Telomeres and Related Repetitive DNA in the Mouse Genome

Jacqueline Starling

PhD

University of Edinburgh

1992

Abstract

This project was designed to isolate and characterise interstitial telomere repeat containing loci from the human and mouse genomes and to investigate the nature of the mouse telomere. Cloning of the internal telomere repeat loci proved to be extremely difficult and so alternative methods such as restriction enzyme analysis, hybridisation analysis, inheritance studies, and mapping within recombinant inbred and backcross mouse strains were employed to characterise these regions within the mouse genome. Similar methods were used to characterise mouse telomeres. From these experiments it was shown that, in the mouse, *Trypansoma*-like $(TTAGGG)_n$ telomere repeats are present at the telomeres and at interstitial sites. Within both of these regions, the $(TTAGGG)_n$ repeats are present within distinct genetic loci that are stably inherited through subsequent generations. New variant generation is observed at both types of loci, takes place at a significantly higher frequency at the telomeric compared to interstitial loci and occurs during gametogenesis. It is possible that the higher rate of new variant generation at mouse telomeres compared to internal sites may relate to their position within the mouse genome. Restriction enzyme and hybridisation sequence analysis demonstrated that both classes of loci are composed of telomere-related repeats and that an undefined simple repeat may also be present. Direct sequencing is required before the nature and organisation of simple repetitive DNA within these loci can be determined. Mapping of the interstitial, $(TTAGGG)_n$ telomere repeat containing loci within the BxD RI and *Mus spretus* / C57Bl/6 backcross mice demonstrated their presence within the protermini of chromosomes 9, 13 and X. It remains to be determined whether this distribution is functionally significant. However, these loci provide an end to the genetic map of these chromosomes and hence, will be extremely useful for genome mapping. Characterisation of the telomeric and interstitial $(TTAGGG)_n$ containing loci will require their isolation from the mouse genome, allowing their sequence organisation to be determined and the mechanism of new variant generation at these loci to be elucidated. In addition, such studies may reveal how the terminal $(TTAGGG)_n$ arrays function as telomeres and reveal if a function exists for internalised telomere repeats within the mouse genome.

Declaration

This is to certify that :

(a) the thesis has been composed by the named author.

(b) the research work has been carried out by the named author unless otherwise stated in the text.

Acknowledgements

I would like to thank my supervisors, Dr's N. D. Hastie and R. C. Allshire for their help and advice, Professor H. J. Evans for allowing me to work in the MRC Human Genetics Unit, the Medical Research Council for my PhD studentship and Pizza Hut for additional financial support.

Thanks to David Porteous for many helpful discussions, John Maule, John Gosden, Diane Hanratty and Judy Fletcher for technical help, Sheila Christie for always being able to provide enzymes at extremely short notice and the 'boys' at the animal unit for looking after my mice. In addition, I would like to thank Helen and Sheila in the library for finding many of the references used to write this thesis and Norman, Sandy and Douglas for producing all of the photographs to a higher standard than seemed possible from the material provided.

I greatly appreciate the help provided by Elaine, Amanda, Doug and Peter with proofreading and Annette for coming to the rescue and sticking all 270 of the photographs into this thesis. I would also like to thank Robin for his help and advice over the content of this thesis.

Finally, I would like to thank Ruth, Leslie, Maureen H., Suzie and Donald for their friendship and encouragement during my time in Edinburgh and my parents, brothers and sisters for their belief in my ability to finish my PhD and for providing a haven away from the stresses and strains of the scientific world.

Contents

Figures

Tables

Abbreviations

A	adenine (when representing DNA bases)
ARS	autonomously replicating sequence
BAP	bacterial alkaline phosphatase
BxD RI	C57Bl/6 x DBA/2 recombinant inbred mouse strains
bp	base pairs
C	cytosine (when representing DNA bases)
C-terminal	carboxyl terminal
dATP	deoxyadenosine triphosphate
DMS	dimethyl sulphate
DNA	deoxyribose nucleic acid
ds DNA	double stranded DNA
dUTP	deoxyuracil
EDTA	ethylenediaminetetra-acetic acid
EGTA	ethylene glycol tetra-acetic acid
FITC	fluorescein isothiocyanate
G	guanine (when representing DNA bases)
IPTG	isopropyl-1-thio-β-D-galactosidase
Kbp	kilobase pairs
Kd	kilodalton
MNase	micrococcal nuclease
MPTS	g-methacryloxypropyltrimethoxysilane
Mwt	molecular weight
N7	nitrogen at position 7 of guanine residues
N-terminal	amino-terminal
OD	optical density
3'OH	3' hydroxyl terminus
Oligos.	oligonucleotides
O/N	overnight
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PEG	polyethylene glycol
PFGE	pulsed field gel electrophoresis
PNK	polynucleotide kinase
PRINS	oligonucleotide primed *in situ* DNA synthesis
RAP 1	repressor/activator site binding protein

rDNA	ribosomal RNA
RNA	ribose nucleic acid
RNase	ribonuclease
SDP	strain distribution pattern
SDS	sodium dodecyl sulphate
SSC	salt and sodium citrate
ss DNA	single stranded DNA
SV40	Simian virus 40
T	thymine (when representing DNA bases)
TEMED	N,N,N'N'-tetramethylethylenediamine
u	units of enzyme
vol.	volume
X-gal	5-bromo-4-chloro-3-indole-β-galactopyranoside

Chapter 1

Introduction

1.1 Brief Outline of the Organisation of Genomic DNA within the Nucleus

The genetic material, DNA, comprising eukaryotic genomes, when present as an extended molecule, greatly exceeds the packaging capacity of its cellular compartment. For example, in humans, 1.8m of genomic DNA must be contained within a nucleus with a diameter of ~6μm. Therefore, the DNA molecules of eukaryotes are packaged into a DNA/protein structure, called chromatin, and are thus compressed to a size allowing accomodation within the nucleus. Chromatin is further organised by its division into a species specific number of defined units or chromosomes. Further condensation of chromatin at mitosis and meiosis allows their visualisation using microscopy.

Normal mitotic cell division gives rise to two diploid daughter cells each containing a full complement of parentally derived chromosomes, while meiosis leads to haploid products containing only one copy of each chromosome. The process of division of chromosomes between two cells is facilitated by structures both within the nucleoplasm and each chromosome. One such element, the centromere, has for a long time been associated with the control of chromosomal inheritance. The role of the centromere is to serve as a site of attachment of the chromosomes to the spindle apparatus. The spindle is a collection of microtubules that provide a framework for the orientation and movement of chromosomes during mitosis and meiosis. Attachment of the chromosomes to the spindle at the centromere is required for movement of the chromosomes to opposite poles of the cell. Experiments looking at plasmid inheritance in yeast have shown that the absence of a centromere within these molecules results in their loss from many of the products of cell division (reviewed by Blackburn and Szostak, 1984). A second element involved in chromosome function resides at the extremities of chromosomes and is described as the telomere. This has been shown to possess functions in protection of the chromosome against degradation, DNA replication and, more recently, in chromosome segregation at cell division. For example, abnormalities involving this structure have resulted in increased recombination, chromosome loss and senescence phenotypes (Yu *et al.*, 1990). From these observations it appears that both centromeres and telomeres have essential roles in the maintenance of eukaryotic genomes.

Investigation of the structure and sequence composition of centromeres and telomeres has revealed much information concerning the mechanisms by which they function within the nucleus (reviewed by Blackburn and Szostak, 1984). Analysis of these structures from

different organisms has also provided an insight into the degree of conservation of function. This project was designed to characterise the repetitive DNA found within the telomeres of different mouse strains and provide an insight into how these structures might relate to previously characterised telomeres from other organisms. In addition, the presence of telomere-like, repeated sequences within the human and mouse genomes was also investigated to establish whether they still possess the properties of telomeres, or, if they resemble some form of repetitive DNA when present at an interstitial site.

1.2 Telomeres

1.2.1 General Outline

1.2.1 (a) Initial observations

In 1938, Müller described the inert nature of natural chromosome ends to fusion reactions during crossing over between Drosophlinae chromosomes in meiosis. In 1954, Müller and Herskowitz, by studying the behaviour of broken chromosomes ends after irradiation of *Drosophila* spermatozoa, concluded that these ends, exposing the adhesive faces of interstitial genes, do not automatically take the form and properties of natural chromosome termini. Additionally, natural termini do not join to each other or to broken ends and hence, become interstitial regions by chromosome fusion. Therefore, he concluded that it would be appropriate to distinguish any monopolar chromosome end by the term 'telomere'. Subsequently, telomeres have been defined as the physical ends of eukaryotic chromosomes or, in molecular terms, the simple repeats found at DNA termini and the proteins that bind specifically to these sequences *in vivo* (Zakian, 1989).

1.2.1 (b) Possible cellular functions

From the many investigations into the nature of telomeres, their possible functions within the cell have been assigned to four main areas.

(1) Completion of DNA replication at chromosome ends

The synthesis of DNA is performed by the DNA polymerase enzymes in the 5'-3' direction from a free 3'OH group. The 3' OH group is provided by an RNA primer complementary to the template strand. The two strands of a double-stranded DNA molecule run antiparallel to each other, ie, 5'-3' and 3'-5', and are termed the leading and lagging strands, respectively. In order to solve the problem of synthesising the lagging strand with enzymes that only function in the opposite direction, it is replicated discontinuously from RNA primers in the 5'-3' direction. Upon completion of replication, the primers are removed and the resultant gap

filled in by a DNA polymerase (reviewed by Stryer, 1981). However, as stated by Watson, removal of an RNA primer from the very end of the lagging strand will leave a small region of unreplicated DNA due to the absence of a binding site for the next primer. It is thought that the telomeres provide a means to replicate the ends of chromosomes fully and therefore stop the otherwise unavoidable, progressive loss of genetic information during subsequent cell divisions (Watson, 1972).

(2) Stability of chromosomes

The ability of the telomere to provide a protective 'cap' for the end of the chromosome was first demonstrated by Müller and McClintock. This stabilises the termini against end-to-end fusions and exonucleolytic degradation, both of which are characteristic of broken chromosome ends, and allows normal segregation of the chromosomes during cell division (Müller, 1938, McClintock, 1940 and 1942, Müller and Herskowitz, 1954). An example of broken end reactivity was provided by studies of the pattern of inheritance of a chromosome 9 in maize, whose short arm terminated in a broken end (McClintock, 1940, 1942). Crossing over during meiosis led to end-to-end fusion of the two sister chromatids producing a dicentric chromosome. This was then ruptured at the meiotic anaphase stage, resulting in broken ends. Upon subsequent gametophytic divisions, fusion and breakage was repeated, resulting in, what is termed, the breakage-fusion-bridge cycle. Chromosome fusion was only observed between broken sister chromatids and not with the intact telomeres of the chromosome 9 homologue (McClintock, 1940, 1942). Broken-end instability has also been observed in the yeast, *S.cerevisiae* (McCusker and Haber, 1981, Mann and Davis, 1983, Haber and Thorburn, 1984). In both organisms, the presence of the broken end led to gene and chromosome loss (McClintock, 1940, 1942, McCusker and Haber, 1981, Mann and Davis, 1983, Haber and Thorburn, 1984).

The stability and stable inheritance of some of the products of chromosome breakage presented the possibility that the cell is able to repair these potentially lethal mutations. In maize, yeast, *Drosophila* and man, cases of broken end healing have been reported (McClintock, 1940, McCusker and Haber, 1981, Haber and Thorburn, 1984, Beissmann *et al.*, 1990b, Wilkie *et al.*, 1990). In both *Drosophila* and man, telomeric repeat sequences have been added to a formerly interstitial site resulting in no further loss of DNA upon subsequent rounds of cell division (Beissmann *et al.*, 1990b, Wilkie *et al.*, 1990). Hence, it would appear that the telomere is essential for maintenance of genetic information within the cell.

(3) Involvement in establishing the 3D architecture of the interphase nucleus

(a) Are the chromosomes in the nucleus randomly distributed or is there some form of organisation?

Rabl, in 1885, proposed the telophase or 'Rabl' orientation for interphase chromosomes. It resembles the polarised orientation of the chromosomes as they enter telophase *ie*, centromeres pulled forward by the spindle and telomeres trailing behind (reviewed by Foe and Alberts, 1985, Gasser and Laemmli, 1987). From this it was suggested that interphase chromosomes occupy fixed territories in the nucleus with their centromeres and telomeres at opposite poles. The chromosomes of syncytial blastoderm stage interphase nuclei of *Drosophila* embryos are orientated in the telophase configuration, with centromeres pointing to the exterior and telomeres to the interior of the nucleus (Foe and Alberts, 1985). A similar arrangement has also been observed in the polytene nucleus of *Drosophila* (Agard and Sedat, 1983, Mathog *et al.*, 1984) and interphase nuclei of the protozoan, *Trypanosoma brucei* (Chung *et al.*, 1990).Therefore, there is evidence for chromosomal territories within the nucleus. It can be predicted that the observed polarity would help to order the large amount of DNA within any one cell, and so may also have an indirect organisational role in gene expression (reviewed by Agard and Sedat, 1983, Gasser and Laemmli, 1987, Chung *et al.*, 1990).

(b) Telomeres mediate telomere-telomere and telomere-nuclear envelope interactions

The mechanism of homologous chromosome pairing, which involves the moving together of matching chromosomes or chromosome segments across substantial distances within the nucleus, is not understood. The observation in most species of a 'bouquet' orientation of chromosomes during prophase of the first meiotic division, in which telomeres are clustered together near the nuclear envelope, led to the possibility of their involvement in bringing homologous chromosomes together at meiosis (Comings and Okada, 1972, Maguire, 1984, Sen and Gilbert, 1988). Associations between both homologous and non-homologous chromosomes have also been observed in *Ornithogalum virens* (Ashley and Wagenaar, 1974). In addition, repetitive sequences have been discovered in the subtelomeric regions of nonhomologous human chromosomes which have been shown to crosshybridise and it has been suggested that they may have a role in chromosome pairing (Brown *et al.*, 1990, Wilkie *et al.*, 1991). It is thought that telomere-mediated association of non-homologous chromosomes could initiate the process of chromosome pairing by bringing all of the chromosomes into close proximity. However, it is unlikely that this is the major mechanism of pairing as, for example, in humans, pairing facilitated by the subtelomeric repetitive sequences would result in the association of non-homologous chromosomes.

(4) Effects on gene expression

The presence of a gene at or next to a telomere has been shown to effect its expression. For example, in *Drosophila*, transposition by P element vectors of the white gene to ~20 non-telomeric sites produced the wild-type, red-eyed phenotype (Levis *et al.*, 1985). However, insertion into the end of the right arm of chromosome 3 or left arm of chromosome 2 led to abnormal expression resulting in a varigated eye colour (Gehring *et al.*, 1984, Levis *et al.*, 1985). This phenotype was shown to be due to a position effect rather than mutation of the gene (Levis *et al.*, 1985). Similarly, in yeast, insertion of pol II genes (URA3, ADE2, HIS3 and TRP1) close to the telomeres resulted in epigenetic dependent gene expression (Gottschling *et al.*, 1990). The opposite effect is observed with *Trypanosoma* surface antigen genes which only show predominant expression over other such genes when placed next to the telomere (reviewed in Pays and Steinert, 1988).

In order to determine how telomeres might facilitate their proposed functions of replication, chromosome stability, spatial organisation and involvement in gene expression, an understanding of their composition is required. Hence, what type of sequences constitute the telomere and how are they organised ? The telomeres of diverse organisms all have a function in maintenance of the genome, is such a common function reflected in their sequence and structure ? What structural proteins are associated with telomeres and are they the same, or separate, from the histone proteins of chromatin ? How is the essential process of replication achieved ? All of the afore mentioned aspects of telomeres have been studied in detail in a number of different organisms.

1.2.2 Isolation and Sequence Composition of Telomeres

1.2.2 (a) Methods of cloning

The hypotrichous and holotrichous ciliated protozoa provided the first information concerning the sequence and structure of telomeres (Blackburn and Gall, 1978, Klobutcher *et al.*, 1981, Pluta *et al.*, 1982). These lower eukaryotes are characterised by two separate, differentiated types of nuclei found within each single celled organism. The germinal diploid micronucleus functions both to maintain genetic continuity between generations and to carry out the sexual cycle or conjugation. During conjugation the micronucleus produces the haploid products of meiosis, which undergo crossfertilisation between pairs of conjugating cells resulting in the formation of the zygotic nucleus. This then divides mitotically and the somatic- or macronucleus develops from one of the resulting daughter nuclei. During its development, the macronucleus becomes DNA-rich and transcriptionally active, in contrast to the transcriptionally inactive micronucleus, and drastic changes take place in its genomic organisation (reviewed by Blackburn and Karrer, 1986). The chromosomes are converted to

defined subchromosomal DNA molecules ranging in size from ~600kb in *Tetrahymena thermophilia* (Preer and Preer, 1979), 300kb in *Glaucoma chattoni* (Katzen *et al.*, 1981) to 2-3kb in *Oxytricha nova* (reviewed by Lawn *et al.*, 1978) (reviewed by Blackburn *et al.*, 1982). These molecules are all linear and so need telomeres at their termini for maintenance and replication. One well characterised example of this is provided by the genes encoding the ribosomal RNA (rDNA) present in many lower eukaryotes. For example, in *Tetrahymena thermophilia* the rDNA genes are amplified from one integrated copy in the micronucleus, (Yao and Gall, 1977), to 9000 copies, or 200 copies per haploid genome, in the macronucleus (Yao *et al.*, 1978). Two copies of the gene are present on each subchromosomal molecule in a palindrome, *ie* head-to-head. Hence, the production of rDNA genes in the macronucleus results in an extra 4500 DNA fragments and so an enrichment for telomeres as compared to the situation in a human cell where there are a total of 46 chromosomes and so 92 termini.

Isolation of *Tetrahymena* telomeres was achieved by first purifying the rDNA molecules intact from the macronucleus (Gall, 1974) and then directly sequencing the termini. The sequencing reaction took advantage of single strand gaps in the termini. Using *E.coli* DNA polymerase I and different combinations of dNTP's, it was demonstrated that a repetitive sequence of the form CCCCAA on one strand and GGGGTT on the complementary strand was present at the *Tetrahymena* termini (Blackburn and Gall, 1978). Further characterisation revealed that this sequence occurred as a tandemly repeated hexanucleotide, written $5'(CCCCAA)_n3'$, the number, n, of these repeats varying between 20-70 on different rDNA molecules. Experiments with T4 DNA polymerase suggested that there were no free cohesive ends on the rDNA. This enzyme will carry out strand elongation from a 3' terminus on gapped or partially single stranded DNA templates (reviewed by Blackburn and Gall, 1978). This implies that there is a special structure at the termini of DNA molecules, perhaps a bound protein or self-complementary hairpin loop, which prevents ligation or fusion with other DNA molecules and so essentially is unclonable by conventional techniques.

A similar method for telomere characterisation was adopted for members of the hypotrichous ciliates, slime molds, flagellated protozoa and sporozoa (Johnson, 1980, Emery and Weiner, 1981, Klobutcher *et al.*, 1981, Pluta *et al.*, 1982, Bergold *et al.*, 1983, Blackburn and Challoner, 1984, Baroin *et al.*, 1987, Forney *et al.*, 1987). In each case, advantage was taken of the findings of Blackburn and Gall of single standed gaps in the repeats and the non-cohesive nature of the termini. In general, telomere sequences were cloned from the rDNA molecules or in *Trypanosoma brucei* from the 100 minichromosomes (Borst *et al.*, 1983), by first treating these ends with Bal 31 or S1 nuclease to remove the terminal unclonable region. Both of these enzymes progressively remove double stranded DNA from the free ends of chromosomes and, by timing the reaction, the number of bases

removed could be calculated. The blunt end was ligated to either a plasmid or bacteriophage vector and the rDNA molecule digested with an enzyme known to leave only the most terminal fragment. The whole molecule was then either circularised or ligated to the other phage arm and propagated in a suitable host cell. The single strand gaps, absent from the hypotrichous ciliate termini, allowed the synthesis, by *E.coli* DNA polymerase I, of probes complementary to the terminal repeats which could then be used to identify positive transformants. The terminal nature of the clones was established using Bal 31. Total genomic DNA from a comparable organism was treated with Bal 31 through a timecourse, digested with a suitable enzyme, size fractionated in agarose and hybridised with the clone. Sensitivity to Bal 31 and, therefore, telomeric origin was characterised by loss of intensity of hybridisation signal with progressive Bal 31 treatment.

Isolation of telomeres from some species of *Trypanosoma spp.* (*T.cruzi*, *T.vivax*), flowering plants, yeast and mammals has proved difficult due to the low copy number of telomeres compared to the total DNA content in the genome. Unlike *Trypanosma brucei*, with 100 minichromosomes and so 200 telomeres per haploid genome of 40 mega base pairs (Mbp) (Borst *et al.*, 1983), the yeast, *Saccharomyces cerevisiae* has 16 chromosomes and so 32 telomeres per ~16Mbp (Carle and Olsen, 1985), the flowering plant, *Aradopsis thaliana*, has 10 telomeres per 70Mbp (Leutweiler *et al.*, 1984), and in humans there are 46 telomeres per 6000 Mbp (reviewed by Alberts *et al.*, 1983). Hence, in each of these species there is a small number of telomeres with a large total, genomic DNA content, making direct cloning from the whole genome extremely difficult. Szostak and Blackburn solved the problem of limited representation and non-ligatability of telomeres by mimicking the rDNA molecules found in lower eukaryotes. They constructed a linear plasmid with the terminal DNA fragments of *Tetrahymena* rDNA molecules that could be propagated in yeast. The rDNA telomeres retained all their structural and sequence characteristics. Removal of one terminus and ligation to yeast genomic fragments allowed selection for constructs containing a yeast telomere as only these could be stably propagated upon transformation into a yeast host (Szostak and Blackburn, 1982). Telomeres were cloned from *Aradopsis* and humans by enriching for restriction fragments containing telomeres, followed by ligation either to a plasmid vector and transformation into an *E.coli* strain or to a yeast artificial chromosome and propagation by complementation in yeast (Moyzis *et al.*, 1988, Richards and Ausubel, 1988, Brown, 1989, Cheng *et al.*, 1989, Cross *et al.*, 1989, Reithman *et al.*, 1989, de Lange *et al.*, 1990).

1.2.2 (b) Sequence characteristics

The cloning and characterisation of DNA termini from such a wide range of organisms has led to the discovery of several general features. They are all made up of tandem repeats of a

sequence that is G-rich on one strand and C-rich on the complementary strand, Table 1.1. The general consensus of the telomeric sequences can be written as 5'[$C_{1-8}(A/T_{1-4})$]3' (Blackburn, 1984). The number of repeat units at different telomeres within an organism was found to vary in all cases except for the hypotrichous ciliates, where a fixed unit number was always present. They also lacked single strand discontinuities which were a common feature in other species. The overall size of the telomere varied quite considerably between different organisms, Table 1.1. The G-rich strand was always orientated in the 5'-3' direction towards the end of the chromosome; the C-rich strand orientated 5'-3' towards the interior of the molecule. The G-rich strand was shown to have extended to form a single strand overhang in the hypotrichous and holotrichous ciliates and the slime mold, *Didymium spp.*, and this may also prove to be a conserved feature (Klobutcher *et al.*, 1981, Henderson and Blackburn, 1989). The repeats were not recognised by restriction enzymes suggesting the presence of simple repeats with minimal sequence variation and the ends were blocked from normal cloning methods. The most striking fact to come from these observations was that the structure and sequence of DNA termini was very similar between evolutionarily diverse species. For example, the $(TTAGGG)_n$ repeats identified at the ends of minichromosomes in *Trypanosoma brucei* have been shown to be present at the telomeres of humans, two species of slime molds and sporozoa, Table 1.1.

1.2.2 (c) Telomere-associated sequences

In addition to the simple repeats which form the very ends of chromosomes, a separate type of repetitive DNA has been found lying immediately adjacent to the telomere in a number of quite separate organisms (reviewed by Zakian, 1989). These regions are normally composed of middle-repetitive elements and are termed telomere-associated sequences (Zakian, 1989). The proportion of these sequences compared to total genomic DNA has been found to vary greatly between different organisms. In the *Secale* species there is a 20% interspecies variation in total chromosomal DNA content which is entirely attributable to telomere-associated heterochromatic regions (Bedbrook *et al.*, 1980). While in *Chironomus*, 1.2%, or ~300kb per telomere, of the genome is composed of these sequences (Saiga and Edström, 1985). In yeast , however, a much smaller amount of telomere-associated repeat is present at most of the chromosome termini (Chan and Tye, 1983a and b). *In situ* hybridisation has been used to demonstrate the telomeric location of telomere-associated sequences and in *Drosophila* and *Chironomus* these have been shown to extend into the ectopic fibres connecting chromosomes (Rubin, 1977, Saiga and Edström, 1985). In both *Drosophila* and man, telomere-associated sequences have been found interstitially, at the pericentric heterochromatin and position 2q11-2q14 of chromosome 2, respectively (Young *et al.*,1983, Traverse and Pardue, 1989, Brown *et al.*, 1990 and Cross *et al.*, 1990).

Table 1.1 Sequence Composition of the Telomeres of a Range of Different Organisms

Organism	Repeat Sequence, 5'-3'	Size of repeat sequence at telomere.	Reference
Holotrichous ciliates :- *Tetrahymena thermophilia*	GGGGTT	0.12 - 0.42kb	Blackburn and Gall, 1978.
Glaucoma chattoni	GGGGTT	> 0.22kb	Katzen *et al.*, 1981.
Paramecium primaurelia	GGGGTT and GGGTTT	0.2kb	Barion *et al.*, 1987.
Hypotrichous ciliates :- *Oxytricha nova* *Oxytricha spp.* *Oxytricha fallax* *Stylonicha pustulata*	GGGGTTTT	20bp of (C_4A_4) 36bp of (G_4T_4) :- 16bp = 3' ss tail.	Klobutcher *et al.*, 1981, Dawson and Herrick, 1982, Pluta *et al.*, 1982.
Euplotes aediculatus *Euplotes crassus*	GGGGTTTT	28bp of (C_4A_4) 42bp of (G_4T_4) :- 14bp = 3' ss tail.	Klobutcher *et al.*, 1981, Roth and Prescott, 1985.
Slime molds :- *Physarum polycephalum*	GGGATT	~600bp	Johnson, 1980, Bergold *et al.*, 1983, Forney *et al.*, 1987.
Didyium iridis	GGGATT	96->420bp	Forney *et al.*, 1987
Dictyostelium discoideum	$(G_nA)_m$:- n = 1-8	m = 18-34 units of variable repeat size.	Emery and Weiner, 1981.
Yeast :- *Saccharomyces cerevisiae*	$(G_{1-3}T)_m$	m = 18-54 units of variable repeat size.	Shampay *et al.*, 1984, Walmsley *et al.*, 1984.
Saccharomyces pombe	$(G_{1-6}C_{0-1}A_{0-1}CAT_{1-2})_m$	200-400bp.	Sugawara and Szostak, 1986.

Flagellated protozoa :- *Trypanosoma brucei, cruzi and vivax.*	GGGATT	>1kb.	Blackburn and Challoner, 1984, Van der Ploeg *et al.*, 1984.
Sporozoa :- *Plasmodium berghei*	GGGATT, GGGGATT, GGGGTT, GGGAGT, GGGAT.	?	Ponzi *et al.*, 1985.
Flowering plant :- *Aradopsis thaliana*	G(G/A)CATTT	~2.5kb	Richards and Ausubel, 1988.
Mammals :- Human	GGGATT, GGGGTT, GGGAGT, *etc*	~5-20kb	Allshire *et al.*, 1988 and 1989, Moyzis *et al.*, 1988, Brown, 1989, Cheng *et al.*, 1989, Cross *et al.*, 1989, Reithman *et al.*, 1989, Brown *et al.*, 1990, de Lange *et al.*, 1990.

In many cases the telomere-associated sequences have been shown to be internally repetitious (Chan and Tye, 1983a, Saiga and Edström, 1985, Pace *et al.*, 1987, Brown *et al.*, 1990, Dore *et al.*, 1990). For example, in *Plasmodium berghi*, the telomere-associated sequences are composed of 2.3kb units, containing 160bp of telomere derived sequence and perfect tandem copies of a 27bp repeat (Pace *et al.*, 1987, Dore *et al.*, 1990). Similarly, in *Chironomus*, the associated DNA is composed of tandem 340bp repeats which are divided into two alternating subrepeats of 60 and 29bp (Saiga and Edström, 1985). The most extensively characterised telomere-associated sequences belong to the yeast *Saccharomyces cerevisiae*. This repetitive DNA was originally isolated as autonomously replicating sequences (ARS) (Chan and Tye, 1980), which have since been divided into two classes, X and Y' (Chan and Tye, 1983a and b). Each of the class X ARS are embedded within a repetitive region, designated X, of variable length and sequence, ~3 to ~3.75kb. The class Y' ARS are embedded within a highly conserved sequence of 5.2kb in length. Both of the X and Y' sequences are located directly adjacent to an element termed 131 of ~1 to 1.5kb (Chan and Tye, 1983a). The arrangement of these sequences at yeast telomeres takes three forms :-

$$T\text{-}Y'\text{-}(131\text{-}Y')_n\text{-}131\text{-}X\text{-}$$

$$T\text{-}(Y'\text{-}131)_{n+1}\text{-}X\text{-}$$

$$T\text{-}Y'_{n+1}\text{-}X\text{-}$$

Where T = telomere and n ranges from 0 to 3 (Chan and Tye, 1983b). Also, located between the X and Y' elements are variable length tracts of $(C_{1\text{-}3}A)_n$ repeats (Walmsley *et al.*, 1984), a feature that has also been observed in *Plasmodium berghi* (Pace *et al.*, 1987, Dore *et al.*, 1990).

Telomere-associated sequences have been shown to be polymorphic in yeast (Chan and Tye, 1983b, Horowitz *et al.*, 1984), *Plasmodium* (Corcoran *et al.*, 1988, Pace *et al.*, 1990), *Drosophila* (Young *et al.*, 1983), *Chironomus* (Saiga and Edström, 1985), *Secale* (Bedbrook *et al.*, 1980) and humans (Simmler *et al.*, 1985 and 1987, Cooke and Smith, 1986, Cheng *et al.*, 1989, de Lange *et al.*, 1990, Brown *et al.*, 1990, Cross *et al.*, 1990, Rouyer *et al.*, 1990, Wilkie *et al.*, 1991). The nature of the polymorphism varies considerably between different organisms. In the yeast genus, *Saccharomyces*, all species contain repetitive sequences homologous to the element Y'-131 but the pattern of bands obtained after restriction digestion and hybridisation vary both within and between the different species (Chan and Tye, 1983b, Horowitz *et al.*, 1984). Variation in size of restriction enzyme fragments has also been observed in different *Drosophila* stocks (Young *et al.*, 1983). Highly variable, telomere-associated sequences have also been isolated from the pseudoautosomal region of the human sex chromosomes (Simmler *et al.*, 1985 and 1987, Cooke and Smith, 1986). One such repetitive element demonstrated >50 alleles within the

population (Cooke and Smith, 1986). Other human telomere associated sequences have been isolated and shown to hybridise to different subsets of chromosomes (Brown *et al.*, 1990, Cross *et al.*, 1990). Their polymorphic nature being reflected in the variation in signal intensity and presence or absence from the relevent chromosomes of different individuals.

Chromosomal size polymorphisms, generated by the presence or absence of telomere-associated sequences, have been observed in *Plasmodium*, *Secale* and humans (Bedbrook *et al.*, 1980, Corcoran *et al.*, 1988, Pace *et al.*, 1990, Wilkie *et al.*, 1991). In *Plasmodium berghi*, chromosome 4 has been shown to vary in size by ~50kb after prolonged *in vivo* multiplication. The extra DNA sequences being composed of tandem copies of the 2.3kb telomere-associated repeat (Pace *et al.*, 1990). In humans, analysis of the short arm of chromosome 16 from 47 different individuals revealed a major size polymorphism upon hybridisation with telomere-associated sequences isolated by Brown *et al.* (1990) (Wilkie *et al.*, 1991). Three alleles were observed when the hybridisation was compared against a known locus, α globin, which was positioned 170kb, 350kb and 430kb from the telomere (Wilkie *et al.*, 1991).

Many possible functions for telomere-associated sequences have been described. In yeast, autonomously replicating sequences are located within the X and Y' elements, which if present at a high density at the telomeres may serve to ensure the initiation of DNA replication (Chan and Tye, 1983a). Their absence from the smallest yeast chromosomes could suggest a role in the stable maintenance of large yeast chromosomes, all of which carry both X and Y' elements (Zakian and Blanton, 1988). However, further analysis is required before a definite function can be assigned to these elements. A second possible role relates to the presence of telomere repeats within the telomere-associated sequences in *Plasmodium* and *Saccharomyces*. Such tracts could be used in the formation of new telomeres if the chromosomes had become broken or degraded. Indeed it is thought that these telomere-associated sequences may have a general protective role, preventing the loss of essential genes from the broken ends of chromosomes (Zakian, 1989). Thirdly, the variability of these sequences suggests a role in recombination mechanisms. For example, in humans, telomere-associated sequences on the sex chromosomes are highly polymorphic and present within a region of obligate recombination between the X and Y chromosomes (reviewed by Pritchard and Goodfellow, 1985). In the *Plasmodium* species' recombination events resulting in transfer of telomere-associated sequences between chromosomes has been implicated in the process of antigenic diversity of the parasite (Corcoran *et al.*, 1988). Finally, a role in the association of chromosomes at mitosis and meiosis has been proposed for telomere-associated repeats. However, in both *Drosophila* and *Chironomus,* such repeats have been visualised within the ectopic fibres connecting chromosomes (Rubin, 1977, Saiga and Edström, 1985), and in humans, sub-families of telomere-associated

repeats have been observed at the termini of non-homologous chromosomes (Brown *et al.*, 1990). The association of non-homologues at meiosis or mitosis could result in the abnormal segregation of chromosomes into the daughter cells. Hence, it seems more plausible that such association events could be involved in the spatial organisation of chromosomes within the nucleus. Further analysis of the structure and function of these sequences is required before a definite function can be determined.

1.2.3 Protein and DNA Conformation at Telomeres

1.2.3 (a) Introduction

Isolation and elucidation of the sequence composition of the telomeres of many different organisms has helped in determining the possible structures that might exist at DNA termini. The presence of G-rich repeats and unclonable termini has led to the idea that specialised structures formed by DNA-DNA interactions might exist at telomeres. Analysis of the chromatin structure at telomeres has revealed the presence of specific, nonhistone proteins. From such findings, ideas have been formulated to explain some of the possible functions of telomeres.

1.2.3 (b) Telomere-associated proteins

The genomic DNA found within the nucleus of the cell is organised into chromatin fibres which, in turn, form the chromosomes. Chromatin has the same design in all eukaryotes and is composed of subunits containing ~200bp of DNA organised by an octamer of small basic proteins, or histones, into a bead-like fibre (reviewed by Kornberg, 1977, Lewin, 1985). There are five classes of histone protein, H1 (not yet identified in yeast, cited by Kornberg, 1977), H2A, H2B, H3 and H4. Two copies of H2A, H2B, H3 and H4 interact directly with the DNA to form the DNA-protein subunit described as the nucleosome. In this structure the proteins form the interior with the DNA coiled around the outside. The fifth histone protein, H1, interacts with each nucleosome subunit and, together with a strand of free duplex DNA, forms the bead-like structure of chromatin (reviewed by Kornberg, 1977, Lewin, 1985). It has been demonstrated that individual nucleosomes can be obtained by digestion of chromatin with the enzyme micrococcal nuclease, MNase. This enzyme cleaves the fibre at the free duplex DNA connecting strands. Nucleosomes prepared by MNase digestion and released from nuclei by lysis in a low ionic strength buffer are insoluble when the salt concentration is increased to 0.1 - 0.2M if still associated with histone H1; H1 free nucleosomes are soluble under these conditions (Olins *et al.*, 1976).

Using the known characteristics of chromatin, several research groups have examined the DNA-protein structure at the termini of rDNA molecules in *Tetrahymena*, *Oxytricha* and

Dictyostelium (Blackburn and Chiou, 1981, Edwards and Firtel, 1984, Gottschling and Cech, 1984, Budarf and Blackburn, 1986). These investigations have led to the discovery of non-nucleosomal interactions with the telomeric DNA. For both *Tetrahymena* and *Oxytricha*, MNase digestion has revealed the majority of the DNA digested into multiples of ~200bp, consistent with each being bound to a nucleosome unit, while the telomeres are present within structures corresponding to the number of terminal repeats. For example, the *Tetrahymena* rDNA telomeric repeats vary in size from 2-800bps, which is reflected in the size of the DNA fragment protected from MNase digestion; in *Oxytricha*, where a fixed number of repeats are present at all rDNA termini, a ~100bp fragment is protected, Figure 1.1 (Blackburn and Chiou, 1981, Gottschling and Cech, 1984). The proteins forming these complexes are not histone derived as demonstrated by their lack of solubility in salt concentrations ranging from 0.075-2M regardless of the extent of MNase digestion. In addition, it has been shown that specific phased nucleosomes extend from the DNA-protein complex at the telomere into the interior of the rDNA molecules, Figure 1.1 (Gottschling and Cech, 1984, Budarf and Blackburn, 1986). From the above observations it would appear that telomeres are packaged by a protein distinct from the histones which could reflect an involvement in the specialised roles proposed for DNA termini.

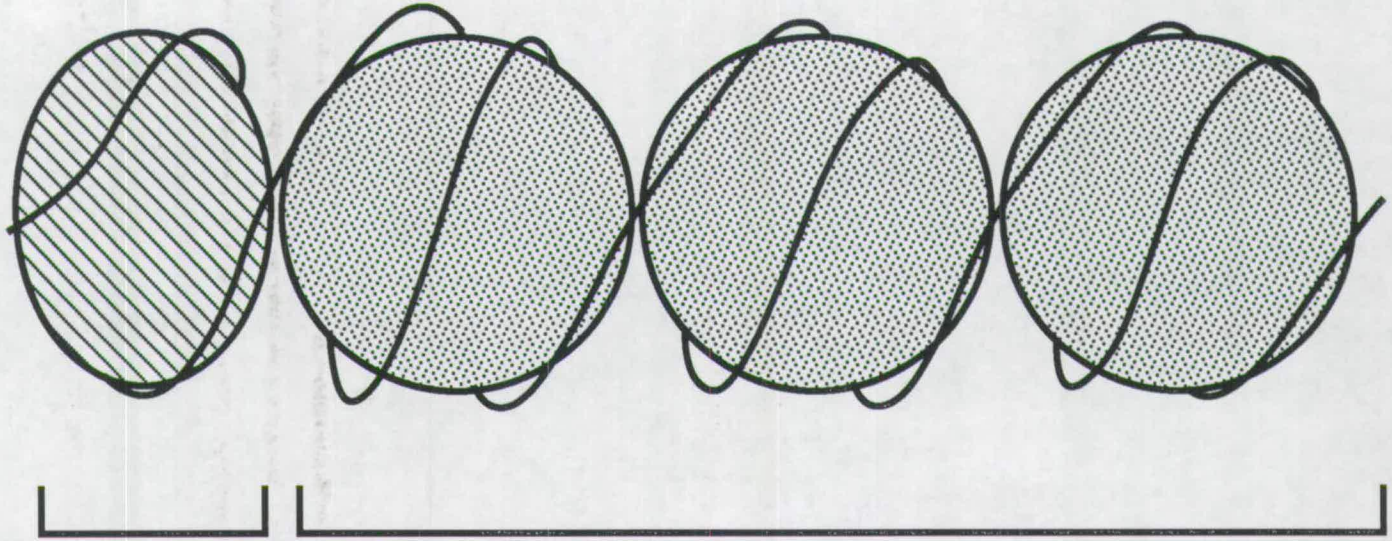
Further characterisation of the protein-DNA complexes at chromosomal termini has led to the isolation of candidate telomere binding proteins in *Physarum*, *Oxytricha* and the yeast, *Saccharomyces cerevisiae* (Cheung *et al.*, 1981, Lipps, *et al.*, 1982, Berman *et al.*, 1986 Gottschling and Zakian, 1986). In *Physarum*, proteins of 5,000 and 13,000Mwt have been found associated with the rDNA termini (Cheung *et al.*, 1981). Similarly, in *Oxytricha*, two immunologically distinct proteins of 55 and 43Kd have been identified binding to the telomeric C_4A_4 repeats (Lipps, *et al.*, 1982, Gottschling and Zakian, 1986, Price and Cech, 1987). An activity binding to poly ($C_{1-3}A$) repeats present at yeast telomeres has also been isolated and shown to bear many similarities to RAP1, Repressor/Activator site binding protein (Berman *et al.*, 1986, Shore and Nasmyth, 1987, Longtine *et al.*, 1989).

The *Oxytricha* and *Saccharomyces* telomere binding proteins have been extensively studied. In both organisms, binding activity is lost upon treatment with proteinase K, 0.1% SDS, phenol or high temperatures suggesting that the interactions are non-covalent in nature (Berman *et al.*, 1986, Gottschling and Zakian, 1986). In addition, the *Oxytricha* telomere-protein complex is resistant to high salt concentrations, chloroform and RNase, suggesting it is tenaciously bound by electrostatic forces such as hydrogen bonds and hydrophobic interactions (Gottschling and Zakian, 1986).

In both *Oxytricha* and *Saccharomyces* the telomere binding proteins have been shown to bind in a sequence specific manner (Gottschling and Zakian, 1986, Price and Cech, 1987, Raghuraman *et al.*, 1989, Raghuraman and Cech, 1989). In *Saccharomyces*, specific

Figure 1.1 Model for the Structure of Telomeric DNA in Oxytricha Macronuclei

The terminal 100bp at each end of the DNA molecules of *Oxytricha* macronuclei is thought to be associated in a telomere/protein complex, as shown in the diagram. Interior to the complex are phased nucleosomes (adapted from Gottschling and Cech, 1984).


Telomere / protein complex
Phased nucleosomes

sequence motifs within the poly ($C_{1-3}A$) repeats are responsible for protein binding (Longtine *et al.*, 1989). The 55Kd protein of *Oxytricha* binds specifically to the telomeric repeats, while the 43Kd protein appears to interact with the 55Kd protein to form the protein complex (Gottschling and Zakian, 1986). The substrate for binding must contain the 3' single strand (G_4T_4) tail and a duplex region that does not necessarily need to contain G_4T_4 . C_4T_4 repeats (Gottschling and Zakian, 1986). Methylation protection experiments have shown that specific G residues within the 3' tail are involved in binding the 55Kd protein and substitution of the bases with either A's or T's prevents formation of the DNA-protein complex (Price and Cech, 1987, Raghuraman *et al.*, 1989).

The *Oxytricha* telomere-protein complex has been studied extensively resulting in the identification of two quite distinct complexes (Raghuraman *et al.*, 1989, Raghuraman and Cech, 1989). These complexes, designated 1 and 3, display different mobilities upon glycerol gradient centrifugation, varying dissociation rates, in that complex 3 exchanged DNA at a more rapid rate than complex 1, and in the set of residues with which they interact in the 3' single strand tail (Raghuraman and Cech, 1989). In addition, complex 1 was found to exist in two forms, as a monomer of ~ 98Kd Mwt and a polymer (Raghuraman and Cech, 1989). Isolation of the gene encoding the 43Kd protein has revealed a N-terminal 1/3 containing the polypeptides most hydrophobic region, while the C-terminal 2/3's resembles the histone H1 (Hicke *et al.*, 1990). It is possible that the hydrophobic region provides a surface for protein-protein interactions with the 55Kd subunit (Gottschling and Zakian, 1986, Hicke *et al.*, 1990). The resemblance of the C-terminal 2/3's to histone 1 suggests a role in association between individual telomere complexes; it may be responsible for the formation of the polymer of complex 1 (Raghuraman *et al.*, 1989, Raghuraman and Cech, 1989).

The characterisation of the telomere binding proteins has implicated them in a number of possible telomere functions within the cell. Firstly, telomeres bound to these proteins are resistant to degradation by exonucleases suggesting a role in the protection of DNA molecules within the cell (Gottschling and Zakian, 1986). Secondly, two distinct, telomere-protein complexes were observed in *Oxytricha*. Complex 3 exchanged telomere-DNA complexes at a more rapid rate than complex 1 and was able to form complexes with sequences that had T residues at the 3' end of the single strand tail (Raghuraman *et al.*, 1989, Raghuraman and Cech, 1989).This suggests a role during telomere elongation (to be discussed later), when partial repeats have been added, this complex might form and protect the nascent, T-ending strands from nuclease degradation. In addition, the rapid exchange rate would ensure binding at such termini but would not prevent the telomerase from completing telomere replication (Raghuraman and Cech, 1989). Thirdly, telomere binding proteins may play a role in the telomere associations observed within many different cell types. For example, the polymer of complex 1 from *Oxytricha* could be involved in the

generation of end-to-end associations of DNA molecules in the macronuclei (Lipps, *et al.*, 1982). Finally, in yeast, a telomere binding activity has been identified as the RAP1 protein. This protein is thought to have many functions within the cell as in its absence senecence results (Conrad *et al.*, 1990). Experiments involving over-expression of the RAP1 gene within yeast cells has resulted in elevated levels of both chromosome loss and recombination suggesting a role in telomere maintenance and chromosome segregation. In addition, when lower levels of the RAP1 protein are present, telomeres become shorter, suggesting a role for the protein in protection against nuclease degradation (Conrad *et al.*, 1990).

1.2.3 (c) DNA structure at telomeres

The possibility of a DNA/DNA interaction at telomeres was first reported in the ciliate, *Stylonicha* (Lipps, 1980). It was shown that concentrated solutions of purified macronuclear DNA aggregated when incubated at room temperature in the presence of sodium ions and was demonstrated by a reduction in the mobility of the aggregated form on agarose gels (Lipps, 1980). From the results of aggregation experiments and electron microscopy, it was proposed that two macronuclear DNA strands fused at their telomeres to form a four stranded structure. Further analysis of a related ciliate, *Oxytricha nova*, revealed the same aggregation of the macronuclear DNA molecules in the presence of sodium ions (Oka and Thomas, 1987). Characterisation of the interactions involved in the aggregation reaction revealed that they occurred in the absence of protein and so involved direct DNA/DNA coherence. Aggregation required the presence of only sodium ions and was stabilised upon the addition of potassium ions. In addition, a 3' single strand tail of $(G_4T_4)_2$ from both interacting termini was required for the aggregates to form (Oka and Thomas, 1987). From the results two models for the cohesion reaction were proposed. Firstly, the macronuclear DNA molecules were antiparallel forming a linear molecule, Figure 1.2a, and secondly, the two participating molecules were parallel to each other giving rise to a 'kink' in the resultant aggregate at the point of interaction, Figure 1.2b (Lipps, 1980, Oka and Thomas, 1987).

Subsequent studies of the DNA/DNA structure at telomeres have involved the use of oligonucleotides (oligos.) representing the telomeric repeats of different eukaryotes. Single stranded oligos. corresponding to the telomeres of *Tetrahymena*, *Oxytricha*, *Trypanosoma*, *Dictyostelium* and *Saccharomyces* have been shown to self-associate by forming non-Watson/Crick, G-G base paired, intramolecular structures. The structures formed below 40^{o}C, at moderate ionic strength and neutral pH and behaved like hairpin duplexes in non-denaturing, polyacrylamide gels. The guanine bases are present in the syn as opposed to the anti-conformation normally present in double helical DNA (Henderson *et al.*, 1987).

Detailed examination of the *Oxytricha* and *Tetrahymena* self-associated structures revealed that certain monovalent cations influenced their mobility upon nondenaturing,

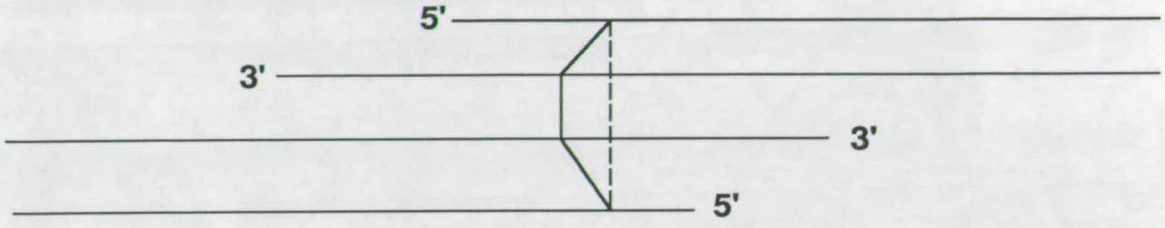
Figure 1.2 Schematic Models of the Aggregated Telomeres of Two Macronuclear DNA Strands in Oxytricha nova

(a) This model shows the antiparallel association of two double stranded telomeres to form a quadruplex structure.

(b) This model shows the parallel association of two double stranded telomeres to form a quadruplex structure which represents the 'kink' observed in the macronucleus.

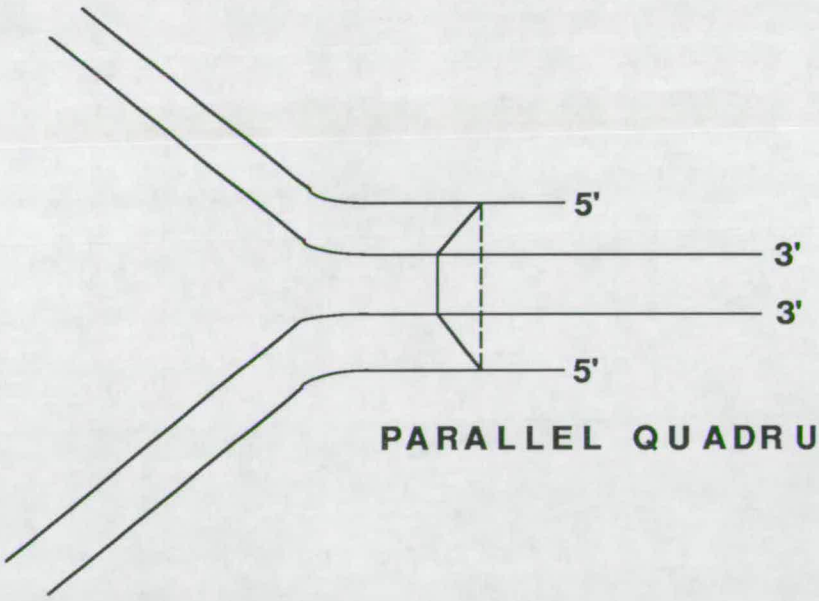
Both diagrams are highly schematic and do not show any detailed structural information (adapted from Oka and Thomas, 1987).

a)



ANTI PARALLEL QUADRUPLEX

b)



PARALLEL QUADRUPLEX

polyacrylamide gels (Williamson *et al.*, 1989). The presence of sodium, potassium or caesium ions converted the *Oxytricha* single stranded oligos., Oxy4 $(dT_4G_4)_4$ and Oxy2 $(dT_4G_4)_2$, to structures with twice the expected electrophoretic mobility. Lithium ions were unable to completely convert Oxy4 to the high mobility form, while all other cations used gave the high mobility form of Oxy2. Two versions of the *Tetrahymena* telomere were also investigated, Tet4 $(dT_2G_4)_4$ and Tet2 $(T_2G_4)_2$. Tet4 had previously been shown to occur as a high mobility, (that of a 12mer), concentration independent form upon electrophoresis. The presence of sodium, potassium or caesium ions was shown to increase the mobility to that expected for a 6mer, while lithium ions only slightly enhanced mobility. For Tet2 all four cations used enhanced mobility to the same extent without any specific monovalent cation effect; the mobility was equal to that of a 6mer (Williamson *et al.*, 1989). In an additional experiment with Tet4 and Tet2, the guanine base analogue, inosine, was substituted for guanine residues. This had no effect upon the observed mobilities of Tet2 in the presence of different monovalent cations, but appeared to interfere with structure formation of Tet4.

The accessibility of the nitrogen at position 7 of the guanine residues within the *Oxytricha* and *Tetrahymena* telomeric oligos. was probed with dimethyl sulphate, DMS. DMS specifically methylates, and piperidine subsequently cleaves, the guanine base at N7. (Maxam and Gilbert, 1980). In Oxy4 and Tet4, piperidine cleavage was observed at all of the guanine residues in the presence of lithium ions or no added salt, in sodium ions all of the G's were protected and with caesium and potassium the two outer G's of the block of 4 were cleaved to a greater extent than those occupying the centre of the repeat. The presence of methylation at the N7 of the guanine residues was also found to interfere with the formation of Tet4 and Oxy4 complexes. Finally, UV-induced crosslinks within Tet4 and Oxy4 reflected an intramolecular, folded structure.

From the results of the above experiments a model was proposed for the molecular structure of DNA at telomeres. In both Oxy4 and Tet4, a double hairpin loop is thought to form, stabilised by 4 G-G base pairs. The G residues would be derived from the ascending and decending strands of the two hairpin loops forming the corners of a square, Figure 1.3. The cavity within the middle of the square would be filled by a monovalent cation which in turn would stabilise the structure. The size of the cavity and size of the cation may explain the monovalent cation effect observed with Tet4 and Oxy4. Lithium would be too small to fill the space and so unable to stabilise the structure, as shown by its lack of effect upon mobility and methylation protection and interference. A sodium ion appears to be able to fill the gap exactly, while potassium and caesium ions are slightly too large, resulting in displacement of G bases and loss of methylation protection. Although *Oxytricha* is known to have only two single stranded telomeric repeats at its 3' end and the structure described would require 4 repeat units, it is thought possible that displacement of two (C_4A_4) repeats might take place.

