transcription factor binding at the X1 box (Kara & Glimcher, 1991). The N-terminal domain of CIITA exhibits transactivation function, whereas the C-terminal
domain confers transcriptional specificity by interacting with other class II transcription factors (Riley et al., 1995; Zhou & Glimcher, 1995). From these data it appears that CIITA co-ordinates class II transcription through its interaction with the DNA-bound RFX complex.

1.3.3. The Class III Region
The ~1Mb class III region (NOTCH4 to BATI) is the most gene dense class, with one every 11Kb. In contrast to the classical class II region, it is heterogeneous in its gene content and, remarkably, it contains no pseudogenes. Most of its genes are not immune-related, but immunologically important products include C4 isotypes, C2 and factor B - all of which are components of the complement cascade. The genes in the HSP70 family - encoding the interferon-inducible heat shock proteins - have been proposed to play a role in antigen presentation because of their ability to bind peptides (Srivastava et al., 1994). At its telomeric end, the class III region encodes several genes involved in the inflammatory response. It has been suggested that this gene cluster constitutes a distinct subregion - the class IV region (Gruen & Weissman, 1997). It includes the genes in the tumour necrosis factor cluster (TNF α and β, LTA and LTB), and a number of monocyte-expressed genes (including B1444, 1C7 and AIF-1).

1.3.4. The Class I Region and its Products
The 2Mb class I region (MIC-B to HLA-F) has a gene every 14Kb (Shiina et al., 1999) and contains the six expressed HLA class I genes: the three classical (HLA-A, B, and C) and three non-classical (HLA-E, F and G). It also contains two class I-chain related genes (MICA and MICB) which, unlike the classical class I molecules, can bind peptide without the assistance of β2-microglobulin. As with the HLA class II isotypes, the three classical HLA class I loci are highly polymorphic. Unlike the class II molecules, however, they
consist of a single, long polypeptide chain. However, they do share the same gene structure: all three loci have a 5' leader exon encoding a short signal peptide followed by 5 or 6 exons encoding the single large α-chain. The first three of these exons encode the extracellular α-subdomains. The next downstream exon encodes the transmembrane region. Finally, a 3' terminal exon or two exons encode the cytoplasmic domains and the 3' UTR.

The class I molecules are large glycoproteins found on virtually all nucleated cells; always non-covalently linked to a small peptide called β2-microglobulin. Endogenous peptides, derived from proteins that have been degraded in the cytoplasm are delivered via the TAP transporter into the endoplasmic reticulum and bind to the groove formed between the α1 and α2 domains. These peptides are presented to CD8+ cytotoxic T-cells. Each type of class I molecule binds a unique set of self-peptides, and a nucleated cell expresses thousands of these molecules, which collectively bind many different endogenous peptides. If the T-cell does not recognize the antigen presented, for example if the self-peptides are replaced by viral-derived peptides, it will mediate the destruction of the infected presenting cell.

1.3.5. The Isochore Structure of the MHC

As discussed in section 1.1.4, the human genome is a mosaic of isochores, and such isochore structures are visible in the MHC (see Fig 1.2). The classical class II region has been shown to be 40.2% GC, and therefore an L2 isochore. At 53% GC, the class III region is a more GC-rich H3 isochore. This high GC status is maintained in the class I
region, which averages ~46% GC and is an H1 isochore. The isochore boundary separating the class II and III regions has been found to contain sequences homologous (~80% nucleotide identity) to the pseudoautosomal boundaries on the human sex chromosomes that separate the sex-specific and pseudoautosomal regions (Fukagawa et al., 1995). The pseudoautosomal boundary-like sequence (PABL) is not at the exact GC% transition point, but Fukagawa et al. (1995) suggest that such sequences are characteristic of the interfaces between isochores and/or chromosome bands. Using their PABL sequence as a probe, they detected many PABL-type sequences in genomic DNA libraries, though not enough to account for every GC% boundary or, indeed, every band boundary (Fukagawa et al., 1996). They suggest that PABLs and PABL-type sequences exist only at certain types of boundary and that they may be just one of multiple sequence elements that define physical and temporal mosaic boundaries.

Figure 1.2. The isochore structure of the MHC. Note the marked difference in GC content of the L2 isochore in comparison to the adjacent regions. The isochore families are indicated along the bottom and landmark genes along the top. Adapted from The MHC Consortium (1999).
The classical class II region also has an isochore boundary at its centromeric end, where it becomes the extended class II region. The extended class II is \(~51\%\) GC (H2 isochore) and the transition is very sharp, but the region does not contain any pseudoautosomal boundary-like sequences (Stephens et al., 1999).

Given that G-bands contain predominantly L isochores, these findings might suggest that the classical class II region localizes to a G-band. However, FISH studies on metaphase chromosomes have mapped class II probes to the R-band 6p21.31 (Senger et al., 1993). As discussed in section 1.1.3, large bands do contain small subbands that are beyond the resolution of conventional cytogenetics. And indeed, high-resolution prometaphase G-banding has identified two thin G-subbands within 6p21.31 (Yunis, 1981), one of which could correspond to the relatively AT-rich classical class II region.

1.3.6. Repeat Elements in the MHC

The MHC houses many dense clusters of the interspersed repeat elements (SINES and LINES) introduced in section 1.1.4.2. Its classical class II region is also peppered with human endogenous retroviruses (from which SINEs and LINES are ultimately derived), reviewed in (Andersson et al., 1998). Indeed, the classical class II region has become a paradigm for studying the influence of retroelements on genomic activity, plasticity and evolution (Becker et al., 1996). Repeat elements make up about 23% of the class II region (Andersson et al., 1998) and contribute to the polymorphism in different DR-haplotypes (since particular retroelements are linked with particular DR-alleles).

Because the MHC class II region is among the most polymorphic and gene-dense regions of the genome, increased knowledge of its retroelements should have implications for understanding of the mechanisms that influence genomic plasticity and organization, and
may provide further potential candidate genes for auto-immune diseases such as type 1 diabetes and multiple sclerosis (Conrad et al., 1997; Perron et al., 1997).

1.4. Aims of this Thesis

This thesis focuses on DNA replication – since the temporal organization of this process is a fundamental feature of eukaryotic organisms. This realization of replication's arch-importance underlines our need to understand its organization at levels of resolution higher than the metaphase chromosome. The means to examine replication timing in interphase cells arrived with the work Howard Cedar and colleagues. Their insights into the capabilities of the FISH method have yielded a simple and effective means of examining the apparent correlations between replication timing, genomic organization and gene expression within the context of the interphase nucleus.

The main aims of this work are as follows:

- To test the validity of FISH as an approach to assaying the replication time of particular genetic loci
- To use FISH to analyze replication in the human major histocompatibility complex (This locus contains interesting "genomic features" and has a gene expression pattern which includes genes that are tissue-specific and/or IFN-γ inducible)
- To test current ideas concerning the temporal organization of mammalian DNA replication at the level of the interphase chromosome
Chapter 2. Materials and Methods

2.1 Materials

2.1.1. Chemicals and reagents

L-agar plates 1.5% w/v bacto agar in LB
Sterilized by autoclaving
Solid media was microwaved and cooled
before adding antibiotic (50mg/ml)

L-broth (LB) 10g bactotryptone
5g bacto yeast extract
10g NaCl to 1litre with distilled water

0.5% DMSO 0.5ml Dimethyl sulphoxide
9.5ml foetal calf serum

20 x SSC 3M NaCl
0.3M sodium citrate pH 7.0

TE 10mM Tris.Cl pH 8.0
1mM EDTA pH 8.0

The following solutions P1 - QF were supplied in the Qiagen-tip 100 (Midi) kit:

P1 solution (resuspension buffer) 50mM Tris.Cl pH 8.0
10mM EDTA
100mg/ml RNase A

P2 solution (lysis buffer) 200mM NaOH
1% sodium dodecyl sulphate (SDS)

P3 (neutralization buffer) 3M potassium acetate pH 5.5

QBT (equilibration buffer) 750mM NaCl
50mM MOPS pH 7.0
15% isopropanol
0.15% Triton X-100

QC (wash buffer) 1M NaCl
50mM MOPS pH 7.0
15% isopropanol

QF (elution buffer) 1.25M NaCl
50mM Tris.Cl pH 8.5
15% isopropanol
49

FISH hybridization buffer
- 50% deionized formamide
- 2 x SSC pH 7.0
- 10% dextran sulphate
- 2% Tween 20

Denaturation buffer
- 70% formamide
- 2 x SSC pH 7.0

2.1.2. Genomic clones

Forty-two cosmid clones covering the MHC were used (excluding controls); with the majority of the work focusing on the 30 clones containing the class II (both classical and extended) and class III regions. The specificity and efficiency of hybridization of each probe was initially checked by FISH on normal human metaphase chromosomes.

The following clones were used (see Fig 2.1): **class I region** (provided by Dr Steve Powis, Centre for Nephrology, Royal Free & University College Medical School, London): E2416, H0924, P1646, P1946, F0132, P1454, N0353, J2331, C0426, I1421, D0716 and G178 (Goldsworthy et al., 1996); **classical class II region** (provided by Professor John Trowsdale, formerly ICRF, London and Dr Stephan Beck, Sanger Centre, Hinxton): MANN3.6, A1, O14, O27, HA14, U15, U10, M4, M27B, F1121, E1448, PAKR and LH1 (Beck et al., 1996; Hanson et al., 1991); **extended class II region** (provided by Dr Jethro Herberg, formerly ICRF, London): A094, B2046, I0332, and P0717 (Herberg, 1998a and 1998b); **class III region** (provided by Dr Duncan Campbell, UK HGMP Resource Centre, Hinxton): C47, E91, D3a, A68, CH122, BF23, K101, KM2, G10s, F9M, M7B, TN82 and TN62 (Sargent et al., 1989; Kendall et al., 1990).

The control clones (see section 3.6 for more details): **late controls** were cosmids from the cystic fibrosis (CF) locus on chromosome 7, and were provided by Professor Douglas...
Figure 2.1. Physical map of the human 4Mb MHC, running from centromere to telomere. The relative positions of the cosmid clones used in this study are shown. * indicates IFN-γ inducible genes. Additional mapping information on chromosome 6 can be obtained from the Sanger Centre, http://www.sanger.ac.uk.
Higgs, Institute of Molecular Medicine, Oxford: cW44 and cJ21a (Rommens et al., 1989; Selig et al., 1992; Smith & Higgs, 1999); an additional late clone isolated from the ICRF chromosome 6 specific library using YAC 886c1 was provided by Dr Jiannis Ragoussis, Division of Medical Molecular Genetics, King's College, London: A9.5 (Davies et al., 1998). The early control was a cosmid specific for the glycogen phosphorylase gene (PYGM) on chromosome 11 and was also supplied by Professor Higgs, Oxford (Lichter et al., 1990; Selig et al., 1992; Smith & Higgs, 1999).

2.2 Methods

2.2.1. Cell culture

The human lymphoblastoid B-cell line, AHB, was supplied by the ICRF's Cell Production Laboratory. Originally from a Dutch female, this is an EBV-transformed cell line which retains two normal copies of chromosome 6. Unusually, it is homozygous for the DR3 haplotype in the DR52 group, see Figure 2.2, below.

![Variation in organization of the DRB genes in the different DR-haplotypes. Adapted from Trowsdale (1995).](image-url)
AHB was cultured in a 30ml suspension of RPMI-1640 with 10% foetal calf serum (FCS) at 37°C, in a humidified atmosphere of 95% air, 5% CO₂. When it reached approximately 10⁶ cells/ml, usually every third day, it was sub-cultured. Cells were spun at 300 x g for 5 minutes. The pellet was resuspended in 2ml of fresh medium and divided between two new flasks each containing 29ml of fresh medium. This gave approximately 5 x 10⁵ cells/ml.

MRC5 human lung fibroblasts had a normal male karyotype (its DR haplotype has not been determined). They were cultured as monolayers in RPMI-1640 with 10% FCS and 2% glutamine at 37°C, in a humidified atmosphere of 95% air, 5% CO₂. Cells were passaged when the cells reached confluency, usually every third day. Cells were detached with 1ml trypsin/versene, rinsed with fresh medium and divided between to new flasks at 30 - 50% confluency.

For normal chromosomes, 0.5ml of normal blood was cultured for 72 hours in 10ml RPMI-1640 with 10% FCS and 0.1ml reconstituted phytohaemagglutinin (PHA) (Life Technologies). Both the normal blood and the AHB cultures were incubated with colcemid (Gibco) at final concentration of 0.05 mg/ml 30 minutes before harvesting.

To identify S-phase cells, cultures were incubated with 100μM bromodeoxyuridine (BrdU) (Sigma), 30 minutes (normal blood and AHB) or 1 hour (MRC5) before harvesting. Initially, a parallel culture from each cell line was cultured without bromodeoxyuridine. It was harvested as described below except ice-cold 70% ethanol was used as the fixative. This preparation was sent to the ICRF's FACS laboratory to estimate the percentage of S-phase nuclei compared to the percentage determined by BrdU incorporation.
2.2.2. Harvesting and slide making

Cells were harvested by spinning at 300 x g for 5 min, removing the supernatant and resuspending in 10ml of 0.075M KCl. After incubation at 37°C for 15-20 minutes, cells had 0.5ml of ice-cold 3:1 methanol:glacial acetic acid added as a pre-fixative. Cells were pelleted again and the KCl removed. The pellets were resuspended in the residual hypotonic KCl and fix was carefully added up to a volume of 10ml. The cells were pelleted again and left in a second 10ml of fix over-night at 4°C. The following day, cells were pelleted and washed a further three times before being dropped onto slides.

Slides were soaked in methanol and wiped clean prior to use. Fixed cells were dropped onto wet, clean slides and air dried. The quantity and quality of the metaphases were checked using phase microscopy. If the chromosomes were insufficiently spread, the evaporation of the fixative would be slowed by drying the slide over a water-bath. The density of the cells was checked and adjustments made by the addition or removal of fixative. Slides were stored with desiccant at room temperature and used within one month.

2.2.3. DNA extractions for FISH analysis (Midipreps)

Cosmids were prepared using Qiagen-tip 100 midiprep columns according to the manufacturer's instructions, with modifications. A single bacterial colony was used to inoculate a starter culture of 2-5ml L-broth/antibiotic. This was incubated for ~8 hours at 37°C with vigorous shaking. The starter culture was then diluted (1/500 to 1/1000) into 100ml selective L-broth and incubated overnight at 37°C with vigorous shaking. Bacteria were harvested by centrifugation at 6000 x g for 15 minutes using a Beckman JA-10 rotor.

The pelleted bacteria were resuspended in 4ml P1 buffer leaving no cell clumps. P2 buffer was then added (4ml) and mixed gently by inverting the tube several times. This was left at
room temperature for 5 minutes. The P3 buffer was pre-chilled and 4ml used to neutralize
the P2. Again, this was mixed thoroughly but gently, and incubated on ice for 15 minutes.
The lysate was then spun in a Beckman JA-17 rotor at 20 000 x g for 30 minutes at 4°C.
The supernatant was transferred to a fresh tube and spun again for 15 minutes. During this
spin the Qiagen-tip was equilibrated by applying 4ml of buffer QBT and allowing the
column to empty by gravity flow.

The supernatant was applied to the equilibrated Qiagen-tip and allowed to enter the resin by
gravity flow. The Qiagen-tip was washed with 2 x 10ml of the QC buffer, and then the
cosmid DNA eluted using 5ml QF buffer (sometimes this was heated to 65°C prior to use).
The DNA was precipitated by adding 10ml cold absolute ethanol and incubating at -20°C
overnight. The DNA was pelleted by spinning at 15 000 x g for 30 minutes at 4°C, washed
with 2ml room temperature 70% ethanol, and spun again. The pellet was allowed to air dry
and then resuspended in 50-80μl TE pH 8.0.

2.2.4. Interferon induction of the MHC class II region
Various genes, including the HLA-class II genes, can be induced in non-immunological
cells using the cytokine interferon-γ (IFN-γ). This was performed on normal lung
fibroblasts, MRC5. Briefly, after sub-cultured (1:4) cells had settled down, they were
incubated for 24 h with 200U (10U/ml) of recombinant human IFN-γ (R&D Systems).
Two flasks were used to extract messenger RNA (see 2.2.4.1), one flask was used to
check for expression of the DRA class II molecule on the cell surface (see 2.2.4.2), and the
remaining flask was sub-cultured further for either 48 or 72 hours without IFN-γ present.
2.2.4.1. mRNA analysis of \( HLA-DRA \)

Messenger RNA was prepared using the TRIzol Reagent kit (Gibco, BRL) according to the manufacturer’s instructions. Between 5-10 \( \times 10^6 \) trypsinized or pelleted cells were suspended in 1ml TRIzol reagent. Cells were not washed before addition of TRIzol as this increases the possibility of mRNA degradation. The samples were incubated at room temperature for 5 minutes to permit the complete dissociation of nucleoprotein complexes. After the addition of 0.2ml of chloroform, the sealed tubes were shaken vigorously by hand for 15 seconds and incubated at room temperature for a further 2-3 minutes.