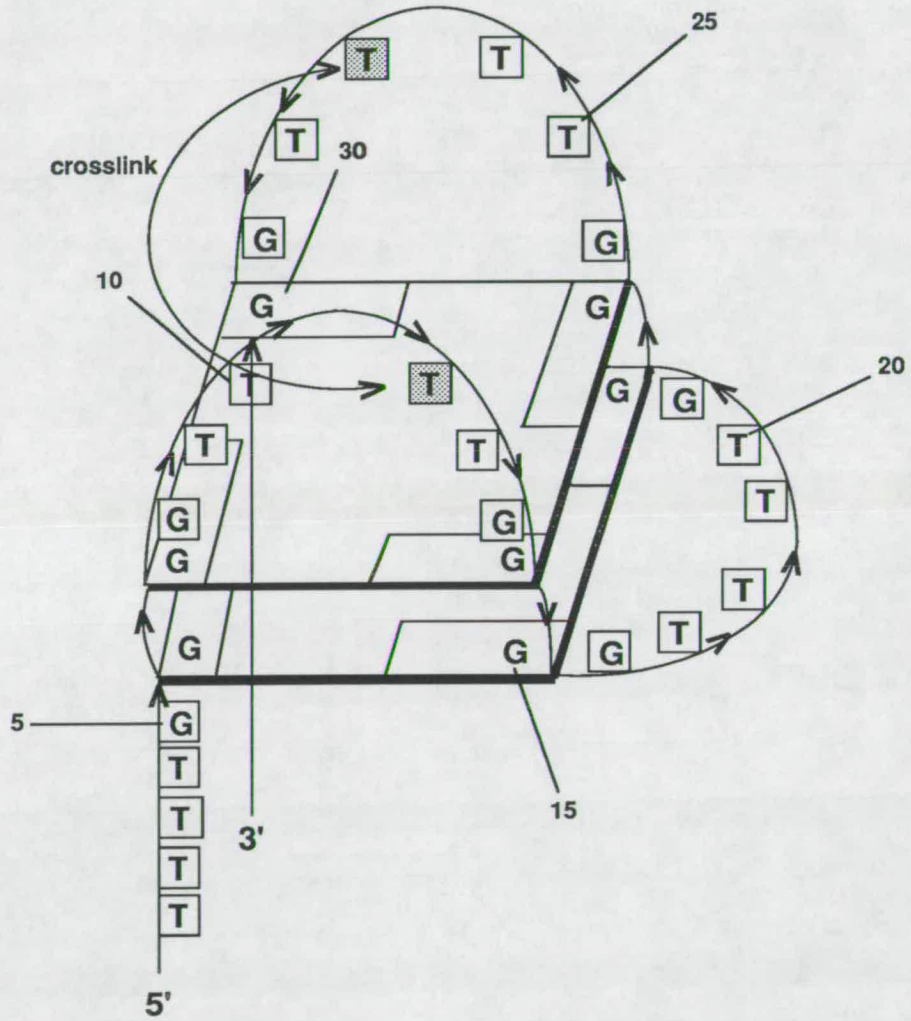
Figure 1.3 Model for Oxy-4

The model shows two G-quartets stacked over each other that are connected at adjacent corners by loops of d(GTTTTG). Above the diagram is a letter representation of the *Oxytricha* telomere repeats present at each macronuclear terminus. Each numbered letter relates to the bases numbered in the model and the 3' and 5' orientation of the strands is indicated. The arrowheads on the loops connecting each corner of the G-quartet show the direction of the DNA strand which moves from 5' (internal) to 3' (external). The position of a UV-cross-link is indicated at the top of the model by the arrows. Variations of the model involve three or four stacked G-quartets with correspondingly shorter loops (taken from Williamson *et al.*, 1989).

5 10 15 20 25 30

GGGG TTTT GGGGTTTT GGGGTTTT GGGGTTTTGGGG3'
CCCC AAAACCCC AAAACCCC AAAA5'

Oxy-4



The double hairpin loop formed by the quartet structure may play a role in stabilisation of the DNA termini after addition of repeats by telomerase during replication and before binding of the telomere proteins (to be discussed later) (Shippen-Lentz and Blackburn, 1990, Morin, 1991).

The lack of a specific monovalent cation effect and base specificity in the formation of the Tet2 and Oxy2 structures suggests a different conformation from that seen with the 4 repeat unit oligos. One possible explanation is an involvement in telomere-mediated, chromosomal associations within the cell. The hairpin duplexes from two chromosomes could associate and become stabilised by G-G base pairs in a G quartet, Figure 1.4. In experiments looking at the association of synthesised oligos. mimicking the telomere at rDNA molecules from *Tetrahymena*, *ie* a duplex region and 3' single strand tail of telomeric repeats, dimers were shown to form in a salt dependent manner. The dimers also demonstrated methylation protection patterns resembling those of the G quartet structure (Sundquist and Klug, 1989). Such a structure may also explain the aggregation of *Oxytricha* macronuclear DNA, Figure 1.2b (Lipps, 1980, Oka and Thomas, 1987).

It has not yet been determined whether these structures have a role in telomere function or if, indeed, they are actually present *in vivo*. A role in recognition of the telomere prior to replication by telomerase is unlikely as it has been shown that oligos. representing telomeres but unable to form G-strand structures are still able to function as substrates for repeat addition (Henderson *et al.*, 1990). In addition, it has been shown that folded forms of telomeric DNA known to serve as primers for the *Oxytricha nova* telomerase *in vitro* when in their unfolded form, are not recognised by the enzyme (Zahler *et al.*, 1991). Likewise, an involvement in the telomere binding protein / DNA complex is unlikely as the methylation protection patterns of the hairpin duplex and protein bound telomeres are very different (Williamson *et al.*, 1989). However, it is possible that these DNA associated structures play roles in the association of telomeres within the cell and also in the phase after repeat replication and before protein binding. The formation of such structures could have a role in the promotion of dissociation of the newly synthesised repeat from the telomerase RNA template (to be discussed later, Shippen-Lentz and Blackburn, 1990). Alternatively, G-quartets may contribute to the regulation of telomere length *in vivo* by directly inhibiting the binding of telomerase to its substrate (Zahler *et al.*, 1991).

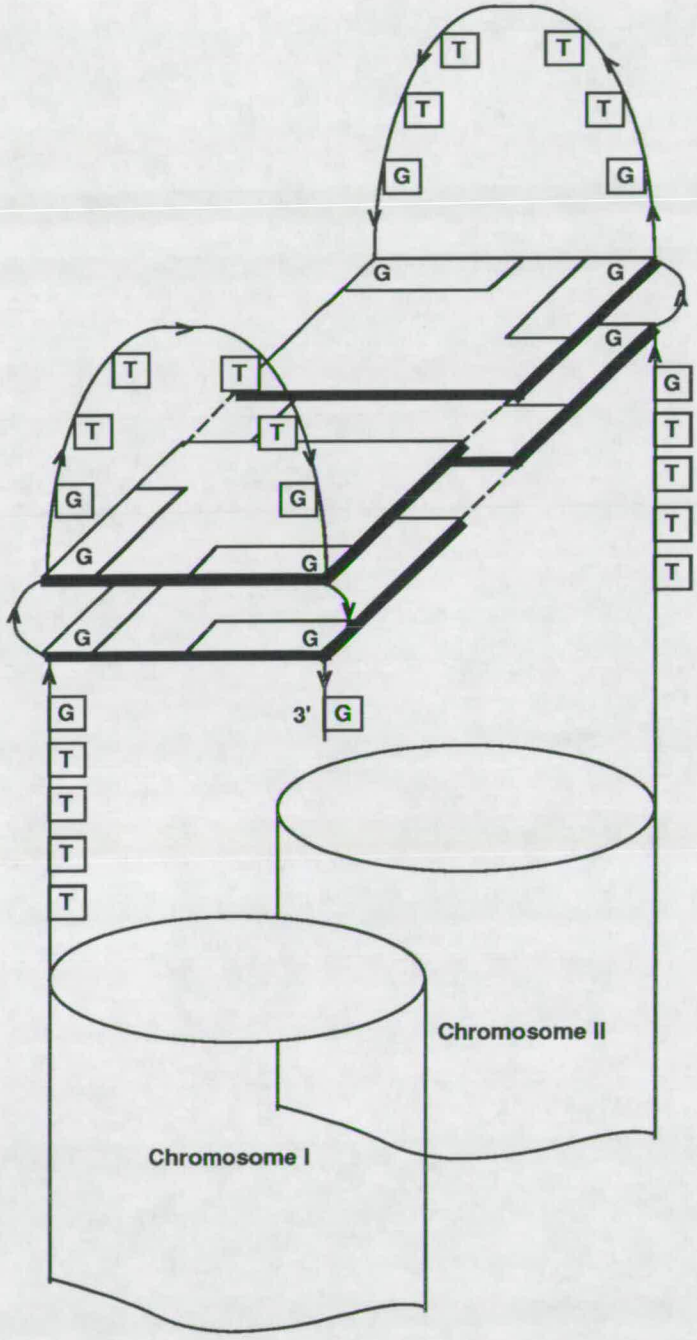
1.2.4 Telomere Replication

1.2.4 (a) Introduction

The most well documented function for telomeres within the cell is their involvement in replication of the ends of DNA molecules. As has already been discussed, the DNA

Figure 1.4 Model for the Association of Natural Oxytricha Telomeric DNA

The model shown in Figure 1.3 has been divided vertically down the middle into two, two-repeat molecules. Monovalent cations could then induce the dimerisation of two telomeric ends to form structures containing G-quartets (taken from Williamson *et al.*, 1989).


T
T
T
T
G
G
G
G
G
T
T
T
T
G
T
T
T
T
T
G
G
G
G
G
G
G
3'
G
T
T
T
T
Chromosome II
Chromosome I

polymerase enzymes are dependent upon a 3' OH group for replication. This is supplied by an RNA primer which upon removal from the 5' terminus results in a region of unreplicated DNA. Consecutive rounds of cell division would result in a progressive loss of coding sequences from the cell's genome by at least a primers length of DNA, *ie* 8-15bp. Hence, a special mechanism must exist at telomeres for the completion of DNA replication (Watson, 1972).

The requirement for a special mechanism of replication at telomeres has resulted in the proposal of many different models. These have involved inter- and intra-molecular complexes and recombination induced models, all initiated by recognition of the repeats found at all telomeres (Cavalier-Smith, 1974, Bateman, 1975, Heumann, 1976Dancis and Holmquist, 1979, Szostak, 1983, Formosa and Alberts, 1986, Morin and Cech, 1988). Most of the earlier proposed models for telomere replication have now been replaced by mechanisms involving either recombination or *de novo* repeat addition to DNA termini by a novel telomere terminal transferase enzyme.

1.2.4 (b) Recombination-mediated telomere replication

Telomere-replication via a recombinational pathway provides an attractive model, as it could account for a number of unusual properties of telomeres. For example, the telomeres of *Trypanosoma* and *Tetrahymena* have been shown to grow and shrink, a feature that could be explained by gain and loss of DNA sequences by unequal exchange during recombination (Bernards *et al.*, 1983, Larson *et al.*, 1987). In addition, the presence of terminal repeats and single strand nicks could promote telomere-telomere recombination.

Several experiments have been performed to visualise possible recombination events involving telomeres. Linear plasmids have been constructed containing a natural *Tetrahymena* rDNA telomere and sequences homologous to yeast Y' elements at opposite ends of the molecule. Upon transformation and propagation in yeast, the Y' element terminus was shown to have acquired additional Y' elements together with $C_{1-3}A$ telomeric repeats. The result suggested that the broken end of the plasmid had been 'healed' by a recombination event with a yeast chromosome end via the Y' telomere-associated repeats (Dunn *et al.*, 1984). In a second experiment, intact linear plasmids with *Tetrahymena* telomeric repeats at both termini, were shown to be involved in recombination events between the end of the plasmid and chromosomal telomere-associated repeats. The reaction was initiated via homology between terminally added $C_{1-3}A$ repeats on the plasmid; a common occurrence when such plasmids are replicated in yeast cells; and $C_{1-3}A$ repeats within the chromosomal X - Y' junction. The reaction resulted in the addition of 1 - 4 Y' elements to either end of the linear plasmid (Dunn *et al.*, 1984). Both events required the RAD 52 gene product, which is needed for the double strand nicking of DNA taking place

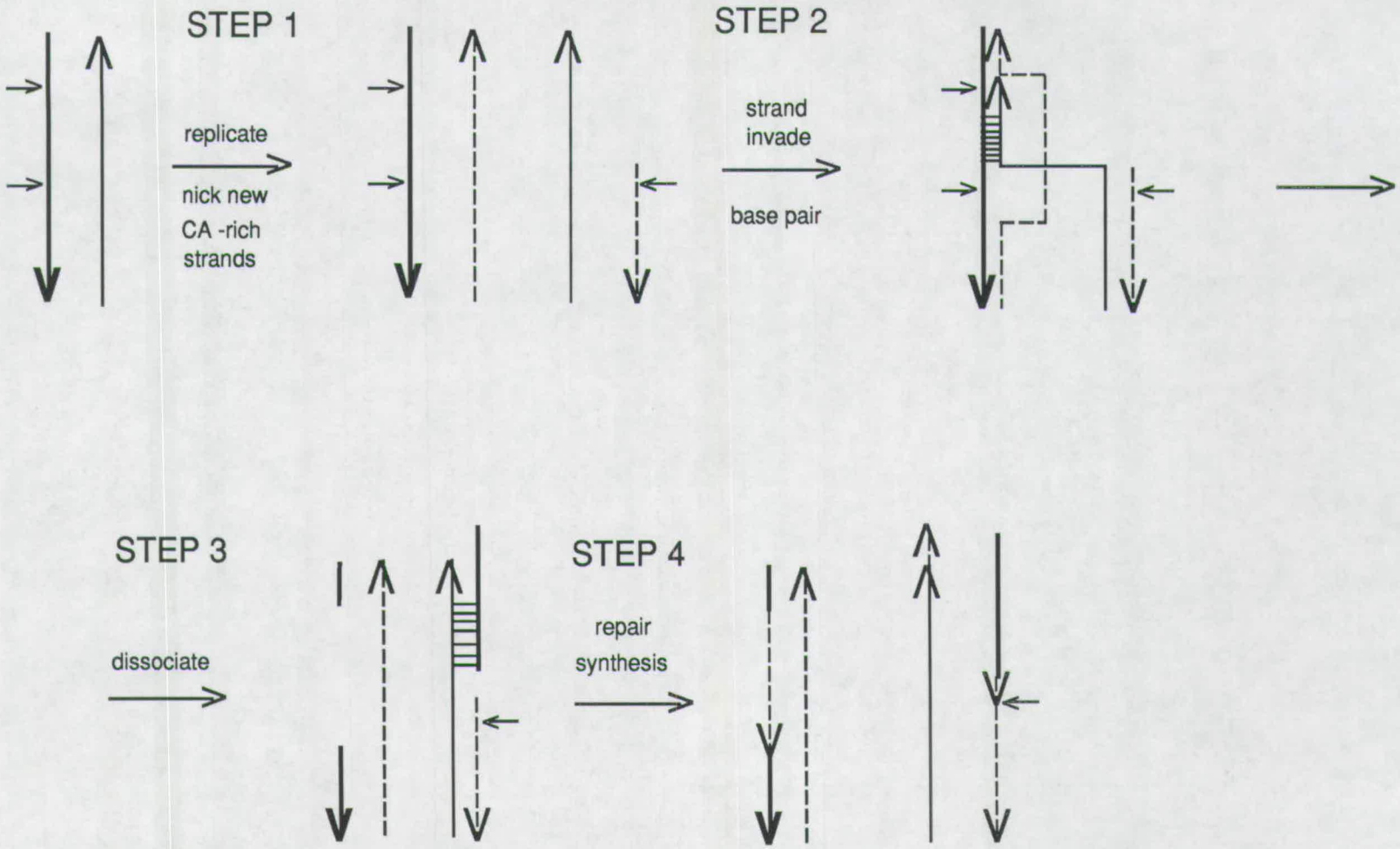
during the recombination event (Dunn *et al.*, 1984). It has been suggested that this recombination process may play a role in distributing the telomere-associated sequences to all chromosome ends within a species which, in turn, fits in with a recombination mediated mechanism for the healing of broken chromosome ends (Dunn *et al.*, 1984, reviewed by Zakian, 1989).

Possible models of recombination-mediated telomere replication have also been shown to occur in the absence of the RAD 52 gene product (Pluta and Zakian, 1989, Wang, *et al.*, 1989, Wang and Zakian, 1990). In these experiments, linear plasmids containing a natural telomeric fragment from *Tetrahymena* and a variable number of *Oxytricha*-like repeats at opposite ends, were constructed. Upon transformation and propagation in yeast, *Tetrahymena* C_4A_2 repeats were added on to the *Oxytricha* C_4A_4 repeat, test end. The expected addition of $C_{1-3}A$ yeast telomeric repeats was also observed on both ends of the plasmid (Pluta and Zakian, 1989). Further characterisation of the reaction showed that very little homology was required; the C_4A_2 repeat was able to recognise the C_4A_4 repeat and the extent of pairing could be as little as 12bps (Wang *et al.*, 1989, Wang and Zakian, 1990). The added C_4A_2 repeats were shown to be positioned distal to the C_4A_4 repeats at the test end suggesting that inter- or intramolecular telomere-telomere recombination had occurred either between or within plasmid molecules, respectively. Rad 52 independent telomere recombination could then be visualised in terms of a model for telomere replication, Figure 1.5. After replication, the 3' overhang, produced by primer excision at the end of the chromosome, would be able to strand invade and base pair with a complementary region on another strand. Subsequent replication would proceed using the donor telomere as a template. After dissociation, the terminus carries a single strand tail, composed of new repeats, that serves as a template for primase and conventional DNA polymerase mediated replication of the complementary strand. Removal of the RNA primer would still leave a gap at the 5' end of the newly replicated strand, but no sequence information would be lost. The addition of the yeast telomeric repeats to the ends of the input plasmids can be predicted to occur in a similar way with the chromosomal telomeres acting as templates (Wang and Zakian, 1990).

From the described model, several roles for telomere-telomere recombination have been proposed. As already mentioned, recombination could be involved in telomere replication. However, if this was the case recombination events would be expected involving authentic telomeres. No transfer of C_4A_4 or C_4A_2 DNA to yeast chromosomes was observed during propagation of the relevant plasmids in yeast. These results could be explained if yeast telomeres were more likely to recombine with each other than with plasmid termini. This is possible as chromosomal telomeres and internal stretches of $C_{1-3}A$ are more abundant than plasmid termini, and also, chromosomal telomeres have a much greater

Figure 1.5 Recombination Model for Telomere Replication

Each strand of a DNA duplex is represented by a line:
Thin lines indicate the GT-rich strand
Thick lines indicate the CA-rich strand
Dashed lines indicate newly replicated DNA
The 3' end of each strand is indicated by an arrowhead
Nicks in the CA-rich strands are indicated by horizontal arrows
Step 1 : after DNA replication and subsequent removal of the RNA primer, a 5' terminal gap is produced on each duplex. Specific nicks are made on the newly-replicated CA-rich strand.
Step 2 : the free 3' tail (thin line) invades the sister duplex and base-pairs in an out-of-register configuration with the complementary strand (thick line), displacing its complementary strand (dashed line).
Step 3 : dissociation of the newly base-paired region is facilitated by nicks present on the CA-rich strand of the donor (thick line).
Step 4 : the resulting internal gap and terminal 3' gap on each duplex is filled by repair synthesis (dashed line) and remaining nicks are sealed by DNA ligase. Replication by this mechanism results in two fully replicated duplex molecules, one of which (duplex on the right) is longer (thereby providing a molecular explanation for how telomeres can increase in length) and has sequence information obtained from the thick duplex on the left (taken from Pluta and Zakian, 1989).


STEP 1
replicate
nick new
CA -rich
strands
STEP 2
strand
invade
base pair
STEP 3
dissociate
STEP 4
repair
synthesis

homology to each other than to C_4A_4 or C_4A_2 DNA. Telomere-telomere recombination could provide a salvage pathway that acts only on those telomeres with very short stretches of $C_{1-3}A$ DNA; its occurrence at natural chromosomes may be too rare to detect by conventional methods such as Southern hybridisation (Wang *et al.*, 1989, Wang and Zakian, 1990). Identification of natural chromosome participation in such a recombination mechanism is required for further assessment of the involvement of this model within telomere functions.

1.2.4 (c) Telomere terminal transferase-mediated telomere replication

The second, and most thoroughly investigated, mechanism for telomere replication and maintenance was first observed in experiments where linear plasmids containing *Tetrahymena* telomeres were transformed into the yeast, *Saccharomyces cerevisiae*. Propagation and subsequent analysis revealed that the plasmid termini had acquired yeast telomeric repeats. This led to the idea that telomere replication involved the nontemplated addition of telomeric repeats onto the ends of chromosomes (Shampay *et al.*, 1984). Further evidence for such a process was provided by the observation that restriction fragments containing telomeric repeats displayed size heterogeneity, appearing as a fuzzy band upon appropriate hybridisation (*eg* Blackburn and Gall, 1978, Allshire *et al.*, 1988, *etc*). In addition, the gradual increase in the number of telomeric repeats when *Trypanosomes* and *Tetrahymena* were kept in log phase growth, suggested a progressive addition of DNA to the ends of these chromosomes (Bernards *et al.*, 1983, Larson *et al.*, 1987).

The discovery of an activity in *Tetrahymena* cell free extracts that added tandem TTGGGG repeats onto synthetic telomere primers led to the proposal that a novel telomere terminal transferase enzyme, or telomerase, was involved in the addition of telomeric repeats necessary for the replication of chromosome ends in eukaryotes (Greider and Blackburn, 1985). Such activities were later shown to be present in the cell free extracts of *Oxytricha nova* (Zahler and Prescott, 1988), *Euplotes crassus* (Shippen-Lentz and Blackburn, 1989) and human HeLa cells (Morin, 1989). Analysis of the repeat addition reactions led to the discovery of a number of common properties between the telomerase enzymes from different organisms. Firstly, all enzyme-containing extracts were able to prime the synthesis of telomeric repeats specific to the derivative organism. In addition, the extracts could also add these repeats to oligonucleotides representing the telomeres of different organisms (Greider and Blackburn, 1987, Morin, 1989, Shippen-Lentz and Blackburn, 1989). Secondly, the repeats were added in a progressive manner as single nucleotide bases, rather than as the units which make up the telomere (Greider and Blacburn, 1987, Shippen-Lentz and Blackburn, 1989). Thirdly, pausing in the addition reaction was detected, for example, in *Tetrahymena*, there is a pause in repeat addition at the end of the guanine

bases, before the next thymine residues, and also, after these two bases have been added (Greider and Blackburn, 1987).

The diversity of telomeric repeat sequences recognised by the different telomerases led to the analysis of the factors involved in primer recognition. In *Euplotes*, HeLa cells and *Tetrahymena* two main elements were important for the initiation of repeat synthesis. Firstly, a guanine-rich telomeric sequence was required for repeat synthesis. Synthesis was not observed with the cytosine-rich strand, random sequence oligonucleotides or blunt-ended, double-strand, telomere-like oligonucleotides. It is possible that a special structure is present at the guanine-rich strand, perhaps similar to the previously described hairpin loops, which is recognised by the telomerase enzyme (Sundquist and Klug, 1989). However, it has been shown that oligonucleotides composed of telomere repeats but unable to form G-strand structures are still able to function as substrates for repeat addition (Henderson *et al.*, 1990, Zahler *et al.*, 1991). Secondly, the nucleotide base at the 3' end of the primer specified the first nucleotide to be added (Greider and Blackburn, 1987, Shippen-Lentz and Blackburn, 1989). For example, using *Tetrahymena* cell free extracts, primers ending with two guanine bases were extended by the addition of two more G bases before new, complete, *Tetrahymena* repeats were synthesised (Greider and Blackburn, 1987). Hence, one repeat unit was completed before further synthesis.

Analysis of the cell free extract activity responsible for the synthesis reaction demonstrated that it contained protein and RNA components. Hence, in all cases the extracts were susceptible to inactivation by RNase (Greider and Blackburn, 1985, Zahler and Prescott, 1988, Morin, 1989, Shippen-Lentz and Blackburn,1989) and in *Tetrahymena*, sensitivity to high temperatures and proteinase K was also observed (Greider and Blackburn, 1985). Isolation of the RNA component revealed sizes of 159 and 191 nucleotides in *Tetrahymena* and *Euplotes* respectively, and sequencing of the two species revealed the presence of a sequence complementary to the G-rich repeats, *ie* 5'CAACCCCAA3' for *Tetrahymena* and 5'CAAAACCCCAAAACC3' for *Euplotes*. The genes responsible for the RNA components of the enzymes have been isolated from *Euplotes* and *Tetrahymena* and shown to contain the sequence of the RNA elements (Greider and Blackburn, 1989, Shippen-Lentz and Blackburn, 1990).

A model for the replication of telomeres by the telomerase has been described, which in both *Tetrahymena* and *Euplotes* involves elongation and translocation of the primer along the RNA moiety of the telomerase enzyme (Greider and Blackburn, 1989, Shippen-Lentz and Blackburn, 1989). Hence, the first step in the reaction is the recognition of the primer followed by hybridisation of the 3' nucleotides to the 3' portion of the CAACCCCAA or CAAAACCCCAAAACC sequence. The primer is filled out to complete the last full repeat and translocation then repositions the primer so that the 3' nucleotides are again hybridised

ready for elongation, Figure 1.6. Such a mechanism would explain the repeat unit periodicity of the addition reaction with the pauses corresponding to dissociation and reannealing of the RNA primer. After the elongation reaction, translocation must occur to allow the continuation of telomere repeat synthesis. Hence, each translocation step must involve unpairing of the newly formed DNA-RNA helix. It has been proposed that unpairing is facilitated by the known ability of telomeric guanine-rich sequences to form intramolecular, non-Watson-Crick G-G base pairs (Shippen-Lentz and Blackburn, 1990). Evidence for such a model is derived from experiments where primers complementary to the telomerase RNA template up to the start of the repeat region were shown to prime addition of only one and a half repeats. Such a small amount of repeat would be unable to fold into a stable hairpin and so the lack of further extension could be attributed to an inability of the primer to dissociate from the RNA template (Shippen-Lentz and Blackburn, 1990). The same principle could be involved in the inability of a *Tetrahymena*-like single unit repeat oligo. to act as a primer for repeat addition (Greider and Blackburn, 1987).

All of the experiments described implicate the telomere complementary sequence in the RNA moiety of the telomerase enzyme as the template for telomere synthesis *in vitro*. Evidence that they function as templates *in vivo* has been supplied by mutation of the CAACCCCAA sequence within the cloned *Tetrahymena* telomerase gene (Yu *et al.*, 1990). Hence, the gene was altered by site directed mutagenesis to give three different mutants, CAACCCCCAA, CAACCTCAA and CGACCCCAA, where the underlined base has been altered. These were subsequently transformed into *Tetrahymena* cells on a high-copy-number vector. Analysis of the resultant subclones showed that the telomeres contained repeats derived from the input, mutated sequence for CAACCCCCAA and CGACCCCAA. No repeat addition was primed from the CAACCTCAA mutated RNA gene. In addition, cells containing the mutant genes demonstrated striking morphological changes and senescence when these genes were over-expressed in *Tetrahymena* cells. Transformants containing CAACCTCAA had the most extreme senescence phenotype. Before death, the cells had rounded and irregular shapes and were much larger than the wild-type cells. In addition, the rDNA telomeric fragment derived from these cells was significantly shorter than expected. For cells containing the CAACCCCCAA and CGACCCCAA mutant genes, the telomeric rDNA fragments had longer mean lengths and broader size distributions than in control cells. Phenotypically, the cells grew extremely large, with very irregular cell and nuclear shapes. Macronuclear and cell divisions were greatly impaired, while DNA replication was relatively unaffected. Hence, all three alterations of the telomere-complementary sequence in the telomerase RNA caused impairment and eventual arrest of cell division in *Tetrahymena*; reversion of this phenotype occurred upon loss of the mutant telomerase RNA gene (Yu *et al.*, 1990).

Figure 1.6 Model for the Elongation of Telomeres by Telomerase

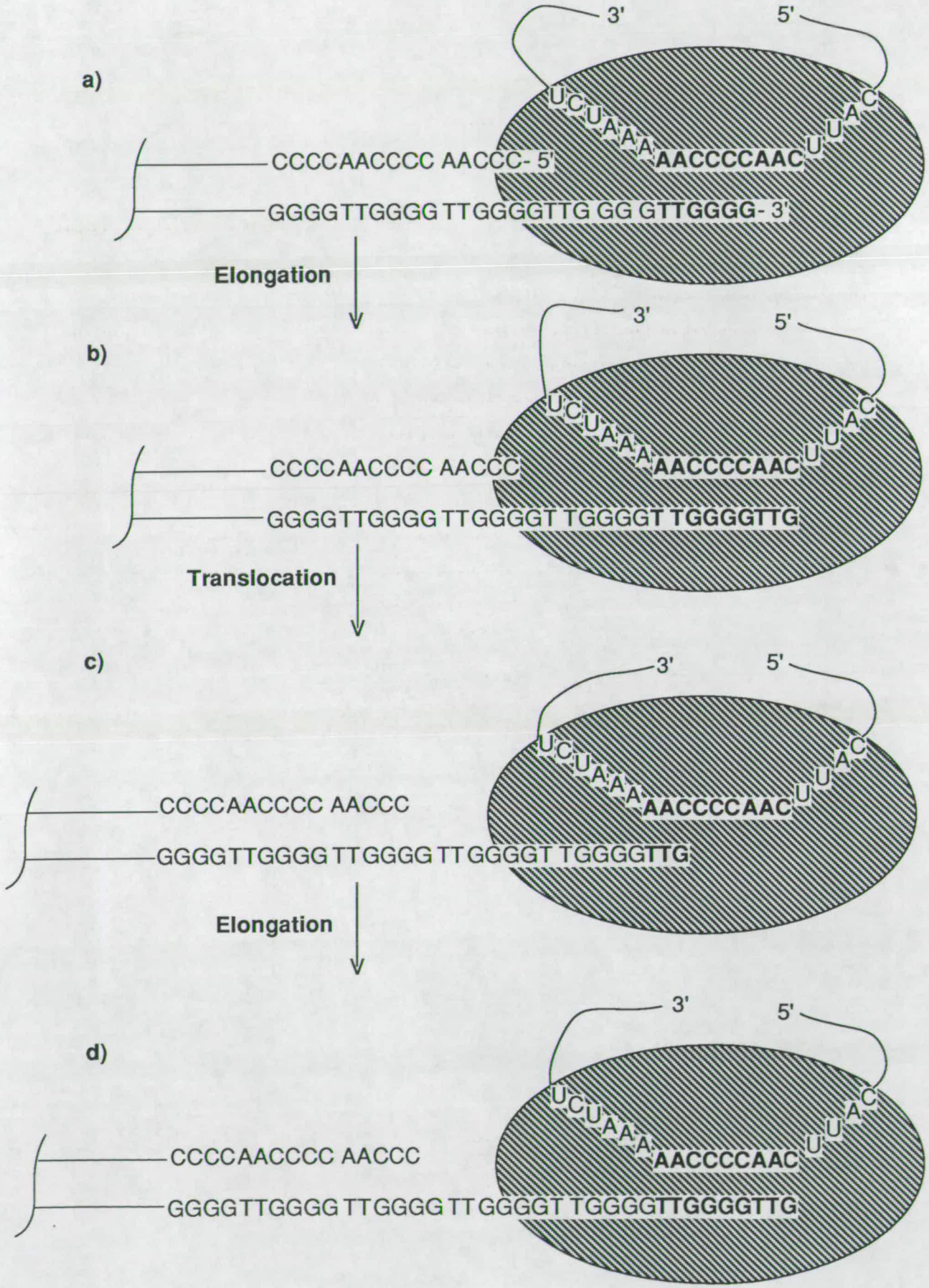
The *Tetrahymena* telomere is shown containing a 13-base overhanging TTGGGG strand.

(a) After recognition of the TTGGGG strand by telomerase, the most 3' nucleotides are hybridised to the CAACCCCAA sequence in the RNA.

(b) The sequence TTG is then added one nucleotide at a time.

(c) Translocation then repositions the 3' end of the TTGGGG strand such that the most 3' TTG nucleotides are hybridised to the RNA component of telomerase.

(d) Elongation occurs again, copying the template sequence to complete the TTGGGGTTG sequence. This mechanism explains how oligonucleotides with 3' ends terminating at any nucleotide within the sequence TTGGGG are correctly elongated to yield perfect tandem repeats of $(TTGGGG)_n$ (taken from Grieder and Blackburn, 1989).


a)
3'
5'
UCUAAA AACCCCAAC UUAC
CCCCAACCCC AACCC- 5'
GGGGTTGGGG TTGGGGTTG GG GTTGGGG- 3'
Elongation
b)
3'
5'
UCUAAA AACCCCAAC UUAC
CCCCAACCCC AACCC
GGGGTTGGGG TTGGGGT TGGGGT TGGGGTTG
Translocation
c)
3'
5'
UCUAAA AACCCCAAC UUAC
CCCCAACCCC AACCC
GGGGTTGGGG TTGGGG TTGGGGT TGGGGTTG
Elongation
d)
3'
5'
UCUAAA AACCCCAAC UUAC
CCCCAACCCC AACCC
GGGGTTGGGG TTGGGG TTGGGGT TGGGGTTGGGGTTG

From these results it can be established that telomerase uses the telomeric sequence within the RNA as a template for telomere synthesis *in vivo*. The presence of an altered template led to longer or shorter telomeres and cell senescence. Hence, it appeared that the dynamic equilibrium between elongation and shrinkage had been disturbed. For elongation, it has been suggested that sequence alteration prevents a 'size-regulator' from binding and hence, repeat addition will continue. Evidence for such a size-regulator is provided by experiments in yeast. Here an increase in the number of $C_{1-3}A$ yeast telomere repeats led to an overall increase in telomere length, suggesting the dilution of a telomere size regulating factor (Runge and Zakian, 1989). A candidate for such a negative regulator is the tightly binding, sequence-specific telomere protein described in the ciliate *Oxytricha* (Gottschling and Zakian, 1986). Binding of the protein to the termini of macronuclear DNA is thought to maintain the strictly defined number of repeats found in all hypotrichous ciliates (Zahler and Prescott, 1988). In *S. cerevisiae* and *Tetrahymena*, the RAP 1 and telomere binding proteins respectively, are thought to play a role in telomere size regulation (Conrad *et al.*, 1990, Blackburn and Chiou, 1981). Lack of repeat addition in the presence of the mutated RNA gene CAACCICAA could reflect an inability of the enzyme to use dATP, required for copying the mutated template sequence. Alternatively, the mutated C nucleotide could have some other function in the activity or formation of the telomerase complex. The impaired cell division and senescence phenotypes of cells containing mutated RNA genes suggests that telomeres are involved in sequence-specific telomeric DNA-protein interactions which must be maintained to ensure proper cell division. It is proposed that these protein-DNA interactions are either weakened in the transformants with altered DNA sequences, or eventually lost in transformants with shortened telomeres. Combining these results with the observations of telomere associations in, for example, *Oxytricha* and *Drosophila*, it is obvious that telomeres play an essential role in chromosome segregation and nuclear division.

1.2.5 Factors Involved in the Formation of New Telomeres

Müller and McClintock first demonstrated that the telomere provides a protective 'cap' for the end of the chromosome. This stabilizes the termini against end-to-end fusions and exonucleolytic degradation, both of which are characteristic of broken chromosome ends, and allows normal segregation of the chromosomes during cell division (Müller, 1938, McClintock, 1940 and 1942, Müller and Herskowitz, 1954). Characterisation of the structure and function of telomeres from many different organisms has revealed a number of occasions where new telomeres are formed upon broken DNA 'ends'.

During the development of the macronucleus in the holotrichous and hypotrichous ciliates several rounds of endoreduplication of the genome occur, with transient formation of

the polytene chromosomes. This step is followed by extensive DNA elimination of the micronucleus-specific sequences generating, for example, ~10^7 DNA molecules in *Oxytricha* and ~9000 in *Tetrahymena* (reviewed by Blackburn and Karrer, 1986, Klobutcher *et al.*, 1984). The mechanism for such chromosome fragmentation has been investigated. Initially, it was proposed that the telomeric repeats found at the termini of macronuclear DNA molecules were also present at the same sites within the micronuclear DNA. Upon fragmentation, the repeats could act as signals for further telomere repeat addition (Katzen *et al.*, 1981). However, in *Oxytricha* , it has been shown that no such repeats exist in the micronuclear chromosomes and so the telomere must be added directly to an originally interstitial site during or after the excision process (Boswell *et al.*, 1982, Klobutcher *et al.*, 1984). Similar results have also been obtained with *Paramecia* (Forney and Blackburn, 1988) and *Ascaris lumbricoides* (Müller *et al.*, 1991). While in *Tetrahymena* it has been shown that chromosome fragmentation seems to be directed by 'chromosome breakage sites' (cbs), which have been implicated in the formation of a terminal complex that mediates telomerase recognition (Yao *et al.*, 1990). In all cases, telomeric repeats are absent from the broken termini but repeat addition and hence, telomere formation takes place (Boswell *et al.*, 1982, Klobutcher *et al.*, 1984, Forney and Blackburn, 1988, Müller *et al.*, 1991, Yu and Blackburn, 1991).

Addition of species-specific telomeric repeats to aberrant 'broken' chromosome ends in *Plasmodium falciparum* and humans has been observed (Forney and Blackburn, 1988, Pologe and Ravetech, 1988, Wilkie *et al* ., 1990). Telomerase has been implicated in telomere repeat addition to the non-telomeric sequences exposed by the breakage reaction. In a human patient with α-thalassemia, a large deletion at the tip of chromosome 16 has resulted in the loss of the telomere. DNA sequence analysis has revealed the addition of $(TTAGGG)_n$ repeats to the normally interstitial breakpoint (Wilkie *et al.*, 1990). Assessment of the ability of the human telomerase to extend primers derived from the breakpoint has revealed several interesting features of this enzyme (Morin, 1991). There appear to be two sites of the primer interaction with telomerase. The first requires recognition of the 3' end of the primer by the template, with the homology required being as little as 2-4 nucleotides. Repeat addition proceeds from this site. The second site interacts with the upstream region of the primers. It is very G-rich and its presence affects primer utilisation, resulting in larger products than for primers containing only the sequence specific site (Morin, 1991). This observation fits in with the idea that such a G-rich region might be able to promote translocation of the telomerase along the template by involving the newly replicated repeats in a G-G base pair hairpin duplex (Shippen-Lentz and Blackburn, 1990). Alternatively, the G-rich region may remain bound to the second site causing hairpin formation of newly

replicated repeats and promote replication by reducing the dissociation rate from the second site (Morin, 1991).

In *Drosophila* , viable terminal deletions have been observed which lose DNA sequences upon each round of cell division without obvious deleterious effects upon the phenotype of the organism (Levis, 1989, Beissman *et al.*, 1990a and 1990b). However, healing events of such termini have been observed which involve the addition of HeT telomere-associated sequences; this differs from the same event in humans and *Plasmodium falciparum*, where simple telomere repeats are added to the broken ends. The actual mechanism for healing in *Drosophila* has not yet been determined. However, the addition of telomere-associated repeats suggests that telomerase might not be involved. Secondly, it is unlikely that homologous recombination has taken place as no reciprocal product has been isolated from the relevant *Drosophila* clones (Beissman *et al.*, 1990b).

From the analysis of 'broken' chromosome healing, it appears that the human telomerase requires very little telomere repeat sequence for extension, unlike previously reported (Morin, 1989). The inability of the appropriate telomerases to extend oligonucleotides formed of only one repeat, *eg* TTGGGG for *Tetrahymena* (Greider and Blackburn, 1985), may reflect an inability to extend such small pieces of DNA rather than a lack of sequence recognition. The results also show that telomeric simple repeats are all that is required for stable maintenance of chromosomes during replication and cell division (Wilkie *et al.*, 1990). In addition, the fact that linear plasmids with telomere repeats proximal to 12-600bp of non-telomeric sequences were still capable of seeding the addition of telomere repeats distal to these unique sequences in yeast, suggests that the putative yeast telomerase is still active when bound just internal to the terminus (Murray *et al.*, 1988). Similar results have been obtained with the *Tetrahymena* telomerase enzyme (Harrington and Greider, 1991). Sequence analysis of the chromosome 16 breakpoint revealed C residues at the breakpoint and GT repeats leading into the chromosome. It is thought that the structure could support repeat addition by telomerase (Wilkie *et al.*, 1990). The presence of unique DNA at the different terminal breakpoints in humans, *Paramecia*, *Oxytricha*, *Plasmodium and Tetrahymena* and telomere repeat addition directly to these sites suggests a common mechanism of telomerase-mediated DNA addition. Hence, telomerase appears to be essential for maintenance of the genetic information within the cell.

1.2.6 Abnormalities in Telomere Structure and Function

Analysis of the phenotypic effects of telomere-related abnormalities has provided an insight into the structure, function and genetic control of telomeres within the cell. The idea that telomeres provide a protective 'cap' for the chromosomes was provided by Müller and

McClintock (Müller, 1938, McClintock, 1940 and 1942, Müller and Herkowitz, 1954). It was observed that natural termini do not fuse with each other or with broken ends (Müller, 1954). Since these initial observations, the fusion of telomeric chromosome ends is still thought to be a very uncommon phenomenon. However, end-to-end associations of chromosomes, implying apparent telomeric fusions either between single or double chromatids, have been reported in cultured human cells exposed to the SV40 virus (Yerganian *et al.*, 1962), from various benign conditions such as Thiberge-Weissenbach syndrome (Dutrillaux *et al.*, 1978) and from patients with cancer such as malignant fibrous histiocytoma (Mandahl *et al.*, 1985), renal tumours (Kovacs *et al.*, 1987), squamous cell carcinoma (Pathak *et al.*, 1988) and B cell malignancies (Fitzgerald and Morris, 1984, Saltman *et al.*, 1989). In most cases, telomeric fusions were observed in 20-30% of the cells and appeared to involve random associations of a variety of chromosome pairs which were different in each cell. A possible explanation for such telomere fusions can be obtained from observations that untransformed human fibroblasts show loss of telomere repeat sequences upon serial passage (Harley *et al.*, 1990). In addition, reduction in the size of telomeres has been observed in colorectal carcinomas when compared to normal mucosa and when fetal tissues and sperm are compared to adult colonic mucosa and blood; the fetal tissues and sperm have larger telomeres than adult blood and colonic mucosa such that a progressive size decrease is observed with increasing age (Hastie *et al.*, 1990).

It has been suggested that the greater telomere size in human sperm compared to blood could be due to a lack of telomerase activity in somatic tissues (Cooke and Smith, 1986). In maize, broken chromosomes are only healed in sporophytic (zygotic) tissues and not in the endosperm (terminally differentiated), suggesting again that telomerase activity is lacking in differentiated cells (McClintock, 1940 and 1942). A lack of telomerase could explain the senescence phenotype of fibroblasts (Harley *et al.*, 1990). Untransformed fibroblast cells sustained in culture have been shown to arrest eventually at a specific stage in the cell cycle, G1, before S phase and show an increase in chromosomal abnormalities, 90% of which are dicentrics attached at their telomeres (Sherwood *et al.*, 1988, Harley *et al.*, 1990). Hence, the observation of shortened telomeres within colon carcinomas, the increase in abnormalities of pre-senescent fibroblasts associated with the loss of terminal sequences, and observations of random telomere-telomere associations in some forms of cancer, suggests that end-to-end chromosome fusions could play a role in the genetic instability associated with tumorigenesis. In addition, the fusion events may result from the loss of long stretches of G-rich repeats found at the ends of all linear chromosomes (Sherwood *et al.*, 1988, Harley *et al.*, 1990, Hastie *et al.*, 1990).

Yeast is a single-celled, as opposed to a multicellular, organism and would be expected to require telomerase-like activity for normal growth and maintenance of genetic continuity

throughout life. Several mutant strains have been identified which alter the length of the $C_{1-3}A$ sequences at telomeres. The *TEL* 1 and *TEL* 2 genes of telomeres, when mutated, lead to an overall shortening of telomeres, which has no effect upon cell viability or growth rates (Lustig and Petes, 1986). A different cell cycle effect is observed in *cdc* 17 mutant yeast strains. *CDC* 17 is a conditional lethal cell division cycle gene where temperature sensitive mutations produce stage specific arrest of the cell cycle at the permissive temperature (Hartwell, 1973). A deficiency for the *CDC* 17 gene product increases the frequency of mitotic recombination and chromosome loss, suggesting an involvement in DNA metabolism (Hartwell and Smith, 1985). There is also an increase in the number of telomeric $C_{1-3}A$ repeats and an elevation in recombination events at the most centromere distal genetic interval of the chromosome (Carson and Hartwell, 1985). The *CDC* 17 gene product has been identified as a DNA polymerase, however, its role in aberrant telomere elongation has not yet been defined (cited by Lucchini *et al.*, 1990). It is unlikely that the extra telomere repeats are alone responsible for the observed phenotype as the introduction of extra telomeric sequences into yeast cells, resulting in chromosomal telomere elongation, do not present with the same abnormalities (Runge and Zakian, 1989).

A second mutation, defining the new gene described as *EST* 1, (for ever shorter telomeres), has been identified in yeast which gives rise to a senescence phenotype in the absence of the gene product. Eventual cell death results from the progressive loss of sequences essential for telomere function (Lundblad and Szostak, 1989). The *est* 1 strain displays phenotypes consistent with a defect in some aspect of telomere replication in that loss of the gene product results in an immediate and progressive decrease in telomere length. The senescence phenotype manifests itself some generations after a decrease in telomere length has been observed. Finally, *est* 1 strains demonstrate chromosome loss. The simplest explanation for the observed phenotypes suggests that the *est* 1 strain has a defect in the non-templated telomere repeat addition reaction (Lundblad and Szostak, 1989). Isolation and characterisation of the *EST* 1 gene product has revealed sequence motifs suggesting that *EST* 1 encodes a protein component of an RNA-dependent polymerase in yeast, which on the basis of the phenotype of *est* 1 mutants is likely to be a component of an essential yeast 'telomerase' (Lundbald and Blackburn, 1990). The phenotype of *est* 1 mutants is analogous to that seen in human fibroblasts grown in culture and cells derived from colonic carcinomas (Harley *et al.*, 1990, Hastie *et al.*, 1990). Hence, it is possible that untransformed, somatic cells do lack a telomerase activity.

Additional telomere-related abnormalities have been observed in other yeast cell systems. As previously discussed, overexpression of the RAP 1 telomere binding activity in yeast cells, results in elevated levels of both chromosome loss and recombination. It has been suggested that both of these effects could be due to the excess protein preventing

dissociation of telomere-telomere and telomere-nuclear envelope interactions before cell division. This could result in chromosome non-disjunction and hence, loss from daughter cells and promote recombination events (Conrad *et al.*, 1990). Mutagenesis of the RNA moiety of the *Tetrahymena* telomerase resulted in senescence of the host *Tetrahymena* cells due to what appeared to be an inability to divide. From these phenotypes it was proposed that telomeres are involved in sequence-specific telomeric DNA-protein interactions that were either weakened or eventually lost in the transformants (Yu *et al.*, 1990).

Investigations into mutations of telomere structure and function has allowed further characterisation of their roles within the cell. Firstly, the simple repeats present at all telomeres so far analysed are essential for maintenance of linear DNA molecules within organisms. Secondly, the telomerase enzyme, found within diverse organisms, is required for replication of the telomeres. Its absence results in the progressive loss of telomere repeats, which in turn, leads to chromosome fusion, breakage and recombination. Such events often lead to senescence of cells and could be involved in the process of carcinogenesis. Thirdly, the proteins that bind to telomeres may involve the linear DNA molecules in telomere-telomere and telomere-nuclear envelope interactions. These may be important in organisation of the total nuclear DNA and alignment of chromosomes before cell division. Loss of this organisation could affect the integrity of cell division leading to unequal separation of the nuclear material to the daughter cells. Hence, telomeres appear to have essential roles in replication of the whole genome, prevention of chromosome fusion and ensuring correct cell division.

1.3 Interstitial Telomere Repeats

1.3.1 Additional Genomic Locations of Telomere-Related Repeats

In addition to the very end of chromosomes, telomere repeats have been demonstrated within other regions of the genomes of a wide variety of different species. In both the yeast *Sacharomyces cerevisiae* and *Plasmodium berghei*, telomere-like repeats have been found interspersed within the subtelomeric repeats (Walmsley *et al.*, 1984, Ponzi *et al.*, 1985, Pace *et al.*, 1987). Experiments showing that the *Tetrahymena thermophilia* telomeric repeat crosshybridised to human telomeres also revealed that this probe recognised discrete bands which varied between different, unrelated individuals and were insensitive to the nuclease Bal 31 (Allshire *et al.*, 1988, Hastie and Allshire, 1989). Therefore, it appeared that *Tetrahymena*-like telomere repeats were located both terminally and within human chromosomes (Allshire *et al.*, 1988). Further studies of these 'internal' loci have led to the isolation of *Tetrahymena*-like telomere repeats from conventionally constructed DNA

libraries, indicating their internal location within the genome (Wells *et al.*, 1990). *In situ* hybridisation analysis of the isolated repeat probes to human metaphase chromosomes suggested that the majority of the 'internal' loci were clustered proterminally (Wells *et al.*, 1990). One exception to this general rule was provided by a locus at the interstitial location of 2q11-2q14 (Allshire *et al.*, 1988, Wells *et al.*, 1990, IJdo *et al.*, 1991). Additional *in situ* hybridisation analysis using the *Trypanosoma* telomeric sequence $(TTAGGG)_n$ upon chromosomes from a number of different vertebrate species has demonstrated the existence of these repeats at telomeric, centromeric and intersitial sites (Meyne *et al.*, 1990).

1.3.2 Possible Functions of 'Internal' Telomere Repeats

The presence of telomere repeats at 'internal' sites within the genomes of different species raises the question of their effect upon the genome. Do they retain the properties of functional telomeres or do they adopt the charateristics of repetitive DNA ?

1.3.2 (a) Are internalised telomere repeats still functional telomeres ?

As has already been discussed, the characteristic repeated DNA and associated proteins of telomeres provide a protective 'cap' at the extreme terminus of chromosomes. This stabilises the termini against end-to-end fusions and exonucleolytic degradation, which are characteristic of broken chromosome ends, and allows normal segregation of the chromosomes during cell division (Müller, 1938, McClintock, 1940, 1942, Müller and Herskowitz, 1954). In addition, the telomere ensures complete replication of the genome at each cell cycle, may be involved in establishing the 3D architecture of the interphase nucleus and has effects upon gene expression (Section 1.2.1). Studies into the protein and DNA conformation at telomeres have revealed that non-histone proteins bind to the termini and the telomeric repeats form unique structures *in vitro* (Section 1.2.3). All of the above facts raise the question of what properties telomere repeats possess when located interstitially within the genome. Are they still capable of forming telomeres and are they still associated with telomere specific proteins ? Finally, are they able to form secondary structures ? All three features could lead to instability at the site; resolution of the repeats into a functional telomere could result in breakage of the chromosome at that site. There are two experimental examples of such resolution and subsequent chromosome breakage. Firstly, the introduction of head-to-head *Tetrahymena* telomere repeats into the *MAT* locus of the yeast chromosome III resulted in its breakage due to the resolution of the inverted repeats into two new telomeres (Murray *et al.*, 1988). Secondly, it has been shown that vectors containing human telomeric DNA, when transfected into mammalian cells, were able to integrate into, and break, the chromosome at that site (Farr *et al.*, 1991). Hence, in both cases, the introduction of telomere repeats into a chromosome have led to its breakage via their

resolution into functional telomeres (Murray *et al.*, 1988, Farr *et al.*, 1991). There are also examples where telomere internalisation may have occurred spontaneously within the cell. For example, as already described, *in situ* hybridisation and cloning has demonstrated that *Tetrahymena* and *Trypanosoma*-like telomere repeats exist within the region of 2q11-2q14 on chromosome 2 (Allshire *et al.*, 1988, Wells *et al.*, 1990, IJdo *et al.*, 1991). This is the site of fusion of two ape chromosomes to form the human chromosome 2 (Yunis and Prakash, 1982) and demonstrates that telomere repeats have remained at this site. Internalisation of telomeres is also thought to have occurred in the translocation chromosome t(6;19) found in the lymphocytes of a mentally retarded woman (Drets and Therman, 1983) and the *Sxr* mutation in the mouse (Hunt and Burgoyne, 1987). The latter two sites of fusion were weak, having a tendancy to break; the Sxr fusion point exhibiting both the appearence and behaviour of a fragile site (Dret and Therman, 1983, Hunt and Burgoyne, 1987). This tendancy to break could suggest that the fused termini are resolving into functional telomeres. The region 2q11-2q14 on the human chromosome 2 also contains two rare fragile sites (Berger *et al.*, 1985). From these observations it has been suggested that fragile sites may be internalised telomeres (Hastie and Allshire, 1989). However, cloning of the Fragile X mental retardation fragile site, Xq27.3, and sequencing have revealed that it is composed of a CGG repeat (Verkerk *et al.*, 1991). This obviously raises some doubt about a connection between fragile sites and internalised telomeres, although until other fragile sites have been cloned and sequenced a possible connection cannot be dismissed. Whether or not such a connection exists, internalised telomeres may still represent a subset of genomic loci that are susceptible to breakage.

1.3.2 (b) Do 'internal' telomere repeats adopt the properties of repetitive DNA ?

Eukaryotic genomes contain three distinct classes of repeated DNA, described as satellite, interspersed repeated and minisatellite DNA. Of the three classes, the human internalised telomere repeats appear to resemble the minisatellite DNA.

Minisatellite DNA

(i) Isolation

The minisatellites (Jeffreys *et al.*, 1985b) or variable number tandem repeats (VNTR) (Nakamura *et al.*, 1987) form a class of repetitive sequence within eukaryotic genomes. The first minisatellite was isolated by chance from a library of random segments of human DNA (Wyman and White, 1980). Analysis of its presence within a Mormon pedigree demonstrated that at least eight different alleles were present, revealed as a series of *Eco* RI restriction fragments ranging in size from 14 to 29kb. In addition, ~75% of individuals were heterozygous possessing two different length alleles (Wyman and White, 1980, de Martinville *et al.*, 1982). Such highly variable regions have also been isolated from the human

insulin gene region (Bell *et al.*, 1982), the α-globin gene cluster (Higgs *et al.*, 1981, Proudfoot *et al.*, 1982, Goodbourn *et al.*, 1983, Reeders *et al.*, 1985, Jarman *et al.*, 1986), the *c-Ha-ras* gene (Capon *et al.*, 1983), the X/Y chromosome pseudoautosomal region (Simmler *et al.*, 1987) and the human myoglobin gene (Weller *et al.*, 1984). In each case, the variable region was shown to consist of tandem repeats of a short sequence and polymorphism resulted from allelic differences in the number of repeats present.

(ii) General characteristics

After these initial observations it was shown, using a probe containing pure repeats of a 33bp repeat element derived from the human myoglobin gene, that minisatellite loci were scattered throughout the human genome (Jeffreys *et al.*, 1985b). The repeat length of each minisatellite region was generally half, the same or double the length of the 33bp probe from the human myoglobin gene. In addition, each repeat unit was shown to contain a unique 10-15bp core region derived from the 33bp probe sequence. The minisatellite loci vary between 2-20kb in human genomic DNA digested with frequently cutting enzymes. As the mean fragment size of human DNA digested with such enzymes is ~0.3kb, minisatellite repeats must be devoid of most restriction enzyme sites (Jeffreys *et al.*, 1985b). Hybridisation analysis using minisatellite repeat probes revealed that these loci were highly variable between unrelated individuals to the extent that few fragments were shared, resulting in an individual specific DNA fingerprint (Jeffreys *et al.*, 1985c).

The pattern of inheritance of minisatellite loci was investigated in several large pedigrees (Jeffreys *et al.*, 1985b and 1986). In one such pedigree, the transmission of *Hinf* I minisatellite fragments within 54 individuals spanning four generations was analysed. Most of the loci were transmitted from each parent to only some of the offspring, establishing that most of the fragments were present in the heterozygous state and that heterozygosity was approaching 100%. It was also established that inheritance was Mendelian in that the heterozygous bands were transmitted on average to 50% of the offspring, or, in other words, all of the fragments in the offspring could be traced back to one or other parent and then in turn to their parents and therefore provided a set of stably inherited genetic markers (Jeffreys *et al.*, 1985b).

The isolation of locus-specific minisatellites has provided further information concerning the polymorphic nature of these regions (Wong *et al.*, 1986 and 1987). One such locus was shown to have 77 different alleles containing 14 to 525, 37bp repeat units per allele in a sample of 79 individuals. All of the alleles except, the shortest, were rare and the resulting heterozygosity was very high at ~97% (Wong *et al.*, 1986). Four other loci were also isolated and shown to detect extremely variable Mendelian loci with heterozygosities ranging from 90 to 99% (Wong *et al.*, 1986 and 1987).

(iii) Distribution amongst different species

The human minisatellite probes have been shown to cross-hybridise to mouse DNA and to detect multiple variable loci. The resultant DNA fingerprints varied substantially between, but relatively little within, an inbred strain. Using BxD recombinant inbred strains it was shown that 8 out of 13 hypervariable loci could be regionally assigned to mouse chromosomes and that the assigned loci were autosomal, dispersed and preferentially associated with centromeres or telomeres. One of the minisatellites was shown to be very complex, with alleles 90kb or more long and with internal restriction enzyme cleavage sites which produced a haplotype of cosegregating fragments (Jeffreys *et al.*, 1987). The presence of minisatellites in other eukaryotes cross-hybridising to the human probes 33.6 and 33.15 has also been observed, for example in dogs, cats and *Drosophila* (Jeffreys and Morton, 1987) suggesting that similar sequences exist within other species.

(iv) Stability of minisatellites

Minisatellite loci detected by probes 33.6 and 33.15, while possessing almost 100% heterozygosity, were shown to be stably inherited through several generations, demonstrating germ line stability (Jeffreys *et al.*, 1985b and 1986). Analysis of minisatellites within different tissues from one individual revealed that somatic stability was also present (Jeffreys *et al.*, 1985c). However, within the pedigrees analysed, new variant alleles were demonstrated (Jeffreys *et al.*, 1985b and 1988). Using probes detecting single hypervariable loci (Wong *et al.*, 1986 and 1987) it was demonstrated that the mutation rate generating new length alleles varied from locus to locus and rose to 5.2% per gamete for the locus showing the greatest level of instability. Mutations were shown to arise sporadically and different offspring in a sibship never shared a common mutant allele. Hence, from these initial observations it was thought that there was no significant level of germ line mosaicism for mutant alleles and that new mutants were preferentially generated in the latest stages of gametogenesis (Jeffreys *et al.*, 1988). However, analysis of tumour tissue and cell lines has revealed the occurrence of somatic mutation at minisatellite loci. DNA analysis of lymphoblastoid cell lines revealed the presence of mutant minisatellite alleles which were absent from the derivative individual and also the offspring of the said individual. In tumour cells, the appearance of the mutant allele was accompanied by a decrease in intensity of one of the constitutional alleles (Thein *et al.*, 1987, Armour *et al.*, 1989). New, mutant, minisatellite alleles have also been detected early in mouse development, again demonstrating that mutation events are not restricted to the germ line but can also arise in the soma (Kelly *et al.*, 1989).

(v) Generation of minisatellite variation

The mechanism generating minisatellite variabilty has not yet been defined. Sequence analysis of a cloned subset of minisatellite repeats has revealed a common 'core' sequence

which is similar to the Chi sequence, a signal for generalised recombination, of *E.coli* (Jeffreys *et al.*, 1985b). There has been speculation that the core sequence might serve as a recombination signal to promote unequal crossing over at minisatellites and hence generate new length alleles. *In situ* hybridisation of the 33.15 minisatellite probe to human metaphase chromosomes showed a clustering of autoradiographic grains, principally at or around chiasmata; autosomal sites where crossing over had taken place (Chandley and Mitchell, 1988). In addition, there is evidence of clustering of minisatellites near the ends of human chromosomes (Royle *et al.*, 1988, Wells *et al.*, 1989, Nakamura *et al.*, 1988). It has been proposed that the preferential localisation of minisatellites near to the ends of human chromosomes reflects a function in the processes of synapsis and recombination associated with these regions (Royle *et al.*, 1988). This suggests that meiotic and mitotic recombination might be involved in generating minisatellite variability. However, analysis of the markers that flank new mutant minisatellite alleles has failed to show evidence for unequal crossing over between homologous chromosomes for 12 different mutants at two different VNTR loci (Wolff *et al.*, 1988 and 1989). This suggests that either unequal exchange between sister chromatids and/or replication slippage may be the primary source of minisatellite mutation. Replication slippage could explain small changes in length of ~4-10 repeat units in the generation of new length alleles (Jeffreys *et al.*, 1988 and 1990). Larger length changes have been predicted to be independent of replication and therefore are thought to occur by unequal sister chromatid exchange (Jeffreys *et al.*, 1988). However, the exact mechanism of new length variant generation in minisatellites has not yet been defined.

(vi) Use of minisatellites in human genetics

Isolation and characterisation of minisatellites from the human and mouse genomes has revealed that they provide both individual specific, in humans, and strain specific, in mouse, fingerprints (Jeffreys *et al.*, 1985b, 1985c and 1987). The inheritance of minisatellite alleles can be followed in large human pedigrees, suggesting a useable degree of germ line and somatic stability (Jeffreys *et al.*, 1985b and 1986). In addition, most minisatellite loci were shown to have very high levels of allele heterozygosity ranging from 90-99% (Jeffreys *et al.*, 1985b). From these results a role for minisatellites in human gene mapping has been suggested (Jeffreys *et al.*, 1985b and 1985c, Nakamura *et al.*, 1987). The large number of alleles within the population for most minisatellite loci would provide a set of highly polymorphic markers for mapping genes involved in human genetic disease (Jeffreys *et al.*, 1985b and 1985c, Nakamura *et al.*, 1987).

The subset of minisatellites identified by Jeffreys *et al.* (1985b) have been applied to forensic and immigration problems (Gill *et al.*, 1985, Jeffreys *et al.*, 1985a). For example, the individual specific fingerprint identified by the minisatellite probes 33.6 and 33.15 has been used to identify a male suspect in a rape case (Gill *et al.*, 1985). In addition, the inheritance of

minisatellite alleles between a mother and her putative son has been used to establish their relationship (Jeffreys *et al.*, 1985a). Minisatellite fingerprints could also be used for identification of different mouse strains (Jeffreys *et al.*, 1987). At present, the exact function of minisatellites within the genome has not been determined, nevertheless, they have provided a useful set of polymorphic markers in forensic science, paternity testing and in general individual identification.

(vii) Comparison of minisatellites and internalised telomere repeats

Characterisation of internalised telomere repeats within the human genome is still in its early stages. Observations that the *Tetrahymena thermophilia* telomeric repeat crosshybridised to human telomeres and also interstitial, discrete bands that varied between different individuals provided the first evidence of their presence within the human genome (Allshire *et al.*, 1988, Hastie and Allshire 1989). These internalised telomere repeats ranged in size from 2-9kb, were devoid of frequently cutting restriction enzyme sites suggesting long stretches of simple repeats and showed a pattern of bands unique to each individual studied (Hastie and Allshire, 1989). All of these features are characteristic of minisatellites. In addition, the individual specific pattern of hybridising DNA fragments suggests a similar level of variability to that seen with minisatellites. If internalised telomere repeat loci take the form of repetitive DNA then their characterisation may outline any further similarities with minisatellite DNA. In addtion, if they are polymorphic, they may provide a new set of informative markers for human gene mapping.

1.4 Origins of Research Project

This project originated from observations by Allshire *et al.* (1988) that the *Tetrahymena thermophilia* telomeric repeat crosshybridised to human telomeres (Allshire *et al.*, 1988) In addition, the probe recognised discrete bands which varied between different, unrelated individuals and were insensitive to the nuclease Bal 31. Therefore, it appeared that *Tetrahymena*-like telomere repeats were located both terminally and within human chromosomes. Further investigation of the human telomere revealed that it was predominantly composed of *Trypanosoma*-like telomere repeats (Table 1.1) (Allshire *et al.*, 1989, Brown, 1989, Cheng *et al.*, 1989, Cross *et al.*, 1989). The probe used to detect these repeats was also found to crosshybridise to discrete bands and a non-resolvable, very high molecular weight region within total genomic mouse DNA (Allshire *et al.*, 1989). This project was designed to isolate internalised telomere repeats from both the mouse and human genomes. Characterisation of the isolated loci would allow their genomic position to be established, any connection of internalised telomere repeats with regions of fragility to be determined and if, instead, they represent hypervariable, repetitive loci, a new set of markers

for genome mapping. A second line of investigation was to charaterise the non-resolvable, very high molecular weight region of $(TTAGGG)_4$ hybridising DNA in mouse genomic DNA. This region could correspond to the mouse telomere and its characterisation will determine how it compares with the telomeres of other species that have already been investigated.

Chapter 2

Local C57Bl/6 mice were obtained from Harlen Olac in Oxford, are derived from the strain C57Bl/6J and originate from the Jackson Laboratories in the USA.

Hull C57Bl/6 mice were obtained from Bantin and Kingman, are derived from the strain C57Bl/6J and originate from the Jackson Laboratories in the USA.

Materials and Methods

2.1 Mouse Strains Used for Analysis

Total spleen and kidney DNA's from *Mus musculus*, *Mus domesticus*, *Apodemus sylvaticus* and the inbred *Mus musculus* derived strain AKR/J mice were a gift from Dr.R. Hill. The *Mus spretus* DNA was a gift from Drs' N. Copeland and N. Jenkins. The *Mus caroli* mice were a gift from Dr. J. Ansell. The *Mus musculus*, *Mus domesticus*, and *Apodemus sylvaticus* mice were trapped wild in Greece, Denmark and locally, respectively. The *Mus spretus* and *Mus caroli* mice were trapped and then held in the laboratory for a period of time. The inbred strains used were DBA/2, C57Bl/6 and AKR/J. DBA/2 mice were obtained from colonies in Hull and Oxford. C57Bl/6 mice were obtained from both local and Hull colonies.

The C57Bl/6 x DBA/2 recombinant (BxD RI) inbred strains and the C57Bl/6 x *Mus spretus* backcrosses were used for mapping studies. Total genomic DNA from the BxD RI strains was obtained from Dr. B. A. Taylor and the backcross DNA from Drs' N. Copeland and N. Jenkins.

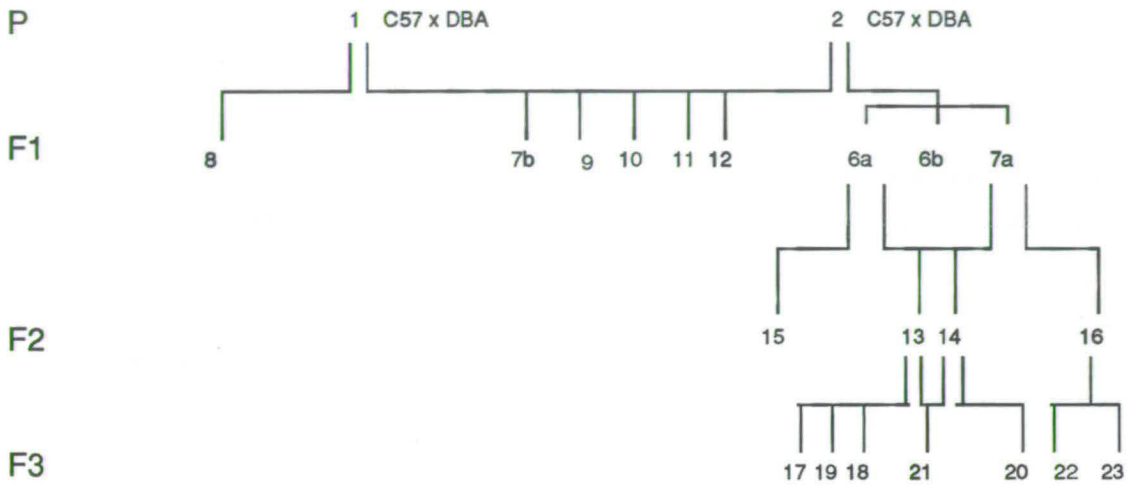
2.2 System for Cross-Strain Mating between C57Bl/6 and DBA/2 Inbred Mice

2.2.1 Outline of Breeding Procedure

Cross strain matings were set up between ≥8 week old male Hull DBA/2 and local female C57Bl/6 mice. At ~8 weeks old the F1 (first filial) progeny from these pairs were either brother x sister or cross-pair mated at random. From a total of 14 F1 progeny, 4 brother x sister and 4 cross-pair, matings were arranged; some of the F1 mice were bred against two different mates. Breeding within each pair was allowed to occur until several litters had been obtained; a total of 178 male and female F2 (second filial) progeny were recorded. The F2 progeny from two of the F1 pairs were then either brother x sister or cross-pair mated in order to obtain the F3 (third filial) generation. Two brother x sister and two cross-pair matings were arranged. From the progeny of this stage of breeding, cross-pair and brother x sister matings were set up in order to generate the F4 (fourth filial) progeny, as outlined in Table 2.1. All of the matings were conducted at random; selection of the progeny from certain pairs was also performed without bias.

The animals involved in the breeding experiment were analysed for the inheritance of both telomeric and internal $(TTAGGG)_n$ containing loci. Hence, after breeding had been completed, the mice were killed by cervical dislocation, the spleen removed, and total genomic DNA extracted by the method to be described.

Schematic Representation of C57Bl/6 x DBA/2, F1, F2 and F3 Matings



Key : No's = Pairs of mice

Table 2.1 Breeding Plan of Cross Strain Matings Between C57Bl/6 and DBA/2 Inbred Mouse Strains

Pair Number	Progenitor Animals of Pair Plus Matings From Which Derived			
	Inbred mice	F1	F2	F3
1	C57Bl/6xDBA/2	NA	NA	NA
2	C57Bl/6xDBA/2	NA	NA	NA
6a	C57Bl/5xDBA/2	P2 BxS	NA	NA
6b	C57Bl/6xDBA/2	P2 BxS	NA	NA
7a	C57Bl/6xDBA/2	P2 BxS	NA	NA
7b	C57Bl/6xDBA/2	P1 x P2	NA	NA
8	C57Bl/6xDBA/2	P1 BxS	NA	NA
10	C57Bl/6xDBA/2	P1 x P2	NA	NA
11	C57Bl/6xDBA/2	P1 x P2	NA	NA
12	C57Bl/6xDBA/2	P1 x P2	NA	NA
13	C57Bl/6xDBA/2	P2 BxS	P6a x P7a	NA
14	C57Bl/6xDBA/2	P2 BxS	P6a x P7a	NA
15	C57Bl/6xDBA/2	P2 BxS	P6a x P6a	NA
16	C57Bl/6xDBA/2	P2 BxS	P7a x P7a	NA
17	C57Bl/6xDBA/2	P2 BxS	P6a x P7a	P13 x P13
18	C57Bl/6xDBA/2	P2 BxS	P6a x P7a	P13 x P13
19	C57Bl/6xDBA/2	P2 BxS	P6a x P7a	P13 x P13
20	C57Bl/6xDBA/2	P2 BxS	P6a x P7a	P14 x P14
21	C57Bl/6xDBA/2	P2 BxS	(P6a x P7a) x (P6a x P7a)	P13 x P14
22	C57Bl/6xDBA/2	P2 BxS	P7a x P7a	P16 x P16
23	C57Bl/6xDBA/2	P2 BxS	P7a x P7a	P16 x P16

Where : P = pair

NA = not applicable

BxS = Brother x Sister mating

During the course of the breeding experiment all of the mice involved were ear-tagged to allow identification. This was performed by the staff of the local animal unit. The staff of the local animal unit also carried out the maintenance of the mice.

2.3 Isolation of Total Genomic, Cellular DNA

2.3.1 Isolation of Total Genomic Human DNA from Blood

The blood sample (approx.10mls) was collected into EDTA (1mg/ml) to prevent clotting of the cells. The sample was mixed with an equal volume of 2xC lysis buffer + 4% SDS in a sterile Falcon tube and left to stand for 30 minutes at room temperature. An equal volume (20mls) of phenol was added, mixed by gentle shaking and then centrifuged at 9,000rpm, 16 ^{o}C, for 10 minutes. The top layer was removed and the DNA present precipitated with 2 volumes of ethanol. The DNA was spooled out of the ethanol and resuspended in 3mls of 1xC buffer. RNase was added to a total concentration of 50µg/ml and incubated for 1 hour at 37^{o}C. SDS and Proteinase K to final concentations of 0.5% and 100µg/ml respectively were then added and the whole reaction incubated at 37^{o}C overnight. The next day the DNA was extracted with phenol (1 volume (vol.)), phenol (0.5 vol.) / chloroform (0.5 vol.), and chloroform (1 vol.). The DNA was precipitated with 0.3 vol. of 7.5M NH_4OAc and 2.5 vols. of ethanol and recovered by spooling into TE (1 mM EDTA, 10mM Tris, pH8.0).

2xC lysis buffer + 4% SDS = 0.1M Tris pH7.9, 1mM EDTA, 20mM NaCl + SDS to a final concentration of 4%.

1xC buffer = 2xC lysis buffer diluted in half minus the SDS.

2.3.2 Isolation of Total Genomic DNA from Mouse Tissue

Two different methods were used:

(i) Nuclei were isolated and DNA recovered as described by Hill *et al.*. (1985) The spleen, liver and kidney were removed from a freshly killed mouse and separately homogenised using a Teflon, loose fitting, automatic homogeniser in 10-15mls of buffer A1. Homogenisation was performed until whole nuclei were released from the tissue. The resulting mixture was passed through 2-3 layers of gauze and layered onto 10-15mls of buffer A2. Centrifugation at 12,000r.p.m. for 15 minutes at 4^{o}C in Swing-out HBV rotor of the Sorvall RC-5B centrifuge separated the nuclei from the cellular debris and the resultant pellet was resuspended in 5mls Buffer A. 5mls of 2x Pronase was then added followed by SDS and RNaseA to final concentrations of 0.2% and 100µg/ml respectively, and the reaction incubated at 37^{o}C for 30 minutes. After this, SDS and Proteinase K to final concentations of 1% and 100µg/ml,

respectively, were added and the reaction incubated at 37°C overnight. Next day, the preparations were extracted and the DNA recovered as described in 2.3.1.

Buffer A (10x stock) : 0.6M KCl, 0.15M NaCl, 5mM Spermine, 1.5mM Spermidine, 0.15 M Tris HCl pH7.4, 1.4 M β Mercaptoethanol (added at time of use of buffer)

Buffer A1 : 1x Buffer A, 0.3M Sucrose, 2mM EDTA and 0.5mM EGTA.

Buffer A2 : 1x Buffer A, 1.37M Sucrose, 1mM EDTA and 0.3mM EGTA.

2x Pronase : 100mM Tris, 300mM NaCl, 200mM EDTA, pH10.0

(ii) The spleen, liver, kidney, lungs, heart or brain were removed from a freshly killed mouse, and dropped into buffer (100mM NaCl, 50mM Tris, 10mM EDTA, pH8.0) and processed for 10 seconds with a Silversun homogenizer. Proteinase K to 100μg/ml and SDS to 0.5% final concentration were then added and the reaction incubated at 37°C overnight. Next day, the preparations were extracted and DNA recovered as described in 2.3.1.

For all three methods of DNA extraction described care was taken to ensure minimum shearing of the DNA.

2.4 Localisation of $(TTAGGG)_n$ within Mouse Metaphase Chromosomes

2.4.1 Preparation of Metaphase Chromosomes

A C57Bl/6 male mouse was injected intraperitoneally with 0.1mls of a 1mg/ml solution of colcemid. Colcemid arrests dividing cells in metaphase of the cell cycle when the chromosomes have condensed and are visible as discrete elements. 1.5 hours later the mouse was killed by cervical dislocation and the bone marrow flushed out of the femur by aspiration. This is achieved by first scraping the bone clean and then cutting off the epiphyses with strong scissors. The bone marrow is flushed out using a 1ml pipette and fine needle filled with ~0.75ml of isotonic saline solution, ie PBS. This method was repeated for each femur and the marrow washed into a sterile petri dish. The bone marrow cells were then pelleted by centrifugation at 800 r.p.m. for 8-10 minutes at 16°C. The pellet was resuspended, while whirlimixing, in 0.56% KCl solution, left at room temperature for 20 minutes and centrifuged at 500 r.p.m. for 8 minutes at 16°C. The resultant pellet was resuspended with care while whirlimixing in 10mls of the fixative, 3 parts methanol : 1 part glacial acetic acid and centrifuged at 500 r.p.m. for 8 minutes at 16°C. The fixative step was repeated 3 more times, on the final step the pellet was resuspended in only a few drops of fix. Single drops of the chromosome containing fixative were dropped from a fine pastette onto ethanol cleaned and polished glass slides. Each slide was air dried and stored in the dark at 4°C until required.

2.4.2 Oligonucleotide Primed In Situ DNA Synthesis (PRINS)

The PRINS technique of oligonucleotide primed *in situ* DNA synthesis to mouse metaphase chromosomes was performed by Dr J. Gosden and Mrs D. Hanratty; as described in Gosden *et al.* (1991). In brief, the $(TTAGGG)_4$ oligonucleotide was annealed to complementary sequences present in the DNA of the prepared mouse metaphase chromosomes and then extended using DNA polymerase (Klenow or Taq1) and the constituents of the polymerase chain reaction; the sequence of DNA bases lying next to the annealed primer provided the template for the reaction.The thymine deoxynucleotide was replaced by bio-11-dUTP which was incorporated into the DNA product in place of thymine. Detection of 'hybridisation' was made possible by an antibody reaction. Firstly, either of the two antibodies FITC-Avidin DCS or Texas Red (TR)-Avidin D were reacted with the metaphase chromosomes. Both antibodies will only recognise and react with the Biotin attached to the dUTP residue. FITC and Texas Red are both fluorochromes which fluoresce upon stimulation with the correct wavelength of light. The signal obtained can be intensified by binding additional layers of the fluorochromes. This was achieved by first binding the anti-antibody, biotinylated anti-avidin, to the original signal and then a second layer of FITC-Avidin DCS or TR-Avidin-D to the metaphase chromosomes. The slides were screened on a Leitz Ortholux II UV microscope fitted with Ploemopak filters and dichroic mirrors, and selected cells containing well spread chromosomes and clear signal were scanned with the BioRad Lasersharp MRC 600 Dual Channel Confocal Scanning system attached to the same microscope.

2.5 Bacterial Cell Culture

2.5.1 Media and Additives

All media was sterilised by autoclaving.

L-Broth and Agar

Per litre : 10g tryptone (Difco), 5g yeast extract (Difco), 5g NaCl, 0.5g $MgSO_4$, pH7.2. L agar contains, in addition, 15g agar/litre.

H Agar

Per litre : 10g tryptone, 8g NaCl, 12g agar, pH 7.0 (H-top contains 8g/l of agar). This agar reduces the number of satellite colonies when plating out bacterial transformations where the drug used is bacteriostatic.

2x TY Broth

Per litre : 16g tryptone, 10g yeast extract, 5g NaCl, pH7.0.

SOC Broth

Basic media per litre : 20g tryptone, 5g yeast extract, 0.59g NaCl, 1.86g KCl. Before use $MgCl_2$ and $MgSO_4$ to a final concentration of 10mM and glucose to a final concentration of 20mM were added from filter sterilised stocks.

"Terrific" Broth (BRL Focus 1987)

Terrific Broth per 800mls : 12g tryptone, 24g yeast extract, 4mls glycerol.

KPO_4 Buffer : 0.17M KH_2PO_4 / 0.72M K_2HPO_4.

The KPO_4 buffer and terrific broth were autoclaved separately and before use 100ml of KPO_4 buffer was added to 800ml of broth.This broth gave high levels of plasmid from an overnight culture and was used in preference to L-Broth for the large scale isolation of plasmid DNA.

Glucose/Minimal Agar

Per litre : 15g agar in 900mls distilled water. Autoclave and while still liquid add 100ml 10x M9 salts, 1ml 1M $MgSO_4$. 1ml 0.1M $CaCl_2$, 1ml 1M thiamine and 10ml 20% glucose. The solutions added to the agar were all sterilized.

10x M9 Salts

Per litre : 60g $NaHPO_4$, 30g KH_2PO_4, 10g NH_4Cl, 5g NaCl

Media Additives

When appropriate, the antibiotic Ampicillin, at a concentration of 100μg/ml, was added to media and agar. X-gal (5-Bromo-4-chloro-3-indole-β-galactopyranoside) was used from a stock solution of 30mg/ml in dimethylformamide. IPTG (isopropyl-B-D-thio-galactopyranoside) stock solution was 100mM.

Bacteria were grown at 37°C, with good aeration for liquid cultures.

Bacterial stocks were kept at -70°C in 20% glycerol.

2.5.2 Bacterial Strains Used

The two bacterial strains used were *E.coli* K-12.

DH5α : Derivative of DH1 (Hanahan, 1983). High efficiency transformation strain for pUC-based plasmids. F^-, recA1, endA1, gyrA96, thi-1, hsdR17 (rk^-, mk^+), supE44, relA1, λ^-, mcrA(+), mcrB(-), Δ(argF-laczya)U169, ϕ80dlacZ ΔM15. The deficiency in the recA recombination protein will inhibit recombination within the bacteria and hence, should allow propagation of repetitive DNA.

JM101 : (lac-proAB), thi, supE, {F'traD36, proAB, lac I^qZ M15} (Yanisch-Peron *et al*, 1985). Strain used for the propagation of M13. Since JM101 contains the $lacI^q$ mutation which overproduces the lac repressor, isopropyl-1-thio-β-D-galactosidase (IPTG) must be added to induce β-galactosidase synthesis (Muller-Hill *et al.*, 1968).

2.5.3 Vectors for Cloning in E.coli

pTZ18/19 : This plasmid was used for the construction of a human size fractionated genomic DNA library. It is a 2.9kb plasmid derived from the pUC series of plasmids. These are based upon pBR322 (Bolivar *et al*, 1977) with both a high copy number, and the ampicillin resistance gene (β-lactamase) (Vieira and Messing, 1982). All of the pUC vectors contain a polylinker inserted into the β-galactosidase gene (LacZ). Non-recombinant plasmids are able to synthesize this enzyme, which breaks down X-gal to release a blue indolyl derivative. In recombinant plasmids the lac gene is interrupted by foreign DNA, resulting in colourless or white colonies. pTZ, in addition, contains the bacteriophage F1 origin of replication which allows the production of single stranded DNA. In addition, the universal and reverse sequencing primers are present, positioned on either side of the polylinker.

2.5.4 Vectors Based on the Single-Stranded DNA Coliphage M13

M13 is a single stranded (ss) DNA filamentous phage of *E.coli* that infects cells via the F.pilus. Replication occurs via a double stranded (ds) replicative form (RF) which upon propagation gives rise to ss progeny virions. M13 has been developed as a cloning vector for DNA sequencing (Messing, 1983; Norrander *et al.*, 1983).Two double strand replicative forms of the phage, vectors mp18 and mp19, were developed, which contain a multiple cloning site in opposite orientations. Also present, lying adjacent to the cloning site, is an oligonucleotide

(17mer) which is complementary to the universal sequencing primer (as defined by United States Biochemical Corporation). Annealing of this primer to the ss form allows inserts to be sequenced in both orientations (to be described). The *E.coli lacz* gene under the control of the Lac operon has been incorporated into the genome; the multiple cloning site resides within this gene. Foreign DNA is cloned into the polylinker of the M13 RF. Therefore, in the presence of the gratuitous inducer IPTG, only non-recombinant plaques will produce a blue colour when plated on the colourmetric substrate X-gal. Recombinant phage are visualised as white plaques.

2.6 Transformation and Propagation of Plasmid DNA in Bacteria

2.6.1 Preparation of Frozen Competent DH5α's for Electroporation

1 litre of L-broth was innoculated with 1/100 vol. of a fresh overnight culture of DH5α bacteria, and grown with vigorous shaking at 37°C until the optical density (OD) at a wavelength of 600nm was between 0.5-1. The flask was then chilled on ice for 15-30 minutes and centrifuged in a cold rotor at 4,000 x g_{max} for 15 minutes. The pellet was resuspended in 1 litre of cold autoclaved water and the bacteria collected by centrifugation as above. The pellet was then resuspended in 0.5 of a litre of cold autoclaved water, centrifuged as before, resuspended in 20mls of 10% glycerol, centrifuged and finally resuspended in 2-3mls of 10% glycerol. The cell concentration at this stage was approx. $3x10^{10}$ cells/ml.The suspension of competent cells were then frozen on dry ice and stored at -70°C .

2.6.2 Electrotransformation of Competent DH5α Cells

The competent cells were thawed on ice. In an ice cold 1.5ml polypropylene tube 40µl of the cell suspension was mixed with 1-2 µl (<1µg) of plasmid DNA and left on ice for approx. 1 minute. The mixture of plasmid DNA and cells was then transferred to an ice-cold, 0.2ml electroporation cuvette and pulsed in the Biorad Gene Pulser once at 25 µF, 2.5 kV, with the pulse controller set at 200 Ω. A pulse time constant of between 4.5 to 5 msec indicated a high level of transformation of the bacteria with the plasmid. After pulsing, the suspension of cells was quickly resuspended in 1ml of SOC medium and incubated at 37°C for 1 hour with shaking. The bacteria were then plated out on the appropriate selective medium.

2.6.3 Screening Bacterial Colonies

A modified version of the method described by Grunstein and Hogness (1975) was followed. DH5α bacteria transformed with plasmid DNA by electroporation were grown on Hybond-N (Amersham) nylon membranes which were washed in boiling H_2O and blotted dry in filter

paper prior to use. Washing was performed to remove any chemicals or detergents deposited on the membrane during its manufacture that might inhibit the growth of the bacteria. After washing, the filters were placed on L-agar plates containing ampicillin and X-gal and the DH5α's transferred to the filter surface using an alcohol flamed glass spreader to obtain ≤ 7 colonies per cm^2 after incubation of the filter-plate at 37°C. This proceduregave rise to the master filter. Duplicate filters were placed on top of the master and a weight rolled across to transfer the colonies. The duplicate and master filters were marked with Indian ink for orientation purposes. The duplicates were then grown at 37°C for 4 hours.

Screening of bacterial colonies required the isolation and binding of the bacterial DNA within each colony to the filter. To prevent movement of the bacteria or DNA from their colonial sites during lysis the filter was placed upon a slightly larger sized piece of 3MM Whatman paper soaked in the appropriate solution and this allowed to diffuse into the colony. For isolation and binding of DNA to the filter.

Bacterial lysis :	10% SDS for 2 minutes
Binding of DNA to filter :	Denaturation solution for 5 minutes
	Neutralisation solution for 5 minutes
	Filter washed in 2xSSC to remove cellular debris, blotted dry in filter paper and the DNA crosslinked to the membrane by UV irradiation (to be discussed later).

Denaturation solution : 0.5M NaOH, 1.5M NaCl

Neutralising solution : 3M NaCl, 0.5M Tris.

SSC : 0.3MTri-sodium citrate, 3M NaCl, pH 7.0.

2.6.4 Small Scale Isolation of Plasmid DNA

As described by Wilimzig (1985).

5mls of "Terrific"-broth containing ampicillin was innoculated with one bacterial colony and incubated at 37°C overnight with vigorous shaking. Next day bacteria from 3mls of the culture were collected by centrifugation, drained and resuspended in 500µl of TELT buffer. 50µl of 10mg/ml lysozyme contained in TELT buffer, made as required, was added and then left at room temperature for 2 minutes. The mixture was then boiled for 1 minute, left on ice for 8 minutes and finally centrifuged for 10 minutes. The supernatant was collected and the plasmid DNA precipitated with 2vols. of ethanol at -70°C for 20 minutes.The plasmid DNA was collected by centrifugation for 15 minutes, dried and resuspended in approx. 25µl of TE.

TELT Buffer : 50mM Tris-HCl, 62.5mMEDTA, 0.4% Triton X-100, 2.5mM LiCl, pH7.5.

2.6.5 Large Scale Isolation of Plasmid DNA

Large scale isolation of plasmid DNA was performed as described in Sambrook, Fritsch and Maniatis (1989) which was a modified version of the alkaline lysis method described by Birnboim and Doly (1979).

500mls of "Terrific" broth was innoculated with a single bacterial colony and the culture grown at 37^{o}C overnight with vigorous shaking. Next day the culture was centrifuged at 6K and 4^{o}C for 5 minutes. The resultant pellet of bacteria was resuspended in 20mls of solution I and 10mg/ml lysozyme and incubated at room temperature for 5 minutes. 40mls of Solution II was then added, mixed thoroughly and incubated on ice for 10 minutes. 30mls of Solution III was added, again mixed thoroughly, incubated on ice for 5-30 minutes and centrifuged at 12K, 4^{o}C for 30 minutes. The supernatant was strained through muslin, 0.6vol. of isopropanol added and the plasmid DNA left to precipitate out of the supernatant for 5 minutes at room temperature. The DNA was collected by centrifugation at 8K, 4^{o}C for 15 minutes. The resultant pellet was washed with 70% ethanol, dried and resuspended to 13ml with TE. 1g/ml of CsCl and 1.5ml of 10mg/ml ethidium bromide were added such that the refractive index of the resulting solution was 1.3860. The plasmid DNA was banded by centrifugation in an 18ml polyallomer tube at 40K, 20^{o}C overnight in a vertical rotor; this allows the CsCl gradient to reach equilibrium rapidly). Plasmid DNA is more dense than the bacterial chromosomal DNA, due to plasmid supercoiling constraining the amount of ethidium bromide which can be intercalated into the molecule. The plasmid band was collected under UV light (300nm) with a syringe and the ethidium bromide removed by a series of extractions with butan-2-ol until no colour remained in the aqueous phase. The DNA was precipitated with 3vols. of 70% ethanol, pelleted by centrifugation and then redissolved in TE.

Solution I : 50mM glucose, 25mM Tris.Cl (pH8.0), 10mM EDTA (pH8.0).

Solution II : 0.2M NaOH, 1% SDS.

Solution III : 3M potassium acetate, made up from (in 100ml) : 60ml 5M potassium acetate, 11.5ml glacial acetic acid, 28.5ml H_2O.

2.7 Manipulation of M13 DNA

2.7.1 Preparation of JM101 Cells Competent for Transformation by M13 DNA

JM101 cells were streaked onto glucose minimal medium plates which selects for cells carrying a plasmid encoding a gene involved in proline synthesis; the host cell has a deletion in this gene. The same plasmid is needed for synthesis of the F-pilus required for infection by M13.

1ml of an overnight culture of JM101 was used to innoculate 100ml of 2x TY broth and the cells grown to OD_{550nm} = 0.4. After harvesting the cells were resuspended in 50ml of cold 50mM $CaCl_2$ and left on ice for 20 minutes. The cells were then pelleted and resuspended in 10ml of the $CaCl_2$ solution, and stored at 4°C for up to 4 days.

2.7.2 Transformation of JM101 by M13 DNA

1ng ds recombinant M13 DNA was added to 300µl of competent cells. After 40 minutes on ice the cells were heat-shocked at 42°C for 3 minutes then returned to ice. 200µl of a fresh exponential culture of JM101 was added and the mixture plated in H top agar. For identification of recombinant phage-containing plaques 40µl of 2% X-gal and 40µl of fresh 100mM IPTG were incorporated into the top agar and the whole mixture plated onto H-agar plates.

2.7.3 Preparation of Single-Stranded M13 DNA

100ml of 2x TY was innoculated with 1ml of an overnight culture of JM101 bacteria. An isolated M13 recombinant plaque was added to 1.5ml of these cells and grown at 37°C for 5-8 hours. The cells were pelleted by centrifugation at 10,000g for 5 minutes. The supernatant was removed, then the phage particles precipitated by the addition of 200µl of 20% PEG 6000 in 2.5M NaCl. After 15 minutes on ice the phage were collected by centrifugation and resuspended in 200µl of TE. After phenol (1vol.), phenol (0.5vol.) / chloroform (0.5vol.) and chloroform (1vol.) extractions, the ss M13 DNA was ethanol precipitated.

2.8 Manipulation of DNA by Enzymes

2.8.1 Restriction Endonuclease Digestion

DNA was digested with restriction enzymes obtained from Boehringer Mannheim or New England Biolabs according to the manufacturers' recommended conditions and buffers provided by both companies. Total genomic DNA digests were, in general, carried out overnight with 2-4 units (u) of enzyme /µg of DNA. Spermidine at 5mM was added to all but the digests carried out in the Boehringer Mannheim L and A buffers, both of which had a low salt concentration.Reactions were stopped by freezing at -20°C.

2.8.2 Partial Restriction Enzyme Digestion of Total Genomic Mouse DNA

Total genomic mouse DNA from the C57Bl/6 and DBA/2 inbred strains was subjected to partial restriction enzyme digestion. The partial digestion reaction was performed by making a stock mixture containing the genomic DNA, 5µg per timepoint, enzyme buffer, water and 1u of enzyme per 8µg of DNA. The reaction mixture was incubated at 37°C and aliquots

removed at 0, 5, 10, 20, 40 and 60 minutes. Then an additional 20u of enzyme was added and the whole incubated O/N. The next morning, the O/N aliquot was removed, an additional 20u of enzyme added and the reaction incubated at 37°C for a further 4 hours. The last two steps were performed to ensure complete digestion of the genomic DNA. The DNA fragments from each timepoint were separated by conventional gel electrophoresis as described in 2.9.2 together with control digests of the same DNA; O/N digestion of 5µg of DNA with 20u of enzyme and O/N + 4hours with a total of 40u of enzyme.

2.8.3 Bal 31 Treatment of DNA

Bal 31 is a 3' exonuclease and an endonuclease. The exonuclease activity removes mononucleotides from both 3' termini of duplex DNA molecules, while the endonuclease function degrades the remaining single-stranded DNA. These reactions are dependent upon the presence of calcium and, therefore, can be stopped by the addition of the chelating agent EGTA. Treatment of restriction enzyme digested DNA fragments of known size with Bal 31 results in a progressive decrease in size which can be observed upon separation by agarose gel electrophoresis. Alternatively, in genomic DNA treated with Bal 31 prior to digestion with restriction endonucleases, the size of the terminal restriction fragments would become shorter with increasing time of Bal 31 digestion.

Bal 31 reactions were performed upon *Mus caroli*, DBA/2 and λ DNA digested with *Hind* III. The proportion of enzyme to DNA in each reaction :

DNA	µg/ml of DNA	u/ml of Bal 31
λ Hind III	100	40
Mus caroli	130	52
DBA/2	200	40

The amount of Bal 31 within the reaction was calculated relative to the amount of DNA and adjusted to produce the required result of progressive loss of mononucleotides from the termini of the DNA molecules. The reaction was performed in the manufacturers' recommended buffer at 30°C. Initially, the mix, without the Bal 31 enzyme, was preincubated at 30°C for 10 minutes, one aliquot removed and incubated in the absence of Bal 31 for the duration of the time course. This sample acted as a control (C) for any nonspecific degradation of the DNA during the time course. Subsequently, the Bal 31 was added and aliquots were removed at the required time points. The reaction was stopped by the addition of 0.1vols. of 200mM EGTA pH8.0. Each aliquot was then extracted once each with BRL Ultra-pure phenol equilibrated with TE pH8.0 (1vol.), phenol (0.5vol.) / chloroform (0.5vol.) and chloroform (1vol.). Care was taken to ensure that the genomic DNA remained intact during this process, hence, at each extraction step the phenol / chloroform was allowed to diffuse into the DNA containing layer for 15 minutes and allowed to separate into two phases

for 15 minutes. The λ *Hind* III DNA was extracted by vigorous mixing in the phenol/chloroform and separation of the two phases by centrifugation at 10,000g for 15 minutes. DNA was precipitated with 0.3vol. 7.5M NH_4OAc and 2.5vol. ethanol, spooled into TE for the genomic DNA and pelleted by centrifugation and resuspended in TE for the λ DNA

Bal 31 buffer : 5x stock, 0.1M HCl, 3M NaCl, 75mM $MgCl_2$, 75mM $CaCL_2$, 5mM EDTA, pH7.2.

2.8.4 Dephosphorylation of DNA

The 5' phosphate group was removed from linear DNA molecules when required with bacterial alkaline phosphatase (BAP); obtained from Amersham International. Removal of the phosphate group prevents ligation between the two ends of the molecule upon the addition of the enzyme DNA ligase. 1 unit of BAP was incubated with 10μg of restriction enzyme digested DNA at 65°C for 30 minutes. Proteinase K, to a final concentration of 100μg/ml, was added to remove the BAP and the incubation continued at 37°C for 30 minutes. The DNA was extracted with phenol (1vol.), phenol (0.5vol.) / chloroform (0.5vol) and chloroform (1vol.),precipitated with 0.3vols 7.5M NH_4OAc, and 2.5vols ethanol, pelleted by centrifugation at 10,000g for 15 minutes, the pellet washed in 70% ethanol and the DNA resuspended in TE.

2.8.5 Ligation of DNA Molecules

The enzyme ligase, isolated from *E.coli* infected with bacteriophage T4, catalyses the formation of a phosphodiester bond between the 3'-OH and 5' phosphate groups of DNA (Weiss *et al.*, 1968). This enzyme can therefore be used to join complementary cohesive termini of ds DNA molecules. Ligations were carried out overnight at 15-16°C with 0.5 units T4 ligase in 50mM Tris-HCl pH7.4, 10mM $MgCl_2$, 1mM spermidine, 100μg/ml BSA, 1mM ATP and 10mM DTT.

2.9 Separation of DNA Molecules by Electrophoresis

General methods were as described by Sealey and Southern (1982).

2.9.1 Solutions and Buffers

20x TAE

Per litre : 96.88g Tris, 32.8g NaOAc, 7.44g EDTA, pH8.2 with acetic acid

5x TBE

Per litre : 54g Tris, 27.5g Boric acid, 4.65g EDTA, pH8.3.

10x stop mix

0.2M EDTA (pH8.0), 15% ficoll and Orange G to suitable colour.

2.9.2 Conventional Agarose Gel Electrophoresis

Conventional, horizontal agarose gel electrophoresis was used to separate DNA fragments ranging in size from 0.5-23kb. Agarose (Sigma, Type II) concentrations of 0.8 and 1% were used, depending upon the resolution of fragments required and the gels were made and run in 1x TAE buffer. The gels were run at variable voltages and for differing times for the required separation of fragments. Orange G was added to the DNA samples before loading and provided a visible marker for the migration of small (<500bp) DNA fragments. λ and Phi X DNA digested with *Hind*III and *Hae*III respectively were used as size markers.

2.9.3 Pulsed-Field Gel Electrophoresis (PFGE)

PFGE, using the Chef DR II Biorad pulse field system, was used to separate DNA fragments in the size range ~30-150kb. The DNA fragments were separated in a 0.5cm thick 1% agarose gel, made and run in 0.5x TAE. The conditions for separation were :

Voltage :	6 volts/cm
Pulse time :	5 seconds
Temperature :	10°C
Run time :	dependent upon the separation required but usually 23 hours

The buffer was circulated through a cooling system in order to maintain a constant temperature of 10°C. Bacteriophage λ concatemers were used as size markers, which ranged upwards from ~50kb in multiples of this value. Concatemers were prepared by heat inducing the lysogen N1323 (cI 857ts, Sam 7, lop 8), concentrating and lysing the host cells and then incorporating the phage into agarose plugs. Following the release of DNA by detergent and Proteinase K treatment, concatemers were formed by a prolonged incubation in 0.1M EDTA (pH8.0) at 50°C. The concatemers were made by Mr J. Maule.

2.9.4 Staining and Photography of Gels

The conventional and pulse field agarose gels, after separation of DNA fragments, were stained with a few drops of ethidium bromide (10mg/ml) for 20 minutes. After staining they were placed on a UV transilluminator and photographed using a Polaroid MP4 Land camera fitted with a red filter using Kodak Technical Pan film with an exposure time of 12 seconds. The film was developed in a RG II Fuji X-ray Film Processor.

2.9.5 Polyacrylamide Gel Electrophoresis

The DNA components of sequencing reactions were separated by polyacrylamide gel electrophoresis as described below.

Preparation of sequencing plates

The two glass plates used were cleaned with both detergent and ethanol to remove any debris. One plate was coated with the mixture, 10mls ethanol, 30µl MPTS (g-methacryloxypropyltrimethoxysilane) which adheres the gel to the plate. The second plate was treated with Dichlorodimethylsilane which prevents adhesion. Both plates were coated twice with their respective reagents and allowed to air dry between each step. After treatment, both plates were washed twice with ethanol and air dried. Sequencing spacer strips and an appropriate comb were washed in ethanol. The plates were assembled using cellotape and Bulldog-clips.

Preparation of polyacrylamide gel

A 200ml stock of acrylamide gel solution was made from 84g Urea, 40ml 5x TBE, 30mls 40% acrylamide stock (for 500mls 190g of acrylamide were added to 10g of bis-acrylamide, brought to 500ml with distilled water and stirred with 25g of Amberlite for 30 minutes to deionise) and brought up to 200ml with distilled water.

For a 40ml acrylamide gel solution, 40µl of TEMED (N,N,N'N'-Tetramethylethylenediamine) and 240µl of freshly made 10% ammonium persulphate were added. The solution was then gently mixed and poured with the aid of a glass 25ml pipette between the two glass plates.

Electrophoresis conditions

The sequencing gels were run in 1xTBE at 26 watts, 24 milliamps, 1100 volts for 1.5-4 hours depending upon the sequence information required. Approximately 2µl of each sequencing reaction was loaded per lane. Metal plates were clamped to the front of the plates to maintain an even temperature across the whole gel. Failure to do this resulted in an increase in the rate of migration of those samples within the centre of the gel relative to those located on either side.

After the run, the plates were prised apart and the gel fixed in 10% methanol and 10% acetic acid for 10 minutes, washed under running water for 10 minutes and dried at $80^{o}C$ for 30 minutes. Once cool, the gels were exposed at room temperature overnight under Kodak XAR-5 film.

2.9.6 Preparative Agarose Gel Electrophoresis

This method was used to isolate specific DNA fragments from restriction digests yielding multiple fragments. The resultant DNA was then used for cloning into vectors, radiolabelling or further restriction enzyme digestion as required. The DNA fragments were separated by electrophoresis (as described) in 1% low melting point agarose (BRL Ultra-pure). The gel was stained with ethidium bromide and the required DNA fragment cut out under UV light. The DNA was isloated from the gel slice following the method described by Burmeister and Lerach (1989). To the gel slice equal volume of low salt buffer (0.2M NaCl, 20mM Tris-HCl (pH7.4), 1mMEDTA) was added, mixed and heated at 65°C for 1 hour; this temperature melts the low melting point agarose. Once molten the mixture was equilibrated at 37°C and agarase (Calbiochem) added to 2u/100µl gel. After overnight incubation at 37°C to allow agarose digestion, the DNA was extracted with phenol (1vol.)(ultra- pure, BRL), phenol (0.5vol.) / chloroform (0.5vol.) and chloroform (1vol.), precipitated and resuspended in TE.

Preparative gel electrophoresis was also used for the isolation of a 140kb DNA fragment from a pulse field low melting point gel. The fragment of DNA was left in the agarose plug and subjected to digestion by a second restriction enzyme. The plug was equilibrated in a 10x volume of the buffer recommended for the second restriction enzyme. For the digestion reaction:

Plug + DNA	100µl
Enzyme buffer	From 10x stock
Spermidine	1mM
Triton x100	0.1%
BSA	100µg/ml
β-mercaptoethanol	0.5µl from stock
Enzyme	20 units

The reaction was incubated at 37°C overnight. Next day, the buffer was removed, the plug rinsed in ice cold TE and 100µl of stop buffer (10mM EDTA, 0.5x TAE, and Orange G as required) added. The DNA fragments within the plug were then separated by conventional gel electrophoresis (as described previously, section 2.9.2).

2.10 Alkali Vacuum Blotting

DNA fragments, separated by conventional and pulse field gel electrophoresis, were transferred to Hybond-N nylon membranes using the Vacugene vacuum transfer apparatus (LKB). The method used for transfer was modified from that described by Olszewska and Jones (1988). Vacuum blotting allows rapid transfer of nucleic acids and improved resolution of fragments upon radioactive detection compared to conventional Southern blotting

described by Southern (1975). Transfer of the DNA fragments from the gel was performed using a modification of the alkali method described by Reed and Mann (1985). The use of an alkali transfer solution allows an improvement in the resolution and a 10-fold increase in sensitivity of subsequent hybridisation analyses.

2.10.1 Alkali Transfer Blotting Protocol

The destained agarose gel is placed within the vacublot apparatus as decribed by the manufacturers. A vacuum of 40-60cm.H_2O is set and maintained throughout the blotting procedure. For a conventional gel, 25-30ml of depurination solution is pipetted onto the surface of the gel and left for 4 minutes. This is then removed and replaced by 25-30ml of denaturing solution. After 3.5 minutes this is removed and replaced by 250mls of alkali transfer solution. After 60 minutes, the buffer is removed together with the gel and the Hybond membrane washed in 2x SSC and blotted dry. For transfer of DNA molecules from pulse field gels the time for each step in the blotting procedure was doubled. The transferred DNA was covalently bound to the membrane using a UV Stratalinker 1800 from Stratagene at the recommended values. The membrane was then baked at 80^{o}C for two hours to ensure binding of the transferred DNA.

Depurination solution : 0.25M HCl

Denaturation solution : 1.5M NaCl, 0.5M NaOH.

Alkali transfer buffer : 0.25M NaOH, 1.5M NaCl.

2.11 Radiolabelling Techniques for the Production of DNA Probes

2.11.1 Random Prime Labelling

The reaction for random prime labelling was taken from the method described by Feinberg and Vogelstein (1983 and 1984). Random hexanucleotides are allowed to anneal to the denatured probe DNA and can then act as primers for the synthesis of the complementary strand by the Klenow fragment of E.coli DNA polymerase I from the 3'OH primer termini. Synthesis takes place in the presence of ^{32}P-labelled dCTP and unlabelled dNTPs. The method allows the DNA to be labelled to a high specific activity. This technique was used for DNA inserts that had been isolated from agarose gels.

The Boehringer Mannheim random prime labelling kit was used for all the reactions and the manufacturers' protocol followed as stated. The percentage incorporation of radioactive molecules into the DNA was assessed from the proportion of counts precipitated onto a Whatman GF/A filter by 5% TCA, which quantitatively precipitates oligonucleotides of >20 bases.

Proteins and unincorporated nucleotides were removed from radiolabelled probes to reduce background signal during hybridisation. This was achieved by passing the labelling reaction down a Sephadex G-50 (fine) column, DNA being excluded from the gel matrix (Maniatis *et al.*, 1989).

2.11.2 Nick translation of dsDNA

The protocol for Nick translation was as described by Sambrook, Fritsch and Maniatis (1989) and was used to radioactively label λ DNA. *E.coli* DNA polymerase I incorporates radioactive nucleotides (^{32}P α-CTP) into the 3' OH terminus of nicked DNA. The enzyme has a 5'-3' exonuclease activity which removes nucleotides from the 5' end as it moves, thereby producing labelled DNA with a high specific activity.

The BRL Nick Translation Kit was used for the reactions and the stated method and conditions followed as described. The unincorporated nucleotides were removed as described for random prime reactions.

2.11.3 End-Labelling Oligonucleotides

Oligonucleotides used for sequence detection in genomic DNA were end-labelled. This reaction is performed by the enzyme T4 polynucleotide kinase (PNK) which catalyses the transfer of the γ-phosphate of ATP to a 5'OH terminus of DNA (Sambrook, Fritsch and Maniatis, 1989).

The constituents for this reaction were 20-30ng of oligonucleotide, 1x PNK buffer, 50μCi γ-^{32}P ATP and 10 units of PNK in a total of 20μl. The reaction was incubated at 30°C for a maximum of 40 minutes and the incorporation of ^{32}P checked by mobility of the unincorporated counts in 0.3M ammonium formate through DEAE paper.

PNK buffer : (10x stock) 0.5M Tris-HCl (pH8.0), 0.1M $MgCl_2$, 50mM DTT, 1mM spermidine and 1 mM EDTA (pH8.0).

The oligonucleotides were synthesized by Mrs D. Chambers on an Applied Biosystems DNA Synthesizer 381A.

2.12 Nucleic Acid Hybridisation Techniques

All hybridisation reactions were performed in sealed plastic bags, under water and in shaking waterbaths. No more than two filters were placed in one bag and were arranged with the DNA containing side facing outwards. All filters were prehybridised in the relevant hybridisation solution and all hybridisations were performed for 16 hours.

2.12.1 Hybridisation Conditions for Random Primed and Nick Translated Probes

Pre- and hybridisation reactions were carried out in the solution described by Church and Gilbert (1984). Both stages of the reaction were carried out at 68^{o}C. After hybridisation, the filters were removed and washed 4 times at 68^{o}C for 10 minutes in 0.1-2x SSC (as required), 0.1% SDS and 0.1% sodium pyrophosphate. After washing the filters were blotted dry and sealed in Saran wrap.

2.12.2 Hybridisation Conditions for End-Labelled Oligonucleotides

Pre- and hybridisation reactions were carried out in 'Quick Oligo. Hyb. Mix'; 0.1% SDS, 0.1% sodium pyrophosphate, 0.05% BSA, 0.05% PVP, 0.05% Ficoll, 5x SSC. Both stages of the reaction were carried out at 48^{o}C for the oligonucleotides. Washing was performed 4 times for 5 minutes in 4xSSC, 0.1% SDS and 0.1% sodium pyrophosphate at 56^{o}C. After washing, the filters were blotted dry and sealed in Saran wrap.

2.12.3 Autoradiography

Autoradiography was carried out using X-ray film (Fuji Medical or Kodak X-OMAT) in cassettes which contained intensifying screens. Films were exposed for varying times, depending upon the intensity of hybridisation, at -70^{o}C. The X-ray films were developed using a RG II Fuji X-Ray Film Processor.

2.13 Dideoxy Sequencing of ds and ss DNA using "Sequenase"

2.13.1 Sequencing Protocol

DNA sequencing was performed following variations of Sanger's dideoxy chain termination sequencing method (Sanger *et al.*, 1977), using either ssDNA in M13 or dsDNA in pTZ. This method involves the synthesis of a DNA strand complementary to a single strand template from an oligonucleotide primer annealed to the said template. Synthesis is catalysed by a DNA polymerase. The reaction is terminated by the incorporation of a nucleotide analog, the 2',3'-dideoxynucleoside 5'-triphosphates (ddNTPs), which lack the 3'OH group necessary for chain synthesis. A radiolabelled nucleotide is also included in the synthesis to allow the visualisation of DNA chains after separation by polyacrylamide gel electrophoresis.

Sequenase, an altered form of T7 DNA polymerase (Tabor and Richardson, 1987), was used for chain synthesis. Its properties include high processivity and low 3' to 5' exonuclease activity. The method of sequencing described by the manufacturers of the enzyme (United States Biochemical Corporation) was used. This differs from the method described by Sanger *et al.* (1977) in two ways. Firstly, the primer is annealed to ssDNA and sequenase

adds nucleotides complementary to the template during which deoxynucleotides at limiting concentrations are used, resulting in random chain termination. Secondly, during the termination step, high concentrations of deoxynucleotides are used with one ddNTP, again allowing chain termination at random. Both alterations are designed to generate termination at every base so allowing the determination of the complete sequence. After sequencing the reactions are denatured before polyacrylamide gel separation.

All sequencing reagents were provided by the manufacturer and sequencing reactions were performed exactly as described. When double stranded sequencing was performed, the DNA was first dried and denatured for 10 minutes in 40µl of 200mM NaOH at room temperature. 4µl of 3M sodium acetate (pH5.5) and 88µl of ice cold ethanol was added and incubated at -70^{o}C for 15 minutes. The denatured DNA was centrifuged at 10,000g for 15 minutes, the pellet washed in 70% ethanol and air dried. The sequenase reaction was then followed using twice the recommended concentration of primer for the annealing reaction. All other steps were exactly as per manufacturers instructions.