Samples were centrifuged at 12,000 \( \times g \) in a 4°C centrifuge for 15 minutes. After centrifugation, the RNA remains exclusively in the colourless upper aqueous layer. This aqueous layer was transferred to a fresh tube and the RNA precipitated by mixing with 0.5ml of isopropyl alcohol. The samples were incubated at room temperature for 10 minutes before centrifugation at 12,000 \( \times g \) for 10 minutes at 4°C. The RNA precipitate forms a gel-like pellet on the side and bottom of the tube.

After removing the supernatant, the RNA pellet was washed once in at least 1ml of 75% ethanol. The sample was vortexed, before its final 12,000 \( \times g \) centrifugation. The air-dried RNA pellet was resuspended by incubation at 55-60°C for 10 minutes in RNAse free water or 0.5% SDS solution.

2.2.4.2. Immunofluorescence analysis of class II expression

To test for expression of the class II molecule HLA-DRA, cells were incubated with an anti-DRA antibody and analysed using FACS. Briefly, in the case of MRC5, cells from one confluent medium flask were trypsinized and rinsed in RMPI. After a 5 minute spin at 300 \( \times g \), they were resuspended in 1ml of RPMI (no FCS). This was divided in to 2 x 0.5
ml. To one was added 200μl of FITC-labelled anti-HLA-DRA antibody (Becton Dickinson), the other was left as a negative control. Both were put on ice for 20 minutes. After this incubation, both were made up to 10ml with PBS and spun again. After resuspension in 0.5 ml PBS, the cells were fixed by adding 0.5 ml 1% formaldehyde. The samples were analysed on a FACsCalibur (Beckton Dickinson).

2.2.5. Fluorescence in situ hybridization

1-1.5mg of each probe was labelled with biotin-14-dATP (BRL Bionick kit) or digoxigenin-11-dUTP (Boehringer) according to the supplier’s instructions. To ensure that the probe was in the preferred size range (150 - 500 bp), 5μl of the labelled mixture was run on a 0.8% agarose gel. The probe was not purified: when needed, it was ethanol precipitated with salmon sperm DNA and human Cot1 DNA (Gibco BRL).

Between 200-300ng of probe was combined with 5μg of Cot1 DNA and precipitated using 2 volumes of 100% ethanol and incubation at -20°C overnight. The precipitated DNA was spun out and air dried. It was then resuspended in 11μl of hybridization buffer, denatured at 85°C for 5 minutes, and incubated at 37°C for 15-30 minutes to block repetitive elements in the probe.

Slides were equilibrated to room temperature, then artificially aged in a microwave. Slides were denatured with 100μl denaturation buffer at 74°C for 1.5 - 2 minutes and dehydrated through an ethanol series. The pre-annealed probe was then applied to the slide under a coverslip and incubated over two nights in a moist chamber at 37°C.
Standard post hybridization washes were 50% formamide, 2 x SSC (pH 7) at 42°C followed by 1 x SSC or 2 x SSC at 42°C. Slides were rinsed with 4 x SSC, 0.05% Tween 20 (SSCT) and pre-incubated with SSCT plus 5% low fat dried milk (Marvel) (SSCTM).

For the detection, all antibodies were diluted in SSCTM and 100μl of each was used at 37°C for 30 minutes. Biotin labelled probes were detected with 5 mg/ml avidin-FITC DCS (Vector Labs) followed by 5 mg/ml biotinylated anti-avidin (Vecor Labs) and a further round of avidin-FITC. The bromodeoxyuridine in the S-phase nuclei was detected by incubating for 20 minutes with 6mg/ml anti-bromodeoxyuridine (Boehringer) followed by a further 20 minutes with 20mg/ml anti-mouse Ig-rhodamine (Boehringer). Between antibodies, the slides were rinsed three times in SSCT on a shaking platform at room temperature for 3 minutes. Finally, they were washed twice in phosphate buffered saline for 3 minutes and dehydrated in an ethanol series, 3 minutes each. Metaphase spreads and nuclei were counter-stained blue with 0.2μg/ml DAPI (4',6'-diamidino-2-phenylindole) and mounted in Citifluor antifade solution (Citifluor Ltd).

Slides were visualized with a Zeiss Axioplan fluorescence microscope equipped with an automated filter wheel, Plan-Neofluar 100x oil immersion objective (Zeiss), 10x eye objectives and a 0.5–2x Optivar (Zeiss), giving between 1,000–2,000x magnification. A dual band pass filter (Vysis) was used for simultaneous visualization of FITC (green, probe signal) and rhodamine (red, S-phase nuclei). When collecting data, slides were scanned under the rhodamine filter alone and the hybridization signals in positively-stained S-phase nuclei were examined under the FITC filter alone or the dual-band pass filter. Some images were captured using a CCD camera (Photometrics) and image processing software (IPLabs).
Chapter 3. Using FISH to Assay Replication Time

3.1. Introduction

In the late 1980s, the introduction of fluorescence in situ hybridisation (FISH) revolutionized the mapping of DNA clones (Landegent et al., 1987; Lichter et al., 1988). By 1992 its potential for investigating the replication behaviour of these mapped regions was realized. In their seminal paper, Cedar and co-workers reasoned that the proportion of doublet FISH signals observed in interphase nuclei could be used to infer replication time (Selig et al., 1992). The advantage of their approach – in addition to its relative simplicity – is that it allows the study of replication in unamplified regions in asynchronous cells. This approach let us analyze large chromosomal segments in their in situ context: the interphase nucleus. For examples see, (Kitsberg et al., 1993a and 1993b; Boggs & Chinault, 1994; Bickmore & Carothers, 1995; Haaf, 1997; Strehl et al., 1997; Smith & Higgs, 1999; Squire et al., 2000).

3.2. FISH Patterns in Interphase Nuclei

In normal diploid nuclei, an unreplicated locus gives a singlet (S) hybridization signal while a replicated locus is characterized by a doublet (D), see Figure 3.1. The ratio of these two patterns in an asynchronous cell population indicates when the locus under investigation replicates. Clearly, later replicating sequences spend longer in an unreplicated state and have a higher proportion of singlets than their earlier replicating counterparts. (see Appendix A for calculation of mean replication times). Some nuclei house a singlet and a doublet (SD), where one chromosome has replicated the locus ahead of its homologous partner.
Figure 3.1. The FISH patterns (FITC, green) seen in normal diploid nuclei (rhodamine, red). (a) both homologues are unreplicated at the locus of interest (SS). (b) one homologue has replicated at that locus ahead of its partner (SD). (c) the locus of interest has replicated on both homologues (DD). Here S-phase nuclei were identified by pulse-labelling with BrdU and captured at 2,000x magnification.

Before the advent of FISH, replication time was estimated by Southern blot analysis of newly replicated, BrdU-labelled DNA isolated from S-phase cells that had been fractionated using elutriation or flow cytometry (for examples see, Little et al., 1993; Spack et al., 1992). This is time consuming and technically demanding but, since it was the 'gold standard' of its time, it was necessary for Selig et al. (1992) to use it to validate the FISH approach, which they did using the CF locus on human chromosome 7.

3.3. Development of the FISH Protocol

3.3.1. The requirements for good quality FISH

FISH requires high quality slides, with minimal cytoplasmic material, a reasonable density of flattened nuclei and numerous metaphase spreads (to assess hybridization efficiency).
To reduce background from the glass of the slides, they are 'blocked' with a solution of dried milk (e.g. Marvel) and washed between antibodies in X4 SSCT.

To increase the likelihood of specific hybridization, the nick translation of the DNA probes is optimized to give fragments in the desired size range (150 - 500 bp) and with the maximal amount of incorporated label. Too short, and the background signal increases on metaphases and nuclei because the fragments hybridize non-specifically. Too long, and the fragments stick to the slide itself rather than penetrate the nuclei.

Since the human genome contains many repetitive elements (see section 1.2.4.2.), any clone is likely to contain a fraction of Alu repeats and so produce some undesirable hybridization. To suppress this unwanted signal, probes are pre-annealed with an excess of unlabelled competitor Cot-1 DNA (Landegent et al., 1987; Lichter et al., 1988). Each probe is tested with varying amounts of Cot-1 to achieve the optimal signal:noise ratio. Finally, the stringency of the post-hybridization washes are different for different probes. The probes in this work required either X1 SSC or X2 SSC at 42°C or 37°C.

3.3.2. Scoring criteria

Slides were scanned under a dual Texas Red/FITC filter to identify S-phase nuclei, but nuclei were not scored if there was (i) weak and inefficient (<90%) hybridization on metaphases and in nuclei, (ii) high non-specific background within nuclei or (iii) high non-specific background over the entire slide. An individual nucleus on an otherwise acceptable slide was not scored if (i) it was damaged, (ii) it had too few signals, i.e. 1 or 0, (iii) it had multiple 'signals' or (iv) if its FISH signals could not be confidently assigned to one of the groups SS, SD or DD. Tallies were kept of nuclei with 0 or 1 signal (to indicate the efficiency of interphase hybridization) and of un-scored nuclei (to indicate the
overall quality of the experiment). To eliminate scorer bias, slides were scored 'blind'. The scorer would know the cell line but not the probe. Jill Williamson, a former scientific officer in the laboratory, and I performed all the hybridizations and were the only people to do the scoring (with the exception of the pilot study, see section 3.5).

3.4. Replication Patterns in S-phase Nuclei

The optimization steps described above were already part of the Human Cytogenetics Laboratory’s FISH protocol. In addition to this standard approach, S-phase nuclei were identified by incubating the cultures with bromodeoxyuridine (BrdU) before harvesting. The BrdU was incorporated into cycling cells and identified with rhodamine-labelled antibodies directed against the bromo-group (Kitsberg et al., 1993a). The percentage of cycling cells observed down the microscope was in agreement with that estimated by FACS analysis, at between 25-30%.

In 1992, O’Keefe et al. used fluorescence-, confocal laser- and immunoelectron microscopy to define five distinct patterns of DNA replication based on the nuclear distribution of incorporated BrdU (O’Keefe et al., 1992). Similar patterns are seen in a variety of cell types (Nakayasu, 1989; Neri et al., 1992; Sparvoli et al., 1994); therefore, the morphological patterns in the cells we intended to use were visualized for comparison. Asynchronous cultures were pulse-labelled with BrdU, which was subsequently detected in fixed nuclei using anti-BrdU. The distribution of the BrdU, and therefore of the newly replicated DNA, was examined by fluorescence microscopy. The five replication patterns were recognizable in both the foetal lung fibroblasts (shown in Fig 3.2 a-e) and B-cells. Since the cytological aspect of the assay performed as expected, it is reasonable to assume that these cells traverse S-phase in a manner typical of mammalian cells.
In early S-phase, replication starts at several hundred small foci (see section 1.2.3) distributed throughout the nucleoplasm, excluding the nucleolus (the nucleolus is clearly outlined in Fig 3.2.a). As S-phase continues, foci decrease in number but increase in size; replication of euchromatin decreases and the peripheral regions of heterochromatin begin to replicate (replication of the nuclear periphery is illustrated in Fig 3.2.b-c). By the end of S-phase, replication is confined to a few very large replication foci (Fig 3.2.d-e).

The previous studies which demonstrated a distribution of replication foci in different cell types, included one performed in pea root cells (Sparvoli et al., 1994). Such similarity across different cell types and completely unrelated organisms suggests that this replication distribution is a basic feature of the temporal and spatial organization of DNA replication in higher eukaryotes, see also (Berezney & Coffey, 1975; Zink et al., 1998 and 1999).
3.5. Pilot Study

The use of FISH to assay replication time was still relatively new when this study began. Because of this, the analytical precision (reproducibility) of the FISH approach needed to be established.

Three separately prepared AHB cultures were used. The same cosmid (I1421) from the MHC class I region was hybridized separately to two slides from each culture and the number of S-phase nuclei with the SS, SD or DD patterns were scored in 250 consecutive nuclei (i.e. 500 chromosomes). Different investigators have scored different numbers of nuclei, from 100 (Kitsberg et al., 1993a), through 100-200 (Selig et al., 1992), to >200 (Bickmore & Carothers, 1995). Dr Peter Sasieni of the ICRF’s Mathematics, Statistics and Epidemiology laboratory suggested 500 chromosomes per slide. He also devised a simple formula for estimating the mean replication time of a locus. The number of replication events that had occurred (i.e. the number of doublets, D) was divided by the number of potential sites for replication (i.e. the total number of S-phase chromosomes observed):

\[
\frac{SD+2DD}{2(SS+SD+DD)}
\]

This estimates the probability of observing a doublet on a randomly chosen S-phase chromosome. Whilst this formula assumes that replication at a given locus can be thought of as instantaneous, it recognizes that it may not be synchronous on both the homologues.

Table 3.1 (below) shows good precision for cosmid I1421 under these experimental conditions. Based on our samples of 250 nuclei, the greatest intra-culture range was only 6%, seen in culture A. The two most extreme inter-culture estimates were 0.54 and 0.60, again a difference of only 6%. This probe showed very little variation across different
slides and different cultures, therefore, 500 chromosomes was sufficient to give a precise indication of when in S-phase a locus replicates. The chi-squared test for the independence of values in the shaded rows and columns of table 3.1, yielded a p-value 0.057. Given then that the difference between the slides is not statistically significant, we can reasonably assume that this approach to assaying replication time yields data from which we can estimate the mean replication time of a locus.

<table>
<thead>
<tr>
<th></th>
<th>Culture A</th>
<th>Culture B</th>
<th>Culture C</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Slide 1</td>
<td>Slide 2</td>
<td>Slide 1</td>
<td>Slide 2</td>
</tr>
<tr>
<td>SS Nuclei</td>
<td>85</td>
<td>70</td>
<td>78</td>
<td>83</td>
</tr>
<tr>
<td>SD Nuclei</td>
<td>60</td>
<td>59</td>
<td>72</td>
<td>39</td>
</tr>
<tr>
<td>DD Nuclei</td>
<td>105</td>
<td>121</td>
<td>100</td>
<td>128</td>
</tr>
<tr>
<td>Total Nuclei</td>
<td>250</td>
<td>250</td>
<td>250</td>
<td>250</td>
</tr>
<tr>
<td>Proportion of</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Doublet</td>
<td>0.54</td>
<td>0.60</td>
<td>0.54</td>
<td>0.59</td>
</tr>
</tbody>
</table>

Table 3.1. Nuclei counts and estimates of the probability of observing a doublet signal for cosmid 11421 in B-cells.

3.5.1. Doublets without replication

Occasionally, some BrdU-negative nuclei appeared to contain doublets. The question had first been raised by Bickmore & Carothers (1995): are doublet signals always indicative of replication? On metaphase chromosomes, where chromatid numbers can be assessed, close examination reveals a small percentage of chromatids with 'doublet' FISH signals. These false positives are thought to be due to either spatial separation of the denatured chromatin strands or chromatin de-condensation (Houseal & Klinger, 1994). To estimate the extent of this 'doublet' phenomenon on interphase chromatin, I returned to the slides from the pilot study and counted the number of doublet chromosomes in BrdU-negative
nuclei (non-S-phase nuclei with clear doublets on both homologues were not scored as these were considered to be G2 nuclei). The results are shown in Table 3.2.

<table>
<thead>
<tr>
<th></th>
<th>Culture One</th>
<th>Culture Two</th>
<th>Culture Three</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slide 1</td>
<td>0.54</td>
<td>0.54</td>
<td>0.55</td>
</tr>
<tr>
<td>Slide 2</td>
<td>0.60</td>
<td>0.59</td>
<td>0.56</td>
</tr>
<tr>
<td>BrdU +ve nuclei</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BrdU -ve nuclei</td>
<td>0.07</td>
<td>0.08</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>0.07</td>
<td>0.05</td>
<td>0.06</td>
</tr>
</tbody>
</table>

Table 3.2. The proportion of doublet signals observed in BrdU +ve and BrdU -ve nuclei using probe c11421.

The above table illustrates that a small percentage (from 6% to 8%) of doublets can be seen without replication. Some replication may have taken place before cytologically detectable amounts of BrdU had been incorporated, but the fact that many BrdU-positive nuclei have no replication on either homologue would argue against this. This phenomenon is most likely caused by over-denaturation of the chromosomal DNA. Indeed, in the early stages of this work, some experiments were uncountable because of multiple signals in all nuclei. This occurred with freshly prepared, 'young' slides in which the chromosomal DNA is more easily and extensively denatured, so much so that the individual strands of the DNA duplex can be spatially resolved. I found artificially ageing the slides in a microwave the most effective way of reducing this problem.

3.5.2. Replication without doublets

Observing 'doublets' in non S-phase nuclei prompted the reciprocal question: Does a replicated locus always give a discernible doublet? Without the aid of confocal microscopy, the nuclei can only be examined in two of their three dimensions. Consequently, signals that are separated in three-dimensional space may be superimposed when viewed from a single perspective down the microscope.
To estimate the percentage of doublets that might appear as singlets, I captured ten DD
nuclei (20 doublet loci), with the spots in each doublet at varying distances apart. All
images were captured as originals (i.e. not pseudo-coloured) at x1000 magnification then
transferred to Adobe Photoshop from where they were printed at the same size. The
distances between the centres of each spot in the doublet pairs was measured.

The observed distance between the centres of the spots in each doublet is, at most, 2-3
times the diameter of one of the spots. Spots whose centres are less than one radius apart
may not, therefore, appear as a doublet but as a figure of eight. Using these data Dr.
Sasieni calculated that if their spots are only two diameters apart then, at most, 16% of
doublets could appear as singlets (asin(1/4) = 0.16*p/2). However, since this is assumed
to be a random process it should occur to the same extent for all probes and should not
affect the assessment of their relative replication times.

3.6. Early and Late Replicating Controls
The pilot study had shown that inter- and intra-experimental data were precise for the
probe II421. To ensure mean replication times were independent of the scorer and that the
assay was working in accordance with the accepted standards, control cosmids from
published FISH studies (with one exception) were hybridized and scored. These controls
were originally described by Selig et al. (1992) who showed their replication times to be
consistent with those times determined by elutriation (Furst et al., 1981).

The late controls were cosmids from the cystic fibrosis (CF) locus: cW44 and cJ21a
(Rommens et al., 1989). These were supplied by Professor Douglas Higgs, Institute of
Molecular Medicine, Oxford. Previous FISH analysis shows these probes to be late
replicating in cell lines that do not express the CF protein, for example EBV-transformed lymphoblastoid cells such as that used here (Selig et al., 1992) and see (Smith and Higgs, 1999). An additional late replicating probe, A9.5 had been isolated from the ICRF chromosome 6 specific cosmid library. Unlike the MHC, the region containing this clone was mapped by FISH to the gene-poor G-band, 6p24. (Davies et al., 1998). This probe was supplied by Dr Jiannis Ragoussis, Division of Medical Molecular Genetics, King’s College, London.

The early replicating control was a cosmid specific for the glycogen phosphorylase gene (PYGM), first described by Lichter et al. (1990) and shown to be early replicating by Selig et al. (1992) and see Smith & Higgs (1999). This cosmid (also supplied by Professor Higgs) is early replicating in all cell lines tested to date.

3.6.1. Comparisons with established controls

In Selig’s work (1992), the doublets were scored in all nuclei and corrected to represent those in S-phase. The numbers presented throughout this thesis are from scoring S-phase nuclei directly. Table 3.3 (over) indicates how well FISH in our laboratory agreed with the studies performed by others.

Clones cW44 and cJ21 were late replicating in both B-cells and fibroblasts, whereas the ubiquitously expressed PYGM was early replicating. Both AHB and MRC5 cells showed good agreement with Selig's data for PYGM. In AHB and MRC5 cells, probes cW44 and cJ21a showed slightly earlier replication times compared with cell lines K-562 and Manca; however, they still agreed with the middle/late replication assessment made by Selig et al. (1992).
<table>
<thead>
<tr>
<th></th>
<th>Caco-2*</th>
<th>K-562*</th>
<th>Manca*</th>
<th>AHB</th>
<th>MRC5</th>
</tr>
</thead>
<tbody>
<tr>
<td>cW44</td>
<td>62</td>
<td>21</td>
<td>13</td>
<td>31</td>
<td>33</td>
</tr>
<tr>
<td>cJ21a</td>
<td>66</td>
<td>26</td>
<td>13</td>
<td>29</td>
<td>30</td>
</tr>
<tr>
<td>PYGM</td>
<td>64</td>
<td>68</td>
<td>77</td>
<td>66</td>
<td>63</td>
</tr>
<tr>
<td>A9.5</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>24</td>
<td>26</td>
</tr>
</tbody>
</table>

Table 3.3. Comparing % doublets between established controls and the cell lines used in this thesis. *These data are taken from (Selig et al., 1992). Caco-2 is an epithelial-type cell line derived from a human colon carcinoma that expresses the CF gene; K-562 is a human erythroleukemia line; Manca is a human lymphoma line.

Cosmid A9.5 was late replicating in both AHB and MRC5 cells, confirming its assignment to a genetically inert region of the G-band 6p24. Since it replicates later than both CF clones, and shows good correlation between the two cell types, it was used as the late replicating comparison in subsequent experiments.

Both independent scorers showed good agreement in the estimates of singlet/doublet chromosomes. The maximum difference between estimates for any slide was 8%, but for 3/4 of the control probes scorer differences were ≤6%. This reproducibility is very close to the <5% inter-scorer variation for singlet signal estimates previously reported by Boggs and Chinault (1997).

3.7. Discussion

3.7.1. Why FISH?

Non-FISH replication assays, which rely on physical methods to separate S-phase DNA into fractions, give poor temporal resolution. Whatever the method employed to fractionate S-phase, it is usually divided into 2h fractions (occasionally 1h fractions). This then is the
limit of the temporal resolution, leaving the investigators unable to distinguish differences between loci of less than two hours. Such techniques do not show a clear temporal difference in replication time for a single probe hybridized to the different fractions of nascent DNA. Nor is the difference between two or more probes clear when Southern’s are compared.

Many non-FISH methods use synchronization, which may alter replication behaviour. Synchronization has been superseded in recent years by retro-synchrony, using elutriation or FACs sorting to divided S-phase into fractions based on each cell’s C-value. However, the autoradiographs produced after Southern hybridization to the nascent DNA often have strong bands in more than one fraction. The differences between the fractions are not discernible by eye so densitometry or phosphorescence is used to quantitate the signal and determine which fraction has the peak amount of nascent DNA. After this, investigators can only say in which quarter of S-phase they consider a locus to be mainly replicated. Loci that are physically close along the chromosome therefore tend to be assigned to the same quarter fraction of S-phase, with no clear indication as to which replicates first.

The contiguous cosmids available for the MHC provide a unique resource for a detailed replication analysis. FISH offers the advantages of a method that is quick and simple to perform on unamplified regions in asynchronous cells. Moreover, because it considers replication in the context of individual nuclei, results can be obtained from relatively small numbers of cells – about ten-fold less than elutriation needs. Its advantage over retro-synchrony is that there is no restriction on the cell-type one can examine; elutriation being only really effective with fast growing suspension cells. The analysis of the MRC5 fibroblasts which will be presented later would have been virtually impossible if it had needed to rely on retro-synchrony.
3.7.2. Potential inconsistencies

Since Selig et al. validated the FISH method, there have been some disagreements between this approach and others when comparing the replication of imprinted alleles and non-imprinted alleles or genes that are subject to X-inactivation. Allelic differences in expression are mirrored by differences in replication time, with the expressed allele generally replicating earlier than the imprinted one – so called replication asynchrony. However, some loci display asynchrony with non-FISH techniques and synchrony with FISH, or even reverse patterns of asynchrony, for example the X-linked Xist gene (Torchia & Migeon, 1995; Boggs & Chinault, 1994; Hansen et al., 1995; Xiong et al., 1998). However, the MHC is not imprinted and so this specific problem will not be discussed further. More thorough critiques of FISH are given by Bickmore & Carothers (1995) and Kawame et al. (1995), but some of the more general issues surrounding this methodology are addressed below.

FISH has been criticised because three dimensional nuclei are interpreted from a two dimensional perspective, creating a potential problem in resolving doublets from singlets. However, as indicated in 3.5.2 it is reasonable to assume that this artefact is due to the random orientation of the nucleus on the slide and should affect all probes to the same extent. In addition, probes are measured relative to one another and the controls, rather than being assigned an absolute replication time.

Poor hybridisation efficiency has been accused of producing artefacts in FISH. And as discussed in 3.5.1, doublets can sometimes appear in the absence of replication due to over-denaturation or chromosome decondensation. These relatively trivial problems are addressed by optimising the FISH protocol and by assessing the efficiency of hybridization or the extent of denaturation on metaphase spreads.
The structural conformation of the chromatin – due, in part, to important elements in the primary sequence – could influence how quickly the replication machinery is able to deal with a locus or the rate at which sister chromatids are resolved after replication. How influential this might be is not known, but it does highlight the need to examine genomic regions in some detail with contiguous probes. If FISH identifies regions with remarkable or unexpected replication times given the current ideas about replication timing and gene activity, this should alert us to focus on these regions. Such apparent anomalies might provide clues as to the role of chromatin structure in replication.

In summary, the availability of well-mapped genomic probes and the ability to compare results from different cell lines makes FISH a convenient way to map domains of replication timing at any chromosomal position and to relate it to various patterns of gene expression. There are potential problems, particularly when it is used to measure asynchronous replication. However, optimizing the protocol, blind-scoring and repeat experiments should reduce these problems. If interpreted carefully, with probes assessed relative to each other, FISH offers a simple and effective means of examining the important but poorly understood correlations among replication timing, genomic organization and transcriptional competence in mammalian cells, and provides a valuable approach to understanding these interrelationships at the interphase level.
Chapter 4. Cell Type-Specific Replication Timing Profiles of the Human Major Histocompatibility Complex

4.1. Introduction

The replication of important genes and gene clusters has been studied since the 1980s; for examples see (Goldman et al., 1984; Holmquist, 1987; Dhar et al., 1988; Hatton et al., 1988; Spack et al., 1992). Differences in replication time have been measured across the boundaries of certain chromosome bands (Bickmore & Carothers, 1995; Tenzen et al., 1997; Strehl et al., 1997; Bilyeu & Chinault, 1998). Some studies have measured replication time over extended stretches of DNA within bands (Brown et al., 1987; Dhar et al., 1988; Selig et al., 1992; Smith and Higgs, 1999). With the first draft of the human genome now complete (International Human Genome Sequencing Consortium, 2001; Venter et al., 2001), this latter kind of analysis — over longer contiguous stretches of DNA — is increasingly important for understanding how the temporal organization of replication relates to the structural organization of the sequence and its transcriptional activities.

The human MHC is well characterized and is ideal for such an extended replication timing analysis. It maps to the R-band 6p21.31 but, as discussed in section 1.1.1.3, we are interested in levels of resolution higher than the metaphase chromosome band.

This multi-gene family offers several advantages for the characterization of replication organization. In terms of physical and database resources, overlapping cosmid clones are available covering almost the entire region, and the entire sequence is available from the Sanger Centre, http://www.sanger.co.uk. At 3.6 Mb, it is the second longest contiguous sequence in the human genome and much is known about its sequence organization and the transcriptional control of the classical MHC genes (see section 1.3.5). The MHC encodes constitutively expressed genes throughout its length, with
the highest density in the class III region. It also houses tissue-specific genes, most notably the HLA class II genes – whose expression is normally restricted to antigen-presenting cells such as B-cells and macrophages. The HLA genes are inherited as closely linked sets called haplotypes. The maternal and paternal alleles in these haplotypes are co-dominantly expressed; therefore, this is not a imprinted region and should not display replication asynchronicity, e.g. see Kitsberg (1993a).

Analyzing replication at the MHC's class II/III region isochore boundary would allow a high resolution analysis across a recognizable genomic feature. A PCR analysis of nascent DNA from the class II/III boundary has been performed but – in addition to the temporal limitations of the assay discussed in section 3.7.1 – this analysis used two rounds of cell synchronization and focused on just 160Kb (Tenzen et al., 1997). We wanted to extend this analysis into the rest of the MHC and to avoid the use of cell synchronization.

Whilst the replication behaviour of the mouse H-2 major histocompatibility complex has been characterized and its putative replicon structure identified (Spack et al., 1992), there are no such data for the human MHC. Comparing the MHC between human B-cells and fibroblasts will provide data on cell type-specific replication behaviour at a functionally complex locus.

4.1.1. Hypotheses on the replication behaviour of the MHC

Given our current understanding of the structure and transcriptional properties of the human MHC – and based on the positive correlation observed between transcription and replication time – the following hypotheses were made concerning the MHC's replication time in B-cells and fibroblasts.
The GC-poor classical class II region might behave as GC-poor DNA generally does, and replicate late. Indeed, the classical class II region could be one of the thin G-subbands observed in 6p21.31 (Yunis, 1981). Finding the classical class II region entirely late replicating in both B-cells and fibroblasts might support this observation. However, if the classical class II is a G-band it is an atypical example because of its relatively high gene density (1 every 40Kb). Most of its expressed genes (12 of 18) are involved in processing and presenting foreign antigens to CD4+ T-cells and are restricted to antigen-presenting cells. Since such tissue-specific genes are thought to replicate early in the cells in which they are expressed (Hatton, 1988), we hypothesized that the ~800Kb classical class II region would, regardless of its GC content, be early replicating in B-cells and late replicating in fibroblasts (however, see below).

Virtually the whole MHC should be expressed in B-cells, and we predicted the entire ~4Mb to be early replicating. In fibroblasts, the constitutively expressed class III region should replicate early (class I was not tested in fibroblasts); however, the classical class II region posed a problem. Approximately 700 of its 800Kb would be redundant in fibroblasts, with the exception of the TAP and LMP genes at its centre which are involved in antigen presentation through MHC class I molecules (York & Rock, 1996). Early replication of an expressed gene or gene cluster from a nearby origin can extend into the flanking chromatin (Herbomel, 1990). The TAP and LMP genes are clustered in less than 40Kb, but if they are replicated early from a nearby origin then this could extend into the surrounding chromatin. We were prepared, therefore, for the possibility of an island of early replication in a relative sea of later replication.

The extended class II region belongs to a different isochore fraction and contains both constitutively expressed genes and genes important in antigen presentation through the ubiquitous MHC class I molecules – Tapasin and RXRβ – it was therefore thought to
be sufficiently distinct both in terms of its sequence structure and genetic function to be early replicating in both B-cells and fibroblasts.

In summary, our main hypothesis was that the tissue-specific classical class II region would be entirely early replicating in B-cells and late replicating in fibroblasts (with the possible exception of the \textit{TAP/LMP} genes). This predicts that, in fibroblasts, the classical class II region would be bounded by temporal borders with its relatively earlier replicating neighbours (the extended class II and the class III regions).

4.2. Methods

To determine the relative replication time across the MHC, forty-two cosmids clones from all three classes were hybridized to lymphoblastoid B-cells. In normal lung fibroblasts, we used thirty of these probes covering the class II and III regions. Briefly, S-phase nuclei were identified by pulse-labelling the cells with BrdU thirty minutes prior to harvesting. Biotin-labelled cosmid clones were hybridized to fixed nuclei, then at least 1000 chromosomes were scored for each probe.

In contrast to Chapter 3, the rest of this thesis presents the mean replication times as a percentage of the time spent in S-phase. This is estimated by the percentage of singlets – with earlier replicating loci having fewer singlets than late replicating loci (see Appendix A). The mean replication times for each probe and the average for each class were interpreted relative to one another and to the control cosmids.

4.3. Results

4.3.1. Replication of the MHC in B-cells

4.3.1.1. The MHC is not entirely early replicating in B-cells

Figure 4.1 shows the replication profile for the entire MHC in lymphoblastoid B-cells. We can see that our original hypothesis – that the MHC will be entirely early
Figure 4.1. The replication profile of the entire MHC in B-cells. The data points are the mean relative replication time (expressed as % singlets) with the bars representing +/- 2 standard errors. The internal controls PYGM (early) and A9.5 (late) are indicated by the dotted lines. The approximate positions of a number of landmark MHC loci are indicated below the graph.
replicating in B-cells – is not supported by these data. The extended class II, the class III and the class I regions all replicated relatively early as predicted. Looked at as distinct classes, the mean replication times across their lengths were 0.36 +/- 0.01, 0.34 +/- 0.05 and 0.40 +/- 0.04, respectively. However, the classical class II region confounded our prediction. The first ~200Kb of the classical class II region replicated early (300-500Kb on Fig 4.1); it then underwent a transition to later replication (500-700Kb on Fig 4.1) which was maintained through its remaining 200Kb. As a distinct class, its overall replication time was later than the adjacent classes at 0.51 +/- 0.11. And unlike its neighbours, whose replication times remained relatively constant, the mean replication time of the classical class II region ranged from 0.32 to 0.66, which is reflected in its greater standard deviation.

4.3.1.2. B-cells have a temporal boundary in their MHC

The late replication of the HLA-DQ/DR locus contrasts the early replication seen at the proximal end of the class III region (~1200Kb on Fig 4.1 (though it may be seen in greater detail in Fig 4.6). Finding a temporal boundary between the classical class II and class III regions in B-cells was unexpected, since we had expected the entire classical class II region to be early replicating. The transition from early to late replication timing was coincident with the region of GC/AT transition that separates the isochores making up the two classes (Fig 1.2). One mechanism to account for this progressive change in replication time would be the sequential activation of replicons within the transition region. Another would involve a single replication fork progressing from the last replicon in the proximal class III region into the distal end of the classical class II region.

The duration of S-phase could not be determined in the B-cells because they were refractory to cell synchronization. However, if we accept that mammalian S-phase is between 6-8 hours, we can assign replication times in terms of hours and minutes and
estimate the approximate rate of replication. The last probe of the classical class II (LH1, see Fig 4.1 or detailed in Fig 4.6) and the first in the class III (C47, see Fig 4.6) are separated by ~125Kb and 18% singlets – or 18% of S-phase. This equates to between 1 and 1.5 hours. This suggests that replication has occurred from C47 to LH1 at a rate of about 125 kb/h or between 1.5-2 Kb/min, well within the accepted range of 0.3-6 Kb/min for mammalian replication fork movement (Edenberg & Huberman, 1975). Alternatively, this ~125Kb region could be replicated by numerous minute replicons, with the first initiating around C47 and the rest firing sequentially in the direction of LH1.