2.13.2 Sequence analysis

Sequences were read by eye and homology searches carried out using the University of Wisconsin Genetics Computer Group's software package on the Daresbury computer searching the GenBank/EMBL data.

Chapter 3

Isolation of Internal Telomere-Like Repeats from Human and Mouse DNA

3.1 Introduction

Cloning of loci containing *Tetrahymena* $(TTGGGG)_n$ and *Trypanosoma* $(TTAGGG)_n$ telomere-like repeats from human and mouse DNA respectively, was pursued to determine their genomic locations and establish whether or not these repeats exist at interstitial sites. The isolation of two regions of the repeat containing loci was attempted, starting with telomere-like repeats from the two genomes. By using conventional methods of library construction any telomere-like repeats obtained will not be derived from the telomere. Such libraries require cohesiveness or clonability at both ends of the insert material for its insertion into a cosmid, bacteriophage or plasmid vector. Telomeres by definition do not present a clonable site (Blackburn and Gall, 1978 *etc*, Section 1.2.2(a)). Therefore, the isolation of such repeats from a conventional library would establish their existance internal to the telomere. Secondly, the isolation of the unique sequences flanking the telomere-like repeats was attempted. If these repeats are common to all of the loci, their presence within clones derived from different regions of the genome will mask any unique sequence within that clone. Hence, isolation of unique flanking sequences bordering the repeats may provide markers for individual loci, allowing determination of their position within the human and mouse genomes.

3.2 Screening of Pre-Existing Human and Mouse Total Genomic DNA Libraries

Human libraries constructed in bacteriophage and cosmids were screened with the *Tetrahymena*-like telomere $(TTGGGG)_4$ oligonucleotide probe, while a mouse cosmid library was screened with the *Trypanosoma*-like telomere $(TTAGGG)_4$ oligonucleotide probe. Although no positive clones were obtained from the mouse library, the same was not true with the human libraries. Analysis of the human clones from the bacteriophage library was hampered by problems with propagation and DNA isolation. Telomere-repeat containing clones from the cosmid library contained rearrangements within the unique flanking sequences. This was demonstrated by hybridisation of unique sequences from the cosmid library derived clones to total DNA used in the construction of the library. No common hybridising restriction fragments between the cosmid and total DNA were detected. Hence, screening pre-existing libraries for telomere repeat containing loci was abandoned.

3.3 Construction of a Human Size Fractionated Library for the Isolation of Telomere Repeat Containing Loci

3.3.1 Cloning Strategy

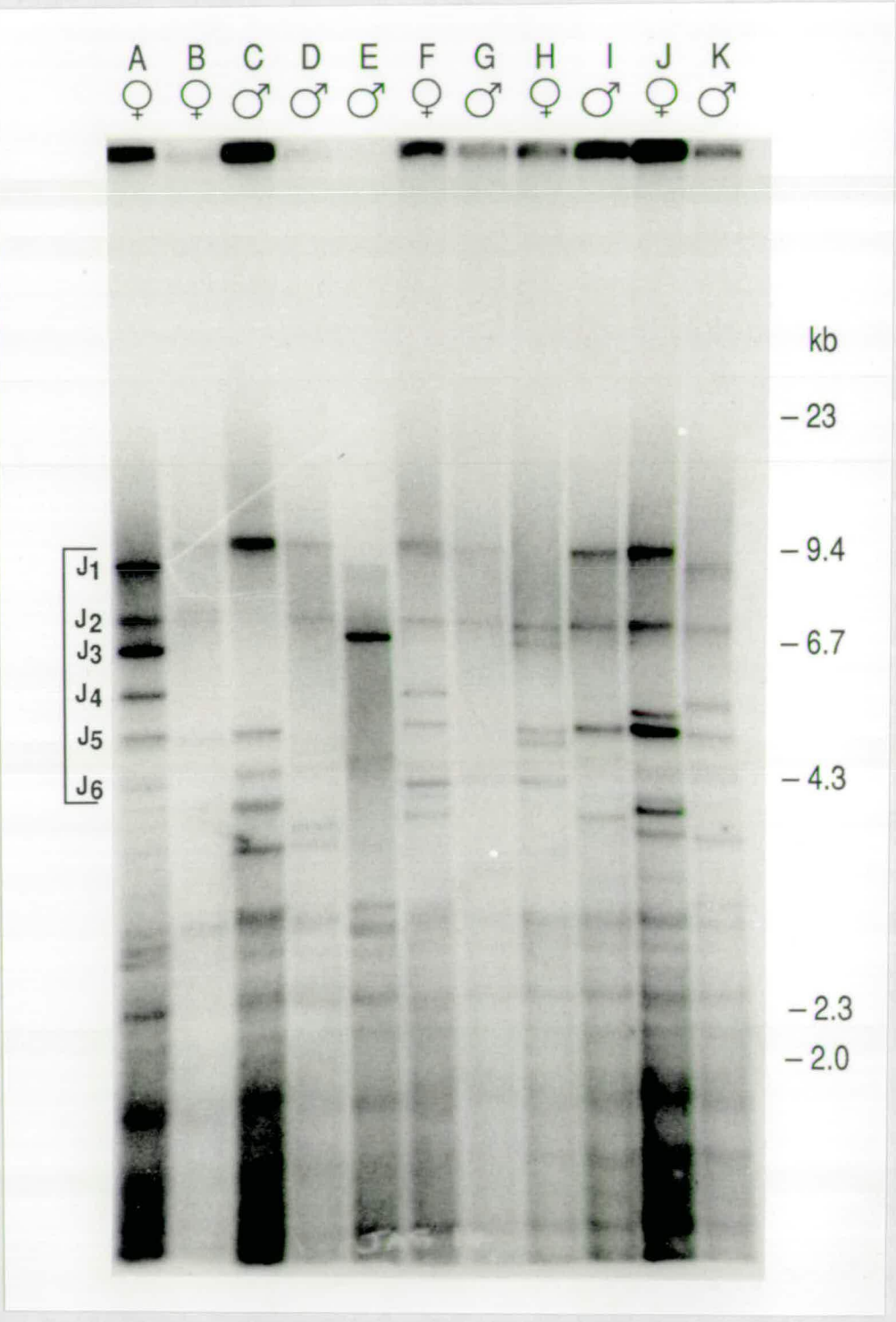
Analysis of cloned DNA from the cosmid library revealed rearrangements in the unique sequences flanking the *Tetrahymena*-like telomere repeats. Cosmid vectors have been constructed to contain the 5kb of DNA with the bacteriophage *cos* site and related sequences for packaging into a phage coat. This process has a size selection system, in which the DNA taken up into the phage head must be between 37-52kb of which 5kb is provided by the bacteriophage DNA plus the essential cos site. Hence, inserts of 32-47kb must be taken up by the vector to allow packaging into the phage head and propagation within a bacterial host. The polymorphic bands seen by Allshire *et al.* (1988) within total human DNA varied in size from ~2-9kb when the genomic DNA had been digested with a 4bp recognition site restriction enzyme. These enzymes cut frequently within the total genomic DNA but do not recognise the $(TTGGGG)_n$ repeat sequence. Therefore, the bands observed will contain mostly $(TTGGGG)_n$ or a related repeat with a small amount of flanking sequence. Cosmid cloning requires a DNA insert of between 32-47kb, so fragments containing $(TTGGGG)_n$ repeats would also be expected to possess a substantial region of flanking sequence available for rearrangement. To circumvent this problem an alternative strategy using plasmid cloning was employed. A library was constructed using frequently cutting restriction enzymes together with size selection for the region of total genomic DNA containing the human, $(TTGGGG)_4$ hybridising, polymorphic fragments. This would be expected to enrich for clones containing the required loci and for contiguous flanking sequence.

3.3.2 Preparation of Human DNA for Cloning

Total genomic DNA obtained from the blood lymphocytes of several different individuals was digested with the restriction enzyme *Sau* 3A, separated in 0.8% agarose, transferred to a nylon membrane and hybridised with the $(TTGGGG)_4$ oligonucleotide probe. Figure 3.1 shows the pattern of hybridisation in different individuals; as can be seen each individual has its own set of bands. Individual A was chosen for isolation of size fractions with $(TTGGGG)_n$ containing fragments. Hence, 500 µg of total genomic DNA was digested to completion with *Sau* 3A and separated by gel electrophoresis in a 0.8% low melting point agarose gel. The gel was stained with ethidium bromide to visualise the separated DNA fragments, and gel slices, approximately corresponding to the bands indicated in Figure 3.1, were removed. The DNA was recovered by digesting the agarose with the enzyme Agarase, and further purified

Figure 3.1 Polymorphic Tetrahymena-like Telomere Repeat Containing Loci within Human Total Genomic DNA from Different Individuals

Total genomic DNA from the blood lymphocytes of individuals labelled A to K was digested with *Sau* 3A, separated by conventional electrophoresis, transferred to a nylon membrane and hybridised with the $(TTGGGG)_4$ oligonucleotide probe. Hybridising fragments, labelled J1-6 in individual A, were used for construction of the size fractionated library.
Hind III digested λ DNA markers are indicated (kb).


A
B
C
D
E
F
G
H
I
J
K
kb
– 23
– 9.4
– 6.7
– 4.3
– 2.3
– 2.0
J1
J2
J3
J4
J5
J6

by phenol/chloroform extraction. The whole procedure was repeated for each fraction in an attempt to eliminate any contaminating smaller sized DNA fragments. Figure 3.2a shows an ethidium stained gel of the resultant fractions from which an estimate of the total yield was obtained (Figure 3.2a legend). This gel was blotted and hybridised with the $(TTGGGG)_4$ probe to determine the presence of this repeat within the fraction. As shown in Figure 3.2b all of the fractions gave a signal upon hybridisation indicating that each of agarose fractions contained at least one of the previously identified fragments.

3.3.3 Preparation of the Vector for Library Construction

The 2.9kb plasmid, pTZ18/19, as described in Section 2.5.3, was used to construct the size fractionated library. It was digested with *Bam* HI within the polylinker to generate a 4bp 3' cohesive single strand end compatible with the ends generated by *Sau* 3A in the size fractionated DNA. One problem associated with generating identical ends is that the plasmid is prone to recircularisation upon ligation. To reduce the frequency of this event the linearised plasmid was treated with *E.coli* alkaline phosphatase (BAP) which removes the 5' phosphate groups from free DNA ends. As DNA ligase requires both a 3'OH and 5' phosphate group to catalyse the ligation of two DNA molecules, recircularistion should be reduced or prevented. Test ligations were performed to determine the efficiency of this reaction. After subjecting the plasmid to dephosphorylation and religation followed by transformation into DH5α bacteria, no ampicillin resistant colonies were observed. In contrast, a high level of transformation was obtained using the nonphosphatased, religated plasmid.

3.3.4 Vector to Human Size Fractionated DNA Ligation Reactions

Test ligations were performed to determine the maximum conditions for vector/insert ligation. Hence, different ratios of plasmid to insert molecules were ligated and transformed by electroporation into DH5α bacteria and the number of white colonies counted. The best ratio was 3.5 insert molecules to 1 pTZ18/19 molecule. Ligation reactions for each of the size fractions were set up according to the determined ratio and transformed into DH5α bacteria. A total of ~10,000 ampicillin resistant colonies were obtained from all six size fractions ligated into pTZ18/19 and transformed into DH5α bacteria.

3.3.5 Isolation of $(TTGGGG)_n$ Containing Recombinant Plasmids from the Library

The library was plated out and screened with the $(TTGGGG)_4$ oligonucleotide probe. 17 positive clones were obtained from all 6 size fractions. These colonies were picked, propagated and recombinant DNA isolated using standard plasmid preparations. The DNA

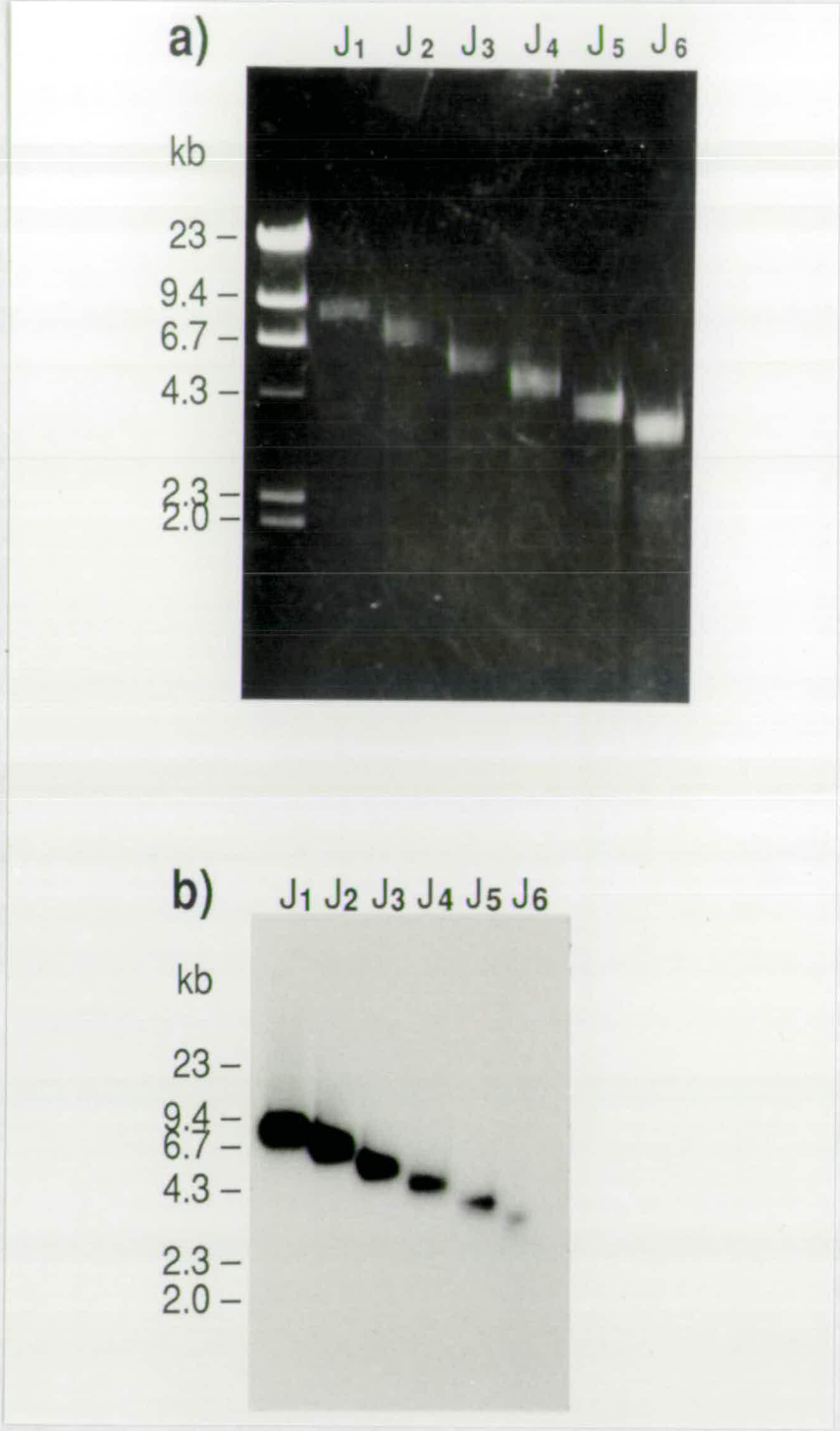
Figure 3.2 Size Fractionated DNA for Library Construction

Human, total genomic DNA from individual A (Figure 3.1) was digested with *Sau* 3A and the fractions J1-6 isolated after two rounds of electrophoretic separation in 0.8% low melting point agarose and Agarase purification.

(a) Ethidium stained 1% agarose gel showing a portion of the DNA isolated from the size fractions J1-6.

(b) The DNA within the gel shown in (a) was transferred to a nylon membrane and hybridised with the $(TTGGGG)_4$ oligonucleotide probe. Each of the J1-6 fractions contained material that crosshybridised with the *Tetrahymena*-like telomere repeats.

Hind III digested λ DNA size markers are indicated (kb).


a)
J1 J2 J3 J4 J5 J6
kb
23 –
9.4 –
6.7 –
4.3 –
2.3 –
2.0 –
b)
J1 J2 J3 J4 J5 J6
kb
23 –
9.4 –
6.7 –
4.3 –
2.3 –
2.0 –

inserted into the plasmid was isolated as an intact fragment using the restriction enzymes, *Bam* HI and *Xba* I. Other enzyme sites within the polylinker cloning site appeared to have been lost during the cloning process, such as the *Eco* RI and *Hind* III recognition sites; all polylinker enzyme sites had been shown to be present within the plasmid prior to library construction. Figure 3.3a shows an ethidium bromide stained gel of digests of the 4 clones obtained from the J6, ~3kb, size fraction. Hybridisation with the $(TTGGGG)_4$ oligonucleotide probe to the digested plasmid clones further demonstrated the presence of *Tetrahymena*-like telomere repeats within the inserted DNA fragments, Figure 3.3b. The insert size obtained for all clones was in the range of ~300-500bp regardless of their original, expected length. This observation together with the loss of restriction sites flanking the cloning site suggests that DNA fragments from the $(TTGGGG)_n$ containing loci are unstable when transformed into the DH5α bacteria.

3.3.6 Hybridisation of Cloned DNA Fragments to Human Total Genomic DNA

Inserts from the isolated clones were purified from low melting point agarose, random primed and hybridised back to total genomic DNA digested with *Sau* 3A. The DNA samples were derived from the original individual's DNA, used for construction of the library, and two other human sources. This should establish whether the clones obtained were derived from the DNA fragments originally visualised in human DNA with the $(TTGGGG)_4$ oligonucleotide probe. The panel of DNA samples was initially hybridised with the $(TTGGGG)_4$ oligonucleotide probe, Figure 3.4a. This was used as a reference for comparison of the insert hybridisation patterns. Hybridisation of the inserts to the panel, when the washing conditions were relatively non-stringent at 2xSSC and 65°C, revealed the same pattern of bands obtained with the $(TTGGGG)_4$ probe, Figure 3.4b. When the hybridisation was repeated and washed at a much higher stringency of 0.1xSSC, the characteristic $(TTGGGG)_4$ pattern was lost leaving a heterogenous smear, Figure 3.4c. Hence, from these results it would appear that firstly, any flanking sequence adjacent to the repeat in these clones was not able to detect the locus from which it was derived. Secondly, the cloned fragment contained $(TTGGGG)_n$ or related repeats that were able to crosshybridise with the $(TTGGGG)_4$ oligonucleotide probe. Thirdly, the heterogenous smear seen upon stringent washing of the hybridised DNA suggests that some other form of repeat is present within the cloned fragments.

3.3.7 Sequence Analysis of Human Size Fractionated Clones

The clones were sequenced from both the universal and reverse primers to establish the type of repeat present and the composition of the flanking DNA. By double-strand DNA sequencing the clones were shown to contain a region of unique sequence together with

Figure 3.3 Isolation of DNA Inserts from (TTGGGG)$_4$ Hybridising, Size Fractionated Library Derived Clones

Plasmid DNA preparations of 4 library clones, J$_6$1, J$_6$4, J$_6$6 and J$_6$7, from the J6 size fraction of individual A (Figure 3.1) were digested with *Bam* HI and *Xba* I to release the inserted DNA fragment.

Contents of (a) and (b) :

Lane	Contents
1	*Hind* III digested λ DNA size marker
2	J$_6$1 plasmid
3	J$_6$1 plasmid digested with *Bam* HI and *Xba* I
4	J$_6$4 plasmid
5	J$_6$4 plasmid digested with *Bam* HI and *Xba* I
6	J$_6$6 plasmid
7	J$_6$6 plasmid digested with *Bam* HI and *Xba* I
8	J$_6$7 plasmid
9	J$_6$7 plasmid digested with *Bam* HI and *Xba* I
10	*Hae* III digested Phi X DNA size marker

(a) Ethidium stained 1% agarose gel containing digests of recombinant plasmids from J6 size fraction of library.

(b) The DNA from the gel shown in (a) was transferred to a nylon membrane and hybridised with the (TTGGGG)$_4$ oligonucleotide probe. Hybridisation to the uncut plasmid and to the ~300bp *Bam* HI / *Xba* I released insert from each clone was detected.

Hind III digested λ DNA and *Hae* III Phi X DNA size markers are indicated (kb).

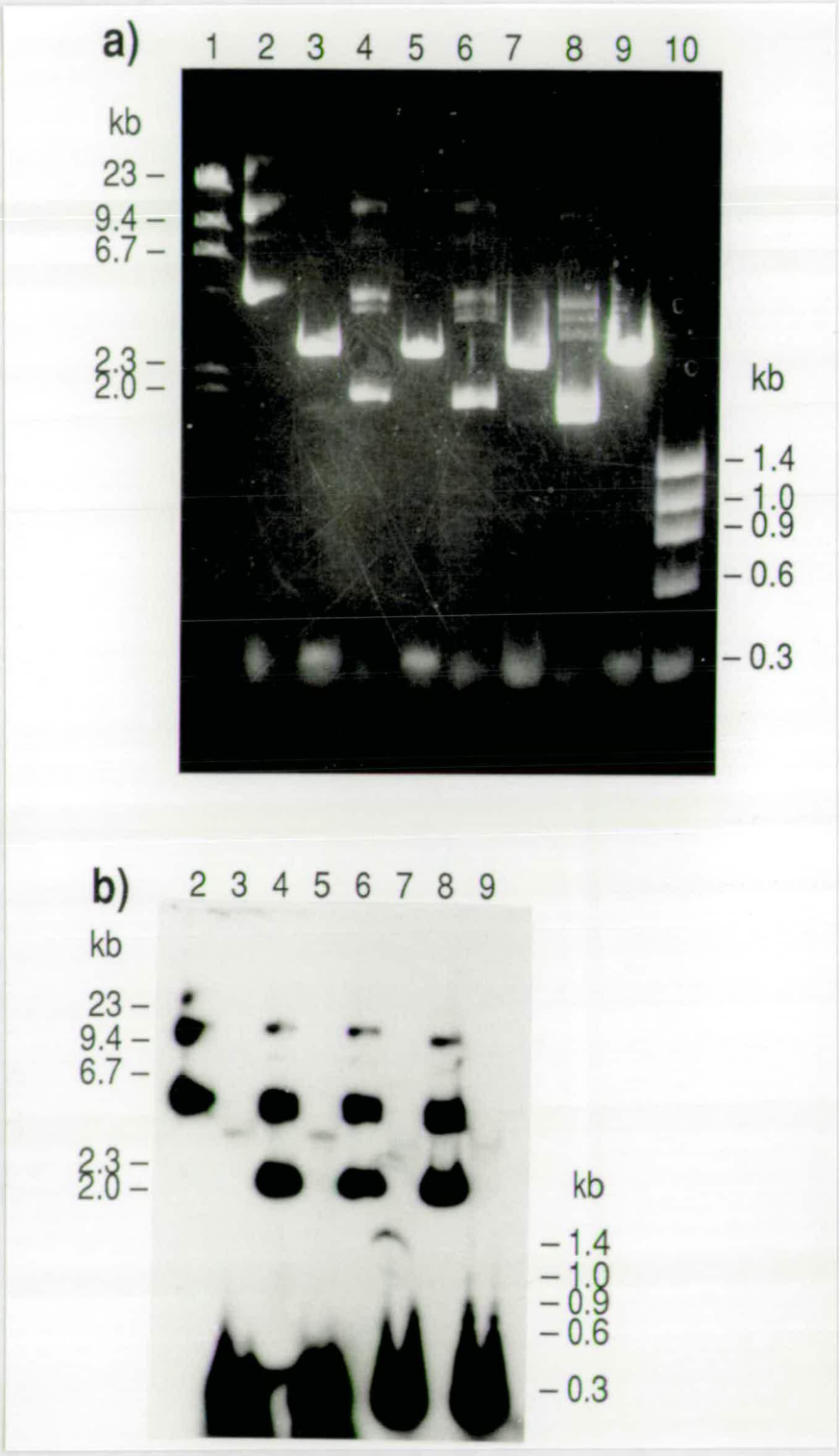

a)
1 2 3 4 5 6 7 8 9 10
kb
23 –
9.4 –
6.7 –
2.3 –
2.0 –
kb
– 1.4
– 1.0
– 0.9
– 0.6
– 0.3
b)
2 3 4 5 6 7 8 9
kb
23 –
9.4 –
6.7 –
2.3 –
2.0 –
kb
– 1.4
– 1.0
– 0.9
– 0.6
– 0.3

Figure 3.4 Hybridisation Pattern of Clone $J_{6}6$ within Total Genomic DNA

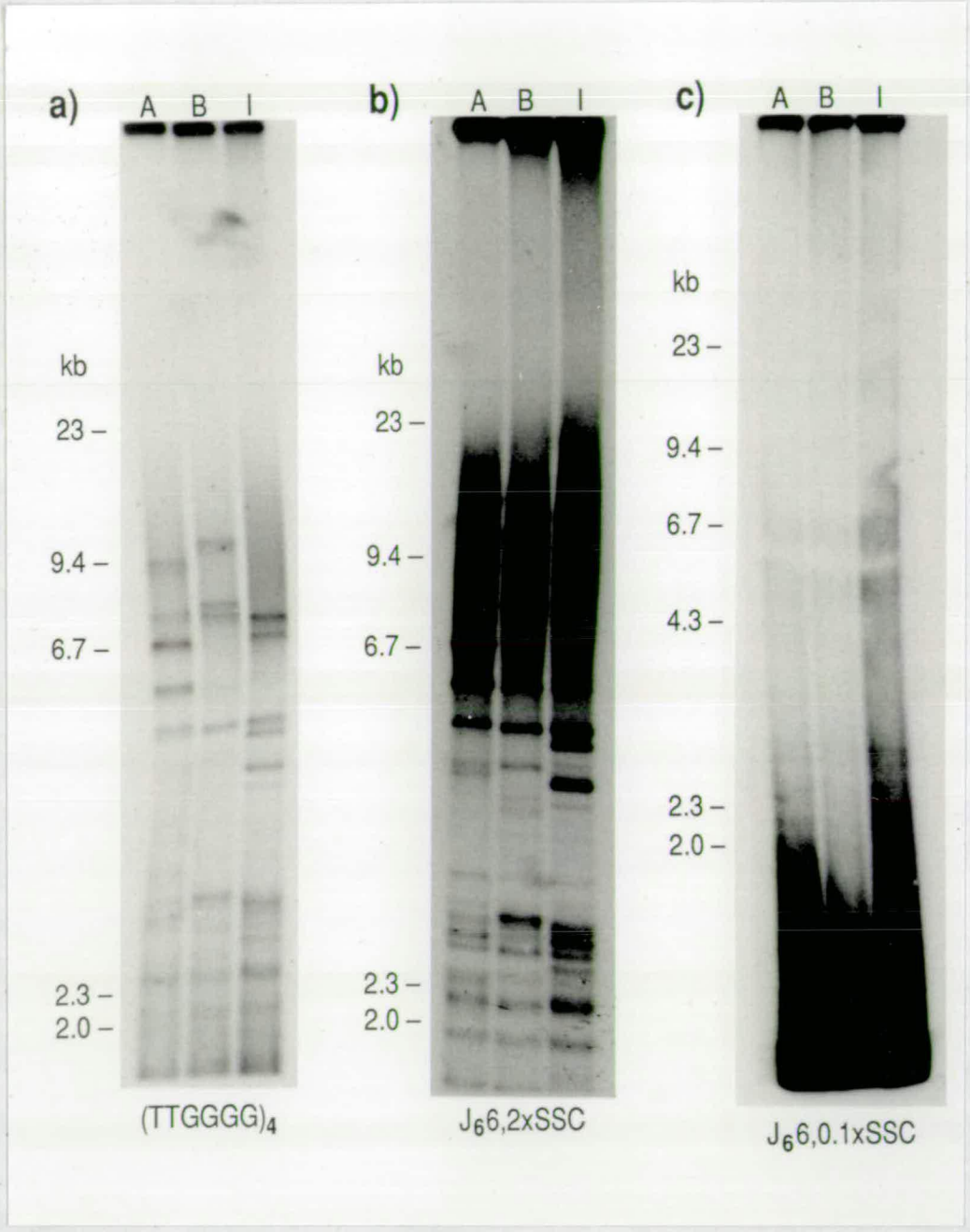
Total genomic DNA from individuals A, B and I (Figure 3.1) was digested with *Hae* III, separated by electrophoresis, and transferred to a nylon membrane. (a), (b) and (c) represent different separations of DNA fragments from the three individuals.

(a) Hybridisation of the $(TTGGGG)_4$ oligonucleotide probe to *Hae* III digested DNA from A, B and I.

(b) Hybridisation of the purified, random primed insert from clone $J_{6}6$ (Figure 3.3) to *Hae* III digested DNA from A, B and I and washed at the relaxed stringency of 2xSSC.

(c) Hybridisation of the purified, random primed insert from clone $J_{6}6$ (Figure 3.3) to *Hae* III digested DNA from A, B and I and washed at 0.1xSSC high stringency.

Hind III digested λ DNA size markers are indicated (kb).


a)
A B I
kb
23 –
9.4 –
6.7 –
2.3 –
2.0 –
(TTGGGG)4
b)
A B I
kb
23 –
9.4 –
6.7 –
2.3 –
2.0 –
J66,2xSSC
c)
A B I
kb
23 –
9.4 –
6.7 –
4.3 –
2.3 –
2.0 –
J66,0.1xSSC

tandemly repeated copies of what appeared to be the hexamer CCCCAA, Figure 3.5. It was extremely difficult to determine the exact sequence of the repeats as their presence led to compression within the sequencing reaction and hence, the appearence of shadow bands within each nucleotide tract, Figure 3.5. Therefore, the inserts were subcloned into an M13 vector. This vector allows single-strand sequencing of DNA and should help to relieve the problem of compression. Unfortunately, the repeats were lost upon subcloning into the M13 vector and so their exact sequence could not be determined. However, the sequence of the flanking region was determined by this method and subsequently analysed through a database to look for regions of homology with other known sequences. Homology was found with the *Alu* family of repetitive DNA (data not shown).

3.4 Construction of a Mouse Size Fractionated Library for the Isolation of Telomere Repeat Containing Loci

The isolation of *Trypanosoma*-like telomere repeats from mouse genomic DNA was also attempted by the technique outlined for human loci. Although, the library construction was successful in that several thousand recombinant clones were obtained upon transformation of DH5α bacteria, no positive colonies were obtained upon screening with the $(TTAGGG)_4$ oligonucleotide probe. The apparent absence of $(TTAGGG)_n$ containing DNA fragments from both a cosmid and a size fractionated library may reflect a complete lack of stability of these sequences upon isolation from the mouse genome.

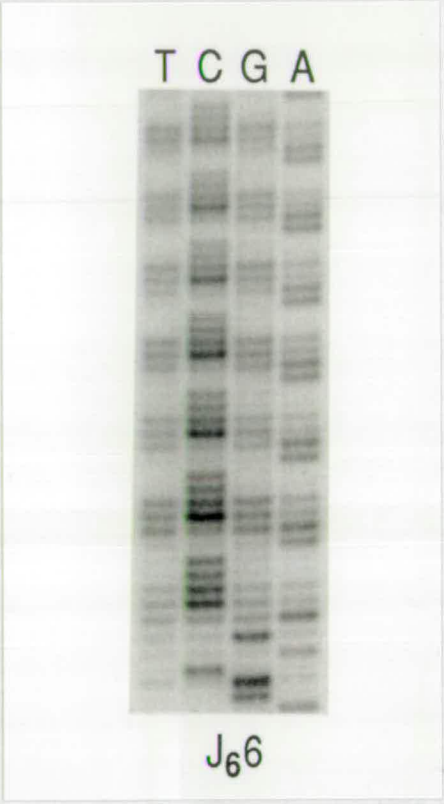
3.5 Summary

A number of strategies have been pursued in an attempt to isolate internal, telomere repeat containing loci from both human and mouse genomic DNA. Although no *Trypanosoma*-like telomere repeat containing clones were isolated from the mouse libraries, the same was not true with the human libraries. In all cases, clones hybridising to the $(TTGGGG)_4$ oligonucleotide probe were obtained and by sequencing it was tentatively shown that these repeats exist at interstitial sites within the human genome. Localisation of the clones within the human genome proved to be impossible due to apparent rearrangements within flanking sequences and general instability of the DNA fragments upon propagation in bacterial systems.

Several observations from attempts to clone $(TTGGGG)_n$ containing fragments have demonstrated their instability within bacterial cloning systems. For example, the insert from the J6 fraction showed a reduction in size from 3kb to 300-500bp. This may have been due to ligation of contaminating fragments of DNA of a smaller size within the fraction. However,

Figure 3.5 Partial Sequence of Clone $J_6 6$

Double strand sequencing of clone $J_6 6$ from the reverse sequencing primer. Lanes T, C, G and A correspond to termination of the sequencing reaction at thymine, cytosine, guanine and adenine deoxynucleotides respectively.


T C G A
J6 6

after a second purification step one would expect such contamination to be minimal. A second indicator of instability arises from the random loss of enzyme sites from the polylinker in pTZ18/19 vector after insertion of the DNA fractions and transformation into bacteria. The original cloning site has been maintained ruling out the possibility of DNA degradation by alkaline phosphatase or any other element involved in the ligation reaction. It does suggest that some form of rearrangement has occurred upon introduction of repeat containing DNA into the plasmid resulting in removal of repeat blocks and rearrangement of the flanking DNA. Many of the clones that were sequenced showed an abrupt transition from tandem repeats to plasmid DNA. This was unexpected as the restriction enzyme used to digest the human DNA, *ie Sau* 3A, cannot cut within the $(TTGGGG)_n$ repeats. Hence, it is possible that the *Tetrahymena* telomere-like repeats together with flanking sequences have been lost from the cloned fragments. Finally, subcloning of the originally isolated clones into the M13 vector led to complete loss of the repeat sequence, suggesting a similar mechanism may have occurred previously.

Instability of repeated sequences upon cloning into cosmids, bacteriophage or plasmid vectors and subsequent propagation in bacteria has been shown to be a fairly common occurrence. Isolation of DNA fragments containing a subset of minisatellites from both human and mouse DNA revealed loss of tandem repeats from the cloned loci (Wong *et al.*, 1986, Kelly *et al.*, 1989). For example, a mouse minisatellite of 7kb was reduced to 400bps upon cloning into the bacteriophage L47.1 (Kelly *et al.*, 1989). Hence, a different approach was pursued for the characterisation of interstitial telomere-repeat containing loci.

Chapter 4

Preliminary Investigation of $(TTAGGG)_n$ Trypanosoma-Like Telomere Repeats within the Mouse Genome

4.1 Introduction

The isolation and localisation of interstitial telomere repeat-like loci, by cloning procedures, from both human and mouse genomes, had proved to be extremely difficult (Chapter 3). Hence, an alternative approach was adopted to establish the genomic location of such loci. One $(TTGGGG)_n$ containing locus had previously been mapped within the C57Bl/6 x DBA/2 recombinant inbred strains which have been characterised by Dr. B.A. Taylor (Taylor, 1978). Linkage was found with the β_2-microglobulin gene on the mouse chromosome 6 (Allshire and Hastie, unpublished); this gene maps to chromosome 11 in the human genome and is situated close to a previously identified fragile site (Allshire and Hastie, unpublished). Hence, it appeared that telomere repeat containing loci could be assigned possible human genomic locations by mapping within mouse inbred strains. However, no additional $(TTGGGG)_n$ containing loci could be mapped by this method and so no further correlation between such loci and specific regions of the genome, such as fragile sites, could be established. Therefore, the possibility of mapping the *Trypanosoma*-like telomere repeat containing loci observed in the inbred mouse strain, DBA/2, (Allshire *et al.*, 1988) was investigated.

After the initial observation that the telomeric repeat from *Tetrahymena thermophilia* crosshybridised with human telomeres and interstitial loci (Allshire *et al.*, 1988), cloning and sequencing of the human telomere revealed that it was composed predominantly of $(TTAGGG)_n$ *Trypanosoma*-like telomere repeats (Moyzis *et al.*, 1989). It was then demonstrated, using a combination of restriction enzymes and oligonucleotides, that three types of repeat existed within the human telomere, *ie* $(TTGGGG)_n$, $(TTAGGG)_n$ and $(TGAGGG)_n$, where the most centromere distal region was composed of tandem $(TTAGGG)_n$ repeats with proximal clusters of $(TTGGGG)_n$ and $(TGAGGG)_n$ (Allshire *et al.*, 1989). The described organisation was later confirmed by sequencing of a cloned human telomere (Brown *et al.*, 1990). In conjunction with this study, the presence of $(TTGGGG)_n$, $(TTTAGGG)_n$ and $(TTAGGG)_n$ repeats was investigated within the genome of the DBA/2 inbred mouse strain (Allshire *et al.*, 1989). It was found that there was very little $(TTGGGG)_n$ and no $(TTTAGGG)_n$ repeated sequences present. However, the $(TTAGGG)_4$ *Trypanosoma*-like telomere probe detected a non-resolvable >23kb region and discrete bands ranging in size from ~2-15kb. It was possible that the discrete bands observed within the DBA/2 DNA might be comparable to the interstitial $(TTGGGG)_n$ repeat containing loci within the human genome (Allshire *et al.*, 1988) Hence, working upon the theory that telomere-like repeats at

interstitial sites within mouse and human genomes might possess similar functions, the characterisation and mapping of the mouse *Trypanosoma*-like repeat containing loci was pursued. In addition, the >23kb unresolved smear of DNA seen within DBA/2 DNA that hybridised with the $(TTAGGG)_4$ oligonucleotide probe (Allshire *et al.*, 1989) was analysed.

4.2 Trypanosoma-like Telomere Repeats within Genomic DNA from Different Mouse Species

The presence of $(TTAGGG)_n$ repeats within the genomes of different mouse strains and species was investigated to ensure that the pattern of hybridisation seen in the DBA/2 genome was not unique to the strain. Hence, total genomic spleen DNA from the inbred strains DBA/2, C57Bl/6 and AKR/J and for the species *Mus domesticus*, *Mus musculus*, *Mus caroli*, *Mus spretus* and *Apodemus sylvaticus* was digested with *Hae* III, the fragments separated on a conventional gel, transferred to a nylon membrane and hybridised with the $(TTAGGG)_4$ probe. In the DBA/2, C57Bl/6, AKR/J, *Mus domesticus*, *Mus musculus* and *Apodemus sylvaticus* mice, discrete bands which differed between each species in the size range 2-10kb and a non-resolvable >23kb region were detected, Figure 4.1a. In the *Mus caroli* DNA the variable bands were smaller in size and the unresolved region ranged from 20-30kb. In *Mus spretus,* the discrete bands ranged from 3-25kb and there was no non-resolvable region, instead a smear of hybridisation was detected between 4.3-6.7kb, Figure 4.1a.

The non-resolvable region was analysed by Pulsed-Field Gel Electrophoresis (PFGE) of *Hae* III digested, total genomic DNA from DBA/2, C57Bl/6, *Mus caroli*, Mus *domesticus*, AKR/J and *Apodemus sylvaticus* mouse strains and species. This technique resolved between 30-150kb, Figure 4.1b. In all, except the *Mus caroli* DNA, discrete bands were detected. In DBA/2 DNA there were several bands between 100-150kb. Fewer bands were detected for the C57Bl/6, *Mus domesticus*, AKR/J and *Apodemus sylvaticus* DNA and they tended to be smaller in size. With the exception of the *Mus caroli* DNA, the majority of $(TTAGGG)_4$ hybridising material fell within the 30-80kb size range. The *Mus caroli* DNA showed hybridisation within the 20-30kb range and this region appeared to be composed of a limited number of bands, Figure 4.1b. Increased resolution would be required to determine the exact composition of this region.

Figure 4.1 Trypanosoma-like Telomere Repeats within the Genomic DNA of Different Mouse Species

Total genomic DNA from several different mouse strains was digested with the restriction enzyme *Hae* III, separated by (a) conventional and (b) pulsed-field, gel electrophoresis, transferred to a nylon membrane and hybridised with the $(TTAGGG)_4$ oligonucleotide probe.

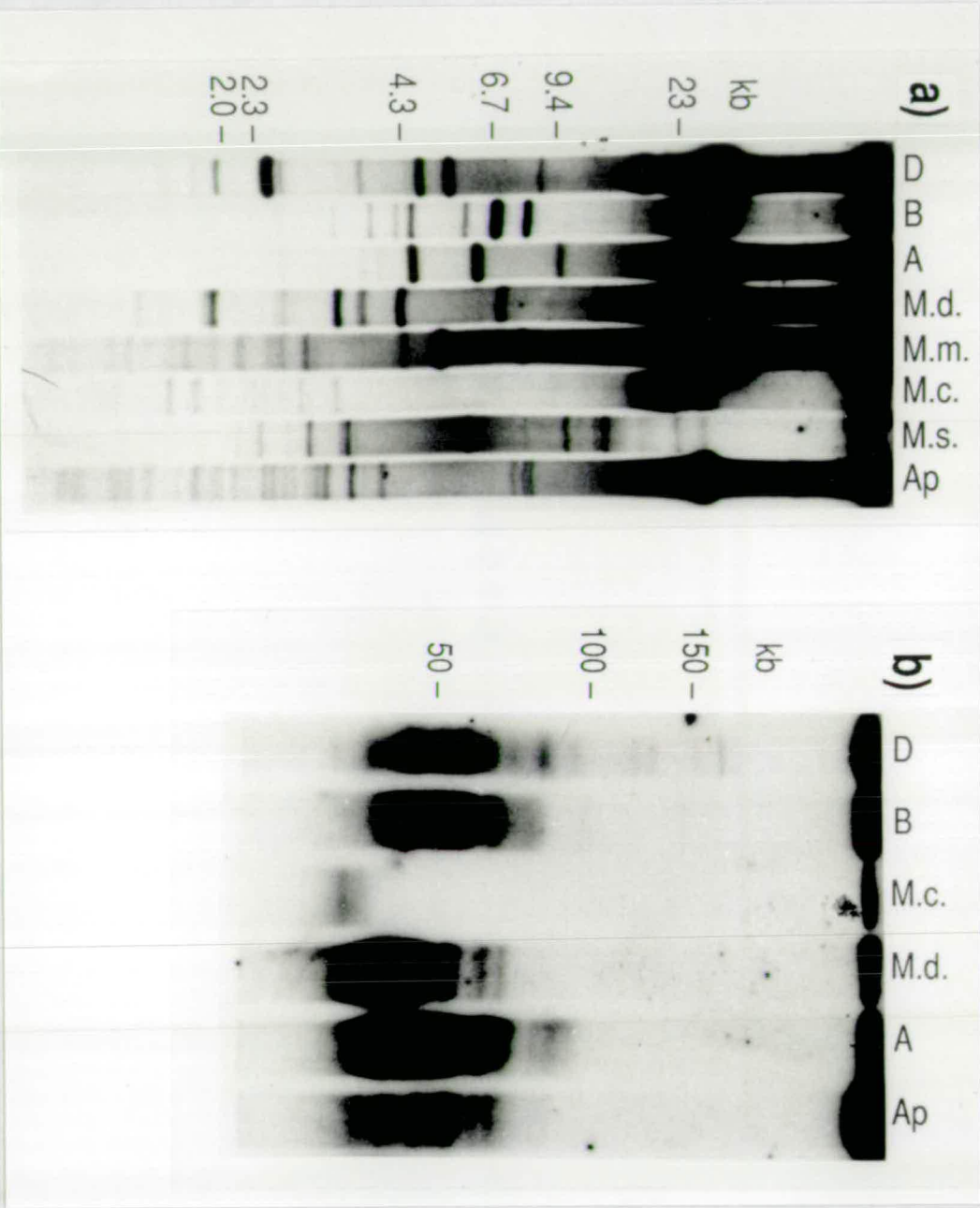
(a) Conventional gel electrophoresis separation of DNA fragments.

D = DBA/2
B = C57Bl/6
A = AKR/J
M.d. = *Mus domesticus*
M.m. = *Mus musculus*
M.c. = *Mus caroli*
M.s. = *Mus spretus*
Ap = *Apodemus sylvaticus*

(b) PFGE separation of DNA fragments.

D = DBA/2
B = C57Bl/6
M.c. = *Mus caroli*
M.d. = *Mus domesticus*
A = AKR/J
Ap = *Apodemus sylvaticus*

Hind III digested λ DNA and λ concatemer size markers are indicated (kb).


a)
kb
23 –
9.4 –
6.7 –
4.3 –
2.3 –
2.0 –
D
B
A
M.d.
M.m.
M.c.
M.s.
Ap
b)
kb
150 –
100 –
50 –
D
B
M.c.
M.d.
A
Ap

4.3 Position of $(TTAGGG)_n$ Repeats Relative to the Telomere within the Mouse Genome

The nuclease Bal 31 was used to establish the presence of $(TTGGGG)_n$ and $(TTAGGG)_n$ repeats within human telomeres (Allshire *et al.*, 1988 and 1989). This enzyme progressively removes deoxynucleotides from the ends of DNA molecules and hence, in total genomic DNA removes sequences from the telomeres. Hence, Bal 31 was used to establish whether the $(TTAGGG)_4$ hybridising material observed in the different mouse strains was located within the mouse telomere. Total genomic DNA from DBA/2, C57Bl/6 and *Mus caroli* mice was treated with Bal 31, digested with the *Hae* III restriction enzyme, the fragments separated by both pulse field and conventional gel electrophoresis, transferred to a nylon membrane and hybridised with the $(TTAGGG)_4$ oligonucleotide probe.

The DBA/2 DNA, upon separation by PFGE, showed a gradual decrease in size of the discrete 80-150kb bands. For example, a 100kb band decreased uniformly in size by ~46kb over a 1.5 hour timecourse. This was a characteristic feature of the whole 30-150kb region resolved by PFGE, Figure 4.2a. Hybridisation with a mouse 196 major satellite (196ms) probe (Pietras *et al.*, 1983), and supplied by A. Mitchell, was used as a control for general DNA degradation within the Bal 31 timecourse. The mouse major satellite maps to the centromere / long arm border of the acrocentric mouse chromosomes (Joseph *et al.*, 1989) and so is present at an interstitial site. Hence, loss of sequence from this region during Bal 31 digestion would indicate non-specific DNA degradation. A smear of hybridisation within the size range of 30-60 kb was observed in all the tracts of the timecourse suggesting that degradation was only minimal, Figure 4.2b. DBA/2 DNA treated in the same way but resolved by conventional gel electrophoresis showed loss of the >23kb region but the discrete bands from ~2-10kb remained intact, Figure 4.3a. The 196ms probe again demonstrated that the DNA had not been subject to general degradation, Figure 4.3b. The same features were seen with Bal 31 treatment of C57Bl/6 DNA (data not shown).

Bal 31 treatment of *Mus caroli* DNA gave a similar result. After one hour of digestion with the nuclease, the 20-30kb region had been reduced to a faint smear, leaving an intensely hybridising smear of DNA at ~6-8kb and the strain specific discrete bands of 1-4kb, Figures 4.4a and b. After 90minutes of Bal 31 nuclease digestion the ~6-8kb hybridising material was reduced to a faint smear of DNA within this size range. As the Bal 31 enzyme progressively removes bases from the end of a DNA molecule, a further decrease in size of the smear would have been expected. The loss of the telomeric smear at ~6-8kb suggests that some other form of simple repeat is present at the border between the telomeric repeats and the subtelomeric DNA.

Figure 4.2 Bal 31 Localisation of Trypanosoma-like Repeats within the DBA/2 Mouse Genome by PFGE

Total genomic DNA was subjected to digestion with Bal 31 for increasing lengths of time, then digested with *Hae* III, the fragments separated by PFGE, transferred to a nylon membrane and hybridised with (a) the $(TTAGGG)_4$ oligonucleotide probe and (b) the 196 mouse major satellite probe. The time, in minutes, for each timepoint is indicated above the relevent track and track C contains DNA which has not been subjected to Bal 31 digestion.

λ concatemer size markers are indicated (kb).

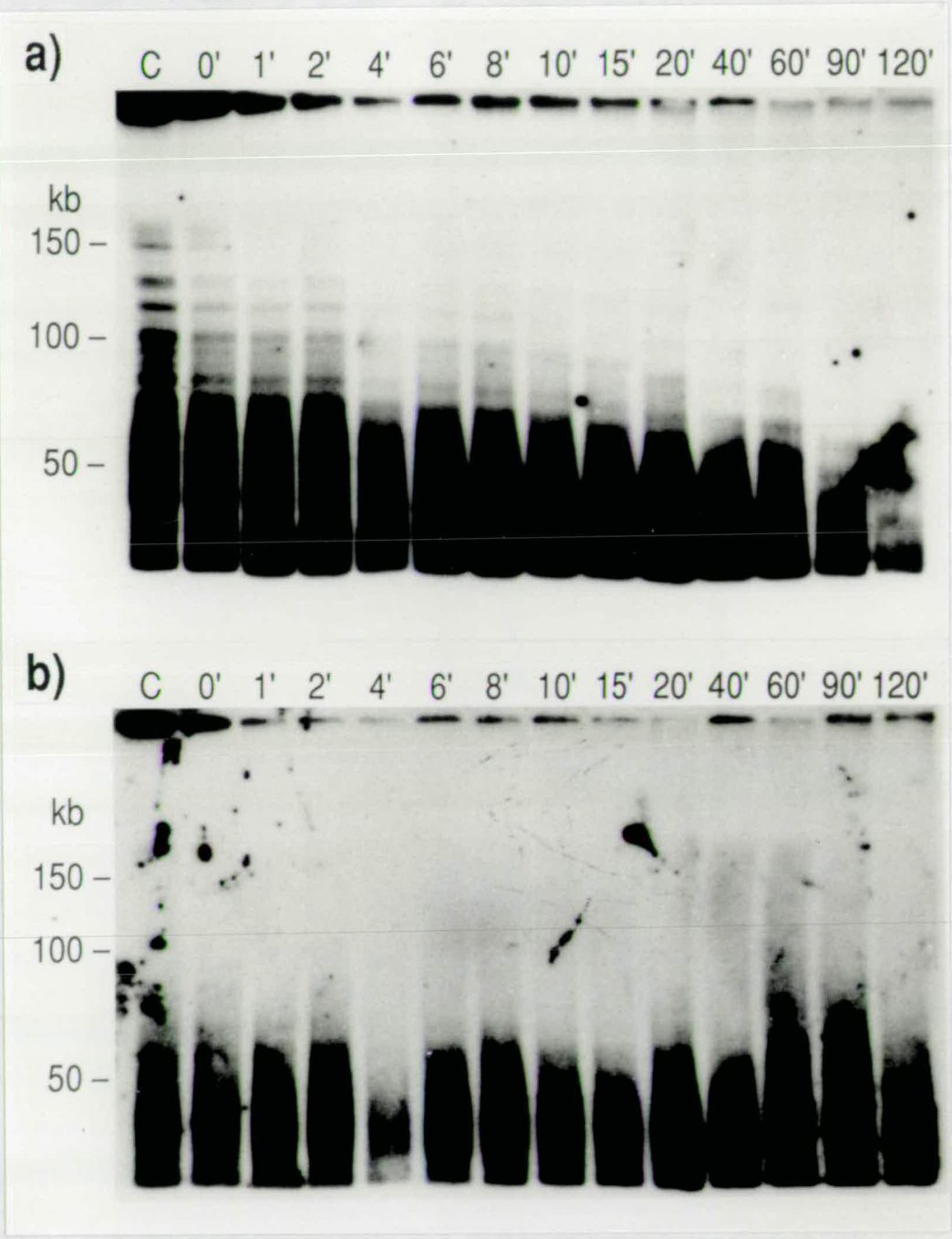

a)
C 0' 1' 2' 4' 6' 8' 10' 15' 20' 40' 60' 90' 120'
kb
150 –
100 –
50 –
b)
C 0' 1' 2' 4' 6' 8' 10' 15' 20' 40' 60' 90' 120'
kb
150 –
100 –
50 –

Figure 4.3 Bal 31 Localisation of Trypanosoma-like Repeats within the DBA/2 Mouse Genome by Conventional Gel Electrophoresis

Total genomic DNA was subjected to digestion with Bal 31 for increasing lengths of time, then digested with *Hae* III, the fragments separated by conventional gel electrophoresis, transferred to a nylon membrane and hybridised with (a) the $(TTAGGG)_4$ oligonucleotide probe and (b) the 196 mouse major satellite probe. The time, in minutes, for each timepoint is indicated above the relevent track and track C contains DNA which has not been subjected to Bal 31 digestion.

Hind III digested λ DNA size markers are indicated (kb).