Within the classical class II, the difference between its early portion (average of probes Mann3.6 to HA14, 0.38 +/- 0.05) and its late portion (FI121 to LH1, 0.62 +/- 0.03) is approximately 1.5 to 2h (see Fig 4.1). Therefore, a single fork may have taken 1.5 to 2h to replicate ~250Kb. If so, ~250Kb in 90-120 minutes (i.e. 2-2.7 Kb/min) is again consistent with the estimate of Edenburg & Huberman (1975). It is important to note that this transition in replication time in the proximal half of the classical class II region is not coincident with the isochore boundary separating the classical class II and extended class II regions. This transition in GC content is approximately 400Kb more centromeric (see Fig 1.2).

To summarize, the class II region's L1 isochore had both early and late replicating portions in B-cells. There was a change from early to late replication from the class III region into the distal end of the classical class II region. A similar change in replication time from early to late was seen across the middle of the classical class II region itself; however, this did not coincide with the isochore boundary known to exist between the classical and extended class II regions. In both instances the observed rate of change in replication time linking the early and late replicating portions was consistent with the previously reported rate for eukaryotic replication forks (Edenberg & Huberman, 1975).
4.3.1.3. B-cells express class II molecules

Since tissue-specific genes are thought to replicate early when they are expressed, there were two possible explanations for the relatively late replication of the HLA-DQ/DR locus. Either the B-cells were not expressing the HLA-DQ/DR molecules, or we had an exception to the rule that tissue-specific gene clusters replicate early when transcribed. Northern analysis and immunofluorescence detection were used to test the B-cells for expression of the DRA gene (encoding the \( \alpha \) sub-unit of the HLA-DR heterodimer) as described in Chapter 2.

![Northern analysis of DRA mRNA in B-cells and fibroblasts. Lane 1 demonstrates the constitutive expression of the HLA-DRA gene in B-cells. Lane 4 shows that HLA-DRA is not transcribed in these fibroblasts (see 4.3.2.2). The ubiquitously expressed \( \beta \)-actin gene was used as a control. The signals in the remaining lanes will be discussed later.](image)
The positive Northern result in lane 1 (Fig. 4.2) shows that the B-cells were actively expressing the \textit{DRA} gene. Figure 4.3 (below) illustrates that the resulting HLA-DR molecule was expressed on the surface of the B-cell. Here B-cells were incubated with a FITC-labelled antibody directed against the \(\alpha\)-subunit of the DR molecule (see Chapter 2).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure4_3.png}
\caption{Immunofluorescence analysis of DRA expression on B-cells. Red: negative control, B-cells without anti-DRA antibody. Blue: B-cells incubated with anti-DRA-FITC are shifted along the logarithmic fluorescence scale due to specific binding of labelled antibody.}
\end{figure}

It is clear from the above results that the B-cells used in these experiments not only transcribed the \textit{HLA-DRA} gene but, as professional antigen-presenting cells, translated it into an expressed DR molecule. Such positive results contrast with the replication time profile for the HLA-DQ/DR locus, which shows it to be relatively late replicating and, therefore, an exception to the rule that tissue-specific gene clusters replicate early when expressed (Hatton \textit{et al.}, 1988).

\subsection*{4.3.2. Replication of the MHC in Fibroblasts}

\subsubsection*{4.3.2.1. The classical class II is late replicating in fibroblasts}

Figure 4.4 shows the replication profile for the MHC in fibroblasts, excluding the class I region. Our hypothesis – that the classical class II region would be later
replicating than its neighbours in fibroblasts – is supported by these data. The probes in the *extended* class II and class III regions had average replication times of 0.44 +/- 0.05 and 0.46 +/- 0.08, respectively. Sandwiched between them, the ~800Kb classical class II region was entirely later replicating, with an average replication time of 0.61
Figure 4.4. The replication profile of the MHC class II and III regions in human fibroblasts. The data points are the mean relative replication time (expressed as % singlets) with the bars representing +/- 2 standard errors. The internal controls PYGM (early) and A9.5 (late) are indicated by the dotted lines.
+/- 0.04. Nor did it have the postulated island of early replication corresponding to the TAP/LMP gene cluster, which provided another example of an apparently transcribed locus that appeared to replicate later than might be expected (Hatton et al., 1988).

The predicted temporal boundaries were present in the fibroblasts. At the class II/III boundary, as in B-cells, there was a change from earlier replication in the class III region to late replication at the distal end of the classical class II region. But unlike B-cells, the position of the first relatively early replicating probe in the class III region has shifted by about 150Kb. Probe D3a is now the first early replicating class III probe and has a replication time similar to probe C47 in B-cells: 0.45 and 0.43, respectively (see Fig 4.6). Assigning approximate 'real' times to the probes LHI and D3a, which are separated by ~250Kb and 24% of S-phase, equates to a single replication fork taking between 1.5 to 2h to cover 250Kb, or 2-3 Kb/min. Since this approximation matches that in B-cells, it suggests that the same replication fork moves from the class III into the distal end of the class II region.

If we had only performed FISH on the fibroblasts we might have concluded that the overall later replication of classical class II region's GC-poor, L1 isochore indicated that it was indeed the small G-subband identified by Yunis (1981). But finding at least part of it early replicating in B-cells makes such a conclusion less clear cut; that is unless one holds the view that there is a causal relationship between GC content and replication time.

4.3.2.2. Fibroblasts do not express MHC class II molecules

Confirmation that the fibroblasts did not express the class II genes was provided by mRNA and expression studies, as in 4.3.1.3 above. Lane 4 in Figure 4.2 and the analysis in Figure 4.5 (below) show that, as expected, the fibroblasts used in these experiments did not express the HLA-DR locus. Since transcription of the class II
genes is co-ordinately controlled (Boss, 1997), we assumed that the fibroblasts did not express any HLA class II molecules (HLA-DP, -DQ or -DR).

**Figure 4.5.** Immunofluorescence analysis of DRA expression on the surface of fibroblasts. Red: negative control, fibroblasts without anti-DRA-FITC antibody. Blue: fibroblasts incubated with anti-DRA are not shifted along the fluorescence scale. The slight shift is due to non-specific binding of the anti-DRA-FITC.

4.3.3. Replication Timing in B-cells Versus Fibroblasts

Figure 4.6 shows the replication profiles of the two cell types overlaid, excluding the class I region. The extended class II region is relatively early replicating in B-cells; though in fibroblasts – contrary to our original hypothesis – its boundary with the classical class II region appears to be later replicating, perhaps under the influence of changes taking place in its later replicating neighbour. The bulk of the class III region (except for its extreme ends) is relatively early replicating in both cell types, though overall it is slightly later in fibroblasts. However, it is the HLA-DQ/DR locus in the classical class II region that presents the most interesting profile.

4.3.3.1. Comparing the classical class II region data

The change in transcriptional status of the classical class II region between the cell types is accompanied by a change in replication behaviour. In Figure 4.6, the first third of the classical class II region (probes Mann 3.6 to U15, Fig. 4.6) replicates earlier in B-cells (0.40 +/- 0.06) than fibroblasts (0.59 +/- 0.04). The replication time of the
Figure 4.6. Comparing the replication profiles in B-cells (black) and fibroblasts (red). Internal controls PYGM (early) and A9.5 (late) are indicated by the lower and upper dotted lines, respectively. Again bars indicate +/- 2 standard errors. * see sections 6.2 & 6.4.1.
remainder (F1121 to LH1, Fig 4.6) is remarkably similar in both cell types, 0.62 +/- 0.03 in B-cells and 0.64 +/- 0.04 in fibroblasts. This is noteworthy because this stretch of the class II houses the tissue-specific HLA-DRA gene which is expressed in B-cells and silent in fibroblasts (Figs 4.2, 4.3 and 4.5).

If one accepts the sensible idea that transcriptionally active, tissue-specific genes replicate early in cells in which they are expressed, then the earlier replication of the HLA class II genes in B-cells comes as no surprise. Of more interest is why the replication time should be so similar (and late) at the very locus where transcriptional activity is different. Replication time at this locus may, therefore, be influenced by factors other than transcriptional activity; including perhaps properties of the underlying chromatin that are shared in common between the cell types.

4.3.3.2. Comparing the MHC class III region data

The constitutively expressed class III region replicates relatively early in both B-cells and fibroblasts, and present us with another interesting phenomenon. The overall pattern of 'peaks and troughs' in the profile is remarkably similar in the two cell types. The difference is, of course, that the entire class III region's profile is shifted towards slightly later replication in the fibroblasts, implying that execution of the temporal programme is slightly delayed in this cell type.

This is the second replication timing analysis to identify a region with a similar replication pattern in different cell lines. Previously, FISH was used to examine replication in the telomeric region of human chromosome 16 (16p13.3), which contains both widely expressed and tissue-specific genes including the α-globin gene family (Smith & Higgs, 1999). Using hybrid cell lines each with a normal copy of chromosome 16 in a mouse erytholeukemia (MEL) background, and a non-erythroid
B-cell line, Smith & Higgs recorded replication profiles that showed a consistently similar pattern of peaks and troughs on every 16p homologue.

Our class III region data show a pattern of peaks and troughs that appears to be cell type-independent. It could, therefore, be an intrinsic property of this constitutively expressed region. Where the pattern differs, such as the tissue-specific classical class II region, this could reflect cell type-specific and/or transcription-specific differences in the organization of the chromatin, which may go on to affect the execution of the replication programme (also see section 5.4.2.). The patterns we record may be a consequence of the specific interactions between chromatin and nuclear components which establish and maintain the precise spatial and temporal requirements of effective DNA metabolism.

4.4. Discussion

4.4.1. The Relationship Between Transcription and Replication

The HLA-DQ/DR locus replicates later than the rest of the class II region regardless of its transcriptional status. This is the first time a cluster of tissue-specific genes has been shown to be late replicating even when expressed, suggesting that DNA replication does not play a crucial role in the regulation of their expression. Whilst this finding contrasts with the some of the 'rules of replication' devised by Schildkraut and co-workers (Dhar et al., 1988 and 1989; Hatton et al., 1988), it does not seriously threaten the idea that transcription is positively correlated with early replication. It is perfectly reasonable that transcribed sequences (and their neighbours) replicate early by default because their de-condensed chromatin allows easier access to the replication machinery (Raghuraman et al., 1997). However, the data presented here do question the extent to which transcription influences replication time and suggest that additional factors may be involved.
A useful way to explore the transcription/replication relationship further is to induce MHC class II expression in fibroblasts using γ-interferon and observe whether such a profound change in transcription causes a switch to earlier replication time. If the ~350Kb HLA-DQ/DR locus becomes earlier replicating in induced fibroblasts then the positive correlation between expression and early replication is preserved. If, however, it remains later replicating this could indicate some other influence on replication timing. The induction of MHC class II expression was carried out and the results are presented in Chapter 5.

4.4.2. The Mouse H-2 Complex on Chromosome 17

The order of DNA replication in the H-2 MHC of the mouse was previously reported by Spack et al. (1992) using density labelled DNA extracted from cells fractionated into four stages of S-phase using centrifugal elutriation. Various H-2 segments were used as Southern hybridization probes to the blotted nascent DNA (though not all the probes were examined in both cell types used). The relative levels of hybridization in each fraction were normalized to the S1 fraction and compared.

In keeping with the findings of other investigators, they showed that the TL region (containing "class I-like" sequences) replicated earlier in T-lymphoma cells, where it is thought to be expressed, than in myelomonocytes, where it is silent. In a result similar to that seen in the human class II region in B-cells, there was a trend from early to late replication from centromere to telomere in the equivalent mouse H-2 I region, whereas the H-2 S region (class III region) maintained a more constant and relatively early replication time. However, the S region was only examined in the myelomonocyte cell line. There has been a previous report of an isochore boundary between the I and S regions (Ikemura & Aota, 1988); however, Spack et al. were unable to provide data for an approximately 400Kb region harbouring this boundary owing to a lack of hybridization probes.
Spack et al. concluded from their analysis that the 2Mb mouse H-2 MHC contains approximately five or six synchronously activated replicons, including some that may be highly asymmetric, or unidirectional. These ideas of synchronicity and asymmetry are interesting from our perspective since the former might explain the pattern of peaks and troughs observed in the replication profiles, and the latter could explain the apparent absence of replicons across the class II/III region boundary; these ideas are explored further in section 5.4.2.

Beyond that given above, a direct comparison with the mouse is difficult since Spack's method is so different from our own and, as they clearly state, theirs was a low resolution approach intended as a first step in the broad definition of the H-2 MHC replicon structure. What we can say is that, with the obvious exception of the TL region, the data collected for both cell types showed areas of similarity, suggesting an organization in the temporal order of DNA replication. Aside from the differences in the classical class II region (as in the mouse TL region), the replication profiles we recorded in B-cells and fibroblasts showed regions of similarity, indicating that they might also have a shared temporal organization. The mouse H-2 region is similarly organized to the human MHC but with some key differences. The H-2 I and H-2 S regions contain a similar number of genes as the equivalent human class II and III regions but are smaller in places (Gasser, 1994). Despite this, and if, as suggested, the H-2 MHC's isochore structure has been conserved during evolution (Ikemura et al., 1988), it would be worthwhile revisiting replication timing in the mouse H-2 MHC using FISH.
4.4.3. The Temporal Change Across the Class II/III Boundary

4.4.3.1. Is there a sharp switch in replication timing?

We detected a change in replication time across the GC/AT transition region separating the class II’s L2 isochore from the H3 isochore of the class III region. This change, seen in both B-cells and fibroblasts, is consistent with that reported by Tenzen et al. (1997) in the same part of the MHC but identified using Southern analysis of nascent DNA from synchronized cells. Whilst both methods identify a temporal change, we would disagree with Tenzen's interpretation of the change and its significance.

Tenzen et al. (1997) used two rounds of synchronization with aphidicolin prior to releasing human myeloid leukaemia cells (HL60) into S-phase and labelling them with BrdU at 1h intervals for 6h. They isolated and quantitated the nascent DNA in each sample using competitive PCR of markers within the class II/III region's boundary. The PCR products were stained with a fluorescent dye and the relative intensities of the signals in each sample were presented graphically. However, even with the data presented in this way, some loci could easily be assigned the same replication time. One has the impression that the synchronization was not a tight, narrow peak passing quickly into and out of each fraction, but was broader and included contaminating cells from other stages of S-phase. This effect may be due to a proportion of the HL60 cells taking longer to enter S-phase after two rounds of aphidicolin.

It is important to note that the 1h interval separating the samples was, by definition, the limit of Tenzen et al.'s temporal resolution. The assay could not, therefore, be expected to distinguish any difference between sequences of less than one hour. FISH does not have such a restriction because in the fixed samples it uses, time is essentially frozen and the entire S-phase is deliberately used to estimate the mean replication time of each locus. The only potential limitation with FISH is the time it takes for a replicated locus to resolve into a pair of fluorescent spots on the sister chromatids.
How long after replication it takes to detect doublets, and whether or not this time is constant for different sequences, has not been formally addressed by any investigators.

Tenzen et al. (1997) reported a difference of one hour between two primer sets just 16Kb apart in the GC/AT transition area separating the class II and III regions (coordinates ~200-230 Kb, Figure 4.7).

Although the primer sets PCD and INT3A appear clearly separated on Figure 4.7, this plot is of the single 1h interval with the highest intensity PCR product, and does not reflect the overlap between adjacent intervals so apparent in the raw data. When one looks at the distribution of the signal across all six 1h intervals and compares the distributions for PCD and INT3A, it is hard to believe that their replication times are really that different. In addition, the difference measured between PCD and INT3A is at the limit of the assay's temporal resolution. Therefore, Tenzen et al.'s results do not
exclude the possibility that PCD and INT3A actually replicate within an hour of each other. Tenzen et al.'s raw data are more consistent with the view that PCD and INT3A have similar replication times because they are only 16Kb apart.

The apparent one hour's difference across 16Kb, was termed by Tenzen et al. a "precise switch" in replication timing. The word "switch" refers to the difference of an hour and "precise" to the suggestion that it occurs across just 16Kb in the region of GC/AT transition separating the isochores of the GC-poor class II and GC-rich class III regions. They suggested that the "precise switch" was evidence that the boundary between the class II and III regions is not merely between different isochores but different cytogenetic bands. Based on the previous observation of a small G-subband within 6p21.31 (Yunis, 1981), it was suggested that a chromosome band boundary existed between the class II and class III regions and – since R-bands replicate early and G-bands late – that this boundary would be "precisely assignable at the nucleotide level by identifying the early-to-late switch point for replication timing" (Fukagawa et al., 1995). Tenzen et al. appear to have interpreted their replication timing data in order to support this hypothesis. Chapter 1 details why the structure, function and organization of chromosomes is more complicated than the popular cliché "GC-rich R-bands and GC-poor G-bands" suggests. We must, therefore, avoid interpreting replication timing data using this and other similarly out-moded ideas.

Aside from the difficulties with the limited temporal resolution, it is difficult to subscribe to the idea – based on such experiments – that something as relatively massive as a chromosome band boundary can be described to within a few specific nucleotides. There seems little sense in higher eukaryotes being so strict about where replication stops when they seem relatively relaxed about where it starts (Vaughan et al., 1990; DePamphilis, 1999). Nor does it makes sense for the replication machinery to replicate the 1Mb class III region and then stop short of the class II region simply because it has a slightly lower GC content. Considering the wide range of functions
and properties attributable to isochores and chromosome bands, their physical boundaries are most likely composed of a number of elements including sequence motifs and GC changes, but also perhaps particular chromatin structures.