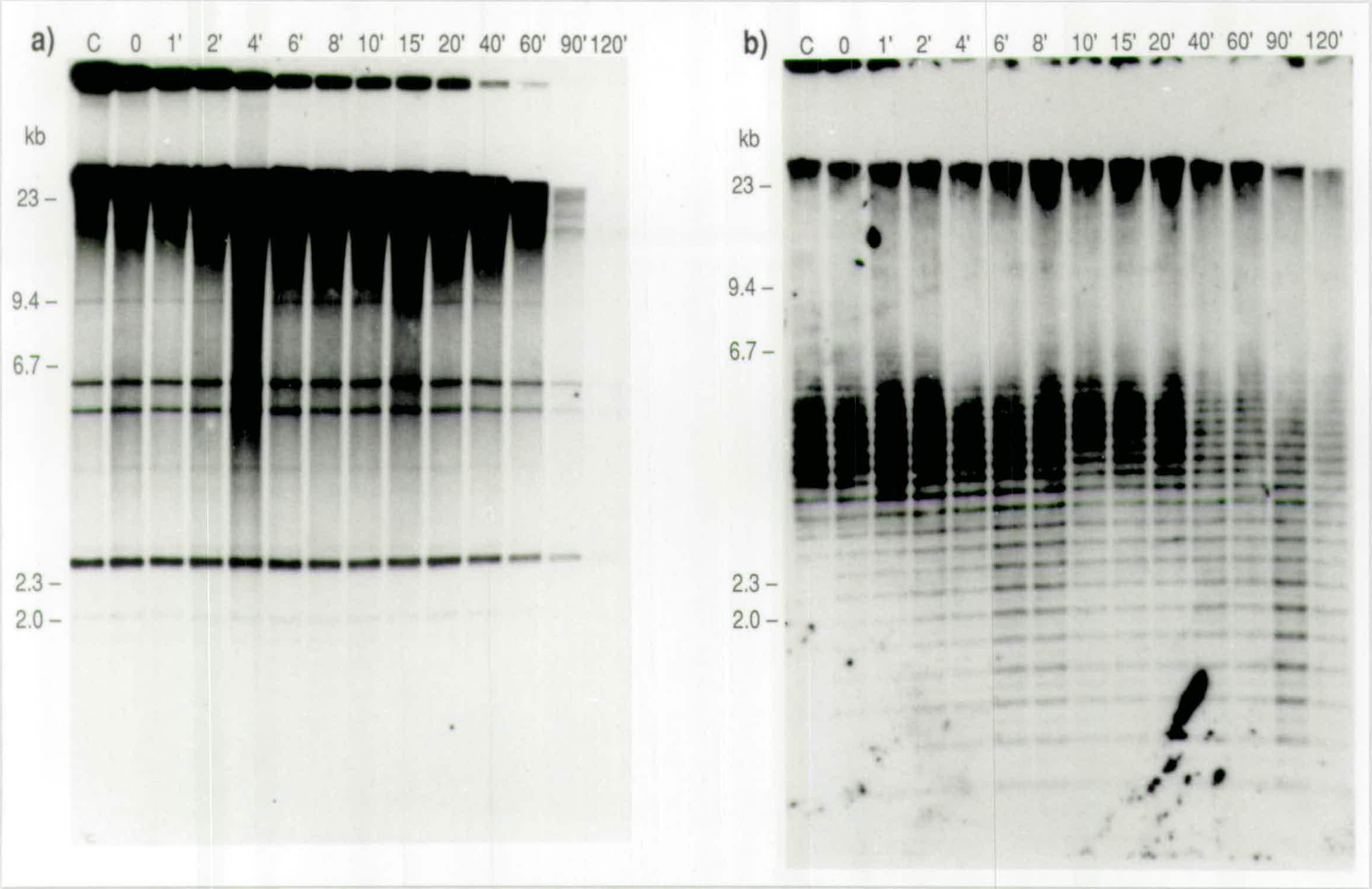
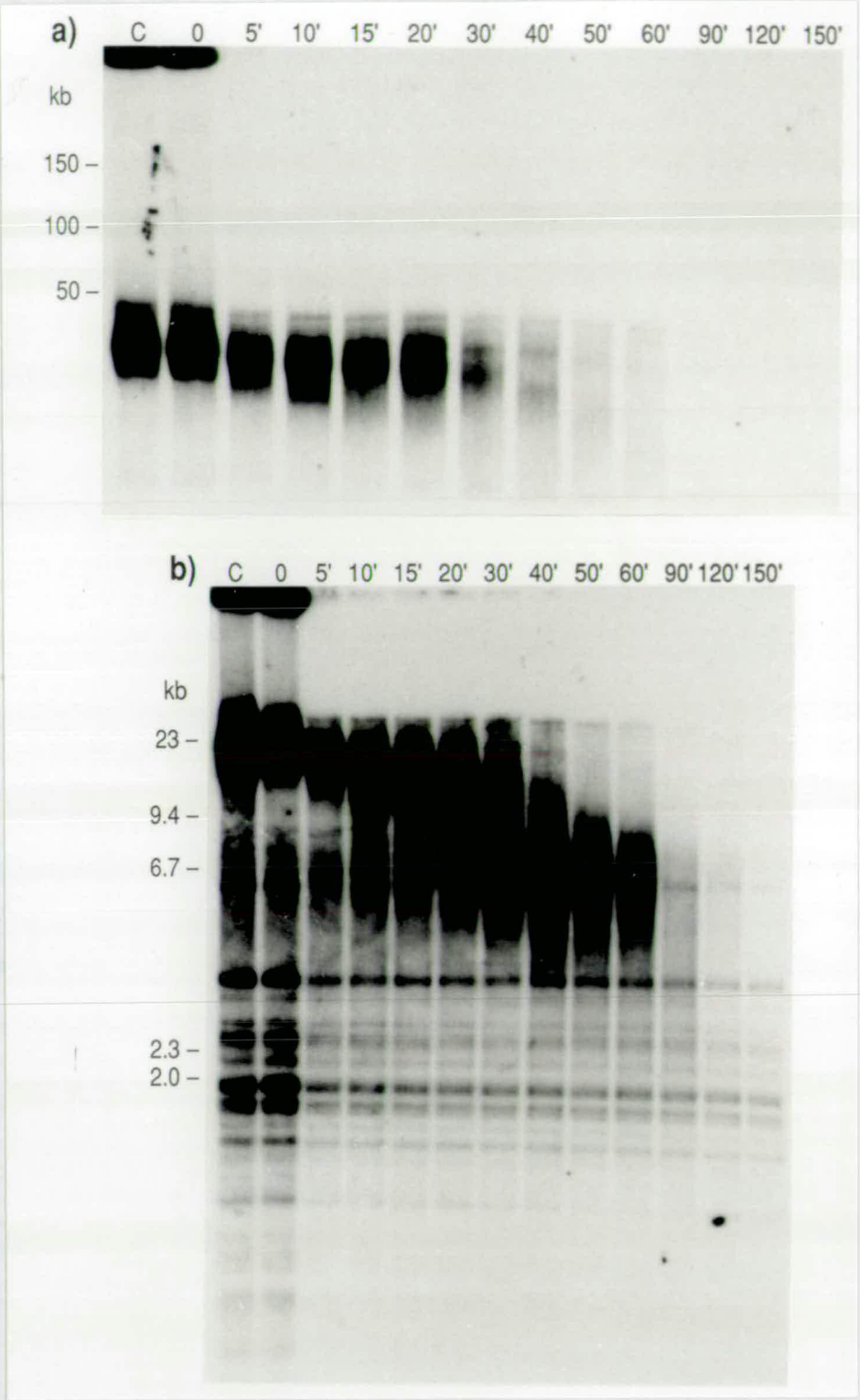

a)
C 0 1' 2' 4' 6' 8' 10' 15' 20' 40' 60' 90' 120'
kb
23 –
9.4 –
6.7 –
2.3 –
2.0 –
b)
C 0 1' 2' 4' 6' 8' 10' 15' 20' 40' 60' 90' 120'
kb
23 –
9.4 –
6.7 –
2.3 –
2.0 –

Figure 4.4 Bal 31 Localisation of Trypanosoma-like Repeats within the Mus caroli Mouse Genome

Total genomic DNA was subjected to digestion with Bal 31 for increasing lengths of time, then digested with *Hae* III, the fragments separated by (a) pulsed-field and (b) conventional gel electrophoresis, transferred to a nylon membrane and hybridised with the $(TTAGGG)_4$ oligonucleotide probe. The time, in minutes, for each timepoint is indicated above the relevent track and track C contains DNA which has not been subjected to Bal 31 digestion. *Hind* III digested λ DNA and λ concatemer size markers are indicated (kb).


a)
C 0 5' 10' 15' 20' 30' 40' 50' 60' 90' 120' 150'
kb
150 –
100 –
50 –
b)
C 0 5' 10' 15' 20' 30' 40' 50' 60' 90' 120' 150'
kb
23 –
9.4 –
6.7 –
2.3 –
2.0 –

The problem of general DNA degradation was further studied by conducting a Bal 31 timecourse upon the *Hind* III fragments of λ DNA, Figures 4.5a and b. For the 6.7 and 9.4kb fragments, a stepwize decrease in size could be followed on both the ethidium stained gel and after hybridisation with nick translated λ DNA, Figures 4.5a and b. The rate of loss of DNA for both fragments was ~160bps/minute. The experiment suggests that firstly, the Bal 31 enzyme is removing deoxynucleotides from the free ends of DNA molecules as predicted. Secondly, the nuclease reaction is progressive, taking place at a steady rate over a period of time. Thirdly, the enzyme does not appear to be degrading the DNA. DNA degradation is often accompanied by a general smearing of random DNA fragments through a track upon gel electrophoresis. This was not observed for either the λ, DBA/2, C57Bl/6 or *Mus caroli* DNA.

Therefore, it appears that the $(TTAGGG)_4$ hybridising, nonresolvable region seen on a conventional gel, and the discrete bands obtained upon PFGE, represent the telomeres of DBA/2, C57Bl/6 and *Mus caroli* chromosomes. The discrete bands of between ~2-10kb are insensitive to Bal 31 digestion over an extended period of time and therefore, represent regions internal to the mouse telomeres of these strains. The presence of these features within the genomes of *Apodemus sylvaticus*, *Mus musculus*, *Mus domesticus*, *Mus spretus* and AKR/J suggests that the presence of $(TTAGGG)_n$ repeats at the telomeres and interstitial sites is a common feature within the mouse genome.

4.4 General Location of $(TTAGGG)_n$ Repeats on Mouse Metaphase Chromosomes

The location of the *Trypanosoma*-like telomere repeats within the mouse genome was also investigated by visualisation, *in situ*, of the $(TTAGGG)_n$ repeats within C57Bl/6 mouse metaphase chromosomes using the PRINS technique, Section 2.4.2. Hybridisation of labelled repeats was detected by excitation of the FITC or Texas Red fluorochromes. Fluoresence and so $(TTAGGG)_n$ repeats were observed at the tips of the long arms of most of the mouse chromosomes, Figure 4.6. The intensity of the signal varied between different chromosome ends which may correlate with the different sizes, and so varying amounts of repeat, of discrete bands seen after PFGE, Figure 4.1b. Alternatively, it may relate to an inaccessibility of some regions of the genome to *in situ* hybridisation. Interstitial sites could also be seen within some of the chromosomes, which may relate to the discrete bands seen with conventional gel electrophoresis, Figure 4.1a. The signal was very weak or absent from the centromeric end of these telocentric chromosomes. This could suggest that there are two classes of telomeres within the mouse genome. The long arm of the chromosome would be expected to contain a long stretch of telomeric repeat while the centromeric region would

Figure 4.5 Bal 31 Digestion of λ Hind III DNA Restriction Fragments

λ Hind III DNA restriction fragments were subjected to Bal 31 digestion for the times indicated above each track in (a) and (b), separated by conventional gel electrophoresis, transferred to a nylon membrane and hybridised with nick translated λ DNA.

(a) Ethidium bromide stained, agarose gel of Bal 31 treated λ DNA.

(b) Autoradiograph of gel shown in (a) hybridised with nick translated λ DNA.

Sizes of the DNA fragments are indicated (kb).

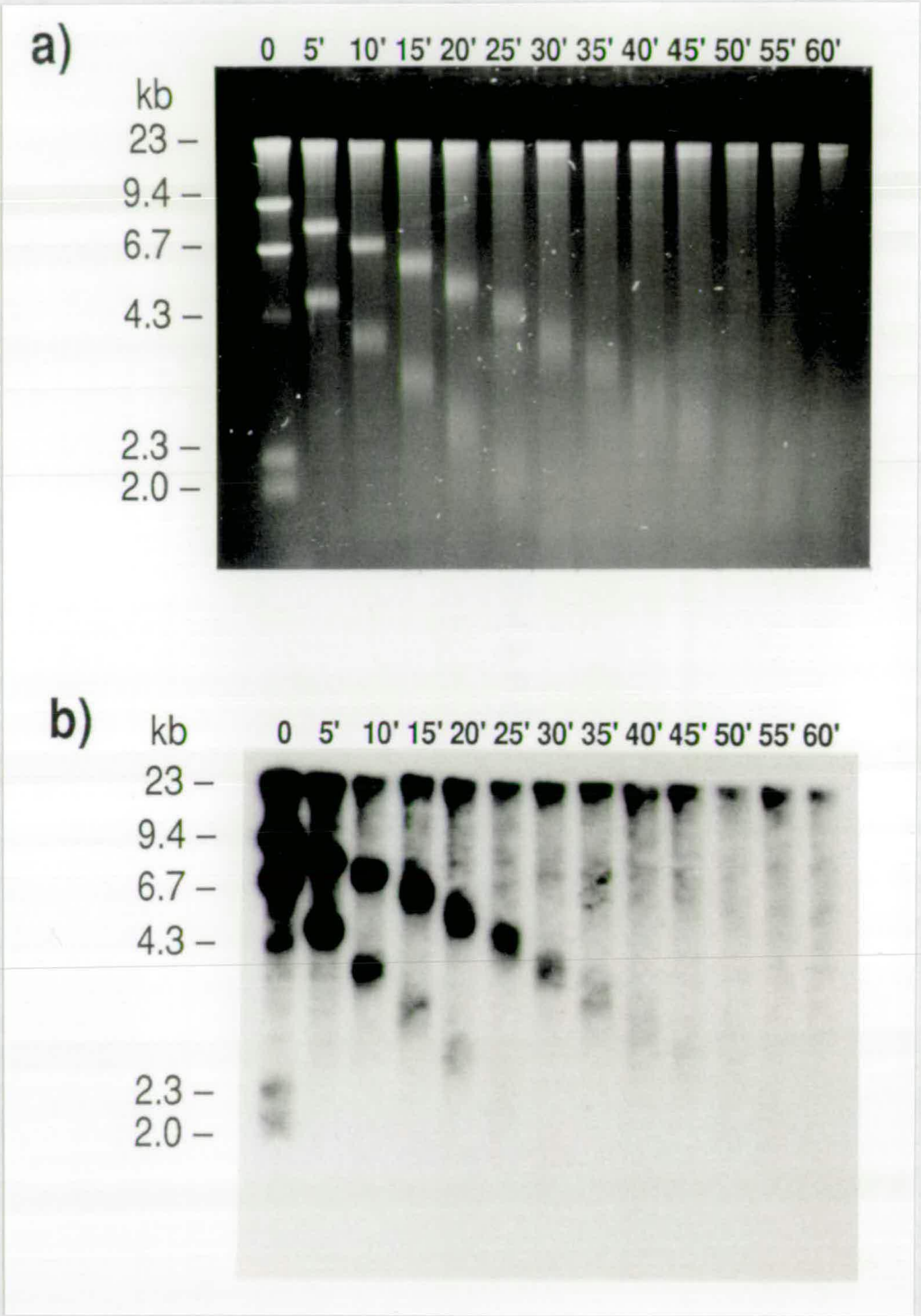
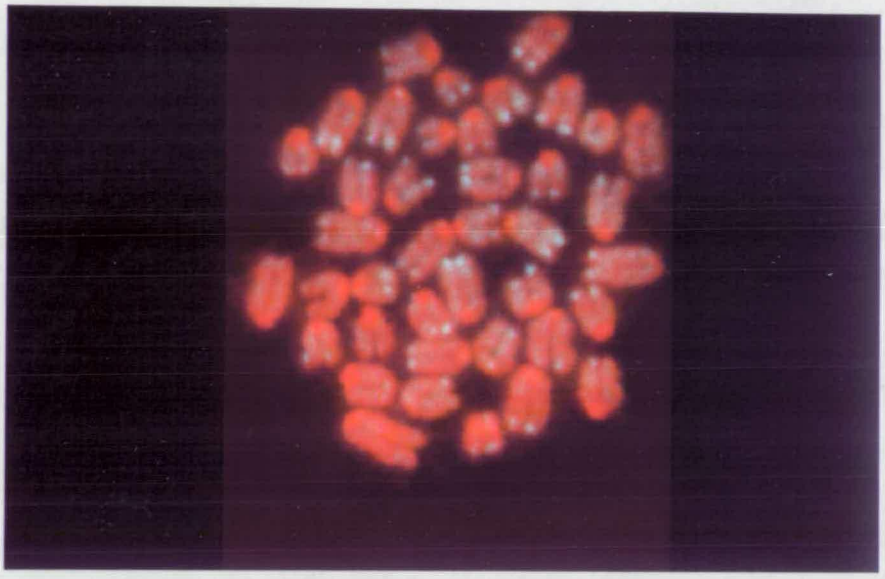
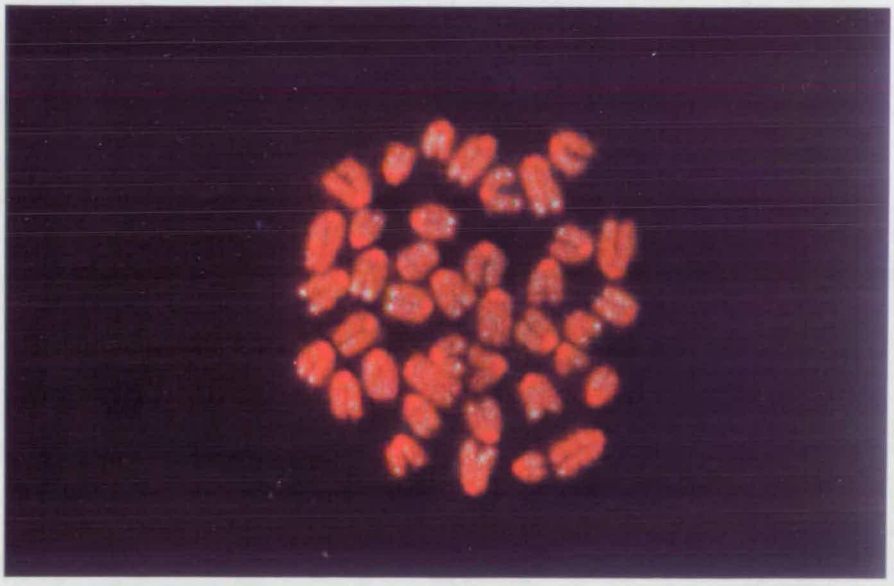

a)
0 5' 10' 15' 20' 25' 30' 35' 40' 45' 50' 55' 60'
kb
23 –
9.4 –
6.7 –
4.3 –
2.3 –
2.0 –
b)
kb
0 5' 10' 15' 20' 25' 30' 35' 40' 45' 50' 55' 60'
23 –
9.4 –
6.7 –
4.3 –
2.3 –
2.0 –

Figure 4.6 Localisation of $(TTAGGG)_n$ Repeats within Mouse Metaphase Chromosomes

Trypanosoma-like telomere repeats were localised within mouse metaphase chromosomes using the PRINS technique. (a) and (b) show two separate metaphase spreads from separate cells. The white 'spots' represent the presence of $(TTAGGG)_n$ repeats within that region of the chromosome. The signal was mostly localised to the tips of the long arms of the chromosomes and also to some interstitial sites.





contain a short stretch of repeats; the amount of repeat would correspond to the intensity of hybridisation. Alternatively, the difference in signal intensity, as described above, might relate to an inaccessibility of the centromeric telomere to the repeat probe.

The results obtained from the *in situ* visualisation of the $(TTAGGG)_n$ repeats within mouse metaphase chromosomes agreed with the results obtained from the Bal 31 experiments, Section 4.3. Therefore, it appears that $(TTAGGG)_n$ repeats are located at the tips and so telomeres of mouse chromosomes, and sites internal to these regions.

4.5 Nature of $(TTAGGG)_n$ Repeats within Different Mouse Strains

The $(TTAGGG)_n$ containing telomeric bands and interstitial loci were seen to differ between different strains of mice, Figures 4.1a and b. Different individuals within mouse strains were analysed to determine the extent of this variability. Initially, DNA from six different individuals from each of the C57Bl/6 and DBA/2 inbred strains were analysed. Hence, total genomic DNA from each animal was digested with *Hae* III, the resultant fragments separated by both pulse field and conventional gel electrophoresis, transferred to a nylon membrane and hybridised with the $(TTAGGG)_4$ probe. The >23kb region, on PFGE, was composed of discrete bands which varied in size range and number to the extent that the hybridisation pattern was different both within and between the strains, Figure 4.7a. The individuals from each strain were obtained from distinct colonies; the C57Bl/6 from a local and Hull colony, and the DBA/2 from colonies in Hull and Oxford, as indicated in Figure 4.7a. The individuals within a colony did resemble each other in the size and number of $(TTAGGG)_4$ hybridising fragments more closely than the same strain of mouse from a different colony, Figure 4.7a. The size range of bands was consistent between different individuals of a species to that previously observed for the different species of mouse, *ie*, the size range for all the $(TTAGGG)_n$ repeat containing bands from the DBA/2 animals was between 30-150kb, while the C57Bl/6 varied from 30-100kb, Figure 4.7a.

The discrete, telomere proximal bands showed far less variability. The banding pattern within a colony was the same for each individual except for slight variations in the intensity of hybridisation to some of the bands. This would be expected to result from differences in the number of telomere repeats present at these loci. As the size of the hybridising DNA fragments remained the same, the increase in signal intensity would be expected to arise from an amplification of the whole fragment. The pattern within a strain but between different colonies differed in the presence or absence of certain bands but the same basic structure was maintained, Figure 4.7b. Hence, the position of $(TTAGGG)_n$ repeats within the mouse genome produces two contrasting features, hypervariability at the telomere and only limited

Figure 4.7 Comparison of (TTAGGG)$_4$ Banding Patterns within the C57Bl/6 and DBA/2 Inbred Mouse Strains

Total genomic DNA from different individuals from separate C57Bl/6 and DBA/2 colonies was digested with *Hae* III, separated by (a) pulsed-field and (b) conventional gel electrophoresis, transferred to a nylon membrane and hybridised with the (TTAGGG)$_4$ oligonucleotide probe.

BL = C57Bl/6 local colony

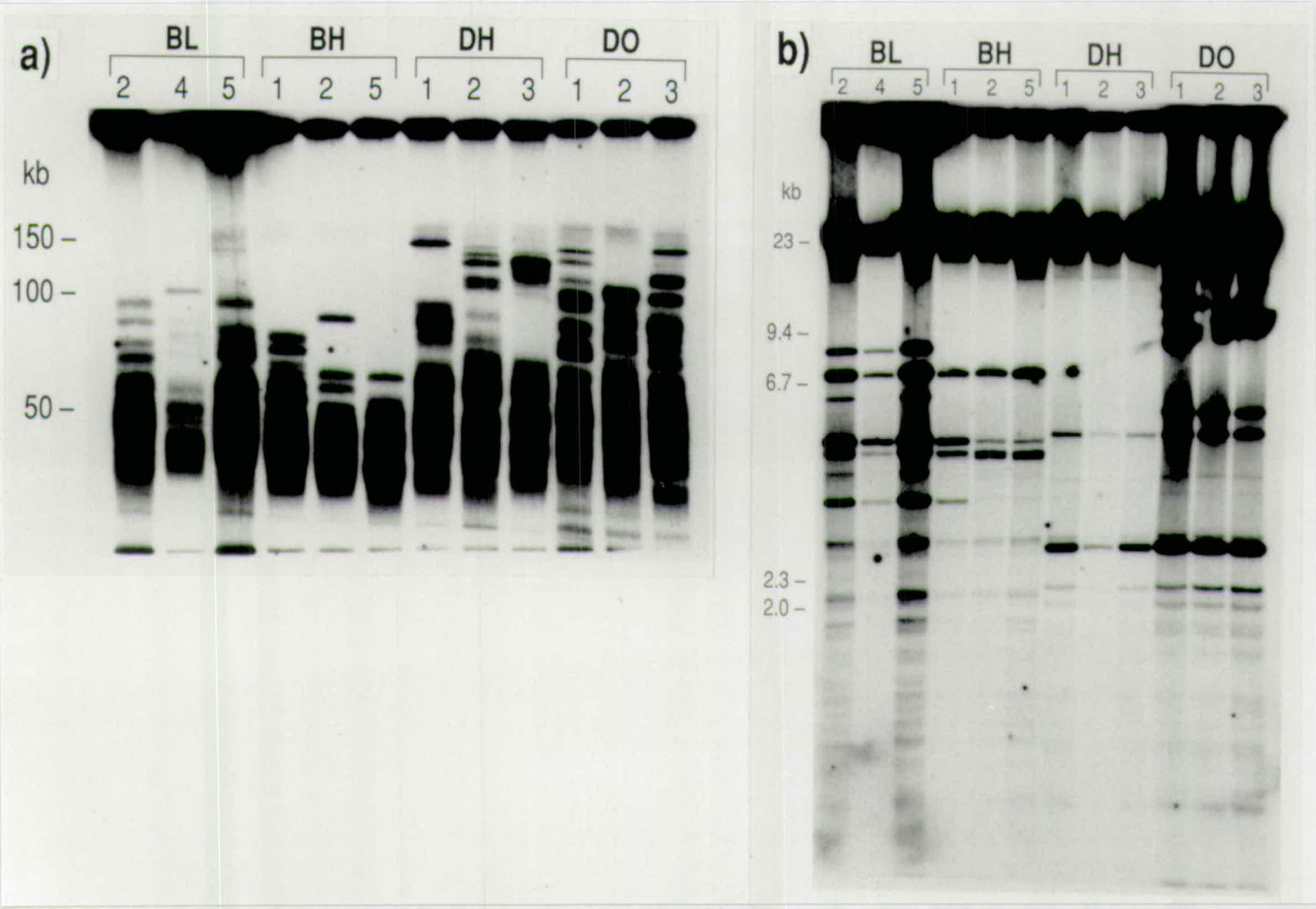
BH = C57Bl/6 colony from Hull

DH = DBA/2 colony from Hull

DO = DBA/2 colony from Oxford

The number above each track refers to a different individual from each colony.

Hind III digested λ DNA and λ concatemer size markers are indicated (kb).


a)
BL
BH
DH
DO
2 4 5 1 2 5 1 2 3 1 2 3
kb
150 –
100 –
50 –
b)
BL
BH
DH
DO
2 4 5 1 2 5 1 2 3 1 2 3
kb
23 –
9.4 –
6.7 –
2.3 –
2.0 –

variation at interstitial sites. The hypervariability of these high molecular weight telomere bands is unexpected as both the DBA/2 and C57Bl/6 strains had been maintained as inbred laboratory stocks. This would be expected to generate a genome in both strains with almost no heterozygosity at any particular locus. This was reflected by the telomere proximal bands.

The characterisation of $(TTAGGG)_n$ containing regions was extended to the *Mus caroli* strain. The telomeric and internal hybridising regions of eight different individuals were studied as before. These mice had not been inbred to the same extent as the DBA/2 and C57Bl/6 strains, having been maintained in the laboratory since capture from the wild. The discrete bands ranging from 1-4kb showed some variation between individuals, although the overall pattern was similar: in some mice one band was missing, in others one additional band was seen, Figure 4.8a. The 20-30kb region of hybridisation varied in size between individuals when analysed by both pulsed-field and conventional gel electrophoresis, Figures 4.8a and b. It was difficult to detect any specific bands within the *Mus caroli* telomere. Different PFGE techniques allowing greater resolution of this region would help to determine the exact nature of the *Mus caroli* telomere.

4.6 Developmental Origin of Variability

The DNA samples used so far in this study had all been extracted from the spleens of the animals. The hypervariability of the telomeric $(TTAGGG)_n$ repeat containing regions may be a result of mutational changes taking place within the stem cells as each tissue within the mouse developes, *ie*, somatic variation. This is in contrast to germ line variation which takes place during gametogenesis. The somatic stability of the internal $(TTAGGG)_n$ containing loci was also assessed. Hence, DNA was extracted from the liver, spleen, kidneys, lungs, heart, brain and leg muscle of different DBA/2, C57Bl/6, DBA/2 x C57Bl/6 F1 and *Mus caroli* mice, digested with *Hae* III, the fragments separated by both pulsed-field and conventional gel electrophoresis, transferred to a nylon membrane and hybridised with the $(TTAGGG)_4$ oligonucleotide probe. Figures 4.9a and b shows the resultant pattern of hybridisation within the DNA derived from the liver, spleen and kidney of a DBA/2, C57Bl/6 and two F1 DBA/2 x C57Bl/6 individual mice. For both the telomere and internal regions no variation in banding patterns between the different tissues was observed, Figure 4.9a and b. This observation was consistent with all tissues tested from these animals. A similar study was conducted in the *Mus caroli* mice and the same result obtained (data not shown). Therefore, it would appear that the different degrees of variation observed in the telomere and internal $(TTAGGG)_n$ containing regions are not due to mutational changes taking place within the tissues as the mice develop and so are most likely taking place during gametogenesis.

Figure 4.8 Comparison of (TTAGGG)$_4$ Banding Patterns within the Mus caroli Mouse Strain

Total genomic DNA from different individuals of the *Mus caroli* mouse strain was digested with *Hae* III, separated by (a) pulsed-field and (b) conventional gel electrophoresis, transferred to a nylon membrane and hybridised with the (TTAGGG)$_4$ oligonucleotide probe. The individuals are numbered from 1-8 and the sex of each is indicated.
Hind III digested λ DNA and λ concatemer size markers are indicated (kb).

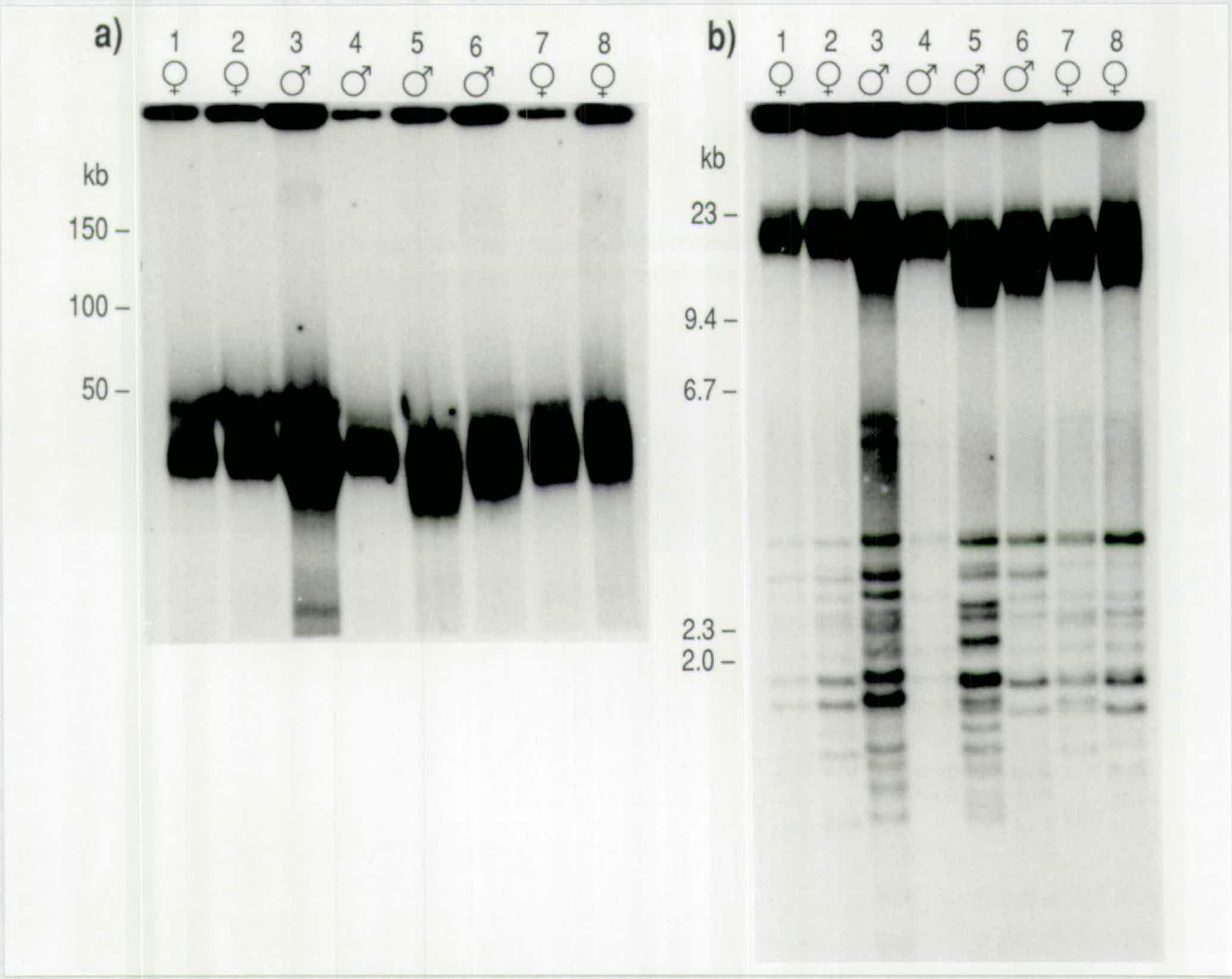
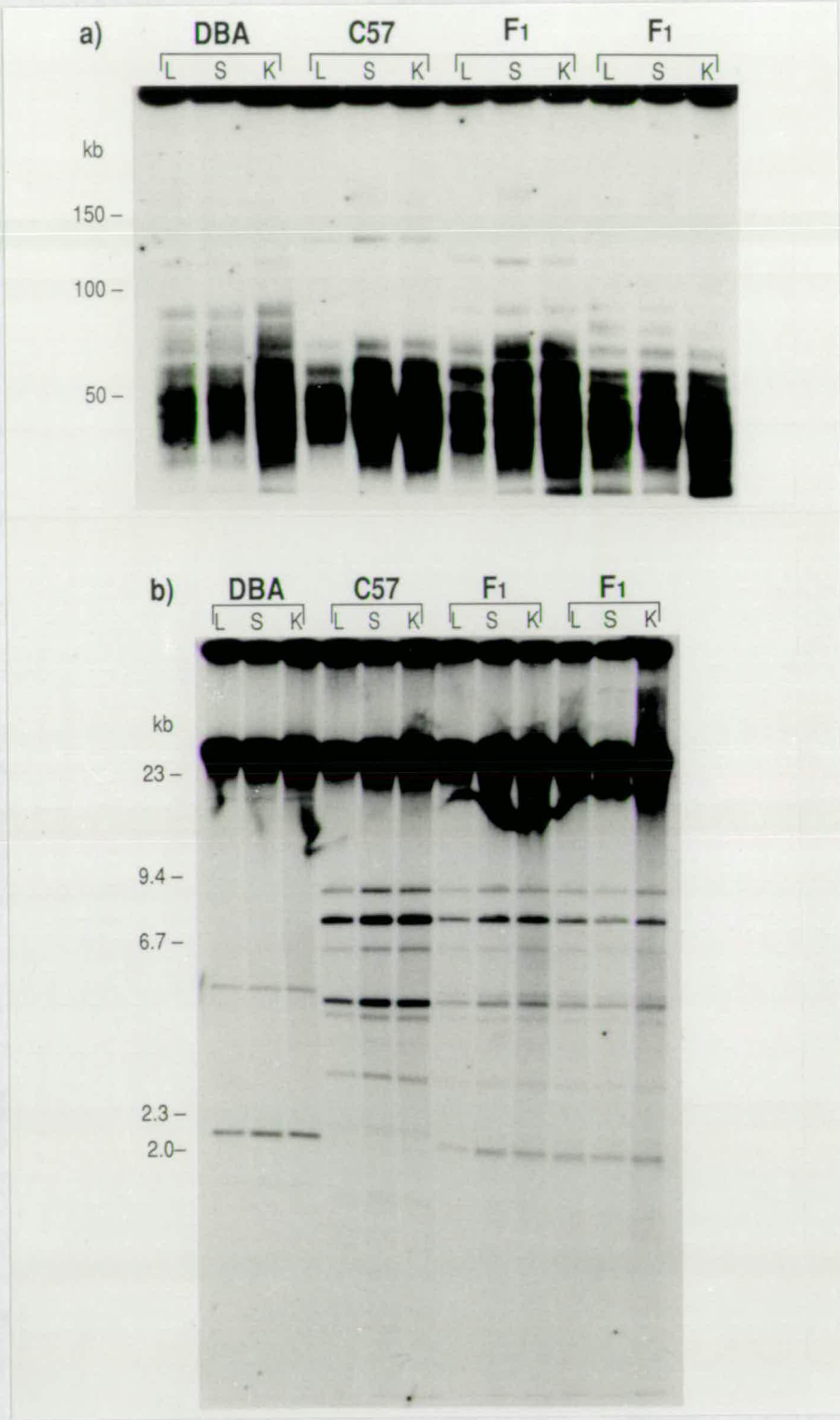

a)
1 2 3 4 5 6 7 8
♀ ♀ ♂ ♂ ♂ ♂ ♀ ♀
kb
150 –
100 –
50 –
b)
1 2 3 4 5 6 7 8
♀ ♀ ♂ ♂ ♂ ♂ ♀ ♀
kb
23 –
9.4 –
6.7 –
2.3 –
2.0 –

Figure 4.9 Comparison of the Pattern of $(TTAGGG)_4$ Hybridising DNA Fragments within Different Tissues of DBA/2 and C57Bl/6 Mouse Genomes

Total genomic DNA from the liver, spleen and kidney of a DBA/2, a C57Bl/6 and two mice derived from a DBA/2 x C57Bl/6 mating was digested with *Hae* III, separated by (a) pulsed-field and (b) conventional gel electrophoresis, transferred to a nylon membrane and hybridised with the $(TTAGGG)_4$ oligonucleotide probe.

DBA	=	DBA/2 genomic DNA
C57	=	C57Bl/6 genomic DNA
F1	=	DBA/2 x C57Bl/6 derived F1 genomic DNA
L	=	Liver
S	=	Spleen
K	=	Kidney

Hind III digested λ DNA and λ concatemer size markers are indicated (kb).


a)
DBA
C57
F1
F1
L S K
L S K
L S K
L S K
kb
150 –
100 –
50 –
b)
DBA
C57
F1
F1
L S K
L S K
L S K
L S K
kb
23 –
9.4 –
6.7 –
2.3 –
2.0–

4.7 Summary

Four main conclusions could be drawn from the preliminary investigation of $(TTAGGG)_n$ *Trypanosoma*-like telomere repeats within the mouse genome. Firstly, there appear to be $(TTAGGG)_n$ repeats located within the telomeres and at interstitial sites of the mouse genome. Secondly, both regions show distinct variation between different strains but only the telomeres vary significantly between the individuals contained within a strain. Thirdly, the hypervariability would appear to originate from germ line events. Finally, the mouse telomeres, with the exception of *Mus spretus*, appear to have a much larger size range, 20-150kb, than any other eukaryotic species examinied to date (Chapter 1, Table 1.1).

From the results obtained, three main lines of study were followed. The mouse telomere appeared to be composed of discrete fragments hybridising to the $(TTAGGG)_4$ probe, each of which may correspond to an individual telomere. If this is the case, it might be possible to follow the inheritance of such loci into a second generation. In addition, the pattern of inheritance of the interstitial loci was followed in C57Bl/6 x DBA/2 cross strain matings to determine if the bands observed upon $(TTAGGG)_4$ hybridisation were allelic. The composite DNA fragments of a locus must be identified if it is to be mapped successfully within recombinant inbred mouse strains. Following the telomere and interstitial $(TTAGGG)_n$ repeat containing loci through several generations of cross strain and subsequently brother x sister mating would allow multicomponent loci to be identified. In addition, an insight into the level of variability of telomeric and interstitial sites could be determined through such matings. A second line of investigation concerns the sequence composition of the $(TTAGGG)_n$ containing mouse telomeric and interstitial sites. Allshire *et al.* (1989) showed, by using a combination of restriction enzymes and oligonucleotides, that the human telomere was composed of three different types of repeat. Due to the large size of mouse telomeres it may be possible that mouse telomeres also contain divergent repeats. Finally, the original aim of this study was to determine the position within the genome of interstitial telomere repeats and to define a possible function for them. Hence, mapping of the $(TTAGGG)_n$ containing loci within the recombinant inbred C57Bl/6 x DBA/2, and C57Bl/6 x *Mus spretus* backcross mice was conducted.

Chapter 5

Inheritance of Telomeric and Internal (TTAGGG)n Containing Loci within DBA/2 x C57Bl/6 Cross Strain Matings

5.1 Introduction

The presence of $(TTAGGG)_n$ *Trypanosoma*-like telomeric repeats at both interstitial and telomeric sites within the mouse genome had been demonstrated by Bal 31 digestion analysis and the PRINS technique of *in situ* localisation to metaphase chromosomes (Chapter 4). The loci were also shown to vary between different species and strains of mice to the extent that each had its own *Hae* III derived banding pattern, Figure 4.1a and b. Examination of the interstitial $(TTAGGG)_n$ containing loci in the inbred mouse strains DBA/2 and C57Bl/6 revealed identical banding patterns between members of a colony but variations when different colonies were compared, Figure 4.7b. For both inbred strains, the different colonies, *ie* local and Hull for C57Bl/6 and Hull and Oxford for DBA/2, had been separated and inbred for a number of years. Hence, the variation in banding pattern for the interstitial loci could be attributed to genetic drift. Inbreeding within each separate colony had been performed to the extent that allelic differences between homologous chromosomes were rarely detected. Hence, within an inbred strain different members would be expected to contain identical gene loci. This was observed for the internal $(TTAGGG)_n$ containing loci. The telomeric loci proved to be an exception to this rule. They were shown to vary considerably both between and within the inbred strains, Figure 4.7a.

The inheritance patterns of the telomeric and internal $(TTAGGG)_n$ containing loci were investigated in a cross-strain mating series between the DBA/2 and C57Bl/6 inbred mouse strains. In summary, the series involved the initial DBA/2 x C57Bl/6 cross, the F1 offspring of which were either brother x sister or cross-pair mated. The resultant F2 offspring were mated in the same way, producing the F3 generation of animals and again, for these, to give the F4 progeny (Section 2.2, Table 2.1). The genomic spleen DNA isolated from the animals at all stages of the breeding series was analysed for the inheritance of internal $(TTAGGG)_n$ containing loci and at the F1 and F2 stages for the telomeric loci. The inheritance patterns of these loci were investigated for a number of reasons :

(a) To determine if the telomeric and internal $(TTAGGG)_n$ containing loci are inherited as distinct genetic loci. This is of particular interest for the telomeric loci. In all other organisms so far analysed the telomeres have been visualised as a non-definable smear corresponding to a high level of variation in the number of repeats present at the chromosome termini.

(b) To establish if the loci are homozygous or heterozygous within the inbred strains. If the locus was homozygous within the inbred strain, it would be inherited into all of the F1 offspring, if heterozygous, and absent from the partner in the mating, into ~50% of the progeny.

(c)To establish if linkage exists between different sized DNA fragments. For both the telomeric and interstitially derived fragments, the pattern of inheritance through several generations will determine if these bands represent separate or multicomponent loci. This will allow segregation analysis and possible genomic mapping of the interstitial loci in the RI lines and Interspecific backcrosses (Chapter 7).

(d) To test if these loci are autosomal or X/Y linked; determined by the percentage inheritance of the loci into male and female offspring (as described later).

(e) To estimate the frequency of generation of new internal and telomeric $(TTAGGG)_n$ variants.

(f) To analyse the frequency of recombination between components of multicomponent loci and establish a genetic order for these components.

5.2 Inheritance of $(TTAGGG)_n$ Containing Loci into the DBA/2 x C57Bl/6 F1 Generation

5.2.1 Inheritance of Internal $(TTAGGG)_n$ Containing Loci into the F1 Generation

DNA from the male DBA/2 and female C57Bl/6 parents and F1 offspring from three separate pairs was digested with the restriction enzyme *Hae* III, the fragments separated by conventional gel electrophoresis, transferred to a nylon membrane and hybridised with the $(TTAGGG)_4$ oligonucleotide probe. A total of 21 F1 offspring from the three pairs were analysed; in all cases each had inherited both sets of parental $(TTAGGG)_n$ containing bands. Figure 5.1 shows one such pair with its male and female F1 progeny. The sizes in kb and a designated number for each clearly analysable band are indicated.

The distribution patterns for the internal $(TTAGGG)_n$ containing loci into the next generation when two homozygous inbred mouse strains had been mated could be predicted by considering Mendelian inheritance. There are two possible distribution patterns for the loci which are dependent upon their presence within either autosomal or X/Y chromosomes.

For example,

(a) If the locus is located upon an autosome and homozygous:

D = DBA/2 autosome

B = C57Bl/6 autosome

Figure 5.1 Inheritance of Internal $(TTAGGG)_n$ Containing Loci into the DBA/2 x C57Bl/6 F1 Generation

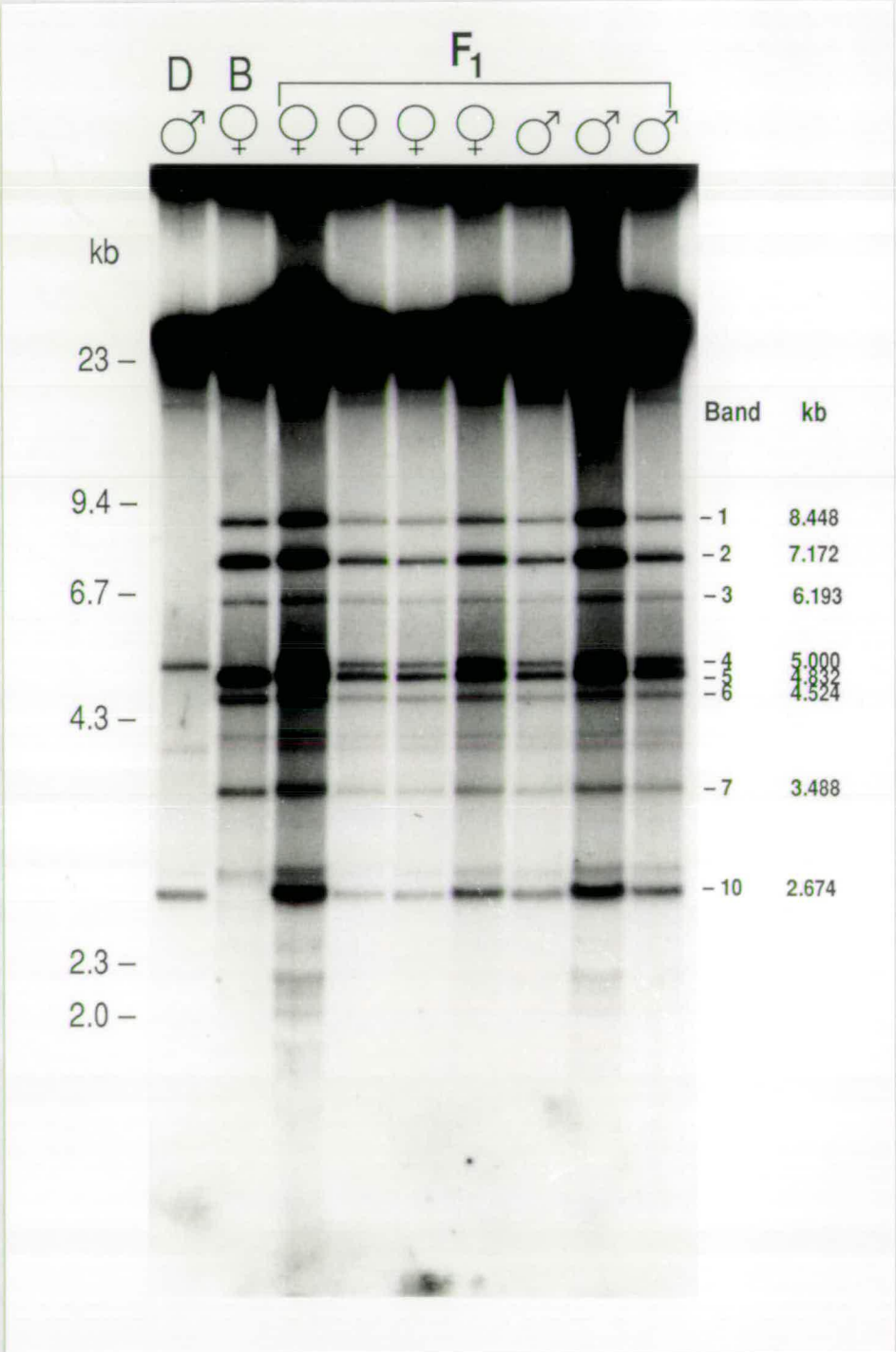
Total genomic DNA from the parents and offspring of a C57Bl/6 x DBA/2 mating was digested with *Hae* III, the fragments separated by conventional gel electrophoresis, transferred to a nylon membrane and hybridised with the $(TTAGGG)_4$ oligonucleotide probe.

Contents of tracks :

D	=	DBA/2 male parent		2348
B	=	C57Bl/6 female parent		2349
F1	=	First generation offspring	female	0001
"	=	" "	female	0002
"	=	" "	female	0003
"	=	" "	female	0004
"	=	" "	male	0005
"	=	" "	male	0006
"	=	" "	male	0007

Numbers and sizes in kb are indicated for each clearly definable C57Bl/6 and DBA/2 hybridising band.

Hind III digested λ DNA size markers are indicated (kb).


D
B
F1
kb
23 –
9.4 –
6.7 –
4.3 –
2.3 –
2.0 –
Band
kb
– 1
8.448
– 2
7.172
– 3
6.193
– 4
5.000
– 5
4.832
– 6
4.524
– 7
3.488
– 10
2.674

	DBA/2 male	x	C57Bl/6 female
Parents	DD	x	BB
F1 offspring	DB DB DB DB		

100% of the offspring will contain the locus.

(b) If the locus is linked to either the X or Y chromosomes:

DX = DBA/2 X chromosome

DY = DBA/2 Y chromosome

BX = C57Bl/6 X chromosome

	DBA/2 male	x	C57Bl/6 female
Parents	DX, DY	x	BX, BX
F1 Offspring	DX, BX DX, BX	DY, BX DY, BX	

If the locus were present upon the C57Bl/6 X chromosome, 100% of the female and male offspring would inherit the corresponding band. Alternatively, if located upon either of the DBA/2 X or Y chromosomes, 100% of females and 0% of males or 100% of males and 0% of females, respectively, would inherit the locus.

From the inheritance of both sets of parental bands into all of the F1 offspring, it could be concluded that the internal $(TTAGGG)_n$ containing loci were only present upon the autosomes in the DBA/2, and on either the X chromosome or autosomes in the C57Bl/6 inbred mice analysed in this study. The question of homozygosity at the loci examined in the inbred strains was answered by 100% inheritance of the corresponding bands into the F1 generation. If one locus was heterozygous in either strain ~50% instead of 100% inheritance into the offspring would have been observed. In addition, no new variant sized bands were detected within the F1 offspring. This again suggests that the interstitial $(TTAGGG)_n$ containing loci were homozygous within the inbred strains. For such loci recombination events would not be seen due to the exchange of identical genetic material between the two, participating chromosomal regions.

5.2.2 Inheritance of Telomeric $(TTAGGG)_n$ Containing Loci into the F1 Generation

DNA from the male DBA/2 and female C57Bl/6 parents and F1 offspring from three separate pairs was digested with the restriction enzyme *Hae* III, the fragments separated by PFGE, transferred to a nylon membrane and hybridised with the $(TTAGGG)_4$ oligonucleotide probe. Again a total of 21 F1 offspring from the three pairs were analysed. The pattern of bands observed between different individuals within a colony had been shown to differ, unlike the case for the internally located loci, Figure 4.7a and b. This suggested that heterozygosity existed within the termini of inbred mouse chromosomes; an event rarely observed in the

rest of the genome. As can be seen for the F1's of the three pairs analysed, Figures 5.2a, b and c, unlike the internally located loci, not one of the offspring inherited a complete set of bands from both parents. In each figure, the inheritance of the DBA/2 and C57Bl/6 specific bands into the F1 offspring is indicated with arrowheads. Those that differ in size compared to the parents, the new variant bands, are marked with a V. As indicated in Figures 5.2a, b and c, an approximate total of 15 new variant bands were seen in the progeny of the three pairs. Some of these bands varied only slightly in size from those of parental origin suggesting that they may represent deviations in the separation of DNA fragments during electrophoresis. Hence, it appeared that the DBA/2 and C57Bl/6 telomeric loci were heterozygous in both strains and were inherited at random into the F1 offspring. In addition, new variants were generated at a high frequency, compared to $(TTAGGG)_n$ containing loci located within the mouse genome. It was difficult to determine the complete inheritance pattern of the telomeric loci as a large proportion fell into the 30-60kb size range, producing an undefinable smear upon hybridisation with the $(TTAGGG)_4$ probe.

A rough estimation of the frequency of new variant generation can be determined from the offspring of the three pairs shown in Figures 5.2a, b and c.

Hence, looking at the F1 cross :

21 mice were analysed in which 38 telomeric DNA fragments could be clearly observed. 15 of the 38 DNA fragments were new variants. Therefore, ~40% of the analysable telomeric DNA bands represented new variants. In any gamete, there are ~40 telomeres (20 chromosomes). If it is considered that all of the telomeres vary at the same rate, it would be expected that each gamete will have, on average, 16 new sized telomeres.

From the above calculations a value for the frequency of new variant generation at telomeres can be calculated :

15 new variant bands were observed in 21 mice,

So 15 ÷ 21 new variants are visible per mouse,

Therefore, ~2 new variants are detected in any 3 mice and 66% of mice or 33% of gametes will display a visible, new variant, telomeric DNA fragment.

5.2.3 Initial Summary of Results

From the patterns of inheritance of telomeric and internal $(TTAGGG)_n$ containing loci demonstrated in the F1 offspring from DBA/2 x C57Bl/6 inbred strain matings several conclusions could be drawn. In general, both the telomeric and internal $(TTAGGG)_n$ containing loci can be observed to be inherited into a second generation and so represent distinct genetic loci. Considering the internal loci, firstly, the strain specific internal loci appeared to be homozygous within both inbred strains of mice. Secondly, the DBA/2 specific internal loci must have been present upon the autosomal chromosomes, while the

Figure 5.2 Inheritance of Telomeric $(TTAGGG)_n$ Containing Loci into the F1 Generation

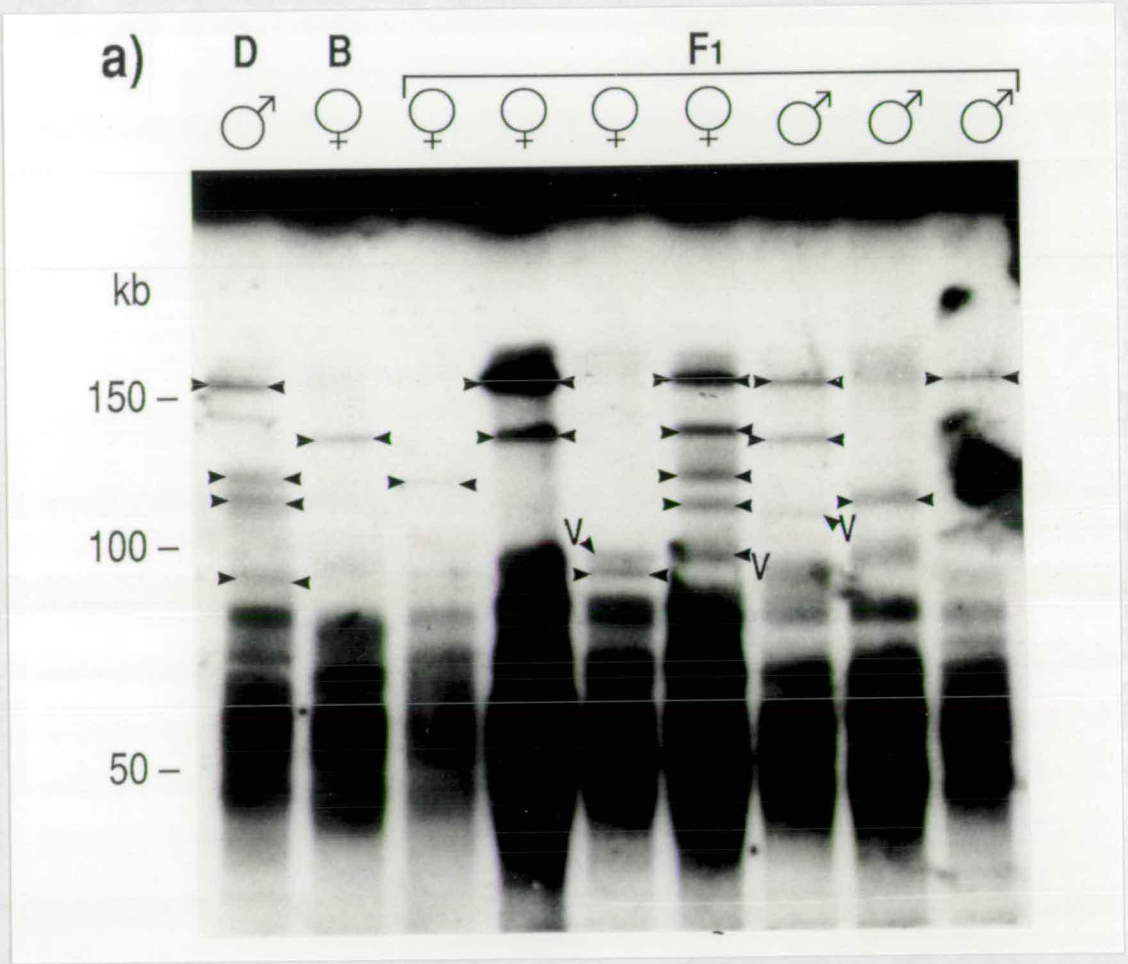
Total genomic DNA from the parents and offspring of three separate C57Bl/6 x DBA/2 matings was digested with *Hae* III, the fragments separated by PFGE, transferred to a nylon membrane and hybridised with the $(TTAGGG)_4$ oligonucleotide probe. Bands that have been inherited from either parent into the offspring are marked with arrowheads and new bands, which vary from those in the parents, with a V

λ concatemer size markers are indicated (kb).

5.2 (a) Pair 2

Contents of tracks :

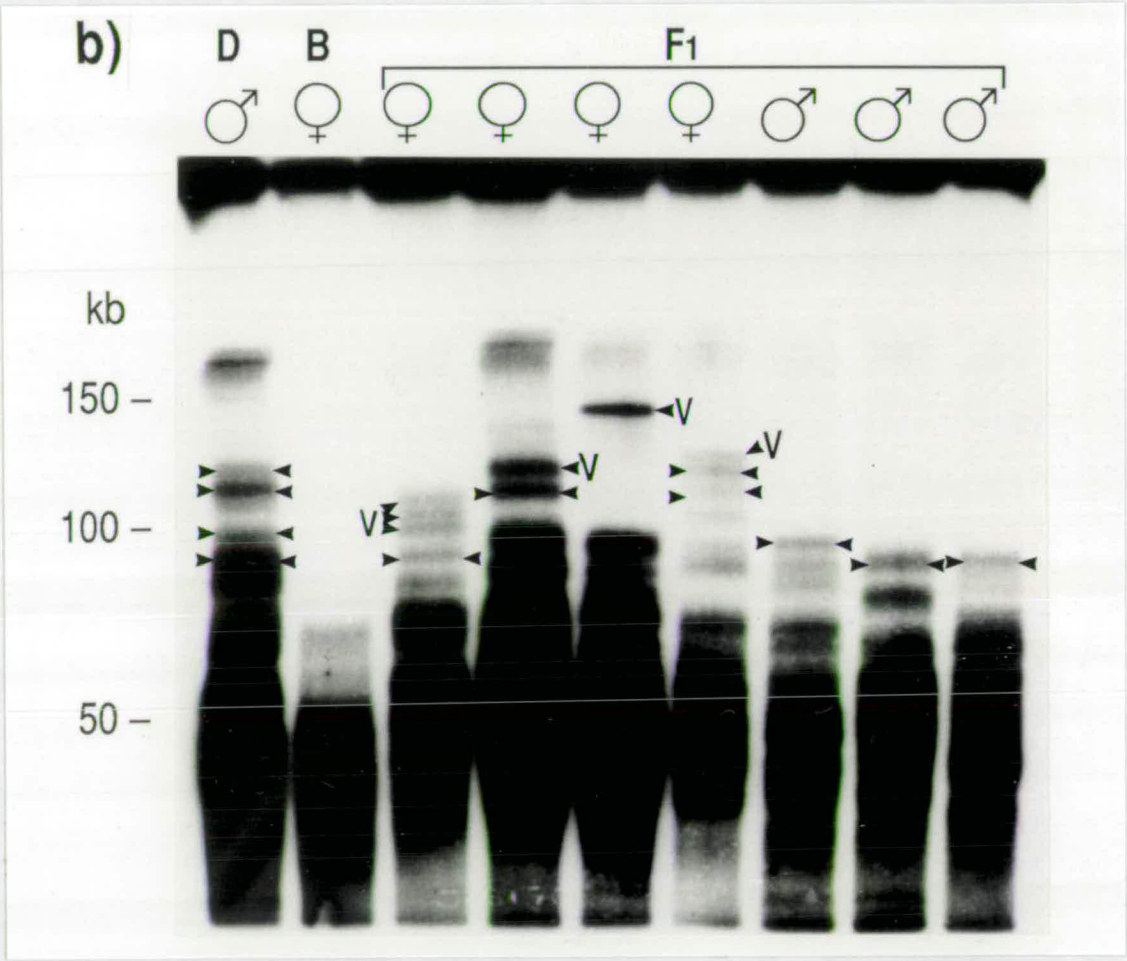
D	=	DBA/2 male parent		2350
B	=	C57Bl/6 female parent		2351
F1	=	First generation offspring	female	0012
"	=	" "	female	0013
"	=	" "	female	0014
"	=	" "	male	0008
"	=	" "	male	0009
"	=	" "	male	0010
"	=	" "	male	0011


a)
D
B
F1
kb
150 –
100 –
50 –
V
V
V

5.2 (b) Pair 1

Contents of tracks :

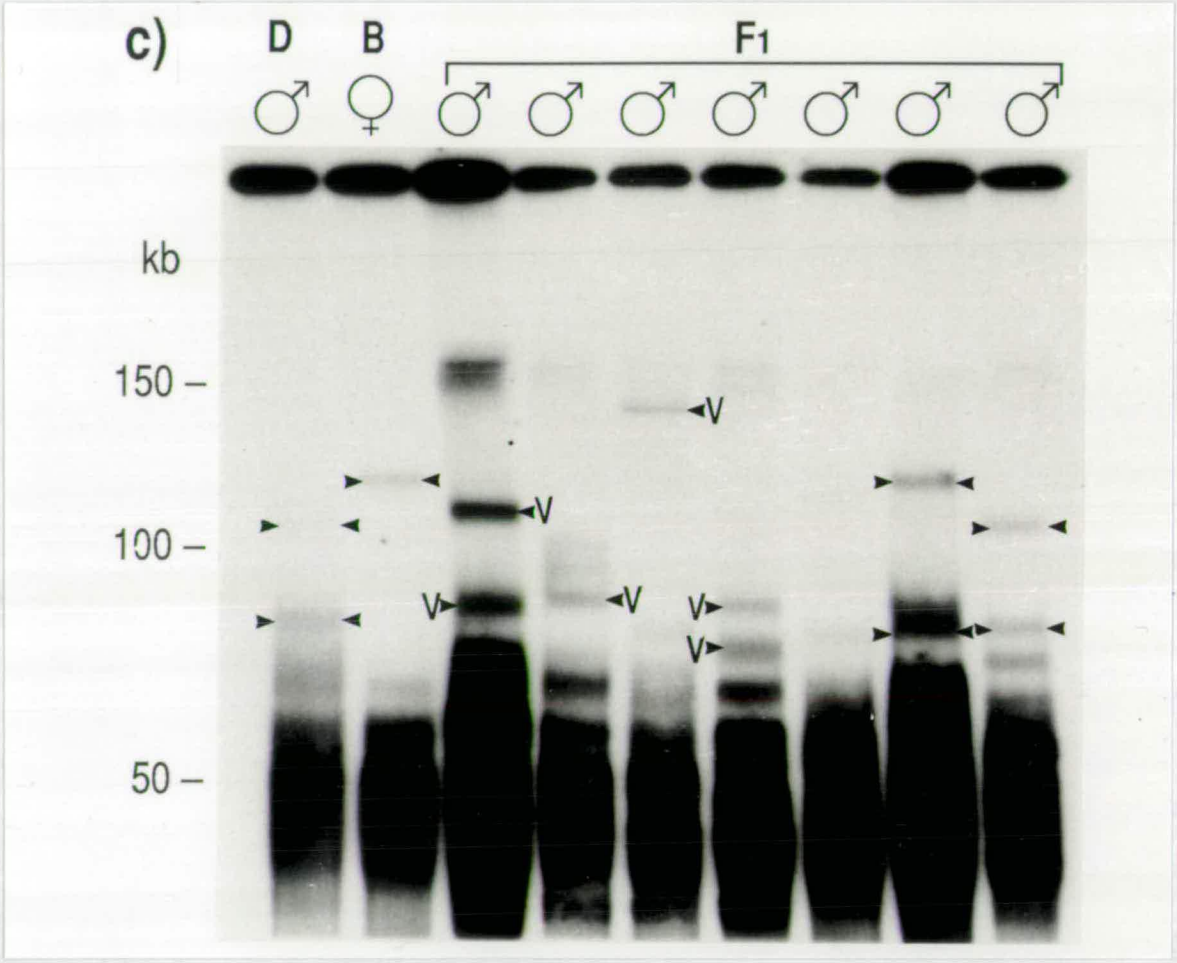
D	=	DBA/2 male parent		2348
B	=	C57Bl/6 female parent		2349
F1	=	First generation offspring	female	0001
"	=	" "	female	0002
"	=	" "	female	0003
"	=	" "	female	0004
"	=	" "	male	0005
"	=	" "	male	0006
"	=	" "	male	0007


b)
D
B
F1
kb
150 –
100 –
50 –
V
V
V
V

5.2 (c) Pair 3

Contents of tracks :

D	=	DBA/2 male parent		2352
B	=	C57Bl/6 female parent		2353
F1	=	First generation offspring	Male	2354
"	=	" "	Male	2355
"	=	" "	Male	2356
"	=	" "	Male	2357
"	=	" "	Male	2358
"	=	" "	Male	2359
"	=	" "	Male	2360


c)
D
B
F1
♂
♀
♂
♂
♂
♂
♂
♂
♂
kb
150 –
100 –
50 –
V
V
V
V
V
V

C57Bl/6 loci could have been present upon either the X or autosomal chromosomes. Thirdly, the internal loci appeared to be passed into the F1 progeny by Mendelian inheritance. The initially homozygous DBA/2 male and C57Bl/6 female, when mated and the offspring analysed, demonstrated that each had inherited a complete set of bands from both parents. Hence, the F1 offspring were heterozygous at each internal $(TTAGGG)_n$ containing locus. The telomeric $(TTAGGG)_n$ containing loci behaved very differently. Firstly, differences in banding patterns were seen between individuals within a colony. Secondly, the loci were passed into the F1 offspring at random so that no one individual inherited a complete set of bands from both parents. This suggested that the telomeric $(TTAGGG)_n$ containing loci were heterozygous within the DBA/2 and C57Bl/6 inbred mouse strains. Thirdly, there was a high frequency of new variant generation at these loci upon mating between the DBA/2 and C57Bl/6 mice which was not observed for the internal $(TTAGGG)_n$ containing regions. The significance and possible reasons for the high frequency of new variant generation at telomeric loci will be discussed in Chapter 8.

The inheritance patterns of both sets of mouse genomic $(TTAGGG)_n$ containing loci were followed into the next, F2, generation. For the telomeric loci, analysis of the F2 offspring would provide further information about their inheritance as distinct genetic loci. The extent of the variability observed in the F1 offspring could be studied further and the inheritance of new variants into a second generation examined. Analysis of inheritance into the F2 generation of the internal $(TTAGGG)_n$ containing loci would allow the demonstration of:

(a) Mendelian inheritance. The expected inheritance of loci from heterozygous parents can be predicted (as described later).

(b) Establishment of linkage, if present, between the $(TTAGGG)_4$ hybridising bands. The C57Bl/6 specific bands 1,2,3,5,6 and 7 and DBA/2 specific bands 4 and 10 (Figure 5.1) may each represent distinct loci within the relevant genome or, alternatively, combinations of the fragments may constitute a multicomponent locus. Analysis of recombination events and determination of their frequency could provide information about the order of DNA fragments within such a locus along the chromosome.

(c) The presence of new variant sized bands generated by recombination events between the C57Bl/6 and DBA/2 genomes.

5.3 Inheritance of $(TTAGGG)_n$ Containing Loci into the F2 Generation

5.3.1 Inheritance of Internal $(TTAGGG)_n$ Containing Loci into the F2 Generation

Male and female F1 mice from either the same (brother x sister) or different (cross pair) litters were mated (Section 2.2, Table 2.1). The DNA from the resultant F2 progeny was digested with *Hae* III, the fragments separated by conventional gel electrophoresis, transferred to a nylon membrane and hybridised with the $(TTAGGG)_4$ oligonucleotide probe. The patterns of loci inheritance were compared to the predicted models for mating between two individuals that were heterozygous at the $(TTAGGG)_n$ containing regions. For example,

(a) If the locus is contained upon an autosome and heterozygous:

D = DBA/2 autosome

B = C57Bl/6 autosome

	F1 male	x	F1 female
F1 Parents	DB	x	DB
F2 Offspring	DD DB	DB	BB

Theoretically, 75% of the offspring will inherit loci specific to either strain, of which 50% will be heterozygous and 25% homozygous for the locus, and 25% will not contain the band representing the locus.

(b) If the locus is contained upon the C57Bl/6 X chromosome and heterozygous:

BX = C57Bl/6 X chromosome

DX = DBA/2 X chromosome

DY = DBA/2 Y chromosome

	F1 male	x	F1 female
F1 Parents	BX, DY	x	BX, DX
F2 Offspring	BX, BX BX, DX	DY, BX	DY, DX

Theoretically, 100% of the females and 50% of the males will inherit the locus. When males and females are considered together, a total of 75% will inherit the locus.

The DNA from a total of 178 F2 offspring (84 males and 94 females) was analysed for the inheritance of DBA/2 and C57Bl/6 specific, internal, $(TTAGGG)_n$ containing loci; numbered as for the F1 generation, Figure 5.1. An example of the inheritance of these loci into the F2 offspring and parents of a brother x sister F1 mating is shown in Figure 5.3. The C57Bl/6 derived $(TTAGGG)_4$ hybridising DNA fragments, designated 1,2,3,5,6,7, appeared to be inherited as two distinct loci, *ie* 1,2,3 and 5,6,7. Both sets of bands were inherited into ~75% of the offspring, and were transmitted as the stated cluster in all but eight individuals for bands 1,2,3, and 3 individuals for bands 5,6,7. In Figure 5.3, two F2 male animals, 0694 and

Figure 5.3 Inheritance of Internal $(TTAGGG)_n$ Containing Loci into the F2 Generation

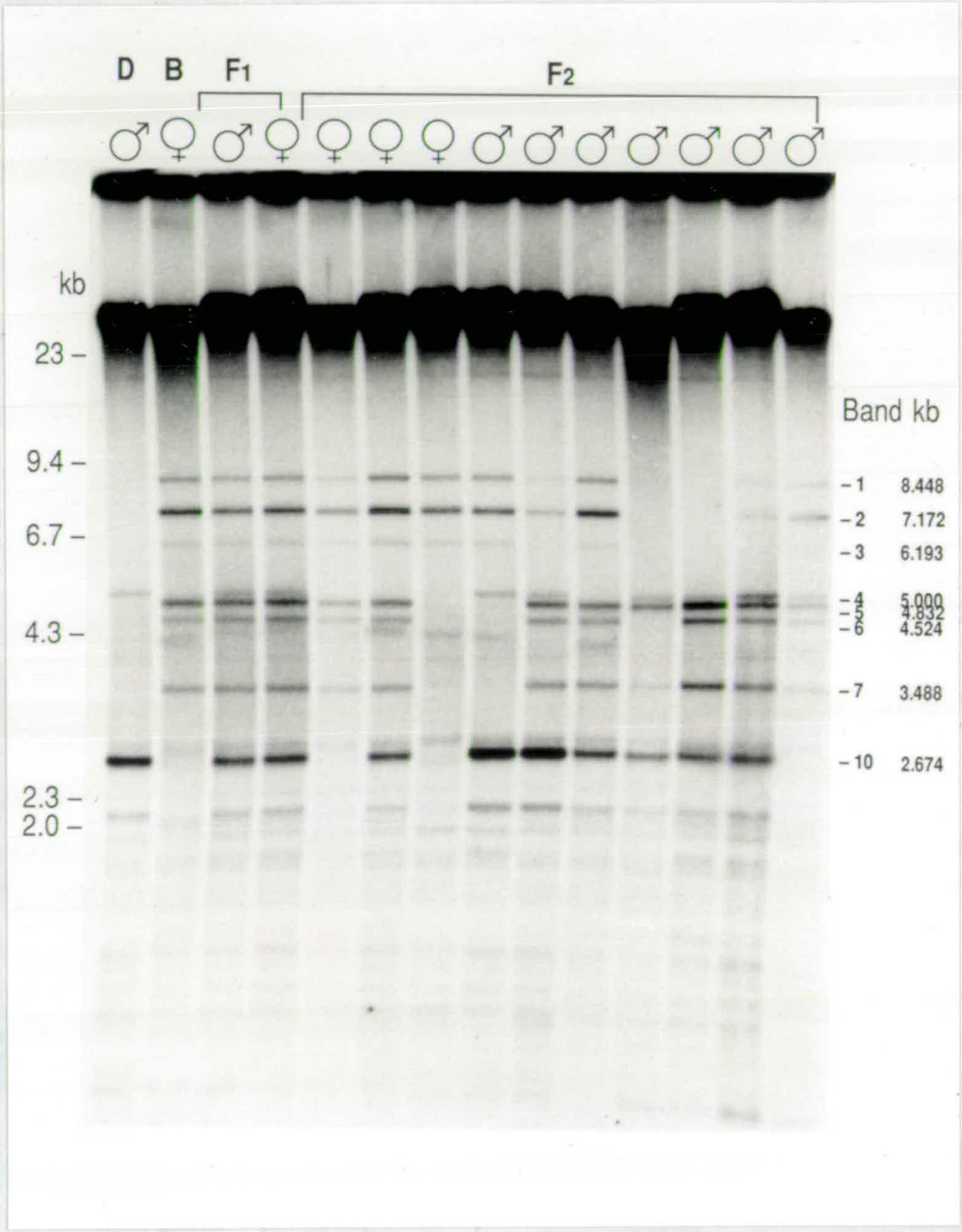
Total genomic DNA from the grandparents, parents and offspring of an F1 x F1 mating was digested with *Hae* III, the fragments separated by conventional gel electrophoresis, transferred to a nylon membrane and hybridised with the $(TTAGGG)_4$ oligonucleotide probe.

Contents of tracks :

D	=	DBA/2 male grandparent		2350
B	=	C57Bl/6 female grandparent		2351
F1	=	First generation offspring	male parent	0008
F1	=	" "	female parent	0014
F2	=	Second generation offspring	female	0680
"	=	" "	female	0681
"	=	" "	female	0682
"	=	" "	male	0689
"	=	" "	male	0690
"	=	" "	male	0691
"	=	" "	male	0692
"	=	" "	male	0693
"	=	" "	male	0694
"	=	" "	male	0695

Numbers and sizes in kb are indicated for each clearly definable C57Bl/6 and DBA/2 derived hybridising band.

Hind III digested λ DNA size markers are indicated (kb).


D
B
F1
F2
kb
23 –
9.4 –
6.7 –
4.3 –
2.3 –
2.0 –
Band kb
– 1 8.448
– 2 7.172
– 3 6.193
– 4 5.000
– 5 4.832
– 6 4.524
– 7 3.488
– 10 2.674

0695, have inherited bands 1 and 2 without 3 and in a separate male, 0692, bands 5 and 7 without 6. The two DBA/2 specific loci examined appeared to be inherited independently of each other and into 75% of the F2 generation. The inheritance into male and female F2 offspring was also assessed to determine the possibility of locus linkage to the C57Bl/6 X chromosome. The only set of $(TTAGGG)_4$ hybridising fragments to show such a connection were the C57Bl/6 derived bands, 1,2,3. As expected, for a C57Bl/6 X-linked locus, 75% of males and females combined, 50% of males alone and ~100% of females alone, inherited bands 1,2,3. Therefore, it appeared that the C57Bl/6 derived bands 1,2,3 were linked to the C57Bl/6 X chromosome. The actual numbers of animals and percentage transmission compared to the predicted values are given in Table 5.1.

The actual values for locus inheritance, in some cases tended to vary from the predicted model, Table 5.1. To determine if the variation was significant to the extent that the predicted result was incorrect, the Chi squared (X^2) test was applied. In general, this test predicts how often deviations from expectations will occur purely on the basis of chance. To calculate this value:

(a) Determine a hypothesis, termed the Null hypothesis, that gives the precise expected result, *eg*, 75% of the offspring will, and 25% will not, inherit the locus.

(b) Calculate the X^2 value from the actual numbers of plus or minus inheritance from the following equation:

$$X^2 = \text{the total of } (O - E)^2 \text{ divided by E over all classes}$$

Where O = the observed numbers

E = the expected numbers

(c) Determine the number of degrees of freedom (d.f.), that is the number of classes within the test that can vary relative to one class :

$$\text{d.f.} = (\text{number of classes} - 1)$$

(d) Calculate the probability (p) of obtaining the observed results if the Null hypothesis is correct using the X^2 and d.f. values and a set of X^2 tables. X^2 tables contain a set of computed X^2 values calculated from known p and d.f. values.

The value of p is expressed as a percentage which if greater than 5% indicates that the stated Null hypothesis is correct. If less than 5% the chance of obtaining the observed values is too low and the original hypothesis has to be discounted.

X^2 values and subsequent probabilities for the 1,2,3, 4, 5,6,7, and 10 loci were calculated, Table 5.2, to reveal how the observed inheritance percentages correlated with the appropriate Null hypotheses. For all the loci except the DBA/2 derived band 4, present in the female F2 offspring, the predicted hypothesis was correct by the X^2 test. For this one exception the probability that the locus was present upon an autosome and inherited into

Extension of Table 5.1 Showing the Individual Inheritance Values of Bands 1, 2, 3, 5, 6 and 7 into the F2 Generation

Band		Total M + F (178)			M (84)			F (94)		
		No	%	Ex. %	No	%	Ex. %	No	%	Ex. %
C57Bl/6	1	141	79	75	47	56	50	94	100	100
"	x	37	21	25	37	44	50	0	0	0
"	2	141	79	75	47	56	50	94	100	100
"	x	37	21	25	37	44	50	0	0	0
"	3	133	75	75	42	50	50	91	97	100
"	x	45	25	25	42	50	50	3	3	0
"	5	133	75	75	71	85	75	62	68	75
"	+	45	25	25	15	18	25	30	32	25
"	6	132	74	75	70	83	75	62	68	75
"	+	46	26	25	16	19	25	30	32	25
"	7	136	76	75	72	86	75	64	68	75
"	+	42	24	25	14	17	25	28	30	25

Key : As for Table 5.1 on the page opposite, *ie* 132.

Table 5.1 Inheritance of Internal $(TTAGGG)_n$ Containing Loci into the F2 Generation

Cosegregating bands		Total males + females (178)			Males (84)			Females (94)		
		No	%	Ex%	No	%	Ex.%	No	%	Ex.%
C57Bl/6	1,2,3	133	75	75	42	50	50	91	97	100
	x,x,x	37	21	25	37	44	50	0	0	0
	1,2,x	8	4.5	NA	5	6	NA	3	3	NA
DBA/2	4	124	70	75	65	77	75	59	63	75
	*	54	30	25	19	23	25	35	37	25
C57Bl/6	5,6,7	132	74	75	70	83	75	62	66	75
	+,+,+	42	24	25	14	14	25	28	30	25
	5,+,7	1	0.6	NA	1	1	NA	0	0	NA
	+,+,7	3	0.7	NA	1	1	NA	2	2	NA
DBA/2	10	125	70	75	60	71	75	65	69	75
	~	53	30	25	24	29	25	29	31	25

Key :

No = Number of animals analysed

% = Percentage of animals analysed

Ex.% = Predicted percentage from Mendelian inheritance

() = Total number of animals in the stated category

NA = Not applicable as aberrant band inhreitance

x = Absence of bands from locus 1,2,3

* = Absence of band 4

+ = Absence of bands from locus 5,6,7

~ = Absence of band 10

Extension of Table 5.2 Showing the Individual X^2 Values and Probabilities for the Hypothesized Pattern of Inheritance of Bands 1, 2, 3, 5, 6 and 7 into the F2 generation

Band	Animals included	Null hypothesis + classes	X^2	d.f.	p%
1	M+F	75% 1 : 25% x	1.68	1	50 - 10
"	M	50% 1 : 50% x	1.19	1	50 - 10
"	F	100% 1 : 0% x	NC	1	NC
2	M+F	75% 2 : 25% x	1.68	1	50 - 10
"	M	50% 2 : 50% x	1.19	1	50 - 10
"	F	100% 2 : 0% x	NC	1	NC
3	M+F	75% 3 : 25% x	0.0075	1	>90
"	M	50% 3 : 50% x	0	1	>90
"	F	100% 3 : 0% x	NC	1	NC
5	M+F	75% 5 : 25% +	0.0075	1	>90
"	M	75% 5 : 25% +	2.62	1	50 - 10
"	F	75% 5 : 25% +	2.8	1	10 - 5
6	M+F	75% 6 : 25% +	0.067	1	90 - 50
"	M	75% 6 : 25% +	1.88	1	50 - 10
"	F	75% 6 : 25% +	2.8	1	10 - 5
7	M+F	75% 7 : 25% +	0.187	1	90 - 50
"	M	75% 7 : 25% +	3.49	1	10 - 5
"	F	75% 7 : 25% +	1.45	1	50 - 10

Key : As for Table 5.2 on the opposite page (133), with the addition that NC = not calculable.

If bands 1, 2, 3, 4, 5, 6, 7 and 10 have their patterns of inheritance into the F2 generation treated separately, it is apparent that bands 4, 5, 6, 7, and 10 are autosomal with the exception of the inheritance of band 4 into the F2 females (discussed in the following text). Bands 1, 2, and 3, by the X^2 test, show X-linked inheritance into the male and female and male F2 offspring. A X^2 value for their inheritance into the F2 female mice cannot be calculated as the calculation for this value in the lack of inheritance class requires division by zero. However, as calcuated in Table 5.1, the presence of bands 1, 2, and 3 in 97% of the female offspring suggests the presence of these bands on the C57Bl/6 X chromosome.

Table 5.2 X^2 Values and Probabilities for the Hypothesized Pattern of Inheritance of Internal $(TTAGGG)_n$ Containing Loci into the F2 Generation

Locus	Animals included	Null hypothesis + classes	X^2	d.f	p%
1,2,3	M+F	75% 1,2,3 : 25% x,x,x	1.27	1	50 - 10
	M	50% 1,2,3 : 50% x,x,x	0.595	1	50 - 10
	F	100% 1,2,3 : 0% x,x,x	0.096	1	90 - 50
4	M+F	75% 4 : 25% *	2.7	1	50 - 10
	M	" "	0.254	1	90 - 50
	F	" "	7.5	1	1 - 0.5
5,6,7	M+F	75% 5,6,7 : 25% +,+,+	0.157	1	90 - 50
	M	" "	3.111	1	10 - 5
	F	" "	1.887	1	50 - 10
10	M+F	75% 10 : 25% ~	2.165	1	50 - 10
	M	" "	0.57	1	50 - 10
	F	" "	1.716	1	50 - 10

Key :

M = Male
F = Female
x = Absence of bands from locus 1,2,3
* = Absence of band 4
+ = Absence of bands from locus 5,6,7
~ = Absence of band 10

75% of females was between 1 and 0.5%, hence dismissing this hypothesis. A possible explanation for the unusual inheritance of band 4 would predict its presence within the pseudoautosomal region of the DBA/2 sex determining chromosomes. The X and Y chromosomes have been shown to contain a homologous region known as the pseudoautosomal region. This region is involved in obligate pairing at meiosis of gametogenesis. Recombination within this region occurs at a higher frequency than for the rest of the genome causing the apparent autosomal inheritance patterns of associated markers (reviewed by Burgoyne, 1982, Pritchard and Goodfellow, 1985). The frequency of recombination within this region has been shown to be 10 fold higher in males than females (Rouyer *et al.*, 1986). Hence, the features associated with the pseudoautosomal region were applied to the inheritance of band 4 into the F2 generation. Firstly, if band 4 was present within the pseudoautsomal region and so, upon both the X and Y chromosomes, it would be expected to be inherited into 50% of females and 100% of males, or 75% of all the F2 animals, if no exchange events had occurred. As shown in Table 5.1 and summarised below, the values of inheritance into males and females varied substantially from the expected.

		Predicted %	Actual %
Male F2's	4	100	77
	*	0	23
Female F2's	4	50	63
	*	50	37
Male + Female F2's	4	75	70
	*	25	30

Where: 4 = Presence of band 4

* = Absence of band 4

Hence, from these results it would appear that band 4 is not present on the X and Y chromosomes. This is to be expected if there is a high level of exchange between the sex determining chromosomes in the progenitor F1 animals.

A second analysis that can be performed to evaluate the possibility that band 4 resides within the DBA/2 pseudoautosomal region involves determining its pattern of inheritance relative to other unlinked loci within the C57Bl/6 and DBA/2 genomes. If loci are unlinked relative to each other and autosomal, or one is X-linked and the other autosomal, inheritance of the loci relative to each other can be predicted by Mendelian inheritance. If band 4 was located within the pseudoautosomal region, it would be expected to be inherited in an autosomal fashion relative to other unlinked markers. In addition, if band 4 is not a real autosomal marker, deviations from the expected values of inheritance might be observed and would possibly provide information about the genomic location of this locus.

The predicted inheritance of two unlinked loci into the F2 generation falls into two categories:

(a) For two unlinked, autosomal loci :

A = presence of chromosome with locus A.

- = absence of chromosome containing locus A.

B = presence of chromosome containing locus B.

~ = absence of chromosome containing locus B.

	F1 male	x	F1 female
F1 parents	(A -) (B ~)		(A -) (B ~)
F1 parents	9/16 or 56.25% = A and B		
	3/16 or 18.75% = A and ~		
	3/16 or 18.75% = - and B		
	1/16 or 6.25% = - and ~		

(b) For two unlinked loci where one is present on an autosome and one is located upon the C57Bl/6 X chromosome :

BX(C) = C57Bl/6 X chromosome with the C locus

DX = DBA/2 X chromosome

DY = DBA/2 Y chromosome

D = presence of an autosome containing locus D

/ = absence of an autosome containing locus D

	F1 male	x	F1 female
F1 parents	BX(C) DY , (D /)		BX(C) DX , (D /)
F2 parents	75% of females = BX(C) , (D)		
	25% of females = BX(C) , (/)		
	37.5% of males = BX(C) , (D)		
	12.5% of males = BX(C) , (/)		
	37.5% of males = DX , (D)		
	12.5% of males = DX , (/)		

The numbers and percentages for the inheritance of band 4 relative to the 1,2,3, 5,6,7 and 10 loci are shown in Tables 5.3 and 5.4. In addition, the segregation percentages for the other combinations of loci are shown. This was used as a control for Mendelian inheritance; if the other loci were unlinked, the actual inheritance values should relate closely to the predicted values. The X^2 test was applied to the inheritance values obtained, Tables 5.5 and 5.6. For both cosegregation studies, inheritance occurred independently confirming the original observation that each locus was present upon a different chromosome in the DBA/2 and C57Bl/6 inbred strains and that they assort independently at meiosis. However, the one exception to this rule was seen with the segregation pattern of the DBA/2 derived loci, 4 and

Table 5.3 Segregation of Two, Unlinked, Autosomal, Internal $(TTAGGG)_n$ Containing Loci into the F2 Generation

Cosegregating loci	Ex.%	Total M + F		M		F	
		(178)		(84)		(94)	
		No	%	No	%	No	%
4 , 5,6,7	56.25	95	53.4	54	64.3	41	43.6
4 , +,+,+	18.75	34	19.1	10	11.9	24	25.5
* , 5,6,7	18.75	39	21.9	16	19	23	24.5
* , +,+,+	6.25	10	5.62	4	4.76	6	6.4
4 , 10	56.25	100	56.2	47	56	53	56.4
4 , ~	18.75	29	16.3	16	19.1	13	13.8
* , 10	18.75	30	16.9	15	17.9	15	16
* , ~	6.25	19	10.7	6	7.1	13	13.8
5,6,7 , 10	56.25	96	54	54	64	42	45
5,6,7 , ~	18.75	38	21	17	20	21	22
+,+,+ , 10	18.75	31	17	6	7	25	27
+,+,+ , ~	6.25	13	7.3	7	8.3	6	6.4

Key :

M = Male

F = Female

Ex.% = Predicted percentage from Mendelian inheritance

() = Total number of animals in the stated category

+,+,+ = Absence of bands 5,6,7.

* = Absence of band 4.

~ = Absence of band 10.

Table 5.4 Segregation of an X-linked and an Autosomal, Internal $(TTAGGG)_n$ Locus into the F2 Generation

Cosegre-gating loci	Total M + F (178)			M (84)			F (94)		
in M and F	Ex.%	No	%	Ex.%	No	%	Ex.%	No	%
1,2,3 + 4 F	37.5	66	37	NA	NA	NA	75	66	70.2
1,2,3 + * F	12.5	28	15.7				25	28	29.8
1,2,3 + 4 M	18.75	32	18.0	37.5	32	38.1	NA	NA	NA
1,2,3 + * M	6.25	12	6.7	12.5	12	14.3			
x,x,x + 4 M	18.75	32	18.0	37.5	32	38.1			
x,x,x + * M	6.25	8	4.5	12.5	8	9.5			
1,2,3+5,6,7 F	37.5	64	36	NA	NA	NA	75	64	68.1
1,2,3+ +,+,+F	12.5	30	16.9				25	30	31.9
1,2,3+5,6,7 M	18.75	37	20.8	37.5	37	44	NA	NA	NA
1,2,3+ +,+,+M	6.25	7	3.9	12.5	7	8.3			
x,x,x + 5,6,7M	18.75	32	18.0	37.5	32	38			
x,x,x + +,+,+M	6.25	8	4.5	12.5	8	9.5			
1,2,3 + 10 F	37.5	69	38.8	NA	NA	NA	75	69	73.4
1,2,3 + ~ F	12.5	25	14				25	25	26.6
1,2,3 + 10 M	18.75	32	18.0	37.5	32	38.1	NA	NA	NA
1,2,3 + ~ M	6.25	12	6.7	12.5	12	14.3			
x,x,x + 10 M	18.75	27	15.2	37.5	27	32.1			
x,x,x + ~ M	6.25	13	7.3	12.5	13	15.5			

Key :

M & F	=	Male and female respectively
Ex.%	=	Predicted percentage from Mendelian inheritance
No & %	=	Number and percentage of animals respectively
()	=	Total number of animals in the stated category
NA	=	Not applicable as not within stated category
* and ~	=	Absence of bands 4 and 10 respectively
+,+,+	=	Absence of bands 5,6,7
x,x,x	=	Absence of bands 1,2,3

Table 5.5 X^2 Values and Probabilities for the Segregation of Two Autosomal Unlinked Internal $(TTAGGG)_n$ Containing Loci into the F2 Generation

Cosegre-gating loci	Animals included	X^2	d.f.	p%
4 , 5,6,7	M+F	1.34	3	90-50
	M	3.35	3	50-10
	F	6.3	3	10-5
4 , 10	M+F	6.49	3	10-5
	M	0.148	3	99.5-97.5
	F	10.25	3	2.5-1
10 , 5,6,7	M+F	1.3	3	90-50
	M	7.64	3	10-5
	F	5.98	3	50-10

Key :

M = Male

F = Female

Table 5.6 X^2 Values and Probabilities for the Segregation of Autosomal and an X-linked Internal $(TTAGGG)_n$ Containing Loci into the F2 Generation

Cosegre-gating loci	Animals included	X^2	d.f.	p%
1,2,3 + 4	F	1.15	1	90-50
1,2,3 + 4	M	0.825	3	99.5-97.5
1,2,3 + 5,6,7	F	2.4	1	50-10
1,2,3 + 5,6,7	M	3.23	3	50-10
1,2,3 + 10	F	0.064	1	90-50
1,2,3 + 10	M	1.46	3	90-50

Key :

M = Male

F = Female

10. The inheritance of the two loci into the female F2 offspring resulted in a X^2 value that discounted the proposed hypothesis, Table 5.5. This reflects the aberrant pattern of inheritance detected previously, Table 5.1. Hence, overall, it would appear that the DBA/2 band 4 is inherited as would be predicted for an autosomal locus. The deviation from the expected result for inheritance of band 4 relative to 10 into F2 females may demonstrate an incomplete autosomal segregation pattern of a locus present within the pseudoautsomal region.

The possibility that band 4 may be present within the DBA/2 X and Y chromosome pseudoautosomal region, suggests that it is being transferred onto the C57Bl/6 X chromosomes which contain the 1,2,3 locus. Hence, determination of the pattern of inheritance of these two loci, relative to each other, into the F2 generation may also provide information about the genomic location of locus 4 within the DBA/2 chromosomes. Hence, the expected inheritance of these loci into the F2 generation when no exchange of markers occurs:

	F1 male	x	F1 female
F1 Parents	BX(1,2,3) , DY(4)	x	BX(1,2,3) , DX(4)
F2 offspring	BX(1,2,3) , DX(4)		BX(1,2,3) , BX(1,2,3)
	DY(4) , DX(4)		DY(4) , BX(1,2,3)

Where : BX = C57Bl/6 X chromosome
DX = DBA/2 X chromosome
DY = DBA/2 Y chromosome

Hence, the expected inheritance percentage values of loci 4 and 1,2,3 into the male and female F2 progeny would be:

Male F2 : 50% 4 , x,x,x
50% 4 ,1,2,3
Female F2 : 50% * , 1,2,3
50% 4 , 1,2,3

Where : * = Absence of band 4
x,x,x = Absence of bands 1,2,3

The actual percentage inheritance of these bands is shown in Table 5.4 and summarised below and the alternative patterns that were actually observed are also shown in Table 5.4.

		Predicted %	Actual %
Male F2 :	4 , x,x,x	50	38.1
	4 , 1,2,3	50	38.1

Female F2 :	* , 1,2,3	50	29.8
	4 , 1,2,3	50	70.2

From the above results it could be seen that the actual percentage inheritance differed dramatically from that expected. If the possibility of recombination events involving the transfer of locus 4 onto the C57Bl/6 X chromosome is considered, the possible alleles segregating at meiosis would be:

For the F1 male : DY(4) , BX(1,2,3) , DY( ) , BX(1,2,3)(4)

For the F1 female : DX(4) , BX(1,2,3) , DX( ) , BX(1,2,3)(4)

The progeny and the expected values of inheritance from such a mating between the two F1 animals and the actual values obtained are shown below:

		Predicted %	Actual %
Male F2 offspring	4 , x,x,x	37.5	38.1
	* , 1,2,3	12.5	14.3
	4 , 1,2,3	37.5	38.1
	* , x,x,x	12.5	9.5
Female F2 offspring	* , 1,2,3	25	29.8
	4 , 1,2,3	75	70.2

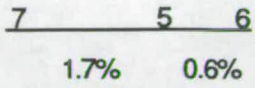
The predicted compared to obtained inheritance values seem to approximately relate, although there is some deviation. In addition, the genotypes of the offspring would be the same as if band 4 were present upon an autosome and its cosegregation compared with the X-linked locus 1,2,3, Table 5.4. Hence, band 4 inheritance can be considered either in terms of its presence upon an autosome or, as a pseudoautosomal locus where the unexpected, observed genotypes are due to exchange of band 4 onto the C57Bl/6 X chromosome. In the latter exchange events, using the inheritance values obtained, the female F2 progeny showed an increase in the number of animals with bands 4 and 1,2,3 and a decrease in those lacking 4 of ~20%. If band 4 is located within the mouse pseudautosomal region, exchange events resulting in its presence within the BX chromosome would be expected to occur at a high frequency. As these events occur seem to occur at a higher frequency in the male than the female, each of the female offspring would be expected to inherit the altered X chromosome from the male leading to a higher frequency with bands 1,2,3 and 4 than those lacking just band 4, where a recombined DBA/2 X chromsome would be inherited alongside a normal C57Bl/6 X chromosome. Male F2 progeny were obtained that lacked band 4 and also, both bands 1,2,3 and 4. This deviation from the expected would arise from the transfer of band 4 onto the C57Bl/6 derived X chromosome and subsequent inheritance of the recombined DX and DY chromosomes. Lack of band 4 with 1,2,3 present, would result from inheritance of the normal C57Bl/6 X chromosome from the F1 female and a recombined DBA/2 Y chromosome from the male.

Hence, having considered all of the possible inheritance combinations, it would appear that although the DBA/2 locus 4 shows some cases of aberrant inheritance into the F2 generation it is not possible to determine whether it is present within the pseudoautosomal region or within an autosome; this is as predicted for loci within the pseudoautosomal region. Therefore, determination of the chromosomal location of locus 4 would require more precise mapping using either recombinant inbred mouse strains, cosegregation with other markers known to be present within the DBA/2 mouse pseudoautosomal region, or by cloning the locus.

5.3.2 Recombination Events Involving the Internal $(TTAGGG)_n$ Containing Loci

From the patterns of inheritance of the DBA/2 and C57Bl/6 internal $(TTAGGG)_n$ containing loci into the F2 generation, it was possible to determine the *Hae* III restriction enzyme DNA fragments constituting each locus. For the DBA/2 specific $(TTAGGG)_4$ hybridising bands, 4 and 10, no other segregating DNA fragments could be clearly observed in the F2 offspring. Segregation analysis of these two bands into the F2 generation strongly suggested that they each represented a separate autosomal locus. Hence, recombination events involving either locus would be extremely difficult to identify as the loss of a band would be extremely difficult to separate from normal random assortment of homologous chromosomes at meiosis. The situation was different for the two C57Bl/6 loci which were each composed of three bands designated 1,2,3 and 5,6,7. Any recombination event occurring within the loci could result in separation and loss of components from the region. The 'lost' component would be transferred to the homologous chromosome involved in the recombination event.

Recombination events resulting in the loss of components from the C57Bl/6 derived loci 1,2,3 and 5,6,7 were observed. For locus 1,2,3, band 3 was lost from 5 males and 3 females out of the 178 F2 offspring. For locus 5,6,7, two types of recombination event were observed, loss of bands 5,6 leaving 7 in 1 male and two females and loss of band 6, leaving 5 and 7, in 1 male. From the number of recombination events within a multicomponent locus it is possible to calculate the frequency of recombination. The percentage recombination frequency was calculated from the number of recombination events compared to the total number of meiotic events involved in the study. Hence, the recombination frequency between bands 1,2 and 3 was 4.5%. Fewer recombination events were observed in the 5,6,7 cluster, between 5,6 and 7 a frequency of 1.7%, and 0.6% between 6 and 5,7. The presence of two different recombination events in the 5,6,7 locus allowed a tentative map of the order of components relative to each other to be established. Hence, for 5,6,7 a possible organisation would be:



From the frequency of recombination between the bands it is apparent that 5 and 6 lie closer to each other than to 7 in the C57Bl/6 genome. For the 1,2,3 locus, all that can be said at the F2 generation stage is that component 3 lies away from the closely linked elements 1 and 2.

The recombination frequencies calculated in the study could not be taken as absolutely correct. Errors can arise from loci which have a higher than normal recombination frequency. Masking of recombination events could have occurred in individuals containing a normal and an altered locus. For example, in an individual inheriting the whole 1,2,3 locus, the additional presence of a locus with only components 1,2 would be difficult to demonstrate. Recombination events within a locus that lead to band separation should result in the inheritance of both products into the offspring. However, bands 5 and 6 were never seen without 7 and band 3 was never seen without 1 and 2. The lack of inheritance could arise from either loss of the *Hae* III restriction enzyme site during the recombination process which releases 5,6 and 3 from the genomic DNA, or, masking by a second complete locus, or, by chance. The recombination events generating the observed products take place at meiosis in gametogenesis. In the female, gametogenesis results in only one viable ovum from the four meiotic products. Hence, the two products of a recombination event have only a 1/4 chance each of appearing in the F2 offspring. In the male, all four meiotic products become viable sperm but many thousands are released for the fertilisation of one ovum. Hence, the combination of all the above mentioned problems together with the analysis of only 178 F2 offspring could have produced the observed result. The problem of masking of recombination events could be solved by matings between F2 offspring. The additional random assortment of homologous chromosomes at meiosis, depending upon the relevant genetic composition of the parents, could lead to the unmasking of further recombination events in the F2 offspring. In addition, further matings might lead to new variant generation and the inheritance of previous recombination events into a new set of offspring could establish gametogenesis as the time of their generation.

5.3.3 Inheritance Patterns of Telomeric $(TTAGGG)_n$ Containing Loci into the F2 Generation

The banding pattern of telomeric $(TTAGGG)_n$ containing loci for two sets of F2 offspring was determined. Total genomic DNA was digested with the restriction enzyme *Hae* III, the fragments separated by PFGE, transferred to a nylon membrane and hybridised with the $(TTAGGG)_4$ oligonucleotide probe. The patterns of inheritance are shown in Figures 5.4a and b. The inheritance of loci from the F1 parents into the F2 offspring and the original

Figure 5.4 Inheritance of Telomeric $(TTAGGG)_n$ Containing Loci into the F2 Generation

Total genomic DNA from the grandparents, parents and offspring of an F1 x F1 pair was digested with *Hae* III, the fragments separated by PFGE, transferred to a nylon membrane and hybridised with the $(TTAGGG)_4$ oligonucleotide probe. Bands that have been inherited from parents into the F2 offspring and from grandparents to parents to F2 offspring are indicated with arrowheads. New bands, which vary in size from those in the parents are marked with a V.

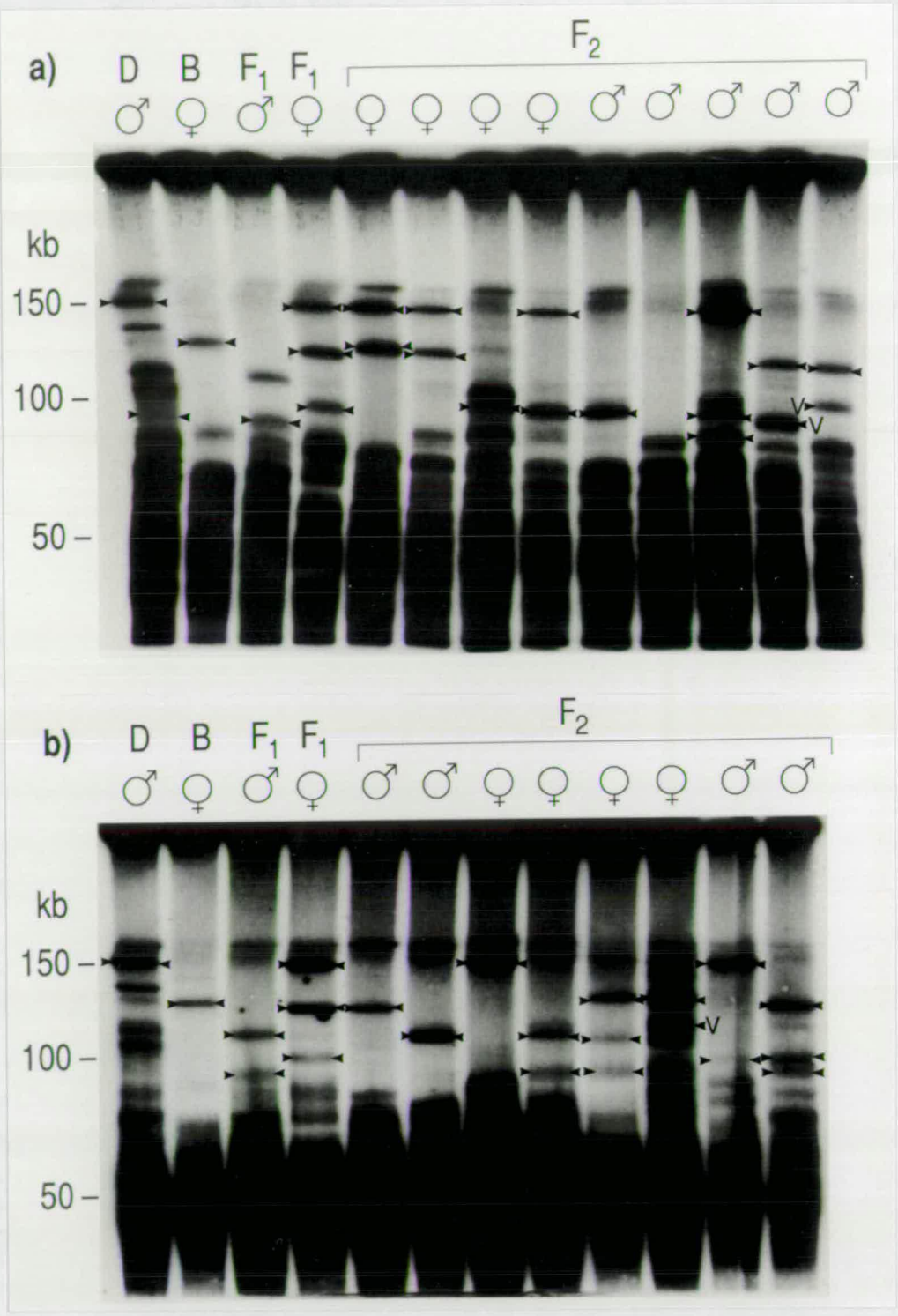
(a) Contents of tracks for band inheritance into first set of F2 offspring :

D	=	DBA/2 grandfather		2350
B	=	C57Bl/6 grandmother		2351
F1	=	First generation father		0010
"	=	First generation mother		0013
F2	=	Second generation offspring	female	0029
"	=	" "	female	0030
"	=	" "	female	0031
"	=	" "	female	0032
"	=	" "	male	0033
"	=	" "	male	0034
"	=	" "	male	0035
"	=	" "	male	0036
"	=	" "	male	0037

(b) Contents of tracks for band inheritance into second set of F2 offspring :

D	=	DBA/2 grandfather		2350
B	=	C57Bl/6 grandmother		2351
F1	=	First generation father		0010
"	=	First generation mother		0013
F2	=	Second generation offspring	male	0038
"	=	" "	male	0039
"	=	" "	female	0017
"	=	" "	female	0018
"	=	" "	female	0019
"	=	" "	female	0020
"	=	" "	male	0015
"	=	" "	male	0016

λ concatemer size markers are indicated (kb).


a)
D
B
F1
F1
F2
kb
150 –
100 –
50 –
V
V
b)
D
B
F1
F1
F2
kb
150 –
100 –
50 –
V

DBA/2 / C57Bl/6 pair, where applicable, are indicated by arrowheads. The telomeric bands were seen to be inherited from the F1 parents into some but not all of the F2 offspring, and not one of the progeny inherited a complete set of F1 derived DNA fragments. Only a limited number of bands were initially derived from the C57Bl/6 and DBA/2 progenitors, as indicated in Figures 5.4a and b. Three new variant sized bands were observed in the two pairs analysed; marked with an arrowhead and V in Figures 5.4a and b. However, it was extremely difficult to determine the exact number of new bands present as it was impossible to follow those fragments within the 50-80kb size range. Also, slight size variations were seen with bands that appeared to have been inherited from the F1 parents. This may have been due to the technique of PFGE used causing aberrant mobilities or, alternatively, the production of new variants that varied only slightly in size compared to the parental bands. However, if the three new variant bands are considered as the only such recombination events, then the frequency of new variant generation can be calculated from the following values :
17 mice were analysed in which 3 new variant bands were detected. Hence, 3 ÷ 17 or 0.176 new variant bands are present per mouse and 17.6% of mice or 8.8% of gametes will contain new variant, telomeric, DNA fragments.

This value is ~four times less than for the F1 offspring. The discrepancy may be a result of new bands possessing only a slight size variation compared to those inherited from the parents. In addition, new variant bands may exist within the 30-80kb non-resolvable region.

The inheritance patterns observed for the telomeric loci contrasted dramatically when compared to that for the internal loci. The internal $(TTAGGG)_n$ containing loci were homozygous in the C57Bl/6 and DBA/2 inbred strains resulting in the inheritance of a complete set of bands from both parents into the F1 offspring. The F1 offspring were mated and upon analysis of the F2 progeny, the internal loci were inherited as predicted by the Mendelian model; each independent locus was inherited at random, with F2 offspring containing either a complete or partial set of the parental DNA fragments. No new variant bands were seen.

From the results of telomeric $(TTAGGG)_n$ containing loci inheritance into the F2 generation, it could be concluded that:

(a) The telomeric $(TTAGGG)_4$ hybridising fragments were distinct genetic loci that were stably inherited into further generations.

(b) They were heterozygous in the DBA/2 and C57Bl/6 inbred strains and inherited at random into the F1 offspring.

(c) New variant sized bands were observed at a higher frequency than seen for the internal loci and these were stably inherited into subsequent generations.

5.4 Inheritance Patterns of Internal $(TTAGGG)_n$ Containing Loci into the F3 and F4 Generations

The patterns of inheritance of internal $(TTAGGG)_n$ into a third (F3) and fourth (F4) generation, produced by either brother x sister or cross pair mating of F2 or F3 offspring respectively, were analysed. This additional study was performed to obtain further information about recombination events involving the C57Bl/6 loci, 1,2,3 and 5,6,7, and to look for recombination events resulting in new variants differing in size from the parental bands. Total genomic DNA from the F3 and F4 offspring was digested with the restriction enzyme *Hae* III, the fragments separated by conventional gel electrophoresis, transferred to a nylon membrane and hybridised with the $(TTAGGG)_4$ probe. Each mating experiment at this stage had to be treated independently, unlike the situation for the F2's. This is due to the random assortment of homologous chromosomes at meiosis producing offspring which are either homozygous or heterozygous for the presence or homozygous for the absence of either of the four loci. Therefore, each F2 or F3 parent could differ from the animal with which it is being mated, unlike the case in all of the F1 animals giving rise to the F2 offspring.

For the generation of F3 offspring, four matings between F2 animals were set up, each involving different F2 parents. Due to the individuality of the crosses, each was considered independently. In Appendix 1, at the end of this chapter, are summaries of band inheritance in Pairs A, B, C and D. As for the F2's, the presence of a band corresponding to either of the four loci is represented by a number *ie* 1,2,3 , 4 , 5,6,7 , 10 and its absence by the symbols x,x,x , * , +,+,+ , ~ respectively. From the numbers of offspring inheriting each locus, together with the percentages and X^2 values it was possible to predict the complete band composition of each F2 parent, Appendix 1.

For the generation of the F4 offspring, matings to establish the nature of recombination events and the new variants within Pair D were set up between three separate sets of F3 animals, *ie* Pairs E, F and G. Again both cross pair and brother x sister matings were analysed and summaries of band inheritance are shown in Appendix 1. Matings were also set up between F3 animals from Pairs A, B and C to determine if the predicted parental band composition was correct and to look for 'hidden' recombination products in the F3 offspring. Two examples of such pairs, *ie* Pairs I and J, are shown in Appendix 1.

5.5 Summary of Recombination Events and New Variant Generation within the Internal $(TTAGGG)_n$ Containing Loci

5.5.1 New Variant Generation

Upon analysis of the F3 offspring, two new variant bands were observed, NV2 and NVQ, Figures 5.5 and 5.6 in Appendix 1; no new variants were detected in the F4 offspring. NV2 was shown to be the only new variant sized band within the F2 generation. An estimate of the frequency of new variant generation could be determined using the same method described for the telomeric loci. Hence, 1 new variant band was observed in 178 F2 offspring, so there are 1 ÷ 178 or 0.0056 new variant bands per mouse, and so, 0.56% of mice or 0.28% of gametes will display new variant sized, interstitial, $(TTAGGG)_n$ containing, DNA fragments.