4.4.3.2. Replication timing studies in other chromosomal regions

Our analysis shows a mean replication time difference of between 1-2 hours across the class II/III isochore boundary region in both B-cells and fibroblasts. The gradient of this difference is consistent with the accepted rate of a mammalian fork (assuming S-phase is between 6 and 8 hours). This suggests that a replication fork initiated within in the class III region has responsibility for copying the class II/III boundary. Indeed, if one holds Figure 4.7 the wrong way up and imagines the data for the entire 150Kb flipped from left to right, one sees a gradient in replication timing more akin to the profiles presented in this thesis.

Another report of a change in replication time across a putative G/R-band boundary was made by Strehl et al. (1997), using competitive PCR on nascent DNA. But instead of synchronization, they used cell-sorting to fractionate S-phase cells based on their total DNA content, or C-value. They showed a change in replication time across human chromosome 13, at the boundary between the R-band 13q14.3 and G-band 13q21.1. However, they concluded that the change from early to late was not precise and did not map to any notable DNA sequences. It is interesting that in their timing analysis, Tenzen et al. make no further claims for the involvement of the pseudoautosomal boundary-like (PABL) sequence, which was ~70Kb away from their replication time switch (Fukagawa et al., 1995) and see section 1.3.5.

The lack of a precise change across regions with different replication times was also reported in a previous FISH analysis of the pseudoautosomal region boundary Xp22.3 and the boundaries of Xp22 at the distal end of the human X chromosome (Bilyeu &
Chinault, 1998). Although they observed replication timing domains that were generally consistent with cytogenetic bands, the investigators found no sharp transition in timing behaviour for either the pseudoautosomal region boundary or the cytogenetic band boundaries. They concluded that the absence of a precise switch in replication time may represent a transition in some aspect of chromatin that is gradual at the molecular level while appearing sharp at the cytogenetic level (Bilyeu & Chinault, 1998).

Based on its overall late replication in fibroblasts and its late replicating portion in B-cells, it might be the case that the classical class II region constitutes the small cytogenetic G-sub-band previously observed in 6p21.31. But we would be reluctant to say that the change in replication time across the class II/III boundary region provides proof of this. Moreover, since we find the "change" to be consistent with the movement of a single replication fork, we can not agree with the conclusion of Tenzen et al. (1997) that there is a "precise switch" in replication time across the MHC class II/III boundary region.
Chapter 5. The Effect of IFN-γ Induction on the Replication Time of the MHC Class II and Class III Regions

5.1. Introduction

As discussed in Chapter 1, the MHC class II molecules (HLA-DP, -DQ and -DR) play a crucial role in the immune response, presenting foreign antigen to CD4+ T-cells. The class II genes are co-ordinately regulated in a tissue-specific manner by the highly conserved proximal promoter they each possess. This promoter usually restricts class II expression to antigen-presenting cells such as B-cells. However, incubation with cytokines such as γ-interferon (IFN-γ) induce non-lymphoid cells such as fibroblasts to express class II molecules.

Although IFN-γ induction of MHC class II in MRC5 cells has been demonstrated (Darley et al., 1993), there are no data concerning its effect on the replication time of the region. With the replication profile in untreated fibroblasts already established, we were in a position to gather such data and make a comparison of the replication profiles for a single cell line in two different transcriptional states.

5.1.1. Hypotheses concerning IFN-γ induction in fibroblasts

It has been well established that transcribed sequences replicate early (Dhar 1988 and 1989; Hatton 1988). However, the previous observation of a cluster of genes in the HLA-DQ/DR locus replicating late in B-cells despite being expressed stimulated interest in the idea that there may be other factors with an influence on replication timing. We thought it safe to assume that IFN-γ would have no effect on the replication time of those regions in untreated fibroblasts that already replicated early (such as the central portion of the class III region) and a positive effect (i.e. cause earlier replication) in those regions that replicated later. However, based on our observations...
in B-cells, we did not expect IFN-γ to cause the late replicating HLA-DQ/DR locus to replicate earlier in induced fibroblasts. Our hypothesis was that IFN-γ induction would confer a B-cell-like replication profile on the fibroblasts. This would include relatively late replication of the HLA-DQ/DR locus, even if it is successfully expressed.

5.2. Methods

Briefly, all but one (M27B) of the thirty probes covering the class II and III regions were hybridised to nuclei from IFN-γ treated fibroblasts. Initial experiments showed that 24h was sufficient to induce expression of \( HLA-DRA \) mRNA; there was also an attempt to 'chase' the IFN-γ from the fibroblasts by harvesting the cells 24 or 48h after the stimulus was removed and replaced with fresh medium (see section 5.3.3). For the FISH experiments, MRC5 cells were incubated with 10U/ml IFN-γ for 24h. S-phase nuclei were identified by pulse-labelling the cells with BrdU thirty minutes prior to harvesting. Biotin-labelled cosmid clones were hybridized to fixed nuclei and at least 1000 chromosomes were scored for each probe. The mean replication times are presented as a percentage of the time spent in S-phase. This is estimated by the % singlets – with earlier replicating loci having fewer singlets than late replicating loci (see Appendix A). The mean replication times for each probe and the average for each class were interpreted relative to each other and to the internal controls.

5.3. Results

5.3.1. IFN-γ incubation times

Before proceeding with the FISH analysis, we assessed the time needed for induction of MHC class II transcription in MRC5 cells. Darley et al. (1997), who had previously shown MRC5 cells to be inducible, incubated with 10 U/ml IFN-γ for 3 three days.
The result in lane 2 on Figure 5.1 shows that 10U/ml for 24h was sufficient for \textit{HLA-DRA} induction; compare this with the non-induced fibroblasts in lane 4.

\begin{figure}[h]
\centering
\includegraphics[width=0.8\textwidth]{figure5_1.png}
\caption{(A repeat of Fig 4.2.) Northern analysis of HLA-DRA mRNA in IFN-\(\gamma\) induced (lanes 2, 3, 5 and 6) and non-induced (lane 4) MRC5 fibroblasts. The 'chase' times indicate how long the cultures were left to continue to grow after removal of the external IFN-\(\gamma\) stimulus.}
\end{figure}

The signals in lanes 2 and 3 are consistent with other studies showing induction of MHC \textit{class II} expression to occur with slow kinetics: HLA-DRA mRNA being first detected 8 to 12h after IFN-\(\gamma\) stimulation, peaking at 24 to 48h, and diminishing after 72h (Vidovic \textit{et al.}, 1990; Bottger, 1988).

\textbf{5.3.2. Induced fibroblasts express MHC class II molecules}

In addition to measuring the mRNA, we ensured that HLA-DRA was expressed on the surface of the fibroblasts. The immunofluorescence detection data presented in Figures 5.2 a and b show this was indeed the case.
Figure 5.2.a. Immunofluorescence analysis of DRA expression on non-induced MRC5 cells (negative control). Red: internal control, fibroblasts without anti-DRA-FITC. Blue: non-induced fibroblasts incubated with anti-DRA are not shifted along the fluorescence scale.

Figure 5.2.b. DRA expression on IFN-γ induced MRC5 cells. Red: internal control, induced fibroblasts without anti-DRA. Blue: induced fibroblasts incubated with anti-DRA-FITC are shifted along the logarithmic fluorescence scale.

It is clear from these experiments that the induced fibroblasts not only transcribed HLA-DRA mRNA but translated it into an expressed cell surface protein. Since the MHC class II genes are co-ordinately controlled (Boss, 1997), the inference from these analyses is that the induced fibroblasts were also transcribing the remaining classical class II genes (HLA-DP and -DQ).
5.3.3. Chasing the IFN-γ induction

We had intended to investigate whether switching transcription on and off again would cause a concomitant switch in replication time from relatively early back to late. The 'chase' experiments— to remove the 24h 'pulse' of IFN-γ and return the fibroblasts to their usual transcriptional state — were performed alongside the initial incubation experiments outlined above. The chase was performed simultaneously because, if successful, the cells would be of a similar passage as the induced cells (albeit slightly older because of the continuing culture needed after the IFN-γ had been removed) and would share a similar 'natural history'. The Northern and immunofluorescence results from the initial chasing experiments are shown in Figures 5.1 (lanes 3 and 6) and 5.3, respectively.

![Figure 5.3. Immunofluorescence analysis of MRC5 cells 48h after removal of the exogenous IFN-γ stimulus. Although the cell numbers are small, it is clear that there is still a peak corresponding to surface expression of DRA.](image)

These results show that the chase was unsuccessful. Although a qualitative examination of the signal in lane 3 (Fig 5.1) suggests a reduction in the HLA-DRA mRNA levels 48h post-incubation, it is not sufficiently reduced for us to consider the fibroblasts to be returned to their native state. Repeat attempts were also unsuccessful, even with the chase period increased >72h. The explanation for this can be found in the work of Hardy & Sawanda (1989) who showed that IFN-γ strongly up-regulates
its own expression in human peripheral blood lymphocytes. The establishment of such an autocrine feedback loop would maintain the expression of \textit{HLA-DRA}, as well as the other 200 or so IFN-\(\gamma\)-inducible genes (Boehm \textit{et al.}, 1997). It is not clear how long cells would have to be cultured after the initial exogenous stimulation for the effects of this endogenous IFN-\(\gamma\) to wear off. Consequently, observations were not collected from these 'chased' experiments.

5.3.4. The MHC class II and III regions in induced fibroblasts

5.3.4.1. Induction fails to make the HLA-DQ/DR locus earlier replicating

Induced fibroblasts had the transcriptional and phenotypic appearance of B-cells, with respect to the MHC class II. We then examined how this class II-positive state would be reflected in the replication timing profile. Figure 5.4 shows the profile for the MHC class II and III regions in induced fibroblasts (superimposed on untreated cells). After induction, part of the classical class II region (from \(\sim\)300-600 Kb, the HLA-DP locus to the \textit{DOB} gene) switched to earlier replication and part of it (the HLA-DQ/DR locus, \(\sim\)700-1100 Kb) remained relatively late.

The mean replication time from \textit{HLA-DP} to \textit{DOB} (probes MANN 3.6 to M4 on Fig 5.4) was 0.45 +/- 0.03 in the induced cells compared with 0.60 +/- 0.04 in untreated fibroblasts. However, this "positive" effect on the late replicating regions in fibroblasts did not extend into the HLA-DQ/DR locus, where IFN-\(\gamma\) failed to illicit earlier replication. The relative replication time averaged across the HLA-DQ/DR locus (probes F1121 to LH1) was the same in both induced and untreated fibroblasts, 0.64 +/- 0.03. These data support the hypothesis that IFN-\(\gamma\) induction of the classical class II and class III regions confers a B-cell-like replication profile in fibroblasts, including
Figure 5.4. Comparing the replication profiles of the class II and III regions in human fibroblasts incubated with γ-IFN for 24h (green) and non-induced fibroblasts (red). The internal controls PYGM (early) and A9.5 (late) are indicated. Again bars represent +/- 2 standard errors. The profile of the B-cells is indicated by the dotted line.
the relatively late replication of the transcribed HLA-DQ/DR locus (see 5.3.4.4, below).

5.3.4.2. Induction causes other MHC sequences to replicate earlier

What of the effect of induction on those sequences that already replicated early in untreated fibroblasts and those (other than the HLA-DQ/DR locus) that replicated late? As outlined above, the previously late replicating 300-600 Kb in the classical class II region replicated earlier. The mean replication time from HLA-DP to the distal end of the TAP/LMP gene cluster (Fig 5.4) was 0.46 +/- 0.03 in the induced cells compared with 0.60 +/- 0.04 in untreated fibroblasts. This is a relative difference of approximately 14% of S-phase, which is equivalent to 50 minutes if S-phase is 6h, or 67 minutes if 8h. Many genes in the class II region are up-regulated by IFN-γ (see Fig 2.1), including the classical class II genes themselves and the TAP/LMP genes; therefore the switch to earlier replication is probably a reflection of this activation. In this sense, the positive correlation between gene expression and early replication has been preserved. However, it makes the persistently late replication of the HLA-DQ and -DR genes that more puzzling. Whatever the nature of the temporal restriction preventing the HLA-DQ/DR locus from replicating earlier, it does not appear to affect the proximal half of the classical class II region.

We also observed a change to earlier replication at the extreme ends of the class III region. The first two probes (from ~1175-1250Kb, Fig 5.4) had a mean replication time of 0.45 +/- 0.01 in induced cells and 0.58 +/- 0.03 in normal fibroblasts. This is a difference of 13% or between 47 and 62 minutes. At the telomeric end, the mean replication time of the last three probes (~1800-1880Kb, Fig 5.4) was 0.45 +/- 0.02 in induced cells and 0.55 +/- 0.04 in normal fibroblasts. This difference of 10% of S-phase equates to between 36 and 48 minutes. Their slightly earlier replication may again be a reflection of the general up-regulation of the region.
The profile in Figure 5.4 (and see Fig 5.5) supports the hypothesis that interferon would have little effect in regions that replicated early in normal fibroblasts. Induction did not alter the replication time of the extended class II region (50-200Kb, Fig 5.4) or the central portion of the class III region (1300-1700Kb, Fig 5.4), which both replicated early in induced and non-induced cells. Their respective replication times were 0.41 and 0.42 in induced cells and 0.44 and 0.41 in normal fibroblasts.

5.3.4.3. The class III region in induced and non-induced fibroblasts
In keeping with the data in section 4.3.3.2 and the observations of Smith & Higgs (1999), the central portion of the class III region (1300-1700 Kb, Fig.5.4) has an almost identical pattern of peaks and troughs in both induced and non-induced fibroblasts (and which is similar to that in B-cells). And as in non-induced cells, its profile in induced cells is maintained at a slightly later replication time; implying that execution of the temporal programme is still delayed in induced fibroblasts relative to B-cells. Nevertheless, induction did cause the extreme ends of the class III region to shift to earlier replication, supporting the prediction of our hypothesis that IFN-γ would have a "positive" effect on those parts of the MHC that were late replicating in non-induced fibroblasts.

5.3.4.4. Induced fibroblasts adopt a B-cell replication profile
The dotted line on Figure 5.1 is the B-cell replication profile presented in Chapter 4. Comparing this with the profile for induced fibroblasts, supports the hypothesis that the general effect of induction is to make the fibroblasts adopt a replication profile similar to that in B-cells. This holds true for the replication time of the HLA-DQ/DR locus. Indeed, the best predictor of replication time in induced fibroblasts is the replication time in B-cells and not that in non-induced fibroblasts. Moreover, the
percentage of singlet carrying chromosomes in non-induced fibroblasts adds no further information as to the percentage in induced cells. Thus, if a probe is relatively early replicating in B-cells it will be early replicating in induced fibroblasts, and if it is late in B-cells it will still be late in induced fibroblasts.

This relationship is further explored in Figure 5.5, below. The difference in the % singlet (S) carrying chromosomes between induced and non-induced fibroblasts (y-axis) is plotted against the difference between B-cells and non-induced fibroblasts (x-axis). In essence, the x-axis looks at the effect of "being a B-cell" on replication time, and the y-axis at the effect of IFN-γ induction.

Figure 5.5. The effect of induction on fibroblasts. The difference in the percentage of singlet carrying chromosomes between fibroblasts with and without interferon is plotted against the difference between B-cells and fibroblasts. See text for details.

The plot shows that (with the exception of M4) the effect of IFN-γ on a probe's replication time is either minimal, or it makes the probe replicate at a time similar to that
in B-cells. That is, probes either stay near the horizontal line of zero effect (dashed box, Fig 5.5) – these are the probes from the central portion of the class III region and HLA-DQ/DR locus (see above sections). Alternatively, IFN-γ induction pulls the probes down towards the diagonal, where the effect of induction on replication time begins to correlate with "being a B-cell". These probes map to the classical class II region (excluding the HLA-DQ/DR locus) and the extreme ends of the class III region.

In the case of M4, it was even earlier replicating than would be predicted by its mean replication time in B-cells. It showed little difference between B-cells and non-induced fibroblasts (approximately -0.05, Fig 5.5), but IFN-γ induction caused a profound change (approximately -0.15), shifting it well beyond the diagonal (this change can also be seen in Fig 5.4). Probe M4 is close to the persistently late replicating HLA-DQ/DR locus, though the significance of this, if any, is unclear.

Considering these replication timing and transcription data, it appears that IFN-γ induction of MHC class II expression in fibroblasts does not cause the HLA-DQ/DR locus to replicate any earlier, but it does cause the fibroblasts to adopt a replication profile similar in overall appearance to that in B-cells.

5.4. Discussion

5.4.1. Why is the HLA-DQ/DR locus unusual?

Since B-cells are a good predictor of the replication time in induced fibroblasts, this suggests that their respective replication profiles are "B-cell-like" specific. Whilst such a positive correlation may explain the general earlier replication observed in induced fibroblasts, it does not explain why the HLA-DQ/DR locus should remain later replicating when transcribed.
The persistently late replication of the HLA-DQ/DR locus does not mean transcription has no part to play in influencing replication time. On the contrary, the positive effect of induction on the rest of the class II region and on the extreme ends of the class III region preserves the relationship between expression and early replication (see probes shifted towards the diagonal in Fig 5.5). However, contrary to the rules governing replication proposed by Hatton et al. (1988), who showed that tissue-specific genes replicate early when expressed, we have identified a cluster of co-ordinately controlled, tissue-specific genes (HLA-DQ and DR) that are relatively late replicating regardless of whether or not they are transcribed.