A second new variant sized band, NVQ, was generated during the production of a total of 68 F3 offspring, Figure 5.6, Appendix 1. An estimate of new variant generation would be : 1 new variant in 68 mice, therefore, 1 ÷ 68 x 100 % or 1.47% of mice or 0.74% of gametes will display a new variant sized, interstitial, $(TTAGGG)_n$ containing, DNA fragments.

The slightly increased frequency of new variant generation for the F3 compared to F2 offspring is probably due to the difference in numbers of animals analysed. The production and analysis of further F3 animals would resolve this question. However, if an average of the two values, 1.015% of mice or 0.51% of gametes, is compared to the values obtained for new variant generation at the telomeric loci of 66% of mice or 33% of gametes and 17.6% of mice or 8.8% of gametes for F1 and F2 offspring respectively, it can be seen that there is an ~65 or 17 fold increased frequency of new variant band formation at mouse telomeres compared to interstitial loci. These values are not entirely correct as new variant generation within the interstitial loci has not been calculated at the F1 progeny stage. However, the values do show a marked elevation in new variant generation at telomeric compared to interstitial $(TTAGGG)_n$ containing loci. Possible reasons for this phenomenon will be discussed in Chapter 8.

5.5.2 Recombination Events within the C57Bl/6 1,2,3 and 5,6,7 Interstitial Loci

Analysis of the F2 offspring revealed recombination events resulting in the loss of components from the C57Bl/6 derived 1,2,3 and 5,6,7 loci. For locus 1,2,3, band 3 was lost from 8 out of the 178 F2 progeny. For locus 5,6,7, two types of recombination event were observed, loss of bands 5 and 6 leaving 7 in 3 animals and loss of band 6 leaving 5 and 7 in a separate F2 animal. No further such recombination events for locus 5,6,7 were observed in the F3 and F4 progeny. However, analysis of the F3 and F4 offspring generated from

matings between a limited number of F2 animals revealed additional recombination events involving the C57Bl/6 1,2,3 locus. One such event was revealed in the F2 female of Pair C which was shown to contain a recombined X chromosome with bands 1 and 2 of the 1,2,3 locus. The presence of this event was masked by the normal 1,2,3 locus upon the second X chromosome within this female. Similar recombination events may have been masked in the same way in other F2 offspring, so preventing a correct estimation of the recombination frequency between bands 1, 2 and 3. However, considering the nine events as the total of this type in the F2 offspring, then the recombination frequency between 1,2 and 3 would be ~5%.

An additional recombination event was observed where band 1 was separated from 2 and 3 and appeared to have been transferred to the DBA/2 Y chromosome of the male of Pair D, Appendix 1. Again this event may also have been present within other F2 animals with band 1 either remaining upon the X chromosome or being transferred to the X or Y chromosome involved in the recombination event, its presence being masked by the coinheritance of a complete, 1,2,3, or partial, 1,2, locus. However, an estimation of the recombination frequency between bands 1 and 2 can be made, being ~0.6%.

Recombination events resulting in the separation of bands 1, 2 and 3 were also observed in the production of the F3 and F4 offspring. In two individuals, an F3 animal from Pair D and F4 animal from Pair J, band 2 was observed without 1 and 3. An idea of the order of bands 1, 2 and 3 relative to each other can be determined by looking at the genetic composition of the parents. Hence, in Pair D, the recombination event would have taken place within the male F2 parent, which contained bands 1 and 2 without 3. While in Pair J, the recombined locus would be present within the female which was heterozygous for locus 1,2,3. Hence, the order of bands would most likely be 2,1,3 or :

2 1 3

0.6% ~5%

Such an arrangement would allow separation of band 3 from 1 and 2 and bands 1 and 3 from 2 by one single crossover event for each case. An exception to this rule was provided by the F2 male of Pair D which appeared to contain a Y chromosome with only band 1 from the C57Bl/6 1,2,3 locus. As it is assumed that the F1 parents of the mouse did not contain chromosomes recombinant within the 1,2,3 locus, it would appear that two recombination events must have occurred to produce the Y chromosome. Alternatively, the prediction of band order could be inaccurrate.

A precise value for recombination events within the 1,2,3 and 5,6,7 loci could be determined by mating each animal from the F2, F3 and F4 generations with a DBA/2 mouse. This mouse strain does not contain either loci, and so the band composition of each sex

chromosome from the afore mentioned offspring could be visualised in its heterozygous state.

5.6 Summary

By following the inheritance of telomeric and interstitial $(TTAGGG)_n$ containing loci into the F1, F2, F3 and F4 generations of an initial DBA/2 x C57Bl/6 mating three observations can be made. Firstly, they were inherited as distinct genetic loci and secondly, the internal loci were homozygous within both the DBA/2 and C57Bl/6 inbred strains as shown by their inheritance into all of the F1 offspring, Figure 5.1. However, the telomeric loci were heterozygous within the inbred strains and hence, inherited into some but not all of the F1 offspring, Figure 5.2 a, b and c. Thirdly, the telomeric bands observed appeared to be inherited independently of each other into the F1 and F2 offspring (Figures 5.2a, b and c, 5.4a and b) suggesting that each represented a distinct locus. The same was true for the DBA/2 interstitial loci represented by bands 4 and 10. In the C57Bl/6 genome, the two main interstitial loci were each composed of three $(TTAGGG)_4$ hybridising bands designated 1,2,3 and 5,6,7.

Analysis of the inheritance of telomeric and interstitial $(TTAGGG)_n$ containing loci into the offspring of an initial DBA/2 and C57Bl/6 mating allowed an estimation of the rate of new variant generation to be determined. For the internal loci, two new variants were observed, NV2 in the F2, and NVQ in the F3, offspring and the overall rate of new variant generation was calculated to be 1.015% per mouse or 0.51% per gamete. For the telomeric loci, the overall rate of new variant generation in F1 and F2 offspring is (66 + 17.6) + 2 = 41.8% per mouse, or (33 + 8.8) + 2 = 20.9% per gamete. From the calculated rates of new variant generation, it could be seen that this process was occuring at a much higher frequency at the telomeric compared to the internal loci. The discrepancy may reflect the positions of each type of locus within the mouse chromosomes (to be discussed in Chapter 8).

The two new variants observed at interstitial loci, NV2 and NVQ, gave a much weaker hybridisation signal compared to bands 1, 2, 4, 5, 6, 7 and 10 but about the same signal as band 3, Figures 5.1, 5.5, 5.6, 5.7, 5.8. It is possible that band 3 represents a single block of $(TTAGGG)_n$ repeats, while the other DNA fragments contain varying numbers of blocks of repeat in tandem with *Hae* III sites distributed at set intervals, which upon cleavage give rise to the identically sized bands. The new variants would, therefore, also represent a single block of repeat. Alternatively, the C57Bl/6 and DBA/2 derived $(TTAGGG)_n$ hybridising fragments could all represent single bands with varying amounts of the *Trypanosoma*-like telomeric repeat. For this to be the case, some other form of *Hae* III site devoid DNA sequence would also need to be present. Variations in hybridisation signal intensity were

observed for all of the telomeric loci and new variant sized bands looked at in this study, Figures 5.2a, b and c and 5.4a and b. This might reflect differences in the amount of $(TTAGGG)_n$ repeat present at mouse telomeres. In addition, variations in signal intensity tended to be incomparable with the size of the telomeric bands such that large DNA fragments could possess a weaker signal than smaller bands, Figures 5.2a, b and c and 5.4a and b. This would suggest that some other form of simple repeat is present, to varying degrees, within the telomeric loci. Investigation into the sequence composition of the telomeric and interstitial loci is addressed in Chapter 6.

Finally, from the patterns of inheritance of the C57Bl/6 and DBA/2 $(TTAGGG)_n$ containing loci into several generations of offspring it was possible to localise the 1,2,3 locus to the C57Bl/6 X chromosome and speculate that the DBA/2 derived band 4 resides within the pseudoautosomal region of the DBA/2 X and Y chromosomes. More extensive mapping of the interstitial loci was carried out in recombinant inbred and backcross mice as described in Chapter 7.

Appendix 1 Summaries of the Inheritance of Interstitial $(TTAGGG)_n$ Containing Loci into the F3 and F4 Generations in Pairs A, B, C, D, E, F, G, I and J.

Pair A (data not shown)

Type of mating: Cross pair

Banding pattern observed in parent:

F2 female 0026: 1,2,3 4 5,6,7 10

F2 male 0016: 1,2,3 * 5,6,7 10

Number of F3's analysed:

Males + Females = 21

Males = 10

Females = 11

Number of F3's inheriting each locus:

	10 males	11 Females
1,2,3	4	11
4	10	11
5,6,7	9	9
10	5	9

Total band composition of F2 parents:

F2 female 0026: BX(1,2,3) , DX 4 , 4 5,6,7 , +,+,+ 10 , ~

F2 male 0016: BX(1,2,3) , DY * , * 5,6,7 , +,+,+ 10 , ~

Recombination events and new variants:

None were observed

Pair B (data not shown)

Type of mating: Brother x sister

Banding pattern observed in parents:

F2 female 0020: 1,2,3 4 5,6,7 10

F2 male 0015: 1,2,3 * +,+,+ ~

Number of F3's analysed:

Males + Females = 9

Males = 6

Females = 3

Number of F3's inheriting each locus:

	6 males	3 females
1,2,3	4	3
1,2,x	1	0
4	2	1
5,6,7	3	1
10	6	3

Total band composition of F2 parents:

F2 female 0020: BX(1,2,3) , DX 4 , * 5,6,7 , +,+,+ 10 , 10

F2 male 0015: BX(1,2,3) , DY * , * +,+,+ , +,+,+ 10 , 10

Recombination events and new variants:

One recombinant was seen in the 1,2,3 cluster resulting in a male F3 with bands 1 and 2 but not 3. This event could have taken place during gametogenesis within either the F1 or F2 parents. In the F1 parent recombination in the 1,2,3 locus could have resulted in exchange of bands 1 and 2 or band 3 to either the DBA/2 X or Y chromosome; both events would be masked in the F2 animal due to the presence of the C57Bl/6 BX(1,2,3) chromosome. Inheritance into the F3 generation from the above mating would result in 25% of the offspring containing the altered locus. Only 9 animals were studied for this pair and one showed the 1,2 locus. A single recombination event in the female F2 parent could also have led to its presence in one male F3 mouse. Hence, the actual origin of the event could not be determined from the results. Analysis of further F3 offspring would establish the exact origin of the recombination event.

Pair C (data not shown)

Type of mating: Brother x sister

Banding pattern observed in parents:

F2 female 0028: 1,2,3 4 5,6,7 ~

F2 male 0021: x,x,x 4 5,6,7 10

Number of F3's analysed:

Males + Females = 18

Males = 14

Females = 4

Number of F3's inheriting each locus:

	14 males	4 females
1,2,3	10	2
1,2,x	4	2

4	8	2
5,6,7	11	3
10	6	2

Total band composition of F2 parents:

F2 female 0028: BX(1,2,3) , X(1,2,x) 4 , * 5,6,7 , +,+,+ 10 , ~

F2 male 0021: DX , DY 4 , * 5,6,7 , +,+,+ ~ , ~

Recombination events and new variants:

A total of 6 of the F3 offspring inherited the recombinant locus 1,2,x, suggesting that the actual event had occurred within the F1 parents. The chromosome containing the altered locus had to have been derived from the F2 female 0028 as the male F2 was devoid of any of these bands. The F2 female presented with bands 1,2 and 3. From analysis of the F3 offspring it was clear that this female contained a normal, 1,2,3, and an altered 1,2,x, X-linked locus.

Pair D (Figures 5.5 and 5.6)

Type of mating: Cross pair

Banding pattern observed in parents

F2 female 0017: 1,2,3 * 5,6,7 ~

F2 male 0023: 1,2,x 4 5,6,7 10

Number of F3's analysed:

Males + Females	=	20
Males	=	10
Females	=	10

Number of F3's inheriting each locus:

	10 males	10 females
1,2,3	6	7
1,2,x	0	2
x,2,x	0	1
x,NV2,x	1	0
1,NV2,x	3	0
4	8	2
5,6,7	10	10
10	10	9
NVQ	0	1

Where NV2 represents a new variant of slightly larger size than the normally inherited band 2, Figure 5.5, F3 animals 0042, 0044, 0047, and NVQ is a second new variant found in only one individual, Figure 5.6, F3 animal 0104.

Figure 5.5 Inheritance Pattern of NV2 into the F3 Generation

Total genomic DNA from the great grandparents, grandparents, parents and F3 offspring of Pair D was digested with *Hae* III, the fragments separated by conventional gel electrophoresis, transferred to a nylon membrane and hybridised with the $(TTAGGG)_4$ oligonucleotide probe.

Contents of tracks :

D	=	DBA/2 great grandfather		2350
B	=	C57Bl/6 great grandmother		2351
F1	=	First generation grandfather		0008
"	=	First generation grandmother		0012
"	=	First generation grandmother		0014
"	=	First generation grandfather		0011
F2	=	Second generation mother		0017
"	=	Second generation father		0023
F3	=	Third generation offspring	female	0040
"	=	" "	male	0041
"	=	" "	male	0042
"	=	" "	male	0044
"	=	" "	female	0046
"	=	" "	male	0047
"	=	" "	female	0048
"	=	" "	female	0049

Numbers and sizes in kb are indicated for each clearly definable C57Bl/6 and DBA/2 hybridising band. The size of the new variant band NV2 is also indicated.

Hind III digested λ DNA size markers are indicated (kb).

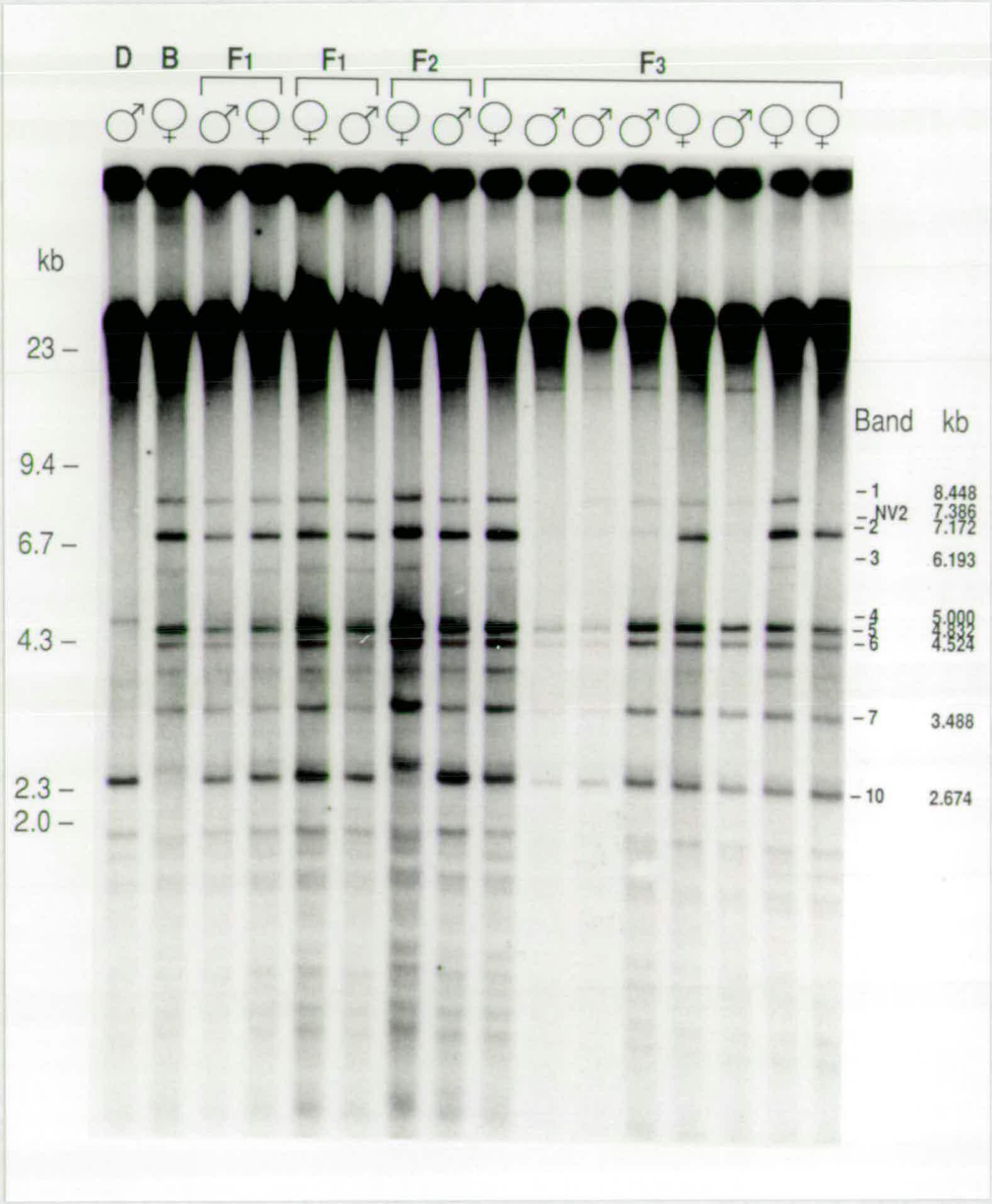

D
B
F1
F1
F2
F3
kb
23 –
9.4 –
6.7 –
4.3 –
2.3 –
2.0 –
Band
kb
– 1
8.448
– NV2
7.386
– 2
7.172
– 3
6.193
– 4
5.000
– 5
4.832
– 6
4.524
– 7
3.488
– 10
2.674

Figure 5.6 Inheritance Pattern of NVQ into the F3 Generation

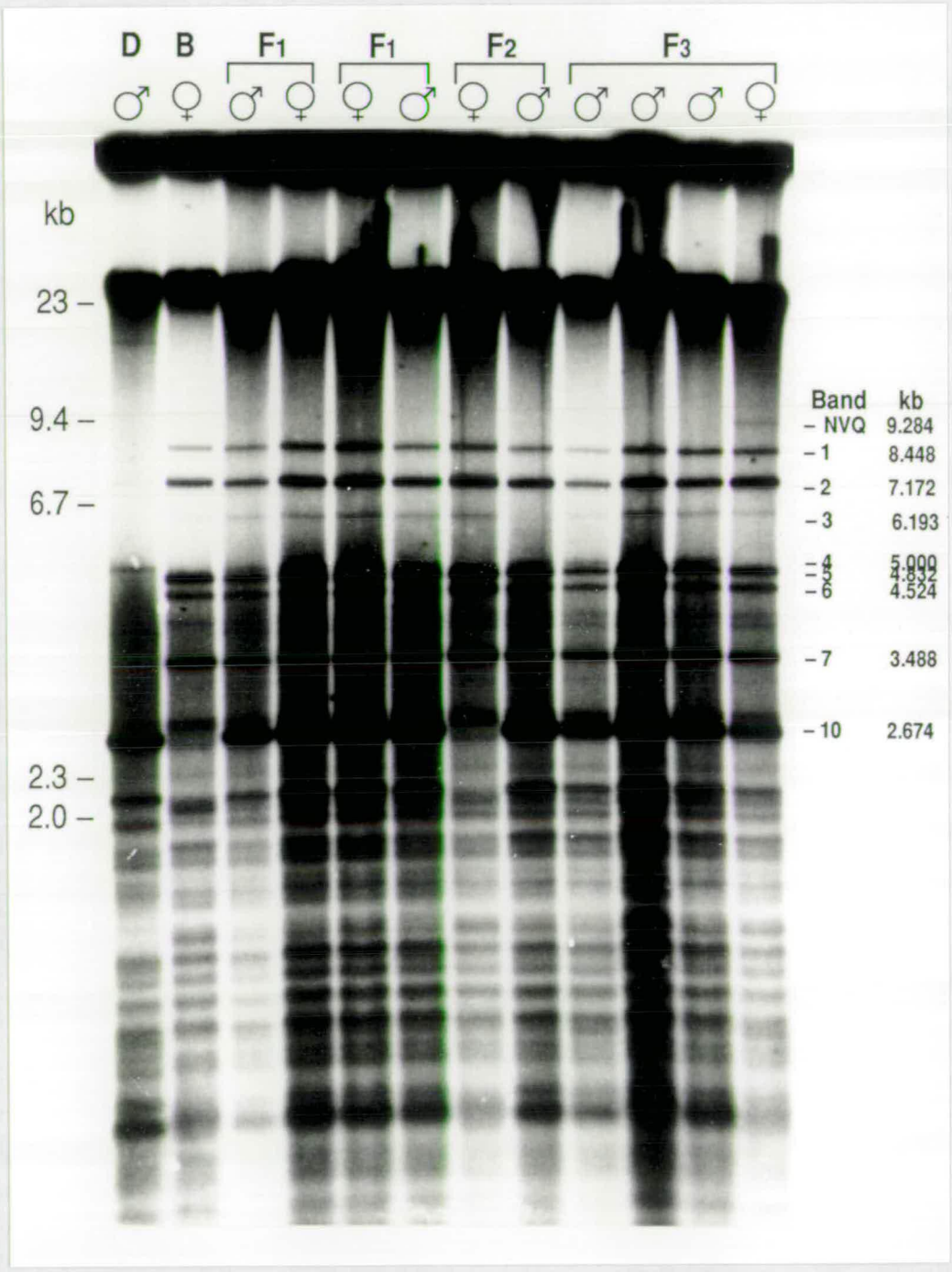
Total genomic DNA from the great grandparents, grandparents, parents and F3 offspring of Pair D was digested with *Hae* III, the fragments separated by conventional gel electrophoresis, transferred to a nylon membrane and hybridised with the $(TTAGGG)_4$ oligonucleotide probe.

Contents of tracks :

D	=	DBA/2 great grandfather		2350
B	=	C57Bl/6 great grandmother		2351
F1	=	First generation grandfather		0008
"	=	First generation grandmother		0012
"	=	First generation grandmother		0014
"	=	First generation grandfather		0011
F2	=	Second generation mother		0017
"	=	Second generation father		0023
F3	=	Third generation offspring	male	0160
"	=	" "	male	0161
"	=	" "	male	0162
"	=	" "	female	0104

Numbers and sizes in kb are indicated for each clearly definable C57Bl/6 and DBA/2 hybridising band. The size of the new variant band NVQ is also indicated.

Hind III digested λ DNA size markers are indicated (kb).


D
B
F1
F1
F2
F3
kb
23 –
9.4 –
6.7 –
2.3 –
2.0 –
Band
kb
– NVQ 9.284
– 1 8.448
– 2 7.172
– 3 6.193
– 4 5.000
– 5 4.832
– 6 4.524
– 7 3.488
– 10 2.674

Total band composition of F2 parents:

F2 female 0017: BX(1,2,3) , X(NV2) * , * {5,6,7 , 5,6,7 or 5,6,7 , +,+,+} ~ , ~

F2 male 0023: X(1,2,x) , DY(1,x,x) 4 , * {5,6,7 , 5,6,7 or 5,6,7 , +,+,+} 10 , 10

Recombination events and new variants:

Several recombination events and two new variant sized $(TTAGGG)_4$ hybridising DNA fragments were observed in the F3 progeny of the pair D mating. Firstly, the C57Bl/6 X-linked locus 1,2,3 was seen to be inherited in this form, as bands 1,2 without 3 and band 2 without 1 and 3. The recombined locus 1,2,x was derived from the male F2 parent 0023. Its observed inheritance into the two F3 females determines its presence within the X chromosome of the F2 parent. If this recombined locus was inherited with the complete 1,2,3 locus from the female F2 parent its presence would be masked. Hence, it is probable that the other F3 female offspring contain the recombined locus. The recombined locus x,2,x was only observed in one female. Again, a true estimation of its frequency within the offspring could not be determined due to masking if the complete 1,2,3 or partial 1,2,x loci were present. Secondly, two new variants were observed, NVQ and NV2. NVQ was observed in only one F3 female and was larger in size than all other $(TTAGGG)_4$ hybridising bands. Hence, it would seem that it was generated within one of the F2 parents. This individual was subsequently mated with a male F3 to determine if NVQ was present upon an autosome or sex chromosome and if it was stably inherited into the next generation, Pair G. NV2 was seen to be inherited into a total of four males, three of which also inherited band 1 from locus 1,2,3, Figure 5.5. The presence of this new variant only within male offspring suggested that it had been derived from the F2 0017 female and that it was present upon one of the X chromosomes. In addition, its presence within more than one of the F3 offspring suggests that it was originally generated within the F1 pair from which the 0017 female was derived. NV2 was not observed within the F2 female 0017 as it varies only slightly in size compared to band 2. However, in this female there is a broader region of hybridisation at the band 2 position, Figure 5.5. This may be caused by the presence of NV2. Hence, its presence was masked when inherited with the 1,2,3 or 1,2,x loci. NV2 was inherited independently of 1,2,3 and 1,2,x, but with band 1 in three of the four males. This suggests that band 1 may also have been present upon the X chromosome containing NV2 or alternatively, on a separate chromosome. However, the inheritance of band 1 into three males suggests that it is present upon the Y chromosome of 0023. Hence, the composition of the X and Y chromosomes of the F2 parental and F3 offspring would be :

Parents : F2 female = BX(1,2,3) , XNV2

F2 male = X(1,2,x) , DY(1,x,x)

Offspring : F3 female = BX(1,2,3) , X(1,2,x) or

X(NV2) , X(1,2,x)

F3 male = BX(1,2,3) , DY(1,x,x) or

X(NV2) , DY(1,x,x)

Unfortunately, two unexpected results were seen in the F3 offspring. One female inherited only band 2 and one male only the new variant, NV2. Both results suggested that either further recombination events had occurred at meiosis within the F2 parents or that the hypothesis of parental band composition was incorrect. To determine which alternative was correct several matings between F3 offspring were arranged. Inheritance into the next generation would help to establish the X and Y linked loci each had inherited and determine the predicted band composition of the parents.

Pair E (Figure 5.7)

Type of mating: Cross pair, Pair A male x Pair D female

Banding pattern observed in parents:

F3 female 0677: 1,2,x * 5,6,7 10

F3 male 0667: x,x,x 4 5,6,7 ~

Numbers of F4's analysed:

Males + Females	=	9
Males	=	7
Females	=	2

Number of F4's inheriting each locus:

	7 males	2 females
1,2,3	0	0
1,2,x	4	1
NV2	3	1
4	3	1
5,6,7	4	2
10	5	1

Total band composition of F3 parents:

F3 female 0677: {BX(1,2,x) or DX(1,2,x)} , {BX(x,x,x,NV2) or DX(NV2)} * , * 5,6,7 , +,+,+ 10 , ~

F3 male 0667: {BX(x,x,x) or DX} , DY 4 , * 5,6,7 , +,+,+ ~ , ~

Recombination events and new variants:

No additional recombination events were observed. The F4 offspring showed patterns of inheritance which allowed predictions of the parental band composition to be formulated. The banding patterns of both F3 parents agreed with the possible offspring derived from mating the F2 parents of Pair D.

Figure 5.7 Inheritance Pattern of NV2 into the F4 Generation

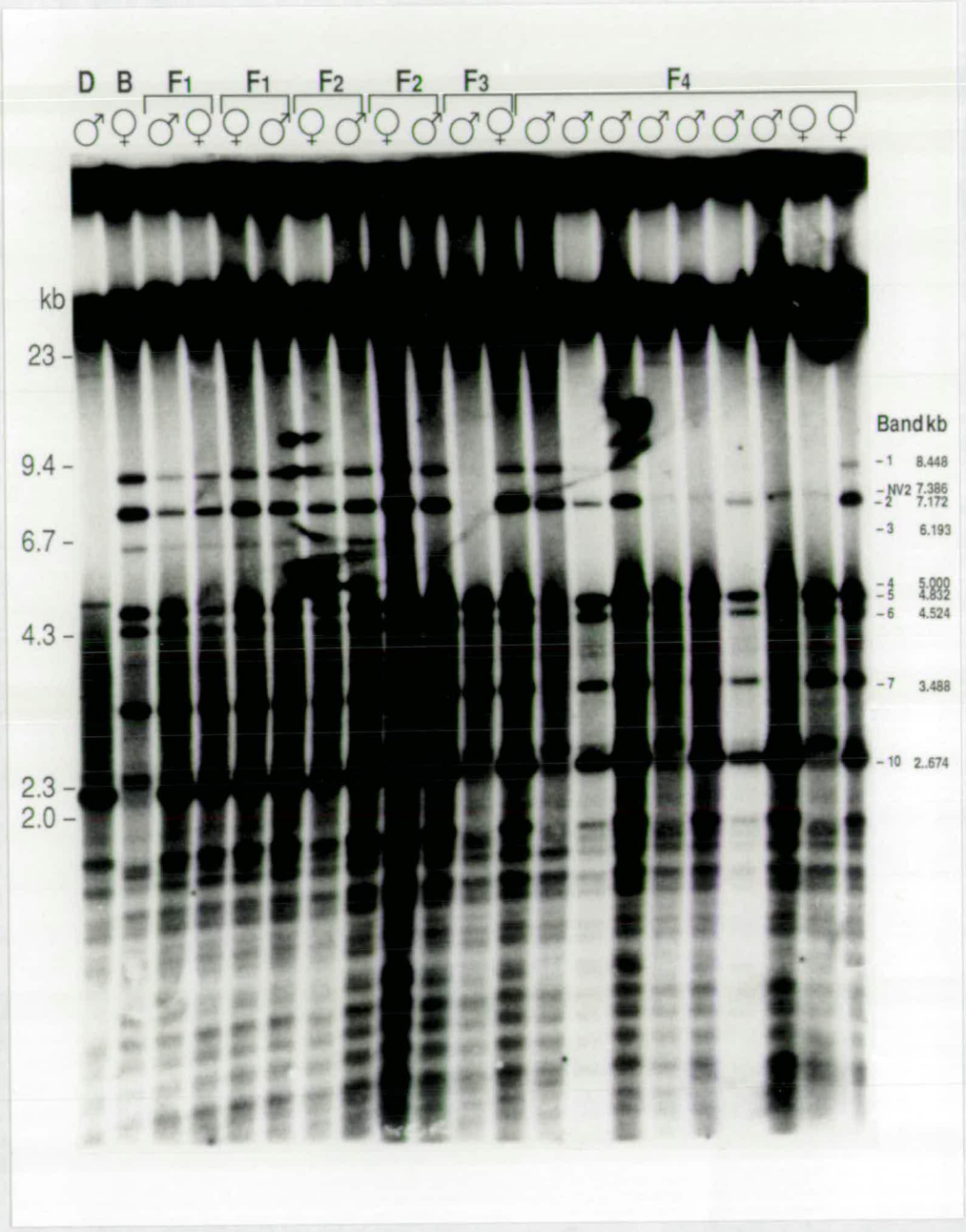
Total genomic DNA from the great, great grandparents, great grandparents, grandparents, parents and F4 offspring of Pair E was digested with *Hae* III, the fragments separated by conventional gel electrophoresis, transferred to a nylon membrane and hybridised with the $(TTAGGG)_4$ oligonucleotide probe.

Contents of tracks :

D	=	DBA/2 great, great grandfather		2350
B	=	C57Bl/6 great, great grandmother		2351
F1	=	First generation great grandfather		0008
"	=	First generation great grandmother		0012
"	=	First generation great grandmother		0014
"	=	First generation great grandfather		0011
F2	=	Second generation grandmother		0017
"	=	Second generation grandfather		0023
"	=	Second generation grandmother		0026
"	=	Second generation grandfather		0016
F3	=	Third generation father		0667
"	=	Third generation mother		0677
F4	=	Fourth generation offspring	male	0200
"	=	" "	male	0201
"	=	" "	male	0202
"	=	" "	male	0203
"	=	" "	male	0204
"	=	" "	male	0205
"	=	" "	male	0206
"	=	" "	female	0207
"	=	" "	female	0208

Numbers and sizes in kb are indicated for each clearly definable C57Bl/6 and DBA/2 hybridising band. The size of the new variant band NV2 is also indicated.

Hind III digested λ DNA size markers are indicated (kb).


D B F1 F1 F2 F2 F3 F4
kb
23
9.4
6.7
4.3
2.3
2.0
Band kb
1 8.448
NV2 7.386
2 7.172
3 6.193
4 5.000
5 4.832
6 4.524
7 3.488
10 2..674

Pair F (data not shown)

Type of mating: Brother x sister, Pair D male x female

Banding pattern observed in parents:

F3 female 0675: 1,2,3 * 5,6,7 ~

F3 male 0669: 1,2,3 * 5,6,7 ~

Numbers of F4's analysed:

Males + Females	=	21
Males	=	10
Females	=	11

Number of F4's inheriting each locus:

	10 males	11 females
1,2,3	5	11
1,2,x	5	0
4	0	0
5,6,7	10	11
10	7	4

Total band composition of F3 parents:

F3 female 0675: BX(1,2,3) , {(BX(1,2,x) or DX(1,2,x)} * , * {5,6,7 , 5,6,7 or 5,6,7 , +,+,+} ~ , ~

F3 male 0669: BX(1,2,3) , DY(1,x,x) * , * {5,6,7 , 5,6,7 or 5,6,7 , +,+,+} 10 , ~

Recombination events and new variants:

No additional recombination events were observed. Again, the F4 offspring analysed suggested that the predictions for band composition in the F2 Pair D parents was correct. Unfortunately, it was not possible to determine if the the male F3 had inherited the DY(1,x,x) chromosome due to masking by the inheritance of either the 1,2,3 or 1,2,x loci.

Pair G (Figure 5.8)

Type of mating: Brother x sister, Pair D male x female

Banding pattern observed in parent:

F3 female 0104: 1,2,3 * 5,6,7 10 NVQ

F3 male 0671: 1,2,3 * 5,6,7 10

Number of F4's analysed:

Males + Females	=	10
Males	=	3
Females	=	7

Figure 5.8 Inheritance Pattern of NVQ into the F4 Generation

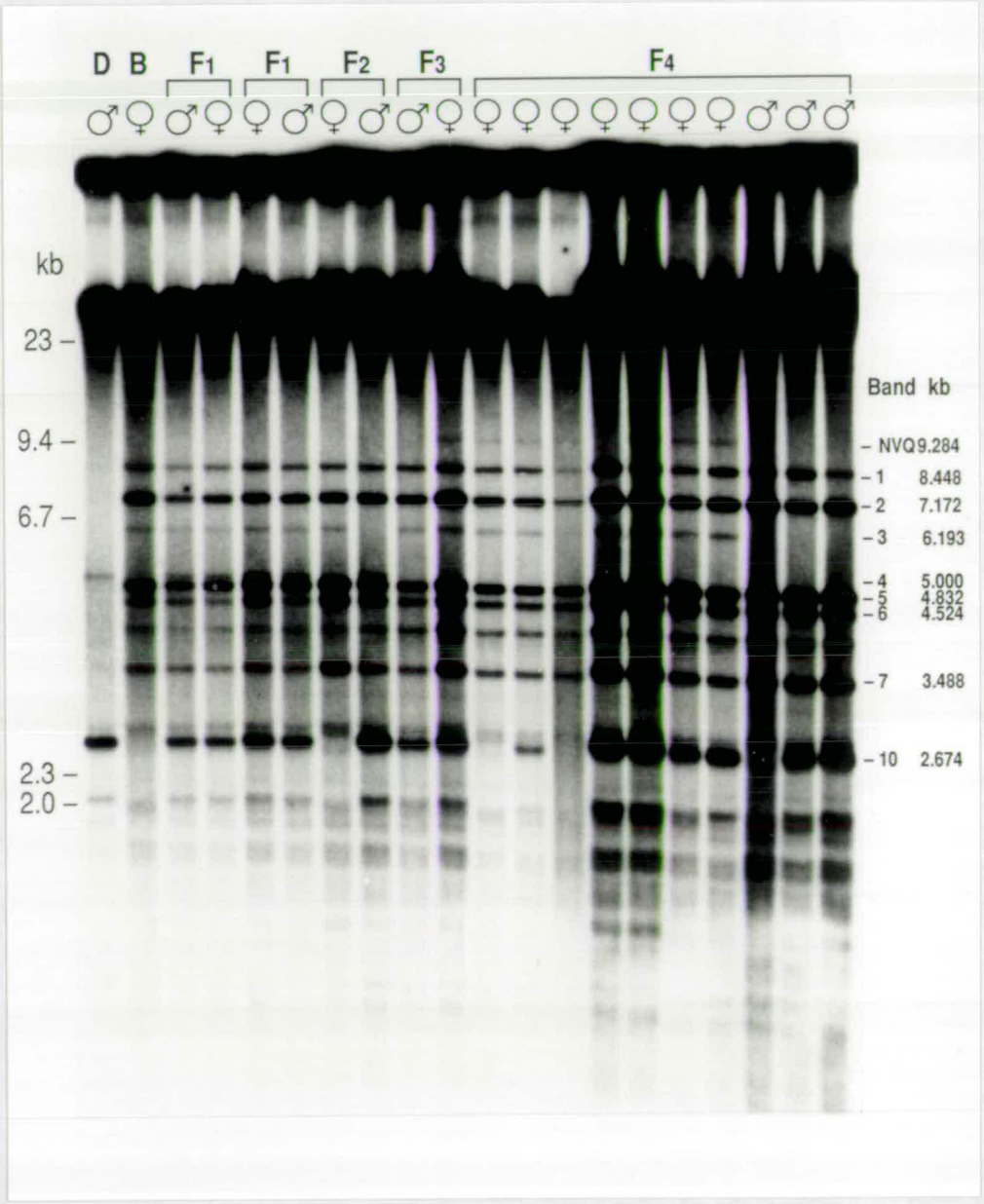
Total genomic DNA from the great, great grandparents, great grandparents, grandparents, parents and F4 offspring of Pair G was digested with *Hae* III, the fragments separated by conventional gel electrophoresis, transferred to a nylon membrane and hybridised with the $(TTAGGG)_4$ oligonucleotide probe.

Contents of tracks :

D	=	DBA/2 great, great grandfather		2350
B	=	C57Bl/6 great, great grandmother		2351
F1	=	First generation great grandfather		0008
"	=	First generation great grandmother		0012
"	=	First generation great grandmother		0014
"	=	First generation great grandfather		0011
F2	=	Second generation grandmother		0017
"	=	Second generation grandfather		0023
F3	=	Third generation father		0671
"	=	Third generation mother		0104
F4	=	Fourth generation offspring	female	0122
"	=	" "	female	0123
"	=	" "	female	0124
"	=	" "	female	0125
"	=	" "	female	0126
"	=	" "	female	0127
"	=	" "	female	0128
"	=	" "	male	0131
"	=	" "	male	0132
"	=	" "	male	0133

Numbers and sizes in kb are indicated for each clearly definable C57Bl/6 and DBA/2 hybridising band. The size of the new variant band NVQ is also indicated.

Hind III digested λ DNA size markers are indicated (kb).


D B F1 F1 F2 F3 F4
kb
23 –
9.4 –
6.7 –
2.3 –
2.0 –
Band kb
– NVQ 9.284
– 1 8.448
– 2 7.172
– 3 6.193
– 4 5.000
– 5 4.832
– 6 4.524
– 7 3.488
– 10 2.674

Number of F4's inheriting each locus:

	3 males	7 females
1,2,3	0	7
1,2,x	3	0
4	0	0
5,6,7	3	7
10	2	5
NVQ	0	6

Total band composition of F3 parents:

F3 female 0104: BX(1,2,3) , X(1,2,x) * , * {5,6,7 , 5,6,7 or 5,6,7 , +,+,+} 10 , ~ NVQ

F3 male 0671: BX(1,2,3) , DY(1,x,x) * , * {5,6,7 , 5,6,7 or 5,6,7 , +,+,+} 10 , ~

Recombination events and new variants:

No additional recombination events were observed. The F4 offspring demonstrated patterns of inheritance which correlated with the predicted band composition of the F2 parents. The new variant NVQ was inherited into the F4 generation, which suggested that it had been produced in meiosis at the F2 stage. It was inherited into 6 of the 7 females and none of the 3 males, Figure 5.8. It seemed that NVQ had been generated upon an autosome as it appeared to be inherited entirely at random. Again, it was not possible to determine if the F3 male 0671 had inherited the DY(1,x,x) recombinant due to the co-inheritance of either BX(1,2,3) or X(1,2,x).

From the results of the F3 matings it appeared that :

(1) The new variant NV2 had been generated in the meiotic events of the relevant F1 animals and was present, but masked by band 2 of the BX(1,2,3) locus within the F2 female 0017.

(2) The F2 female, 0017, had inherited a complete C57Bl/6 X-linked BX(1,2,3) locus.

(3) The F2 male, 0023, had inherited an X chromosome with the recombinant locus 1,2,x.

(4) It could not be determined from the crosses performed if the male 0023 had band 1 alone from locus 1,2,3 present upon its Y chromosome. Its presence would have been masked by the inheritance of an X chromosome containing 1,2,3 or 1,2,x. However, in view of the results obtained for the F3 offspring it seems quite possible for the DY chromosome to have gained the locus 1,2,3 derived band 1. Taken as a whole the results obtained suggest that the predicted band composition of the Pair D parents was, in fact, correct.

The second new variant, NVQ, generated by a meiotic event within one of the Pair D F2 parents, was inherited into 6 out of 7 female, and no male, F4 offspring. It was larger in size, 9.284kb, than any of the other DBA/2 or C57Bl/6 derived $(TTAGGG)_4$ hybridising DNA fragments observed. Hence, its presence was not masked in any way. If the variant had been

located upon either an autosome or X chromosome and heterozygous, inheritance into 50% of both males and females would have been expected. The results obtained suggest that either of these alternatives could be correct.

The predicted patterns of inheritance from the F2 parents were not strictly adhered too. It appeared that two of the offspring had inherited either an X or a Y chromosome which had undergone further recombination events within the F2 parents. One female F3 inherited only band 2 of the 1,2,3 locus suggesting a recombination event between either the X chromosomes of the F2 female 0017 or the X and Y chromosome of the F2 male, 0023. It was not possible to determine from which animal the recombined chromosome was derived. A male F3 inherited NV2 without band 1 or 1,2,3. The chromosome with NV2 was inherited from the F2 female, 0017. Hence, a recombination event must have occurred in the 0023 male at meiosis resulting in loss of band 1 from the DY chromosome.

Pair I (data not shown)

Type of mating: Brother x sister, Pair C male x female

Banding pattern observed in parents:

F3 female 0632: 1,2,x 4 +,+,+ ~

F3 male 0661: 1,2,3 4 5,6,7 10

Number of F4's analysed:

Males + Females	=	10
Males	=	2
Females	=	8

Number of F4's inheriting each locus:

	2 males	8 females
1,2,3	0	7
1,2,x	1	1
4	2	8
10	2	3

Total band composition of F3 parents:

F3 female 0632: BX(1,2,x) , DX 4 , 4 or 4 , * +,+,+ , +,+,+ ~ , ~

F3 male 0661: BX(1,2,3) , DY 4 , 4 or 4 , * 5,6,7 , +,+,+ 10 , ~

Recombination events and new variants.

No new variants were observed. The patterns of band inheritance in the F4's allowed a prediction of the F3 parents internal $(TTAGGG)_n$ containing loci composition which matched exactly that determined for the F2 parents, Pair C.

Pair J (data not shown)

Type of mating: Brother x sister, Pair A male x female

Banding pattern observed in parent:

F3 female 0111: 1,2,3 4 5,6,7 10

F3 male 0666: x,x,x 4 5,6,7 10

Number of F4's analysed:

Males + Females = 27

Males = 10

Females = 17

Number of F4's inheriting each locus:

	10 males	17 females
1,2,3	5	11
x,2,x	0	1
4	6	13
5,6,7	9	14
10	7	13

Total band composition of F3 parents:

F3 female 0111: BX(1,2,3) , DX 4 , * 5,6,7 , +,+,+ 10 , ~

F3 male 0666: DX , DY 4 , * 5,6,7 , +,+,+ 10 , ~

Recombination events and new variants.

Inheritance into the F4 offspring determined that the predictions for the band composition of the F2 parents in Pair A were correct. One recombinant was observed in the 1,2,3 locus being x,2,x. This was observed in a female F4 animal and must have been derived from the 0111 F3 female as the male did not contain any bands belonging to the 1,2,3 locus. From DNA analysis of the F4 offspring and X^2 test calculations it was determined that 0111 was heterozygous for 1,2,3. If band 2 had been present upon the second X chromosome in this female it would have been seen in ~50% of the male and female offspring; it was only observed in one female F4. Hence, it appeared that a recombination event involving the two X chromosomes of the F3 female, 0111, occurred at gametogenesis resulting in an X chromosome containing band 2 separated from 1 and 3.

Chapter 6

Indirect Characterisation of the Sequence Composition of Internal and Telomeric $(TTAGGG)_n$ Containing Loci within the Mouse Genome

6.1 Introduction

Direct cloning of internal $(TTGGGG)_n$ and $(TTAGGG)_n$ containing loci from both the human and mouse genomes respectively, proved to be extremely difficult, Chapter 3. Hence, a more indirect approach was used to determine the type of DNA sequences present at telomeric and internal $(TTAGGG)_4$ hybridising loci within the mouse genome. Several properties of these regions have already been established. They are present both terminally and internally within the mouse genome and can be stably inherited into further generations, defining them as distinct genetic loci. The telomeric loci show a higher rate of new variant generation than those present within the mouse genome to the extent that individuals within an inbred strain differ in the pattern of hybridising bands upon PFGE analysis. All of the above properties were determined using the 4 base pair recognition site restriction enzyme, *Hae* III. A variety of different restriction enzymes were then used to determine, firstly, if the results obtained were reflected by other enzymes or if the loci observed were *Hae* III specific. Secondly, to determine, by partial digestion using 4bp recognition site enzymes, if the interstitially located, $(TTAGGG)_n$ hybridising DNA fragments represented the whole locus or if the locus was composed of tandemly repeated blocks containing the hexamer sequence. Thirdly, if the internal and telomeric loci were composed entirely of $(TTAGGG)_n$ tandem repeats or if they were interspersed with some other form of repeated sequence. Finally, to look for divergence, produced by single point mutations, from the $(TTAGGG)_n$ repeats that might allow recognition and cleavage by a suitable restriction enzyme. The presence of other types of telomere-like repeat was also investigated using the *Tetrahymena*-like telomere repeat probe $(TTGGGG)_4$ and for divergence from the (TTAGGG) pattern, the $(TGAGGG)_4$ probe. It had previously been shown that the human telomere was composed predominantly of $(TTAGGG)_n$ repeats. Internal to this sequence were $(TTGGGG)_n$ and $(TGAGGG)_n$ repeats (Allshire *et al.*, 1989, Brown *et al.*, 1990). The possibility of a similar situation in the mouse telomere was investigated.

6.2 Restriction Enzyme Analysis of Telomeric and Internal $(TTAGGG)_n$ Containing Loci within Mouse Genomic DNA

6.2.1 Strategy

A range of different restriction enzymes were used to study the sequence composition of the internal and telomeric $(TTAGGG)_n$ containing loci. The recognition sites for these enzymes vary between four and six base pairs, and so in the frequency with which they cut within the genome. It is possible to roughly estimate the frequency of such cleavage sites within mouse genomic DNA and so calculate an approximate fragment size for their cleavage products :

For four base recognition site enzymes :

Probability of site = 1 in every 4^4bp

Average size of restriction fragments = 256bp

For six base recognition site enzymes :

Probability of site = 1 in every 4^6bp

Average size of restriction fragments = 4096bp

In addition, if the telomeric and internal $(TTAGGG)_n$ containing loci are composed of blocks of the hexamer, interspersed with short stretches of unique sequence, the more frequently cutting enzymes may reveal this feature. Hence, total genomic DNA from DBA/2, C57Bl/6 and F2 individuals was digested with the restriction enzymes listed in Table 6.1, separated by either pulsed-field or conventional gel electrophoresis, transferred to a nylon membrane and hybridised with the $(TTAGGG)_4$ oligonucleotide probe.

6.2.2 Telomere-Derived $(TTAGGG)_4$ Banding Patterns

The pattern of telomeric bands produced after digestion with different enzymes and hybridisation with the $(TTAGGG)_4$ probe for an F2 individual is shown in Figure 6.1a. The DNA fragments produced varied in size from 30-150kb. The six base recognition site enzyme *Hind* III showed a distribution of bands that were larger in size than for any of the other enzymes. *Bam* HI and *Afl* II also recognise six base sequences. They produced DNA fragments that were more complex than that seen for the remaining enzymes. A general pattern of bands was produced by the four base recognition site enzymes. Exceptions to this pattern were seen for *Mnl* I, *Mse* I and *Hph* I. *Hph* I digestion resulted in loss of the three >100kb bands, *Mnl* I, loss of the 140kb fragment and *Mse* I, a band from the 50-100kb densely hybridising smear, Figure 6.1a.

From the results it appeared that the six and four base recognition site enzymes gave a very similar size range of hybridising DNA fragments. This result suggests that the telomeric loci are relatively devoid of enzyme sites and so must be composed of simple repeats; the

Table 6.1 Restriction Enzymes Used for Analysis of Telomeric and Internal $(TTAGGG)_n$ Containing Loci

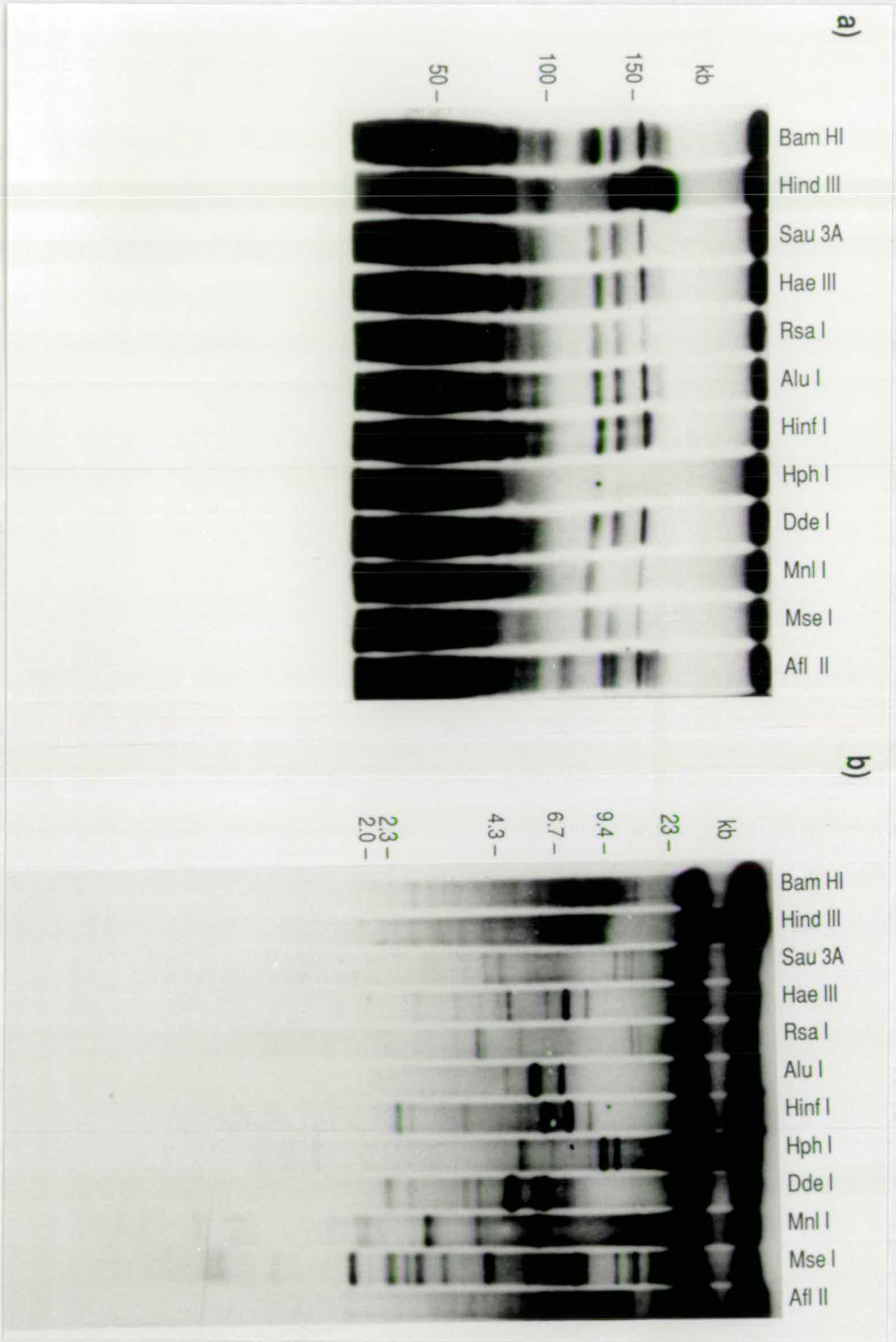
Restriction enzyme	DNA Recognition site	No bps per site
Bam HI	5' G/GATCC 3' 3' CCTAG/G 5'	6
Hind III	5' A/AGCTT 3' 3' TTCGA/A 5'	6
Sau 3A	5' /GATC 3' 3' CTAG/ 5'	4
Rsa I	5'GT/AC 3' 3'CA/TG 5'	4
Alu I	5'AG/CT 3' 3'TC/GA 5'	4
Hinf I	5' G/ANTC 3' 3' CTNA/G 5'	4
Hph I	5' GGTGA$(N)_8$/ 3' 3' CCACT$(N)_7$/ 5'	5
Dde I	5' C/TNAG 3' 3' GANT/C 5'	4
Mnl I	5' CCTC$(N)_7$/ 3' 3' GGAG$(N)_7$/ 5'	4
Mse I	5' T/TAA 3' 3' AAT/T 5'	4
Afl II	5' C/TTAAG 3' 3' GAATT/C 5'	6

Key : / = site of cleavage of DNA

Figure 6.1 General Restriction Enzyme Analysis of Telomeric and Internal $(TTAGGG)_n$ Containing Loci within Mouse Genomic DNA

Total genomic DNA from an F2 mouse was digested with the range of restriction enzymes shown in (a) and (b), the fragments separated by (a) pulsed-field and (b) conventional gel electrophoresis, transferred to a nylon membrane and hybridised with the $(TTAGGG)_4$ oligonucleotide probe.

λ concatemer and *Hind* III digested λ DNA size markers are indicated (kb).


a)
kb
150 –
100 –
50 –
Bam HI
Hind III
Sau 3A
Hae III
Rsa I
Alu I
Hinf I
Hph I
Dde I
Mnl I
Mse I
Afl II
b)
kb
23 –
9.4 –
6.7 –
4.3 –
2.3 –
2.0 –
Bam HI
Hind III
Sau 3A
Hae III
Rsa I
Alu I
Hinf I
Hph I
Dde I
Mnl I
Mse I
Afl II

presence of a sizeable piece of unique DNA would almost certainly result in cleavage by one of the enzymes used. Exceptions were seen in, for example, the fragments produced by *Hind* III. The larger size of fragment must reflect a lack of *Hind* III sites, compared to the other six base recognition site enzymes, in the DNA flanking the telomeric loci. A second exception was seen after digestion with *Hph* I, *Mnl* I and *Mse* I. All three enzymes seemed to cut within the telomeric loci to varying extents resulting in a reduction in size of some of the telomeric loci. All three recognise a sequence that could be generated within the TTAGGG repeats by a single base change, as outlined in Table 6.2. From these preliminary results it appeared that there was some divergence from the predicted $(TTAGGG)_n$ tandem repeats in the telomeric loci. However, a single base change in the $(TTAGGG)_n$ repeating sequence should result in the generation of both an *Mnl* I and *Hph* I enzyme site by the transition of a T to a G nucleotide base (bases involved in bold print and underlined), *ie*:

Hph I site : GGT**G**A

$(TTAGGG)_n$ telomeric repeats: TTAGGGT**T**AGGGTTAGGG

Mnl I site : **G**AGG

Hence, the generation of the same banding pattern with each enzyme would be expected; this was not observed, Figure 6.1a. The different banding patterns may reflect either a second mutation destroying the *Mnl* I site or a different repeat type present in the mouse telomere which could be mutated to give an *Hph* I, but not an *Mnl* I, site. However, it is extremely difficult to establish which type of repeat could have been mutated to generate the *Hph* I site in the absence of the *Mnl* I site. Isolation of DNA from the mouse telomere and subsequent sequencing would reveal the type of divergent telomeric repeat giving rise to the *Hph* I site.