This is the third occasion during the course of this work that the HLA-DQ/DR locus has been observed as late replicating, suggesting that whatever the influencing factor, it is a constitutive property of the locus, immutable from one cell type to another and from one transcriptional state to another. The underlying arrangement of the chromatin – be it for transcription, replication or general structural integrity – is an obvious candidate (see 5.4.2). Do any other physical or genetic features of the HLA-DR/DR locus provide clues as to the reason for its late replication?

5.4.1.1. Linkage disequilibrium in the HLA-DQ/DR locus

As discussed in section 1.3, the highly polymorphic HLA class I and class II genes are inherited as linked sets of alleles termed haplotypes. Because of this, the MHC has been extensively studied for linkage disequilibrium – the non-random evolutionary process responsible for generating the haplotypes, reviewed in Carrington (1999). Recombination tends to take place in certain locations in the MHC, and non-random recombination hotspots have been observed in both the class I and class III regions (Thomsen et al., 1994). In the classical class II region (see Fig 2.1), crossovers have been described throughout the DP-DQ sub-regions, but there are two segments where very little recombination has occurred: HLA-DP to HLA-DNA and, significantly, the
HLA-DQ/DR locus (Cullen et al., 1997). Recombination has never been observed between the HLA-DQA and HLA-DRA genes in all of the large Centre d'Etude Polymorphisme Humain (CEPH) families studied, and this region is therefore inherited as an intact block. It is thought that certain haplotypes of DQ-DR are selected because they provide complementary benefits to the host (Carrington, 1999). Different haplotype lengths (Dunham et al., 1989; Kendall et al., 1991; Zhang et al., 1990) and gene organization (Rollini et al., 1985; Spies et al., 1985) have been suggested as possible structural barriers at this locus, co-selected in order to prevent the disassembly of these beneficial haplotypes.

If natural selection’s priority is to maintain a genomic organization that prevents recombination in the HLA-DQ/DR locus during meiosis, does this organization persist to subsequently influence replication? No-one has yet examined the interrelationship, if any, between linkage disequilibrium and replication timing. We could speculate that the evolutionary processes which shaped the HLA-DQ/DR locus neglected to recruit replication initiation signals (be they sequence- or chromatin-specific, or both). The precepts of natural selection may have since excluded any such additions to the locus, which has subsequently relied on the flanking chromatin for its replication. The B-cell line used here was homozygous for the DRB genes in its DR-haplotype, but we do not know the haplotype for the fibroblasts (see section 1.3 and Fig 2.2). Future work should determine the haplotype of the fibroblasts and it might be worthwhile performing FISH analysis on further cell lines to check whether there is a relationship between late replication at the HLA-DR/DQ locus and the number of HLA-DRB genes in the DR haplotype (see Fig 2.2).

5.4.1.2. MARs in the MHC

There is the possibility that the HLA-DQ/DR locus, or part of it, interacts directly in a non-random way with the nuclear architecture (whether it is defined as matrix, scaffold
or skeleton, see (Craig et al., 1997)), mediated by matrix- or scaffold-attachment regions (MARs/SARs). MARs/SARs are genomic sequence elements thought to delineate the structural and functional organization of the eukaryotic genome. Though originally identified by their ability to bind the nuclear matrix or scaffold (Mirkovitch et al., 1984), binding has not been assigned to any unique sequence element and is dispersed over several hundred base pairs. A recent computational analysis (van Drunen et al., 1999) has identified a bipartite sequence element that is unique to a large group of eukaryotic MARs/SARs. This so called MAR/SAR recognition signature (MRS) consists of two individual sequence elements that are <200bp apart.

Researchers at the Sanger Centre, Hinxton are currently using van Drunen’s MRS to identify MARs within the complete human MHC sequence (Roger Horton, personal communication). The preliminary data for the entire MHC, which Roger Horton has very kindly shared with us, appear to show 436 matches for motif 1 of the bipartite MRS, giving approximately 11.9 MRSs per 100Kb, with a range of 2 to 28/100Kb. Using these preliminary data, we can say that the HLA-DQ/DR locus is unremarkable in that it appears to contain approximately 21 MRS motifs per 100Kb (Roger Horton, personal communication). It remains to be seen which, if any, of these potential MARs can be shown by biochemical means to interact with the nuclear matrix. The Human Cytogenetics Laboratory has started to examine matrix attachment in the MHC and has shown that MARs in the region of the TAP/LMP gene cluster do indeed bind the in vivo nuclear matrix (Rossen Donev, personal communication). Future work should now focus on the binding capabilities of the potential MARs in the HLA-DQ/DR locus.

If the HLA-DQ/DR locus is associated with nuclear matrix via a cluster of MARs, does this impede its progress through the replication machinery? There is no reason to think that this should be the case, since replication has been demonstrated to take place at replication factories that are themselves associated with the nuclear architecture in order to initiate and progress replication e.g. (Hozak et al., 1996). Alternatively, the
supposed association of the HLA-DQ/DR locus with a nuclear sub-structure may retard sister chromatid separation.

5.4.1.3. Are we measuring sister chromatid separation?

Assaying replication with FISH requires adequate separation of the replicated loci on sister chromatids in order to resolve the doublet signal. How long after replication it takes for doublets to be detected, and whether or not this time is constant for different sequences, has yet to be formally addressed (Boggs & Chinault, 1997). The spatial relationship of the new chromatids just after replication is not clear, making exact resolution time impossible to predict. If separation follows closely after replication then replication time is measured, if it is delayed then FISH results may reflect time of chromatid resolution. Such a restriction in resolution may occur between neighbouring domains with very different replication times. In their original paper, Selig et al. (1992) thought this was the case with one of their probes flanking the cystic fibrosis locus. Kawame et al. (1995) have since suggested that structural differences in the organization of the chromatin influence the rate of sister chromatid separation, delaying it and causing replicated loci to appear unreplicated. Could this explain the apparent lateness of the HLA-DQ/DR locus?

Usually with FISH, ~10-20% of nuclei have an asynchronous pattern, where one chromosome has replicated the sequence under investigation ahead of its homologous partner (so called SD nuclei) (Kitsberg et al. 1993a; Boggs & Chinault 1994; Torchia & Migeon 1995). A significant increase in the number of SD nuclei is usually indicative of the asynchronous replication associated with the inactive X chromosome and imprinted autosomal regions (Kitsberg et al. 1993a; Knoll et al. 1994; Boggs & Chinault 1994; Greally et al. 1998). The MHC is co-dominantly expressed and should not be imprinted. However, an increased number of SD nuclei for the HLA-DR/DQ
locus could, as Kawame et al. (1995) suggest, reflect asynchronous sister chromatid separation rather than actual asynchronous replication.

Table 5.1 shows the percentage of S-phase nuclei containing one singlet and one doublet signal for each of the four probes surrounding the HLA-DQ/DR locus (see Fig 5.4) compared to the average number for the rest of the probes in the class II and III regions.

<table>
<thead>
<tr>
<th></th>
<th>SD nuclei</th>
<th>F1121</th>
<th>E1448</th>
<th>Pakr</th>
<th>LH1</th>
</tr>
</thead>
<tbody>
<tr>
<td>B-cells</td>
<td>21 (22)</td>
<td>21</td>
<td>20</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>Fibroblasts</td>
<td>21 (21)</td>
<td>26</td>
<td>21</td>
<td>22</td>
<td>19</td>
</tr>
<tr>
<td>Induced fibroblasts</td>
<td>22 (22)</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>22</td>
</tr>
</tbody>
</table>

Table 5.1. The percentage of S-phase nuclei with SD signals for the probes surrounding the HLA-DQ/DR locus compared to the average of all the other probes.

It is clear from this table - as with the MAR data - that there is nothing unusual about the HLA-DQ/DR locus. Since the HLA-DQ/DR locus appears to resolve into a doublet at the same rate as every other MHC locus tested, we can reasonably conclude that its late replication is not an experimental artefact caused by retarded sister chromatid separation producing false negative singlet signals.

While these data suggest there is a positive correlation between FISH signal and mean replication time, further evidence could be gained by monitoring the conversion of the probes in the HLA-DQ/DR locus from the singlet to doublet state as a function of time through S-phase (Selig et al., 1992). This would require cells separated into different S-phase fractions by retro-synchrony and for each fraction to be assayed by FISH to give cumulative replication profiles. However, we were satisfied with the performance of the FISH method and these corroborating data were not gathered.
5.4.2. Patterns in the genome

The replication profile of the central portion of the class III region had a pattern of peaks and troughs that was virtually identical in both induced and non-induced fibroblasts. In their book *Metaphase Chromosome Bands: Patterns in the Genome* (1997), Bickmore & Craig discuss the cytogenetic banding patterns seen on metaphase chromosomes, and on page 1 they immediately point out:

"It [the chromosome] is ... a structure full of pattern, and pattern in biology is usually a manifestation of underlying mechanisms of control."

With this second clause firmly in mind, let us return to the work of Smith & Higgs (1999) who have suggested an underlying mechanism of control responsible for the pattern of peaks and troughs seen with FISH. In their work with the α-globin locus on 16p13.3, Smith & Higgs recorded replication profiles that were consistently similar on every 16p homologue examined, regardless of whether the homologue was in an erythroid or non-erythroid cell hybrid. They suggested the consistent pattern could be indicative of a stably inherited genetic or epigenetic programme that establishes and maintains the number and position of replication origins and the time in S-phase when they are activated. After showing that a critical genetic sequence/zone was not necessary to initiate replication in this region and by contrasting their findings with those for the β-globin locus (Forrester et al., 1990), they concluded that the pattern of peaks and troughs may be related to the underlying chromatin structure and the chromosomal context in which origins find themselves (Smith & Higgs, 1999).
5.4.2.1. Random initiation in two massive replicons?

Each profile recorded in this thesis has seen a change from early to late replication across the class II/III boundary. However, when a stretch of chromatin, such as the central portion of the class III region, shows no gradient in replication, there are two possible explanations. Either replication initiates synchronously at several adjacent replicons, or replication is accomplished by random initiation events taking place within a long stretch of chromatin.

Taking the latter scenario first: Since the gradient between early and late replication across the class II/III boundary agrees with the rate of eukaryotic replication fork movement, we might suggest two massive replicons replicate the class II and class III regions (and broadly correspond to their isochores). These super-replicons could be initiated - using the class III region as an example - anywhere within the stretch in which adjacent probes have similar replication times. Such a stretch of DNA could be likened to the replication initiation zones first proposed by Vaughan (1990). However, at approximately 500Kb, this class III initiation zone would be much bigger than the largest recorded to date: 55Kb by Dijkwel (1994). Nevertheless, DNA polymerisation could initiate anywhere within this zone, with individual cells using different initiation sites. The randomly chosen initiation sites would be averaged out across the cell population and therefore no gradient would be seen for the central portion of the class III region. Once the replication fork has moved far enough away from the mêlée of the initiation zone, we might resolve its passage into the HLA-DQ/DR locus.

Though some individual replicons can be large, in some cases >1Mb (Yurov & Liapunova, 1977), such massive replicons would be an inefficient way of replicating a region as important as the MHC. Our putative class III super-replicon would need to be at least 1000Kb end to end, and at the replication fork rate we measured it would take approximately four hours to copy the class III region bidirectionally from an ideally situated initiation site. After referring back to Fig 4.1 which includes class I
region data, we must ask where is the other end of the class III super-replicon? The fact that the class I region has a relatively constant early-middle replication time and does not show a gradient of replication timing similar to the one at the class II/III boundary, argues against the presence of a single, giant class III replicon.

5.4.2.2. Synchronously activated adjacent replicons

Given the more generally accepted size for replicons of between 50–300Kb, the 3.6Mb MHC region is likely to house between 12 and 72. Concentrating on the profile for the class III region in B-cells and non-induced fibroblasts (Figs 4.6 and 5.4), the replication time from ~1300-1700Kb is relatively constant in both cell types. The lack of a replication gradient in this 500Kb stretch might be explained by combining Smith & Higgs' hypothesis of a stable temporal programme related to chromatin structure with the (near) synchronous firing of adjacent replicons (Hand, 1978).

The consistency in the pattern of peaks and troughs recorded for contiguous probes in the class III region in the different cell types could be due to a stable epigenetic programme. Because the same programme is faithfully repeated in every cell type, this could indicate that specific chromatin/nuclear interactions are responsible for defining the units of replication. Execution of the temporal programme involves the synchronous activation of adjacent replicons in the class III region, which could explain why there is no gradient in the replication time. Despite the lack of a gradient, we are nevertheless given some indication as to the initiation sites of the most consistently employed and earliest replicating replicons e.g. around 1650Kb and represented by probe G10s (see Fig 4.6, both cell types).

Just as the consistency in the replication profile of the class III region might reflect the temporal organization of a constitutively expressed region, so the differences (see the proximal half of the classical class II region), might reflect the cell type-specific
differences between B-cells and non-induced fibroblasts. We should not expect an exact reproduction of the B-cell's classical class II pattern of peaks and troughs in induced fibroblasts. Though forced into class-II expression, the native chromatin arrangement of the classical class II region would still be fibroblast-specific. Therefore, the classical class II in induced fibroblasts may become earlier replicating because its innate temporal programme is executed sooner, perhaps due to the increased exposure of the region to the processes of DNA metabolism.

5.4.2.3. Asymmetry in the replicon structure of the MHC

If a cluster of synchronously activated replicons replicate the class III region, then why is there a gradient in replication time across the classical class II/III boundary region? Since replication forks can not move backwards in time, the class II/III boundary must be replicated in the direction C47 to LH1, and it follows then that the replicon containing C47 is highly asymmetric. This asymmetry implies that no other replicons are fired in the intervening stretch of DNA. This is not unprecedented. Ermakov et al. (1999) observed a uni-directional replication fork proceeding from early to late replicating domains in the IgH locus in a murine erytholeukemia cell line. They suggested that a single replication fork progressed from the last in a cluster of early activated replicons to the first in a cluster of late activated replicons. In addition, Spack et al. (1992) identified a number of potentially asymmetric replicons in their replication timing analysis of the mouse H-2 MHC complex. Fibre autoradiography has already shown that up to 10% of replicons in higher eukaryotes are unidirectional (Edenburg 1975), and there is evidence to suggest that replication foci contain clusters of replicons that differ in both number and size (Berezney et al., 2000).

But why are there no intervening replicons in the class II/III boundary? Does the nature of the GC transition between the classical class II and class III isochores preclude the existence of origins? Not if we look at the centromeric end of the MHC in B-cells or
induced fibroblasts (see Figs 1.2 & 5.4), where the isochore boundary between the extended and classical class II regions is not coincident with the change from early to later replication time. The extended class II region and its boundary with the classical class II sit in a part of the profile where the replication timing of adjacent probes is relatively constant, indicating synchronous control. The gradient to later replication starts on the proximal side of the \textit{TAP/LMP} genes and finishes at the start of the expressed \textit{HLA-DQ} genes. Thus, the gradient in replication here is more distal than the GC/AT transition.

Finally, what becomes of the replication fork entering the distal end of the classical class II region via the HLA-DQ/DR locus – the only part of the MHC for which cosmid clones were unavailable? However, regardless of what any intervening clones might have to reveal, we would still be left needing to explain why this locus is relatively later replicating than its neighbours.
Chapter 6. Evidence for at Least One Replicon in the MHC Class III Region

6.1. Introduction

DNA replication initiates at thousands of origins in the human genome, creating almost twice as many replication forks in the process. Each fork is on course to fuse with its neighbours and thus form a new contiguous DNA molecule. These thousands of units of replication – initiated at origins and bound by the junctions between neighbours – are called replicons (Hand, 1978). Replicons are generally thought to be between 50-300Kb in length in mammals (Edenberg & Huberman, 1975), though some estimates put them in the region of 500Kb (Bickmore & Oghene, 1996; Verbovaia & Razin, 1997); and there is even evidence to suggest that individual replicons can be >1Mb (Yurov & Liapunova, 1977). It is thought that replication foci (see section 1.2.3) contain clusters of replicons, but that the size and number of replicons in each focus is extremely heterogeneous and a significant proportion of foci are composed of a single large replicon (Berezney et al., 2000). Clusters of adjacent replicons are sequentially activated throughout S-phase, with replicons in the same cluster thought to fire in synchrony. Evidence for the higher-order functional grouping of clusters of replicons comes from the replication banding studies previously described in section 1.2.1.1 (Blumenthal, 1974; Drouin et al., 1994) and see (Blow et al. 2001).

Given the more generally accepted size for replicons of between 50–300Kb, the 3.6Mb MHC region is likely to house between 12 and 72. The experiments in this chapter were designed to provide evidence for at least one of these replicons in the MHC class III region.

6.1.1 Hypotheses on a putative replicon

In each of the replication profiles recorded in this thesis, probe G10s in the MHC class III region had a consistently earlier mean replication time than its flanking neighbours
(KM2 and M7B, Fig 2.1 and asterisks in Fig 4.6). Thus, it appeared that at least 300 Kb of the class III region was replicated from an origin located close to G10s. Based on this observation, we hypothesized that probes KM2 and M7B are in the same replicon as G10s.

This hypothesis predicts that in two-colour FISH experiments (with G10s as one of the probes), the odds of observing a doublet for G10s and a singlet for the co-hybridized probe will be greater if the co-probe is in the same replicon as G10s. In addition, if there is only one G10s doublet in a nucleus and the co-probe has an asynchronous (SD) replication pattern, its doublet is more likely to be on the same homologue as the G10s doublet if the two probes are in the same replicon. These predictions can be tested by comparing the order of replication of G10s and its adjacent probes with the order when G10s is co-hybridized with probes much further away. Our hypothesis predicts that G10s should replicate before probes located in the same replicon but not necessarily before probes in more distant replicons located further along the same chromosome.

6.2. Methods

To confirm the difference in replication time between probes on the same chromosome, their relative order of replication is established by co-hybridizing them in B-cells. This approach allows direct visualization of the linked singlet and doublet signals on both homologues in the same S-phase nucleus. The ratio of these signals should give a reliable indication of their true relative order of replication. This double-labelled FISH protocol was originally reported by Kitsberg et al. (1993b), who used it to confirm the in vivo direction of the DNA replication fork in the human β-globin locus suggested by non-FISH methods.
In our two-colour analysis, pairs of cosmids (with G10s in every pair) were co-hybridized and detected using different fluorochromes. Co-probes were differentially labelled with biotin-14-dATP or digoxigenin-11-dUTP, and each experiment was repeated with the co-probes labelled in the alternate colour. This was to preclude any experimental bias that may be associated with differences in the labelling and detection systems, or in the resolvability of green versus red signals. We selected nuclei containing at least one doublet signal in either colour and scored approximately 175 nuclei on this basis.

Figure 6.1. Assuming G10s was labelled in green and the co-probe in red, below are the possible combinations of observable two-colour FISH signals. With this fluorochrome combination, nuclei A, C, E and G would correspond to the predicted order of replication, whereas B, D, F and H would correspond to the "wrong order of replication".

- If a nucleus contains at least one doublet, it should be at G10s (green), as in A. B should be rare with KM2 & M7B, see below.

- If the co-probe has an SD pattern, its doublet should be on the same chromosome as the G10s doublet, as in C. D should be rare with KM2 & M7B.

- E is similar to A, with G10s replicating before the co-probe on both homologues. Again, F should be rare with KM2 and M7B.

- G is a combination of A and C, with G10s doublets on both homologues. H should be rare with KM2 and M7B.
According to our hypothesis, probes KM2 and M7B are in the same replicon as G10s, whereas probes 027 and A1 are not (see Fig 2.1 and asterisks in Fig 4.6). In B-cells, both 027 and A1 have mean replication times (0.32 and 0.35, respectively) similar to KM2 and M7B (0.34 and 0.36, respectively). That is, on average, 027 and A1 are replicated after G10s. However, probes 027 and A1 map to the classical class II region and are not, therefore, expected to be in the same replicon as G10s. On a given chromosome, G10s (mean replication time, 0.24) should replicate before KM2 and M7B, which are only ~150Kb either side of it (A, C, E & G in Fig 6.1), but it may not always replicate before either 027 or A1, which are >1Mb away (B, D, F & H can be observed). In theory then, the odds of observing the "wrong order of replication" should be lower with probes KM2 and M7B than with either 027 or A1. The null hypothesis (which we hoped to reject) for the co-hybridization experiments was that there would be no difference in the order of replication between the class III region probes flanking G10s and the probes from the classical class II region.

6.3. Results

6.3.1. Rejecting the null hypothesis

We can compare the odds (see Appendix B) of observing G10s replication before the co-probe regardless of whether the FISH signals are on the same or different chromosomes – these are the combined odds of replication order (contrast with section 6.3.2). Since, on average, G10s is earlier replicating than any of the co-probes, we expect the combined odds to be >1. If probes KM2 and M7B are not in the same replicon as G10s, then their combined odds should be similar to those for 027 and A1 (since all four have similar mean replication times). If, on the other hand, KM2 and M7B are in the same replicon as G10s, their combined odds should be much greater (since the odds of observing a G10s doublet and a singlet for the co-probe will be greater if the co-probe is in the same replicon as G10s).
I would like to thank Dr Peter Sasieni, who helped tremendously with the preparation of the following data. Table 6.1 shows the combined odds for the co-probes used in the two-colour FISH analysis. There are two features of note: First, the combined odds for all four probes are not similar (though all are >1). Secondly, the increased estimated odds for KM2 and M7B support the hypothesis that they are in the same replicon as G10s.

<table>
<thead>
<tr>
<th>probe</th>
<th>mean replication time</th>
<th>combined odds</th>
</tr>
</thead>
<tbody>
<tr>
<td>KM2</td>
<td>0.34</td>
<td>3.8</td>
</tr>
<tr>
<td>M7B</td>
<td>0.36</td>
<td>4.8</td>
</tr>
<tr>
<td>027</td>
<td>0.32</td>
<td>1.1</td>
</tr>
<tr>
<td>A1</td>
<td>0.35</td>
<td>1.7</td>
</tr>
</tbody>
</table>

Table 6.1. The combined odds of observing a G10s replication before the co-probe regardless of whether or not the signals are on the same or different chromosomes. G10s has a mean replication time of 0.24 in B-cells.

To ensure that the combined odds for the class III region probes were truly different from those for the classical class II region, we chose the probes with the least difference in their combined odds (and which are the least likely, therefore, to be significantly different) and compared. To test whether the combined odds for KM2 (3.8) differ from those for A1 (1.7), we estimated a combined odds ratio 2.24 and obtained a p-value of less than 0.0001. Therefore, the combined odds of KM2 and A1 are significantly different, and we can reject the null hypothesis that there would be no difference in the order of replication between probes flanking G10s and those from the classical class II region >1Mb away.

These data support the prediction that G10s should replicate before probes located in the same replicon but not necessarily before probes in more distant replicons further along the same chromosome. Whilst it appears that probes KM2 and M7B are in the
same replicon as G10s, the possibility remains that despite our previous experimental results suggesting that all four co-probes have similar mean replication times, the larger odds for KM2 and M7B are due to the fact that these probes are actually later replicating than previously thought.

6.3.2. KM2 and M7B are in the same replicon as G10s

To distinguish between later replication of KM2 and M7B and the existence of a *bona fide* replicon, we looked at the odds of replication order for probes on the same chromosome separately from the odds for probes on homologous chromosomes (Appendix B). Formally, we compared the ipsi- and contra-chromosome odds of observing a G10s doublet with a singlet for the co-probe i.e., the odds of observing the "correct order of replication" on the same chromosome (ipsi) and on both homologues (contra). If there is no replicon then the ipsi- and contra-chromosome odds of observing a G10s doublet and a singlet for the co-probe will be similar. If there is a replicon, then our hypothesis predicts that the ipsi-chromosome odds will be greater.

The ipsi-chromosome odds are:

\[
\frac{1}{\frac{A + D + 2E + G}{B + D + 2F + H}}
\]

where "A" denotes the number of nuclei with pattern "A", etc (see Fig 6.1).

The contra-chromosome odds are:

\[
\frac{1}{\frac{A + C + 2E + G}{B + C + 2F + H}}
\]
Table 6.2. The ipsi- and contra-chromosome odds of observing a G10s doublet with a singlet for the co-probe.

<table>
<thead>
<tr>
<th>probe</th>
<th>mean replication time</th>
<th>ipsi-odds</th>
<th>contra-odds</th>
<th>ipsi/contra ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>KM2</td>
<td>0.34</td>
<td>5.5</td>
<td>3.0</td>
<td>1.8</td>
</tr>
<tr>
<td>M7B</td>
<td>0.36</td>
<td>7.4</td>
<td>3.7</td>
<td>2.0</td>
</tr>
<tr>
<td>027</td>
<td>0.32</td>
<td>1.1</td>
<td>1.1</td>
<td>1.0</td>
</tr>
<tr>
<td>A1</td>
<td>0.35</td>
<td>1.8</td>
<td>1.7</td>
<td>1.1</td>
</tr>
</tbody>
</table>

Table 6.2 shows that probes 027 and A1 have similar ipsi- and contra-chromosome odds (1.1 & 1.1 and 1.8 & 1.7, respectively). This is reflected in their ipsi:contra odds ratio of ~1, indicating that they do not share a replicon with G10s and suggesting that their replication time is not determined by events taking place >1Mb away at the G10s locus. In contrast, the ipsi:contra odds ratio of ~2 for probes KM2 and M7B coupled with their increased ipsi-chromosome odds (5.5 and 7.4, respectively), provides further evidence that they are in the same replicon as G10s.

Despite these positive data, we would have predicted that the contra-chromosome odds for all four probes would be similar – the existence of a replicon should only really affect the ipsi-chromosome odds. The larger (than 027 and A1) contra-chromosome odds for KM2 and M7B suggest that there may be a common trigger for replicons on the two homologues within a nucleus. However, these two-colour FISH experiments cannot distinguish between the firing of one replicon inducing firing of the other replicon on the homologous chromosome and the existence of an "external factor" that triggers both replicons to fire.
6.3.3. G10s and co-probe doublets are on the same homologue

To look further at the difference between the order of replication on the same and on different chromosomes, we looked only at those nuclei in which only one copy of each probe was replicated (i.e., categories C & D in Fig 6.1). We expected replication of one probe on each of the sister homologues to be relatively rare when the probes are part of the same replicon. In contrast, when probes are from different replicons, we expected C and D nuclei to be equally likely. In Table 6.3 we estimate the proportion of nuclei in which we observed an asynchronous (SD) replication pattern for both probes, with the two doublets sharing the same homologue. We also give the p-value for this proportion being equal to 0.5 (i.e. C and D being "equally likely").

When KM2 displayed precisely one doublet, in 89% of nuclei this doublet was on the same homologe as the only G10s doublet (category C, Fig 6.1) – this proportion reached 92% with M7B. For both KM2 and M7B, the estimated proportions are extremely significant and are consistent with the prediction of our hypothesis: that if there is only one G10s doublet in a nucleus and the co-probe has an asynchronous (SD) replication pattern, its doublet is more likely to be on the same homologue as the G10s doublet if the two probes are in the same replicon.

<table>
<thead>
<tr>
<th>probe</th>
<th>mean replication time</th>
<th>ipso proportion (C/(C+D))</th>
<th>p-value (assuming C/(C+D) = 0.5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>KM2</td>
<td>0.34</td>
<td>0.89</td>
<td>2x10^{-14}</td>
</tr>
<tr>
<td>M7B</td>
<td>0.36</td>
<td>0.92</td>
<td>2x10^{-14}</td>
</tr>
<tr>
<td>027</td>
<td>0.32</td>
<td>0.64</td>
<td>0.006</td>
</tr>
<tr>
<td>A1</td>
<td>0.35</td>
<td>0.59</td>
<td>0.068</td>
</tr>
</tbody>
</table>

Table 6.3. The proportion of SD nuclei in which we observed both the G10s and co-probe doublets on the same homologue.
Interestingly, the proportions of 64% for O27 and 59% for A1 suggest that when these probes have an asynchronous (SD) replication pattern, then their doublet is more often than not found on the same homologue as the G10s doublet. Although probe A1 is just outside statistical significance (p-value 0.068), we should note that its proportion of 0.59 is nearer the 0.64 of probe O27 than it is to the theoretical value of 0.5. Since the ipsi- and contra-chromosome odds have already established that O27 and A1 cannot be in the same replicon as G10s, this temporal association between probes that are >1Mb apart suggests the existence of a "field effect" whereby replicons on a given chromosome are not completely independent of each other.

6.4. Discussion

The two-colour FISH experiments in this chapter were designed to uncover evidence for the existence of a putative origin in the MHC class III region around probes G10s, KM2 and M7B. Our data not only support the hypothesis that these probes are in the same replicon, they also give an insight into the behaviour of this replicon in relation to others located in the classical class II region. It appears that on any given chromosome 6, the activation of replicons separated by at least 1Mb can occur simultaneously. These data for the human MHC represent the first demonstration of the near synchronous firing of discrete replicons.

6.4.1. Identification of human MHC replicons

6.4.1.1 The KM2-G10s-M7B replicon

Both the ipsi- and contra-chromosome odds for KM2 and M7B are consistent with the hypothesis that these probes are in the same replicon as G10s. We predicted that G10s should replicate before probes located in the same replicon but not necessarily before probes in more distant replicons further along the same chromosome. Co-hybridization and comparison with probes O27 and A1 from the classical class II region provided evidence that this was indeed the case.
The KM2-G10s-M7B replicon is at least as long as the ~300Kb separating KM2 and M7B. Whilst these experiments do not precisely delineate the borders of the KM2-G10s-M7B replicon, they do give an indication as to how it is replicated. The direction of the replication fork is as follows: KM2 ← G10s → M7B. That is, from an origin initiation region somewhere within or close to probe G10s, a bi-directional replication fork moves outwards towards KM2 on one side and M7B on the other. This movement is confirmed by the mean replication times and relative order of replication recorded for these class III probes in B-cells (see Figs 4.6 & 2.1).

6.4.1.2. The classical class II region (probes O27 & A1)

We did not collect data on the order of replication of probes O27 and A1 and cannot say whether they share the same replicon in the classical class II region. It is difficult to be certain based only on their similar mean replication times in B-cells. If we view their relative positions in the B-cell replication profile, this seems to suggest that they are not in the same replicon (see asterisks on Fig 4.6). However, additional data from further 2-colour FISH experiments would be necessary in order for us to accept or reject this.

6.4.2. Levels of control in replicon activation

From the results presented in this chapter we can suggest at least three determinants of replication time within the nucleus. At the most basic level is the well established replicon; with probes in the same replicon replicating in a reproducible order. In those experiments in which KM2 or M7B was the co-probe, the existence of a small number of nuclei displaying the "wrong order of replication" can be explained by experimental error (e.g. the failure to observe a true doublet) or by variation in the exact location of the replication origin employed by the replicon. This latter explanation would be consistent with non-FISH studies of replication origins which show that within a given
initiation zone there may be a number of potential origins but in any given cell the probability is greatest for a particular site (reviewed in DePamphilis, 1999). That this is a statistical probability and not absolute for every cell, leaves open the possibility that some cells use different origins. What determines which origin is selected is not clear – it might be that in some situations other demands made on the chromatin surrounding G10s make the preferred origin unavailable.

The next determinant of replication time appears to operate across both chromosome homologues. We would have predicted the contra-chromosome odds for KM2 and M7B to be similar to those for O27 and A1. We did not expect initiation of the G10s replicon on one homologue to necessarily trigger the firing of its counterpart on the sister chromosome. We were forced to think again when the larger contra-chromosome odds for KM2 and M7B suggested there was an association between replicons on different homologues. This inter-chromosomal linkage effect means that if a given replicon happens to initiate particularly early, then its counterpart on the sister homologue will be early as well, and vice versa. However, this effect is not necessarily the result of communication between sister chromosomes: it might simply reflect the fact that homologous replicons happen to fire at a similar time in S-phase.

The final – and most interesting – level at which a determining factor appears to be operating is between different replicons on the same chromosome. Such intra-chromosome linkage is evinced by the fact that pattern C in Figure 6.1 is more common than pattern D – even for probes separated from the G10s locus by some distance (i.e. O27 and A1). That is, when only one G10s doublet was present in a nucleus and the co-probe had an SD replication pattern, its doublet was more likely to be on the same homologue as the G10s doublet. We had originally predicted this to be the case only for probes in the same replicon. It appears we can extend this to include probes that are separated by >1Mb of DNA. This linkage along the length of interphase chromosomes, such that replicons on a given chromosome do not appear to be
completely independent of each other, suggests the existence of a "field effect" that causes discrete replicons to initiate at the same time. This is consistent with the current model of DNA replication, in which initiation takes place synchronously at clusters of replicons. We suggest that the KM2-G10s-M7B replicon may – at least in some B-cell nuclei – be in the same cluster of replicon(s) as those represented by probes O27 and A1. The "field effect" which activates them could be the result of a physical association between the replicons in the cluster and the nuclear replication machinery.

Future work should address the following questions. How far does the simultaneous activation extend either side of the replicons identified here? Does it include replicons in the intervening section of DNA? Indeed, does it include the persistently later replicating HLA-DQ/DR locus? Future FISH experiments should examine further combinations of co-hybridized probes and generate a statistical measure of the extent to which they are temporally associated. In this way two-colour FISH could begin to generate the first temporal linkage map. However, such an extensive analysis would require numerous experiments and would not be a trivial undertaking.

6.4.3. The later replication time of the HLA-DQ/DR locus

If replicons in the class II and III regions are fired at similar times in S-phase, why is the intervening HLA-DQ/DR locus excluded? In the current view of in vivo replication, clusters of replicons are thought to associate with a factory of polymerases and co-factors that is in turn attached to the nucleoskeleton, with polymerization taking place as DNA is reeled through the fixed complex e.g. (Jackson, 1990) and reviewed in (Cook, 1991 and 1999). The inference from this model is that each replicon is attached to the replication factory via the chromatin region containing or near to the preferred site of replication initiation. The later mean replication time of the probes in the HLA-DQ/DR locus can still be accommodated in this view of replication.
We have alluded in previous chapters to the possibility of the HLA-DQ/DR locus possessing some unique chromatin properties that make it unusually late replicating. A more prosaic explanation – coming from work on the heterogeneity of replicons in replication foci (Berezney et al., 2000) – is that, in the early stages of S-phase, the HLA-DQ/DR locus is excluded from the replication factory responsible for the MHC class II and III regions. The HLA-DQ/DR locus remains excluded from the replication machinery until it is "reeled in" as a result of the replication of the neighbouring sequences. This model is also consistent with the class II/III boundary region being replicated by an asymmetric replicon – which reels the HLA-DQ/DR locus towards the replication machinery. Once it arrives at the replication focus, the HLA-DQ/DR region may activate its own late-firing origin of replication, but our experiments cannot provide evidence for this.

We can only speculate as to the initial cause of the exclusion of the HLA-DQ/DR locus from the replication focus. It may, for instance, be due to topological constrains unique to the locus. We should, therefore, examine its replication time in additional human cell lines, to assess whether it differs in individuals with different numbers of DRB genes in their DR-haplotype (see Fig 2.2). Alternatively, the exclusion of the HLA-DQ/DR locus may be because it lacks the necessary DNA sequence elements and chromatin structure which would allow it to associate with the replication factory. If this is the case, future work should analyses the available sequencing data to look for those DNA elements thought to play a role in replication initiation. It should also compare the combinations of these sequence elements in different DR-haplotypes.
Chapter 7. Final Discussion

7.1. Review

Chapter 3 established the validity of FISH as a means of assaying replication in a relatively small population of asynchronous cells. Since its inception, the FISH approach to replication timing has been employed in several laboratories (Bickmore & Carothers, 1995; Boggs & Chinault, 1994; Haaf, 1997; Kitsberg et al., 1993a and 1993b; Selig et al., 1992; Smith & Higgs, 1999; Squire et al., 2000; Strehl et al., 1997). We were, therefore, confident in its utility as a simple and effective means of examining the correlations between replication timing, genomic organization and gene expression.

Figure 7.1, below, is a comparison of the relative physical distances sampled in some of the replication timing studies discussed in this thesis. Our analysis of the human MHC presented in Chapter 4 is the longest and most detailed analysis to date (and see contiguous probes in Figure 2.1).

Smith and Higgs (1999), total 325Kb.

Tenzen et al. (1997), total 370Kb.

Selig et al. (1992), total ~1.2Mb

This thesis, 3.9Mb

Figure 7.1. The relative physical distances over which replication timing data have been sampled in previous studies, compared with the work presented here. In this thesis, the entire 3.9Mb was only analyzed in B-cells.
7.1.1. Features in the genome

In Chapter 4 we used the FISH method to estimate the mean replication time of cosmid clones covering the human MHC. We intended to collect replication timing data across a recognizable feature in the primary DNA sequence: the boundary separating the class II region's L2 isochore from the H3 isochore of the class III. In addition, by examining one of the most gene dense loci in the genome and comparing its relative replication profile in B-cells and fibroblasts, we wished to re-examine some of the established ideas concerning replication time and gene expression (Hatton et al., 1988).

The replication timing profile of the MHC region revealed a separation into early replication on the class III side of the isochore boundary, and late replication at the distal end of classical class II region, termed here the "HLA-DQ/DR" locus. We estimated the rate of change in replication time between probes hybridized to the distal end of the classical class II region and the proximal end of the class III region (see Figs 1.2, 2.1 and 4.6). This estimate was consistent with the passage of a replication fork from the class III region into the distal end of the HLA-DQ/DR locus, moving within the accepted rate of between 0.3-6Kb/min (Edenberg & Huberman, 1975). This forced us to disagree with the conclusion of Tenzen et al. (1997) who had previously suggested a "precise switch" in replication time across the GC/AT transition separating the class II region's L2 isochore from the H3 isochore of the class III. This, they suggest, corresponds to the exact boundary delineating the small G-sub-band previously identified in 6p21.31 by Yunis (1981).

This interpretation makes the class II region the small G-sub-band and places the class III region at the very edge of the R-band. However, we find it difficult to accept that something as relatively massive as a chromosome band boundary can be described to within a few specific nucleotides. It might be the case that the classical class II region constitutes the small G-sub-band in 6p21.31, but not for the reason Tenzen suggests
(Tenzen et al., 1997). Our analysis of interphase chromosomes suggest that the "change" in replication time across the class II/III boundary is actually consistent with the in vivo movement of a replication fork from the class III region into the distal end of the HLA-DQ/DR locus.

7.1.2. Replication timing and gene expression

The results in Chapter 4 also show that the replication timing of certain loci in the MHC is different in different cell types. That is, some of the tissue-specific loci in the MHC class II region replicated earlier in B-cells than fibroblasts. This is consistent with idea that tissue-specific loci fall into one of two categories. The first replicate early in S-phase whether they are expressed or not e.g. (Hatton et al., 1988), and the second have a developmentally regulated pattern of replication whereby they replicate early in expressing cells, and late in non-expressing cells e.g. (Dhar et al., 1989). However, the replication timing and expression data for the HLA-DQ/DR locus are not consistent with either of these categories. The persistently late replication of the HLA-DQ/DR locus provides the first evidence for an active gene or gene cluster that is relatively late replicating.

We explored the relationship between replication and gene expression further with the experiments in Chapter 5. Here we incubated human fibroblasts with IFN-γ in order to induce expression of MHC class II loci, including the HLA-DQ and HLA-DR loci. In these experiments, induced fibroblasts adopted a relative replication timing profile similar to that in B-cells. This included the relatively late replication of the HLA-DQ/DR locus (which we were satisfied was not an experimental artefact).
7.1.3. Replication organization

In addition, we noticed in the class III region of the replication profiles that the pattern of peaks and troughs in the data was remarkably similar in each. We are interested in Smith & Higgs' idea that the patterns may relate to the underlying chromatin structure and the chromosomal context in which replication origins find themselves (Smith & Higgs, 1999). Given their high resolution (i.e. contiguous probe coverage), it is possible that these FISH analyses are able to detect subtle differences in mean replication time along the path of the interphase chromatin. The question of whether or not this temporal organization reflects an underlying structural organization is a perennial concern of cytogenetics. Extensive biochemical analyses of the human MHC locus and exploitation of the advances in 3D confocal microscopy should begin to provide some real answers. For our part, we have identified one of the most consistently employed and earliest replicating replicons in the class III region. We sought further evidence for the existence of this replicon in the two-colour FISH experiments presented in Chapter 6.

The two-colour FISH results support the existence of a replicon in the MHC class III region containing probes KM2-G10s-M7B. These data also give the first demonstration using interphase chromosomes of the near synchronous activation of discrete replicons separated by >1Mb. We expanded on a previous model and suggested that – in some B-cells at least – replicons in the class II and III regions share the same complex of replication machinery (Cook, 1991). This can even accommodate the persistently late replication of the HLA-DQ/DR locus – which, in the early stages of S-phase, may be excluded from the replication machinery until it is "reeled in" as a result of the replication of its neighbouring sequences. Such a fate might even be common for chromatin regions separating clusters of synchronously activated replicons. (In this case, the clusters correspond to the replicons in class II and class III regions.)
7.2. Future Work

Workers should continue to employ high resolution interphase FISH analyses in other regions of the human genome. If independent investigators identify timing features with a consistency similar to those reported here (see also Smith & Higgs, 2000), we can be more certain that our data are representative of phenomena that exist independently of our observations.

As well as employing FISH to re-estimate the replication timing profile of the mouse MHC, it should also be used to examine the replication timing of the human MHC classical class II region in plasma cells – these are terminally differentiated B-cells which stop expressing class II molecules. Would the replication profile resemble that of a B-cell, a fibroblast, or neither? And how would the HLA-DQ/DR locus behave?

With the first draft of the human genome completed (International Human Genome Sequencing Consortium, 2001; Venter et al., 2001), future work should focus on the application of more powerful computer programmes with which to examine the DNA for clusters of potentially important sequence elements i.e., eukaryotic origins (e.g. Dobbs et al., 1994). These analyses will be important in helping to refine the definition of a mammalian replication origin, since an unequivocal consensus on this is currently still lacking. Fiber-FISH analysis may offer a relatively simple means of providing candidate clones for these computer analyses, by identifying those with recognizable "replication intermediates" when hybridized to "released chromatin" e.g. see (Rosenberg et al., 1995). Such work could begin with the clones KM2-G10s-M7B, and would make a useful introduction to the FISH method. The Human Cytogenetics laboratory should examine the matrix binding capabilities of the MAR sequence elements identified in the HLA-DQ/DR locus with the same assay used previously in the TAP/LMP region (Roger Horton and Rossen Donev, personal communications).
Support for the role, or otherwise, of the underlying chromatin structure in determining the replication profiles demands a thorough biochemical analysis (DNAse hypersensitivity, etc) of the general configuration of the chromatin throughout the MHC. In the meantime, FISH could examine the replication timing of the HLA-DQ/DR locus in cell lines whose MHC regions have different numbers of (and are heterozygous for) the DRB genes in the DR haplotype. We might also think about whether the later replication of the HLA-DQ/DR locus has anything to do with its immune-related functions.

7.3. Has theory changed?

On the question of whether the classical class II region constitutes the small cytogenetic G-sub-band previously observed in 6p21.31. It might, but the change in replication timing we record across the class II/III isochore boundary region is consistent with the passage of a eukaryotic replication fork. This observation from our interphase FISH analyses is not predicted by Fukagawa et al. (1995), who had earlier hypothesised that such GC/AT transitions mark the boundaries between chromosome bands, and that such boundaries are "precisely assignable at the nucleotide level by identifying the early-to-late switch point for replication timing". Considering the range of functions and properties attributable to isochores and chromosome bands, their physical boundaries are most likely composed of a complex of elements including perhaps sequence motifs, GC changes and particular chromatin structures.

This is the first report of the identification of a cluster of active genes (the HLA-DQ/DR locus) that is relatively late replicating even when expressed. Whilst this finding contrasts with the some of the 'rules of replication' devised by Schildkraut and co-workers (Dhar et al., 1988 and 1989; Hatton et al., 1988), it does not seriously threaten the idea that transcription is positively correlated with early replication. It is
reasonable that transcribed sequences replicate early by default because their de-condensed state allows easier access to the replication machinery (Raghuraman et al., 1997). However, our data do question the extent to which transcription influences replication time and suggest that additional factors may exert an influence (at least in determining later replication).

We can claim another first with our two-colour FISH experiments. These confirmed the presence of a replicon in the class III region and provided the first demonstration that replicons organized over >1Mb of interphase chromatin can be simultaneously replicated (Blumenthal, 1974). We expand on a previous model (Cook, 1991) and suggest that – in some B-cells at least – the replicons in the class II and class III regions replicate simultaneously because they share the same stable unit of nuclear structure, the "replication focus/factory" e.g. see (Jackson & Pombo, 1998).

Our experiments employed the human major histocompatibility complex on 6p21.31. These data can be added to the growing list of observations concerning this remarkable genetic locus.
References


Zhou H., and Glimcher L. H., 1995, Human MHC class II gene transcription directed by the carboxyl terminus of CIITA, one of the defective genes in type II MHC combined immune deficiency, *Immunity* 2:545-560.


Appendix A. Using the FISH Assay to Estimate Mean Replication Time.

The mean replication time of a locus expressed as a percentage of the time spent in S-phase is linked to the probability of replication on a random S-phase chromosome ($\mu$).

It is simply $1-\mu$, where 1 is equivalent to the end of S-phase (i.e. there comes at time at the end of S-phase when the entire genome has replicated and the probability of seeing a doublet at any locus must therefore be 1). $1-\mu$ is also the probability that replication has not occurred (i.e. the proportion of singlet chromosomes). The proportion of chromosomes in an asynchronized experiment that have not replicated and the mean replication time expressed as a percentage of the time spent in S-phase are equivalent, see below.

Ideally, we want to observe the locus of interest on every chromosome within a cell population and, once the cells enter S-phase, use a stopwatch to time precisely when each locus becomes a doublet. In this way we would generate a distribution of actual replication times from which the mean could be read (Figure A.1).

![Figure A.1. Theoretical distribution of replication time measurements](image-url)
If, for example, replication always occurred at time \( t \) way through the S-phase then in the impossible stopwatch experiments, all replication times will equal \( t \) and the mean will equal \( t \). If, on the other hand, the replication of a particular locus takes place at either \( t_1 \) with probability \( p \) or at \( t_2 \) with probability \( 1-p \), see Figure A.2. Then, \textit{on average}, the mean replication time will be \( p*t_1 + (1-p)*t_2 \), hatched areas on Figure A.2.

![Figure A.2. Estimating the mean replication time.](image)

Such stopwatch experiments are, of course, impossible. However, it would be possible to perform FISH experiments on synchronized cells, fractionated into different stages of S-phase by retro-synchrony. In this way we could generate a cumulative curve of the proportion of chromosomes on which a particular locus has replicated by a given time through S-phase. The steepest part of this curve would be equivalent to the highest part of the histogram in Figure A.1. We could read off the median replication time when 50% of the loci had been replicated (see Figure A.3, below).
Figure A.4. Estimating the mean replication time from synchronized experiments
Note that the area "over the curve" (in the hatched box in Fig A.4.) is equal to $1^*t$.
Hence, the area over the curve is equal to the mean replication time.

We can also look again at our example of replication taking place at either $t_1$ with probability $p$ or at $t_2$ with probability $1-p$. From the hatched area on Figure A.5, we can see that the mean replication time is $t_1*1 + (t_2-t_1)*(1-p)$, which simplifies to $p*t_1 + (1-p)*t_2$, the same as the mean replication time in the impossible stopwatch experiments. Hence, the area over the curve is equal to the mean replication time.

![Figure A.5. Estimating the mean replication time from the cumulative curve.](image)

We had to resort to asynchronous experiments. In asynchronous experiments, with probability $x$, a randomly chosen chromosome will be between 0 and $x$ of the way through S-phase. Hence, with probability $t$, it will be at the same time between 0 and $t$ (strictly speaking, less than $t$). Hence with probability $t$, it will not have replicated (see Figure A.4).

In our example in which a locus replicates at $t_1$ with probability $p$ or at $t_2$ with probability $1-p$, we want to know how far into S-phase is a doublet at our locus of
interest. But since these are asynchronous experiments, we can not know this. Instead, we have to observe the locus on many S-phase chromosomes and estimate the mean probability of seeing a doublet.

With probability $t_1$ our locus of interest is between 0 and $t_1$, in which case it will not have replicated (see Figure A.5). With probability $(t_2-t_1)$ it will be between $t_1$ and $t_2$, in which case it will have replicated with probability $p$. This probability is calculated from the number of doublets in the asynchronous experiments, see equation in section 3.5. In this case, the locus will not have replicated with probability $1-p$. Thus the probability that the locus on a randomly chosen chromosome will be a singlet is $t_1 + (1-p)*(t_2-t_1)$, which again becomes $p*t_1 + (1-p)*t_2$, the same as the mean replication time in the above experiments.

To reiterate, $p$ is the probability of observing a doublet at a particular locus in a random sample of S-phase chromosomes, $1-p$ is the probability that replication of the locus has not occurred and equates to the mean replication time expressed as a percentage of the time spent in S-phase (i.e. the area above the curves in Figs A.4 & A.5). What this proportion means in terms of actual hours and minutes through S-phase, cannot be determined from our experiments. This would require knowledge of the approximate length of the S-phase, but to estimate this would require synchronization, which is not possible with this B-cell line.
Appendix B. Calculating the Chromosome Odds (Chapter 6)

Figure B.1. Assuming probe G10s is green and the co-probe is red, below are the possible combinations of observable nuclei. With this fluorochrome combination, nuclei A, C, E and G would correspond to the predicted order of replication (see section 6.1.1), whereas B, D, F and H would correspond to the "wrong order of replication".

If a nucleus contains at least one doublet, it should be at G10s (green), as in A. B should be rare with KM2 & M7B.

If the co-probe has an SD pattern, its doublet should be on the same chromosome as the G10s doublet, as in C. D should be rare with KM2 & M7B.

E is similar to A, with G10s replicating before the co-probe on both homologues. Again, F should be rare with KM2 and M7B.

G is a combination of A and C, with G10s doublets on both homologues. H should be rare with KM2 and M7B.

Combined odds

Below is the estimate for the combined odds of G10s replicating before the co-probe on either chromosome in the same nucleus.

\[
\frac{1}{2A + C + D + 4E + 2G + 2B + C + D + 4F + 2H}
\]

where "A" denotes the number of nuclei corresponding to pattern "A" in Fig B.1, etc.
Ipsi-chromosome odds

The ipsi-chromosome odds examine the replication of G10s and the co-probe on the same chromosome homologue. That is:

No. of nuclei in which G10s replicates before the co-probe on same chromosome
No. of nuclei in which G10s replicates after the co-probe on the same chromosome

This corresponds to:

\[
\begin{align*}
1/ & \quad A + D + 2E + G \\
& \quad B + D + 2F + H
\end{align*}
\]

Contra-chromosome odds

The contra-chromosome odds examine the replication of G10s and the co-probe on different chromosomes homologues in the same nucleus. That is:

No. of nuclei in which G10s replicates before the co-probe on the other chromosome
No. of nuclei in which G10s replicates after the co-probe on the other chromosome

This corresponds to:

\[
\begin{align*}
1/ & \quad A + C + 2E + G \\
& \quad B + C + 2F + H
\end{align*}
\]