Table 6.2 Restriction Enzymes Recognising Divergence in the Telomere Repeat $(TTAGGG)_n$

Restriction enzyme	Recognition sequence	Changes for telomere repeat
Mnl I	5' CCT**C**$(N)_7$/ 3' 3' GGA**G**$(N)_7$/ 5'	5' CCT**A** 3' 3' GGA**T** 5'
Hph I	5' GGT**G**A$(N)_8$/ 3' 3' CCA**C**T$(N)_7$/ 5'	5' GGTTA 3' 3' CCA**A**T 5'
Dde I	5' **C**/TNAG 3' 3' **G**ANT/C 5'	5' **G**TNAG 3' 3' **C**ANTC 5'
Mse I	5' T/TA**A** 3' 3'AAT/**T** 5'	5' TTA**G** 3' 5' AAT**C** 5'

Key : / = site of cleavage of DNA

B = bases involved in mutation event

6.2.3 Internally-Derived $(TTAGGG)_4$ Banding Patterns

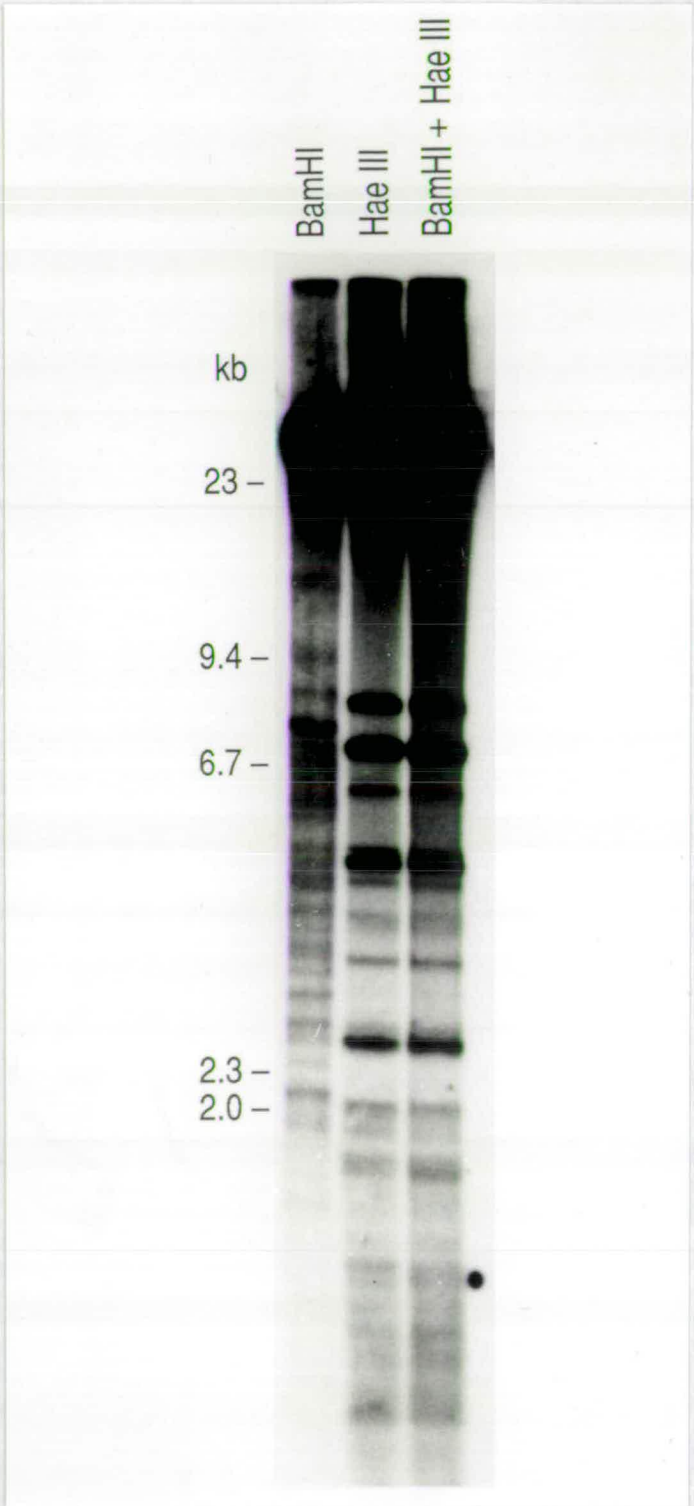
The banding patterns representing the internal $(TTAGGG)_n$ containing loci released from total F2 mouse genomic DNA by different restriction enzymes are shown in Figure 6.1b. The six base recognition site enzymes gave a general smear of hybridisation composed of many different sized fragments. No defined pattern of bands could be determined. In contrast, the four base recognition site enzymes gave enzyme specific fragments representing the internal $(TTAGGG)_n$ containing loci, Figure 6.1b. The banding pattern obtained for the six and four base recognition site enzymes relates to the frequency with which each enzyme site is likely to be found in the genomic DNA, *ie*, for a four base site every 256bp and a six base site every 4096bp. The enzymes with four base recognition sites, due to their frequency within the genome, would be able to cleave very close to the repeat blocks. An example of this is seen with *Hae* III which separates the C57Bl/6 X-linked locus 1,2,3 into its characteristic three bands. The same three bands are not seen in the *Bam* HI digested DNA, but are released from the DNA when it is digested with both *Bam* HI and *Hae* III, Figure 6.2.

From this initial study it was not possible to determine if the internal $(TTAGGG)_n$ containing loci were composed of pure tandem repeats or if there was divergence similar to that found in the telomeric loci. Secondly, the patterns of bands seen with the four base recognition site enzymes differed quite considerably, Figure 6.1b. Hence, the relationship between the bands produced by each enzyme needed to be determined, for example, were they derived from common or distinct loci? By establishing the common pattern of bands

Figure 6.2 Comparison of $(TTAGGG)_n$ Containing Fragments Produced by a 6bp and 4bp Recognition Site Enzyme from Total Mouse Genomic DNA

Total genomic DNA from a C57Bl/6 inbred mouse was digested with *Bam* HI, *Hae* III and both *Bam* HI and *Hae* III, as indicated upon the figure, the fragments separated by conventional gel electrophoresis, transferred to a nylon membrane and hybridised with the $(TTAGGG)_n$ oligonucleotide probe.

Hind III digested λ DNA size markers are indicated (kb).


BamHI
Hae III
BamHI + Hae III
kb
23 –
9.4 –
6.7 –
2.3 –
2.0 –

released from the internal $(TTAGGG)_n$ loci, it might also be possible to determine the presence of repeat sequence divergence with the appropriate enzymes.

6.3 Indirect Sequence Analysis of the Internal $(TTAGGG)_n$ Containing Loci

6.3.1 Strategy

Conventional gel electrophoresis and $(TTAGGG)_4$ hybridisation analysis of mouse genomic DNA revealed enzyme-specific banding patterns for the internal telomere-repeat containing loci. It was important to determine if the $(TTAGGG)_4$ hybridising bands observed were derived from the same or different loci within the mouse genome. Hence, total genomic mouse DNA was digested with different combinations of 4bp recognition site enzymes. If the enzyme-specific banding patterns were derived from different loci a combination of banding patterns would be expected. However, if each pattern was derived from the same locus, a single series of bands would be observed. Such analysis might also reveal which of the enzymes used cut closest to, or within, the repeat blocks.

6.3.2 Indirect Sequence Analysis with Hae III, Alu I, Sau 3A and Rsa I

Initially, total genomic DNA from C57Bl/6 and DBA/2 inbred strains was digested with every combination of *Alu* I, *Sau* 3A, *Hae* III and *Rsa* I four base recognition site restriction enzymes. The resultant fragments were separated by conventional gel electrophoresis, transferred to a nylon membrane and hybridised with the $(TTAGGG)_4$ oligonucleotide probe, shown in Figures 6.3a and b for C57Bl/6 and DBA/2, respectively. By analysing the banding patterns produced, it was possible to determine which of the four enzymes cleaved closest to the repeat blocks and the nature of the DNA sequence in each fragment. In the C57Bl/6 inbred strain genomic DNA, it appeared that *Alu* I was cleaving closest to the repeat blocks giving three main bands ranging in size from ~4-7kb. The *Rsa* I and *Sau* 3A patterns were completely lost in the presence of *Alu* I. *Hae* III, as previously described, gave a characteristic pattern of six bands, all of which were lost upon simultaneous digestion with *Alu* I. However, any genomic digests involving both *Hae* III and *Alu* I also resulted in a reduction in size of the three main *Alu* I bands. Hence, it would appear that located on either side of the three main bands is either an *Alu* I or *Hae* III enzyme site. In the absence of *Alu* I, smaller bands apart from the main loci were seen. Their absence from the *Alu* I digest suggests that either there were no *Alu* I cleavage sites next to these repeat blocks generating large non-resolvable DNA fragments, or that there are several sites within the region resulting in fragments to small in size to be detected.

Figure 6.3 Sequence Analysis of the Internal $(TTAGGG)_n$ Containing Loci with Hae III, Alu I, Sau 3A and Rsa I

Total genomic DNA from (a) a C57Bl/6 and (b) a DBA/2, inbred mouse was digested with every combination of the restriction enzymes *Hae* III, *Alu* I, *Sau* 3A and *Rsa* I, the fragments separated by conventional gel electrophoresis, transferred to a nylon membrane and hybridised with the $(TTAGGG)_4$ oligonucleotide probe.

In (a) and (b) :

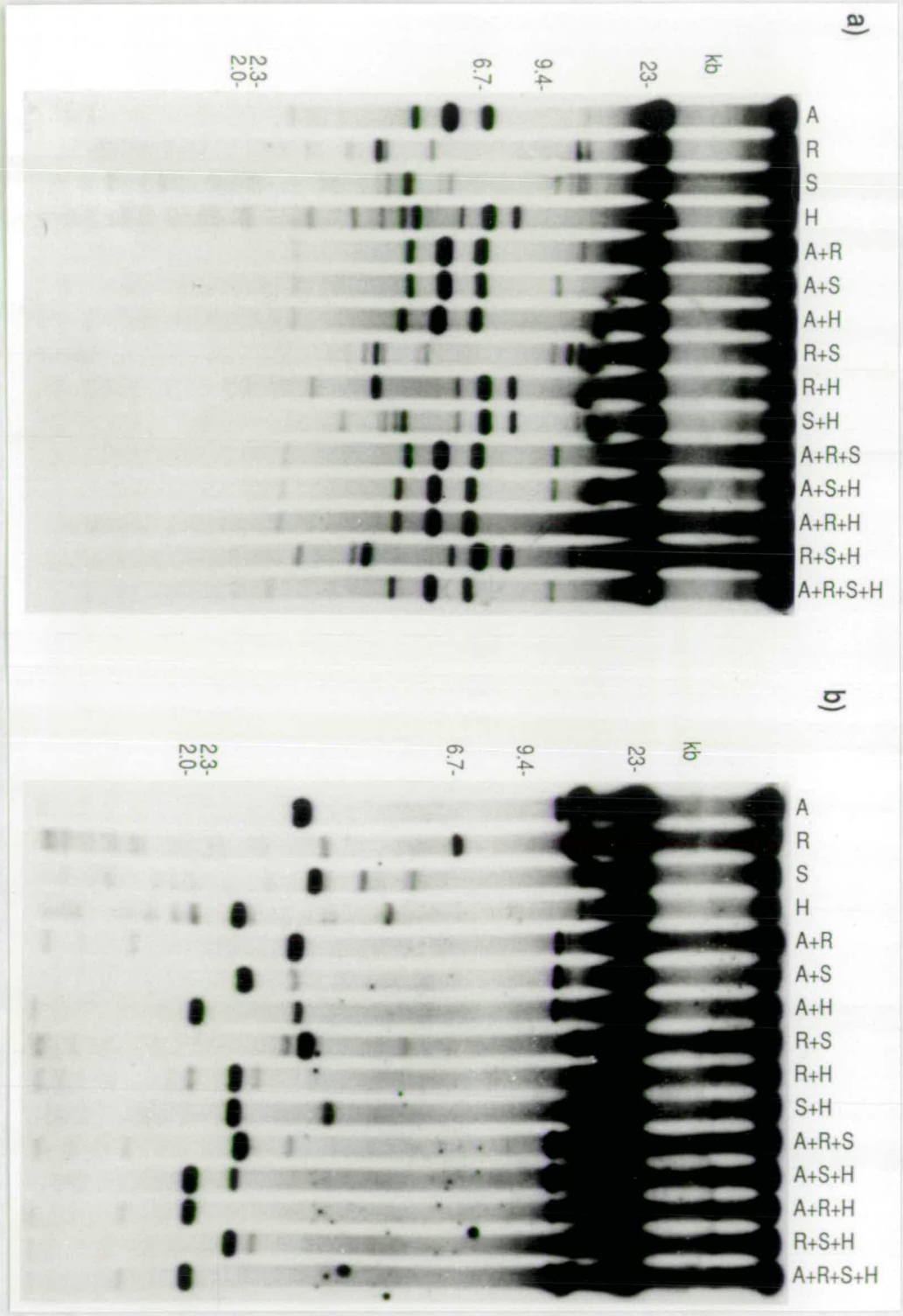
A = *Alu* I

R = *Rsa* I

S = *Sau* 3A

H = *Hae* III

Hind III digested λ DNA size markers are indicated (kb).


a)
kb
23-
9.4-
6.7-
2.3-
2.0-
A
R
S
H
A+R
A+S
A+H
R+S
R+H
S+H
A+R+S
A+S+H
A+R+H
R+S+H
A+R+S+H
b)
kb
23-
9.4-
6.7-
2.3-
2.0-
A
R
S
H
A+R
A+S
A+H
R+S
R+H
S+H
A+R+S
A+S+H
A+R+H
R+S+H
A+R+S+H

For the DBA/2 specific loci a similar result was obtained, Figure 6.3b. Again each enzyme gave its own characteristic banding pattern upon digestion of total genomic mouse DNA. A combination of *Alu* I and *Hae* III revealed the purest form of the repeat containing DNA fragments.

The results suggest that the enzymes used in this study were releasing the same locus from the mouse genomic DNA. If each enzyme was releasing different loci, which were entirely independent of each other, a composite pattern of bands would be seen with the different combinations of enzymes. By looking at the inheritance of these loci into the F2 generation, discussed in Chapter 5, it would be possible to determine which of the enzyme specific bands related to each other. For example, which $(TTAGGG)_4$ hybridising fragments released by *Alu* I, *Rsa* I and *Sau* 3A were derived from the same locus as the *Hae* III derived C57Bl/6 X-linked 1,2,3 locus.

6.3.3 TTAGGG Repeat Analysis with Hae III, Alu I, Mnl I and Hph I

The restriction enzymes *Alu* I and *Hae* III produced the purest form of $(TTAGGG)_n$ containing DNA fragments from the C57Bl/6 and DBA/2 inbred mouse genomes. From this basis the possibility of divergence from the $(TTAGGG)_n$ repeat was investigated using the enzymes *Mnl* I and *Hph* I. Hence, total genomic mouse DNA from the C57Bl/6 inbred strain was digested with every combination of *Alu* I, *Hae* III, *Hph* I and *Mnl* I. The resultant fragments were separated by conventional gel electrophoresis, transferred to a nylon membrane and hybridised with the $(TTAGGG)_4$ probe. The *Mnl* I and *Hph* I restriction enzymes gave their own specific banding patterns, as did *Alu* I and *Hae* III, Figure 6.4. In the absence of *Mnl* I, the same banding pattern previously observed in the presence of either *Hae* III or *Alu* I (Figure 6.3a) was detected. However, the presence of *Mnl* I in the cleavage reaction resulted in the loss of the defined *Hae* III/*Alu* I banding pattern. This was replaced by the *Mnl* I specific pattern with the exception that the smallest, strongly hybridising band of ~1.6kb was absent in the presence of *Alu* I, Figure 6.4. In addition, a series of weakly hybridising, ~2-6kb, DNA fragments were seen in tracts containing DNA digested with *Mnl* I, *Hph* I, and *Hae* III, *Mnl* I, *Hph* I and *Alu* I, *Mnl* I, *Hae* III and *Alu* I and *Mnl* I, *Hph* I, *Hae* III and *Alu* I but were absent when digestion was carried out in the presence of *Mnl* I alone or when present with only one other enzyme, Figure 6.4. This suggests that the ladder of weakly hybridising bands result from incomplete digestion of the DNA, possibly caused by the high concentration of restriction enzymes present.

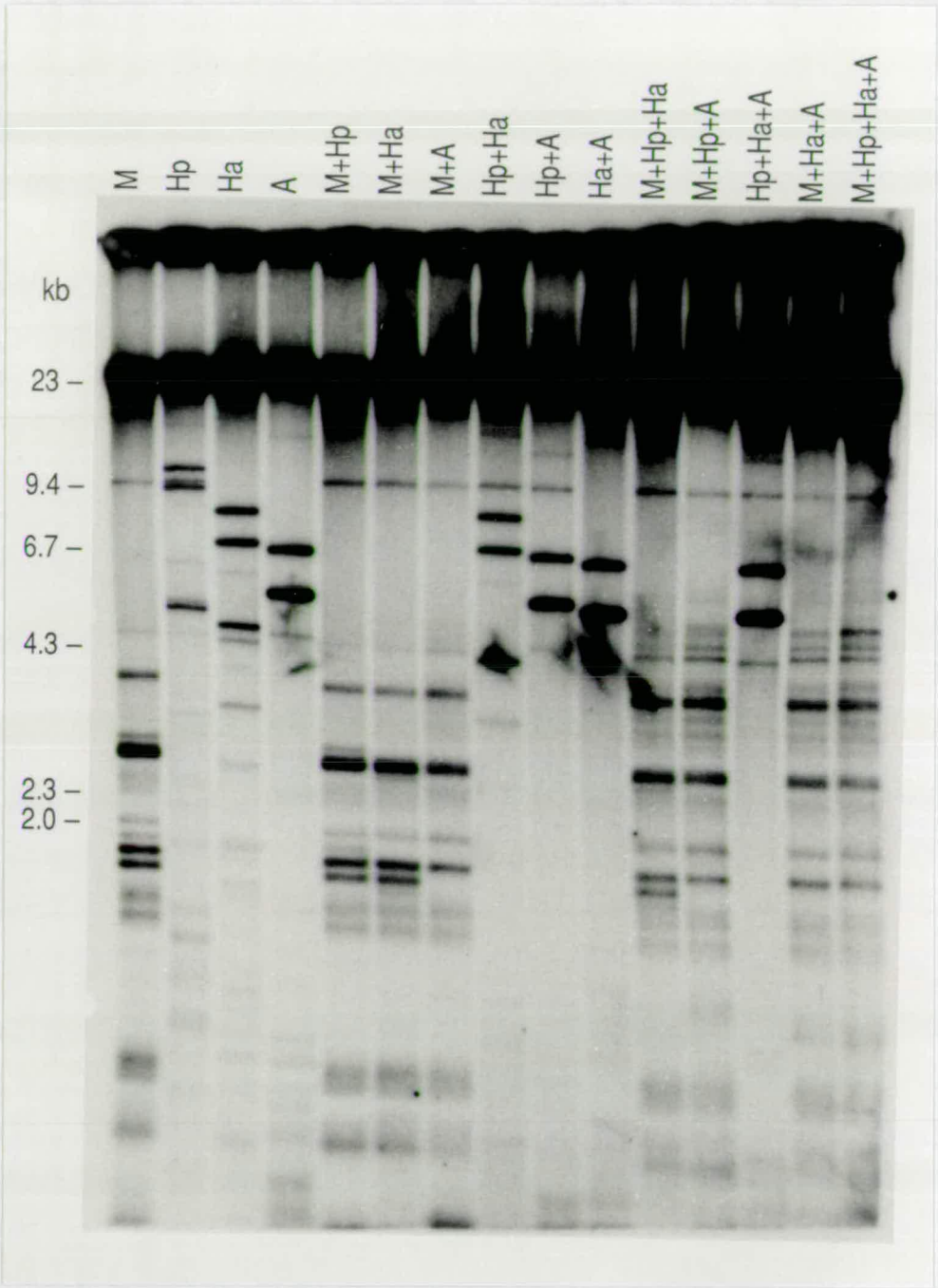
It would appear from the results that *Mnl* I enzyme sites were present within the *Alu* I/*Hae* III derived, 4.3-7kb, $(TTAGGG)_4$ hybridising fragments within mouse genomic DNA. The loss of these bands and presence of smaller hybridising DNA fragments in the *Mnl* I digest suggests that *Mnl* I restriction sites are scattered, at random, through the *Alu* I/*Hae* III bands.

Figure 6.4 Sequence Analysis of the Internal (TTAGGG)$_n$ Containing Loci with Mnl I, Hph I, Hae III and Alu I

Total genomic DNA from a C57Bl/6 inbred mouse was digested with every combination of the restriction enzymes *Mnl* I, *Hph* I, *Hae* III and *Alu* I, the fragments separated by conventional gel electrophoresis, transferred to a nylon membrane and hybridised with the (TTAGGG)$_4$ oligonucleotide probe.

M = *Mnl* I
Hp = *Hph* I
Ha = *Hae* III
A = *Alu* I

Hind III digested λ DNA size markers are indicated (kb).


M
Hp
Ha
A
M+Hp
M+Ha
M+A
Hp+Ha
Hp+A
Ha+A
M+Hp+Ha
M+Hp+A
Hp+Ha+A
M+Ha+A
M+Hp+Ha+A
kb
23 –
9.4 –
6.7 –
4.3 –
2.3 –
2.0 –

The ladder of partially digested, weakly hybridising bands in the *Mnl* I, multiple digest, DNA samples suggests the same conclusion.

From the results obtained using different combinations of four base recognition site enzymes it appeared that the telomeric and internal $(TTAGGG)_n$ containing loci were composed of simple repeats. The presence of *Mnl* I, *Mse* I and *Hph* I sites within the telomere (Figure 6.1a) and *Mnl* I sites within the internal loci (Figure 6.4) suggests that these regions of the mouse genome are composed of repeats which are either predominantly (TTAGGG) or a related repetitive sequence which crosshybridises with the $(TTAGGG)_4$ oligonucleotide probe and that divergence from this repeated sequence occurs within both regions. As seen for the telomeric loci, *Mnl* I and *Hph* I appear to be recognising a different cleavage site to that generated by a T to a G transition giving rise to a TGAGGG repeat. This suggests that some other form of randomly dispersed repeat is present within both the telomeric and internal loci. As previously discussed, it appeared that in the telomeric loci there was the generation of a *Mnl* I/*Hph* I site by a T to a G transition within the $(TTAGGG)_n$ repeats. However, either some other form of repeat divergence must have occurred or the initial mutation event must have occurred within a repeat unit other than (TTAGGG), as an *Mnl* I site was not always generated simultaneously. A very similar situation was seen for the internal $(TTAGGG)_n$ containing loci with the exception that *Mnl* I cleaved within the repeats independently of *Hph* I. This suggests that an alternative form of repeat may be present, at random, within these loci. Divergence within the *Tetrahymena*-like telomeric repeat (TTGGGG) would provide one such alternative. A G to an A base change at the second G deoxynucleotide would result in the creation of a *Mnl* I cleavage site without the simultaneous generation of a *Hph* I site, *ie*

Hph I	GGTGA
(TTGGGG) repeats	GGGTTGGGGTT
Mnl I	GAGG

6.4 Partial Restriction Enzyme Digestion Analysis of Interstitial $(TTAGGG)_n$ Containing Loci within the Mouse Genome

Partial restriction enzyme digestion analysis of the interstitial $(TTAGGG)_n$ containing loci within the mouse genome was performed to determine the organisation of $(TTAGGG)_n$ repeats at these loci. It is possible that these loci could be composed of either tandem blocks of $(TTAGGG)_n$ repeats separated by short stretches of DNA containing frequently cutting restriction enzyme sites, or, alternatively, a single block of $(TTAGGG)_n$ repeats bounded by another form of simple repeat. Such an analysis might help to explain why the new variant sized bands, NV2 and NVQ (Figures 5.5, 5.6, 5.7 and 5.8), gave a much weaker signal upon $(TTAGGG)_4$ hybridisation when compared to bands 1, 2, 4, 5, 6 and 7 and why band 3 of the

C57Bl/6 1,2,3 locus gave a much weaker signal upon hybridisation than bands 1 and 2. Weaker hybridisation could be due to the presence of a single block of $(TTAGGG)_n$ repeats instead of a tandem array, or, fewer copies of the TTAGGG hexamer together with an unrelated simple repeat. If the enzyme-specific DBA/2 and C57Bl/6 $(TTAGGG)_4$ hybridising DNA fragments were derived from tandemly repeated blocks of $(TTAGGG)_n$ repeats, then partial digestion would release a ladder of fragments which varied in length by multiples of the band size seen upon complete digestion of the DNA, Figure 6.5a. Alternatively, if the loci are composed of a single block of $(TTAGGG)_n$ repeats and bounded by random *Alu* I enzyme sites, a ladder of hybridising fragments of all possible sizes would be seen upon partial digestion, Figure 6.5b. Finally, if the loci are composed of a single block of $(TTAGGG)_n$ repeats bounded by copies of an unrelated simple repeat digestion would yield a ladder of bands larger in size than the fragment size seen upon complete digestion of the DNA, but not multiples of this final fragment size, Figure 6.5c.

Total genomic DNA from the C57Bl/6 and DBA/2 inbred mouse strains was subjected to partial digestion with *Alu* I, as outlined in Section 2.8.2, separated by conventional gel electrophoresis, transferred to a nylon membrane and hybridised with the $(TTAGGG)_4$ oligonucleotide probe. For both the C57Bl/6 and DBA/2 DNA a ladder of bands was observed, which was lost upon complete digestion, Figure 6.6a and b; this giving rise to the *Alu* I specific pattern of bands, Figure 6.1b. It was difficult to determine which components of the ladder of bands corresponded to which of the defined *Alu* I restriction fragments in the C57Bl/6 DNA, Figure 6.5a. However, in the DBA/2 DNA the ladder of bands was derived from the strongly hybridising ~3.5kb DNA fragment, Figure 6.6b. This ladder of bands did not represent multiples of the final ~3.5kb band and so suggests that this locus contains a single block of $(TTAGGG)_n$ repeats. From the results obtained from partial digestion analysis of these loci it was difficult to determine whether the ladder of DNA fragments produced were due to random *Alu* I sites or *Alu* I sites within a simple repeat surrounding the block of $(TTAGGG)_n$ repeats, Figures 6.5b and c. However, it seems possible that the differences in signal intensity between different interstitially derived $(TTAGGG)_n$ containing DNA fragments result from size differences of the block of $(TTAGGG)_n$ repeats. The lack of enzyme sites within these loci also suggests that some other form of simple repeat, that does not hybridise with the $(TTAGGG)_4$ oligonucleotide probe, is also present within these DNA fragments. Until the C57Bl/6 and DBA/2 interstitial $(TTAGGG)_n$ containing loci have been isolated from the mouse genome and sequenced, their exact organisation cannot be determined.

Figure 6.5 Schematic Representation of the Organisation of Telomeric Repeats within the Mouse Genomic, Interstitial $(TTAGGG)_n$ Containing Loci

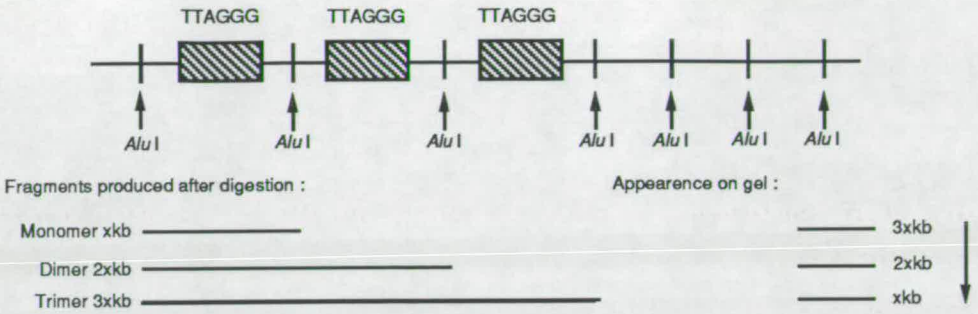
(a) Tandemly repeated blocks of $(TTAGGG)_n$ repeats, separated by short stretches of DNA containing a single *Alu* I restriction enzyme site. Partial digestion of Total genomic DNA would release a ladder of fragments which varied in length by multiples of the band size seen upon complete digestion of the DNA.

(b) Interstitial $(TTAGGG)_n$ containing loci composed of a single block of $(TTAGGG)_n$ repeats bounded by random *Alu* I enzyme sites. Upon partial digestion, a ladder of hybridising fragments of all possible sizes would be seen.

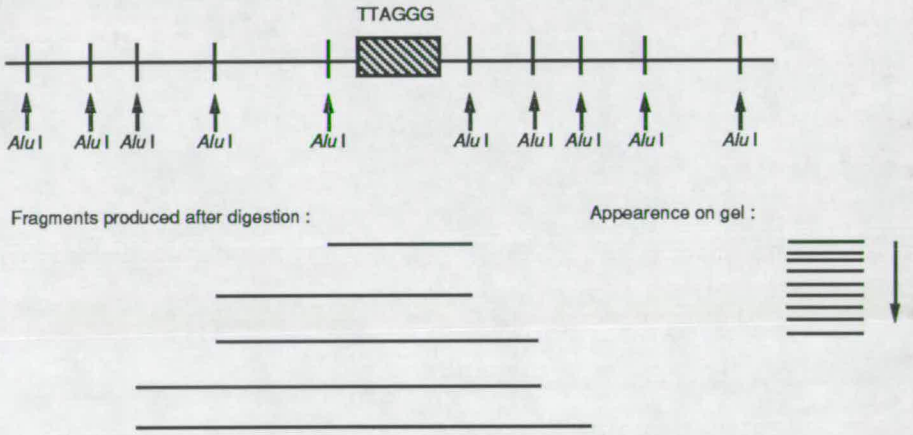
(c) Interstitial $(TTAGGG)_n$ containing loci composed of a single block of $(TTAGGG)_n$ repeats bounded by unrelated, tandem repeats. Partial digestion would yield a ladder of bands, larger than the fragment size seen upon complete digestion of the DNA, but not multiples of this final fragment size.

In each diagram, the predicted appearence of the partial digestion products is shown; the bands run from high to low molecular weight as indicated by the arrow.

a)


TTAGGG
TTAGGG
TTAGGG
Alu I
Alu I
Alu I
Alu I
Alu I
Alu I
Alu I
Fragments produced after digestion :
Appearence on gel :
Monomer xkb
Dimer 2xkb
Trimer 3xkb
3xkb
2xkb
xkb

b)


TTAGGG
Alu I
Alu I
Alu I
Alu I
Alu I
Alu I
Alu I
Alu I
Alu I
Alu I
Fragments produced after digestion :
Appearence on gel :

c)

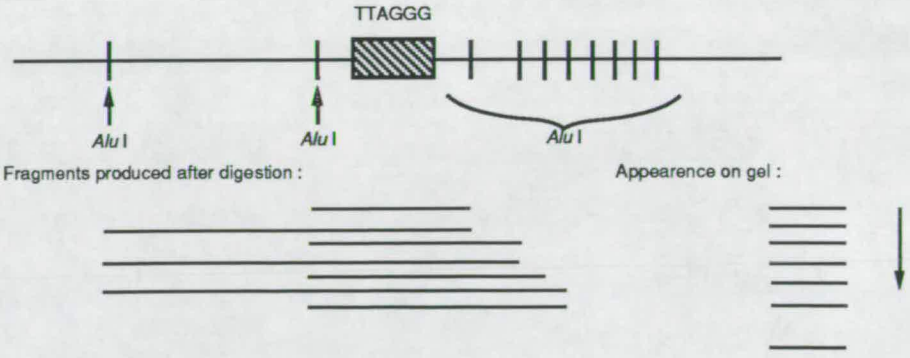

TTAGGG
Alu I
Alu I
Alu I
Fragments produced after digestion :
Appearence on gel :

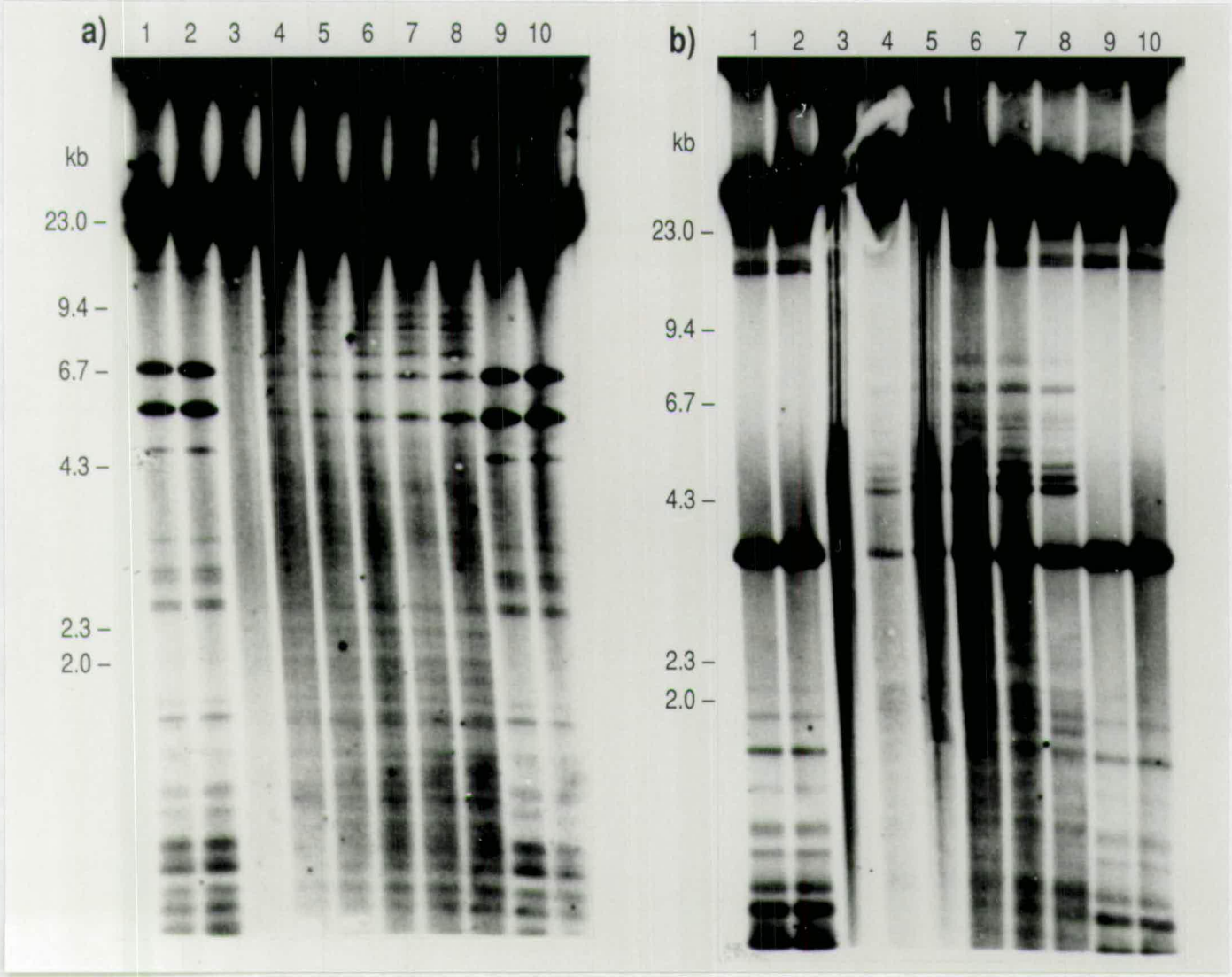
Figure 6.6 Partial Restriction Enzyme Digestion Analysis of Interstitial $(TTAGGG)_n$ Containing Loci within the Mouse Genome

Total genomic DNA from (a) a C57Bl/6 and (b) a DBA/2 inbred mouse was subjected to partial digestion with *Alu* I, as outlined in Section 2.8.2, the fragments separated by conventional gel electrophoresis, transferred to a nylon membrane and hybridised with the $(TTAGGG)_4$ oligonucleotide probe.

Contents of tracks where (a) contains C57Bl/6 and (b) contains DBA/2, total genomic DNA :

Lane	Conditions of digest
1	O/N digestion with 20u of *Alu* I
2	O/N digestion with 20u of *Alu* I + 20u and 4 hours extra the following morning
3	Timecourse : 0 minutes
4	5 minutes
5	10 minutes
6	20 minutes
7	40 minutes
8	60 minutes
9	O/N
10	O/N + 20u of *Alu* I and 4 hours extra the following morning

Hind III digested λ DNA size markers are indicated (kb).


a)
1 2 3 4 5 6 7 8 9 10
kb
23.0 –
9.4 –
6.7 –
4.3 –
2.3 –
2.0 –
b)
1 2 3 4 5 6 7 8 9 10
kb
23.0 –
9.4 –
6.7 –
4.3 –
2.3 –
2.0 –

6.5 Determination of the Presence of $(TTGGGG)_n$ and $(TGAGGG)_n$ Repeats within the Internal and Telomeric $(TTAGGG)_n$ Containing Loci

6.5.1 Strategy

Characterisation of the sequence composition of the human telomere has shown that the most distal and largest portion is composed of pure $(TTAGGG)_n$ repeats. Internal or proximal to this are clusters of $(TGAGGG)_n$ and $(TTGGGG)_n$ repeats (Allshire *et al.*, 1989 and Brown *et al.*,1990). Using the $(TGAGGG)_4$ and $(TTGGGG)_4$ oligonucleotide probes the presence of these repeats within the mouse genome was investigated. Such an experiment would determine the proportion of these repeats relative to the $(TTAGGG)_n$ repeats. Also, it is unlikely that the 24mer oligonucleotide probes used could recognise one or two copies of the repeat alone, hence, the intensity of hybridisation would determine if clusters or randomly dispersed single units of $(TGAGGG)_n$ and $(TTGGGG)_n$ repeats were present within the internal and telomeric $(TTAGGG)_n$ containing loci.

6.5.2 $(TTGGGG)_n$ and $(TGAGGG)_n$ Repeats within the Telomeres and at Interstitial Sites of the Mouse Genome

The $(TTGGGG)_4$ and $(TGAGGG)_4$ oligonucleotide probes were hybridised to the pulsed-field and conventional gel electrophoresis separations of C57Bl/6, DBA/2 and F2 DNA fragments generated with a variety of restriction enzymes. For the telomeric loci a weak hybridisation signal was obtained with both the $(TTGGGG)_4$ and $(TGAGGG)_4$ oligonucleotide probes (data not shown). This suggests that there is either very little of these two types of repeat present at mouse telomeres or that they are scattered as single or double units and so are not recognised by the probes. One exception to this was seen for the DBA/2 derived telomeric loci hybridised with the *Tetrahymena*-like $(TTGGGG)_4$ oligonucleotide probe. The same pattern of bands obtained with the $(TTAGGG)_4$ probe were seen but the intensity of hybridisation was at least 7-fold weaker (data not shown). The result suggests that the telomeres of the DBA/2 mouse strain are composed predominantly of $(TTAGGG)_n$ together with fewer $(TTGGGG)_n$ repeats. Alternatively, this result could be due to crosshybridisation of the $(TTGGGG)_4$ probe to the $(TTAGGG)_n$ repeats. It had previously been shown that the $(TTGGGG)_4$ probe crosshybridised to some extent with $(TTAGGG)_n$ repeats in human and mouse total genomic DNA (Allshire *et al.*, 1989). Hence, it is possible that the signal obtained was due to a combination of real $(TTGGGG)_n$ repeats and crosshybridisation (data not shown).

For the internal $(TTAGGG)_n$ containing loci, very weak hybridisation to some of the $(TTAGGG)_4$ hybridising bands was seen with the $(TTGGGG)_4$ oligonucleotide probe; a

comparison of the hybridisation result for each probe is shown for C57Bl/6 DNA in Figures 6.7a, b and c. Again, this may be due to crosshybridisation or a small amount of the repeat being present within the loci. In addition, both probes detected a smear of hybridisation in the <1kb size range. This smear was larger for both of the six base recognition site enzymes. For the $(TGAGGG)_4$ probe, unique bands could be seen in the *Bam* HI and *Rsa* I digests. Similar DNA fragments were detected by the $(TTGGGG)_4$ probe.

Hybridisation of the $(TGAGGG)_4$ probe was detected in all the genomic DNA digests except when cleaved with *Mnl* I. Considering the recognition sequence of *Mnl* I, GAGG, cleavage at every such site would result in loss of the TGAGGG repeat. The presence of signal within the *Hph* I digested DNA further demonstrates that divergence from the $(TTAGGG)_n$ repeat is not the main cause of *Mnl* I sites. Instead, the most likely cause is divergence from the (TTGGGG) hexamer repeat, which would generate TTGAGG and an *Mnl* I site. Although the $(TGAGGG)_4$ probe contains an additional G deoxynucleotide it seems possible that it would still be able to recognise the divergent repeat; this is suggested by the hybridisation results.

From the above results it appears that there is very little (TGAGGG) and (TTGGGG) repeat within the mouse telomere, in contrast to the case in human telomeres (Allshire *et al.*, 1989, Brown *et al.*, 1990). However, both repeats appeared to be present within the mouse genome. From the general pattern of hybridisation to the internally derived restriction fragments, it is likely that these repeats are dispersed throughout the mouse genome, possibly in a similar way to the human *Alu* repeats.

An alternative method for estimating the general repeat composition of the mouse telomere would involve comparing the intensity of hybridisation of the $(TTAGGG)_4$ probe to mouse telomeres with that seen in the human. Human telomeres have been shown to contain between 5-25kb of $(TTAGGG)_n$ repeats. Hence, total genomic DNA from the mouse and human was digested with *Hae* III, the fragments separated by conventional gel electrophoresis, transferred to a nylon membrane and hybridised with the $(TTAGGG)_4$ oligonucleotide probe, Figure 6.8. The human genomic DNA contained telomeres ranging in size from ~5-15kb. After densitometric analysis, it appeared that the intensity of hybridisation was roughly 10 times stronger in mouse compared to human telomeres. Hence, it suggests that there is between 50-150kb of $(TTAGGG)_n$ or a similar crosshybridising repeat within the mouse telomere.

6.5.3 Summary

The mouse telomere appeared to be composed predominantly of *Trypanosoma*-like $(TTAGGG)_n$ repeats, in common with human telomeres. It had previously been shown that the $(TTAGGG)_4$ probe hybridised to human DNA with very little crosshybridisation to other

Figure 6.7 Hybridisation Analysis to Determine the Presence of $(TTGGGG)_n$ and $(TGAGGG)_n$ Repeats within the Mouse Genome

Total genomic C57Bl/6 DNA was digested with the enzymes indicated in (a), (b) and (c), separated by conventional gel electrophoresis, transferred to a nylon membrane and hybridised with the (a) $(TTAGGG)_4$, (b) $(TTGGGG)_4$ and (c) $(TGAGGG)_4$ oligonucleotide probes.

Hind III digested λ DNA size markers are indicated (kb).

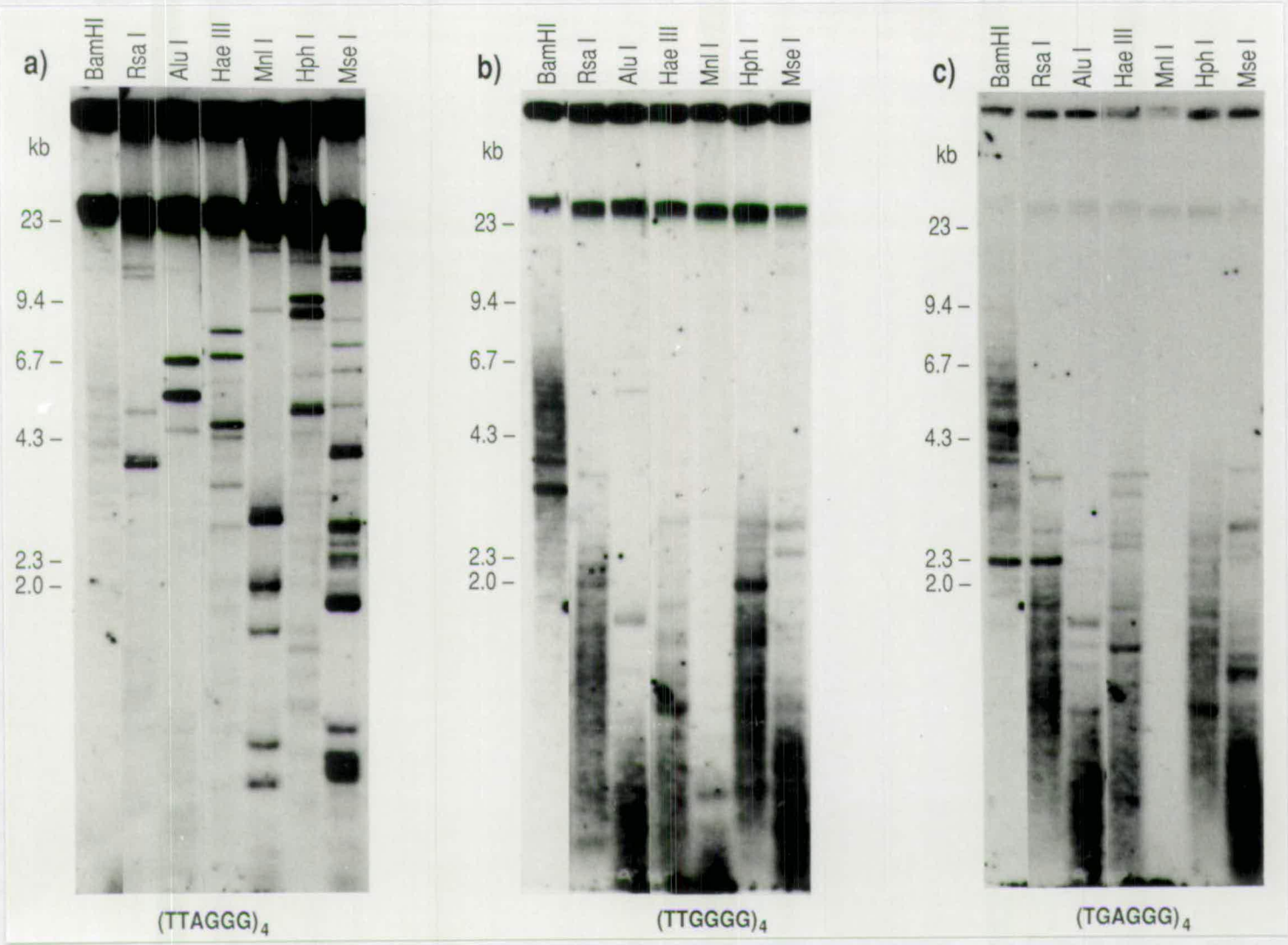
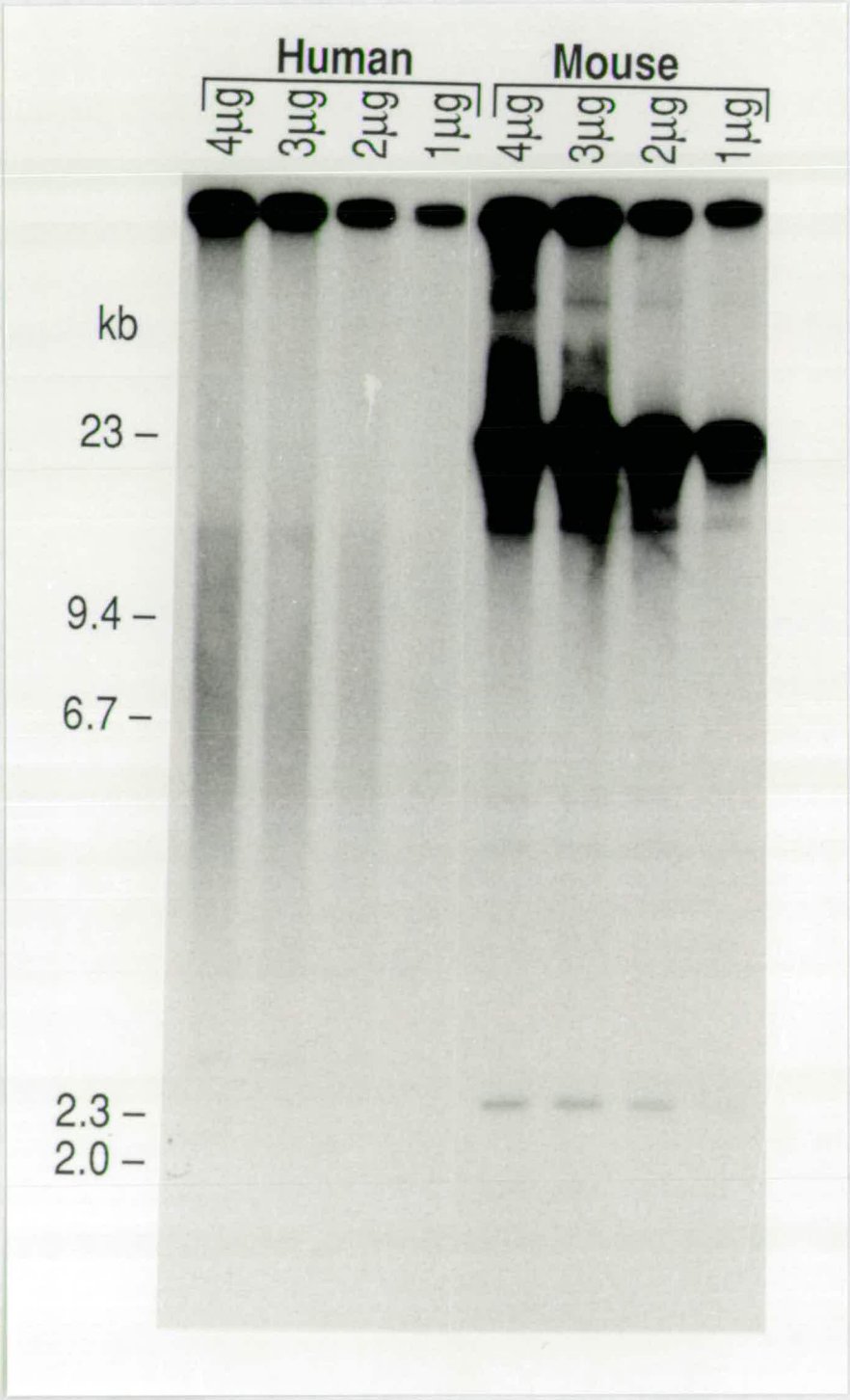

a)
BamHI
Rsa I
Alu I
Hae III
Mnl I
Hph I
Mse I
kb
23 –
9.4 –
6.7 –
4.3 –
2.3 –
2.0 –
(TTAGGG)4
b)
BamHI
Rsa I
Alu I
Hae III
Mnl I
Hph I
Mse I
kb
23 –
9.4 –
6.7 –
4.3 –
2.3 –
2.0 –
(TTGGGG)4
c)
BamHI
Rsa I
Alu I
Hae III
Mnl I
Hph I
Mse I
kb
23 –
9.4 –
6.7 –
4.3 –
2.3 –
2.0 –
(TGAGGG)4

Figure 6.8 Estimation of the Proportion of $(TTAGGG)_n$ Repeat within the Mouse Compared to Human Telomere

Total genomic DNA from a human and DBA/2 inbred mouse was digested with *Hae* III, the dilutions indicated separated by conventional gel electrophoresis, transferred to a nylon membrane and hybridised with the $(TTAGGG)_4$ oligonucleotide probe. The resultant autoradiograph was scanned to determine the extent of hybridisation and hence an estimate of the amount of repeat present within the human and mouse telomeres.
Hind III digested λ DNA size markers are indicated (kb).


Human
Mouse
4µg
3µg
2µg
1µg
4µg
3µg
2µg
1µg
kb
23 –
9.4 –
6.7 –
2.3 –
2.0 –

telomere-derived repeats such as *Plasmodium* (TTTAGGG) and *Tetrahymena* (TTGGGG) (Allshire *et al.*, 1989). Hence, it was unlikely that the $(TTAGGG)_4$ probe was recognising large blocks of either of these two types of repeat. In addition, very little hybridisation was seen with the $(TTGGGG)_4$ and $(TGAGGG)_4$ probes to mouse telomeres. Taken together the results suggest that the mouse telomere is composed predominantly of $(TTAGGG)_n$ or a closely related repeat. The same is true for the internal $(TTAGGG)_n$ containing loci. There is a much weaker hybridisation signal with the $(TTGGGG)_4$ probe suggesting that there is very little of this repeat within the internal loci. Both of these repeats may be dispersed throughout the mouse interstitial DNA, a feature not seen for the $(TTAGGG)_n$ repeats. *In situ* hybridisation of the $(TTGGGG)_4$ and $(TGAGGG)_4$ oligonucleotide probes to the mouse metaphase chromosomes, as outlined in Section 2.4, might help to determine if these repeats are dispersed within the mouse genome.

6.6 Investigation of Divergent Repeats within the Mouse Telomere

6.6.1 Strategy

From the results obtained it appeared that the restriction enzymes *Mnl* I, *Mse* I and *Hph* I had recognition sites within the simple repeats forming the mouse telomere. This suggested that there was divergence from the $(TTAGGG)_n$ repeating sequence and the possible presence of additional types of repeat. Due to the size of the telomeric loci separated by PFGE, it was very difficult to detect small size changes resulting from enzyme cleavage. Conventional gel electrophoresis results in the separation of DNA fragments ranging in size from 2-23kb. Hence, to further investigate the extent of divergence in the mouse telomere which might result in the release of small telomeric fragments, total genomic mouse DNA was digested with each of the afore mentioned enzymes.

6.6.2 Analysis of DNA Fragments Released from the Mouse Telomere by *Mnl I*

Initially, total genomic DNA from C57Bl/6 local and Hull colonies and DBA/2 Oxford and Hull colonies of inbred mice was digested with *Mnl* I. The resultant fragments were separated by conventional gel electrophoresis, transferred to a nylon membrane and hybridised with the $(TTAGGG)_4$ oligonucleotide probe. Differences were observed in the banding pattern obtained for individuals within the C57Bl/6 local and DBA/2 Oxford substrains, Figure 6.9. The same type of difference in banding pattern was not observed for the same genomic DNA when treated with the restriction enzyme *Hae* III (as shown in Figure 4.7). The variant $(TTAGGG)_4$ hybridising fragments could have been derived from the mouse telomeric loci as

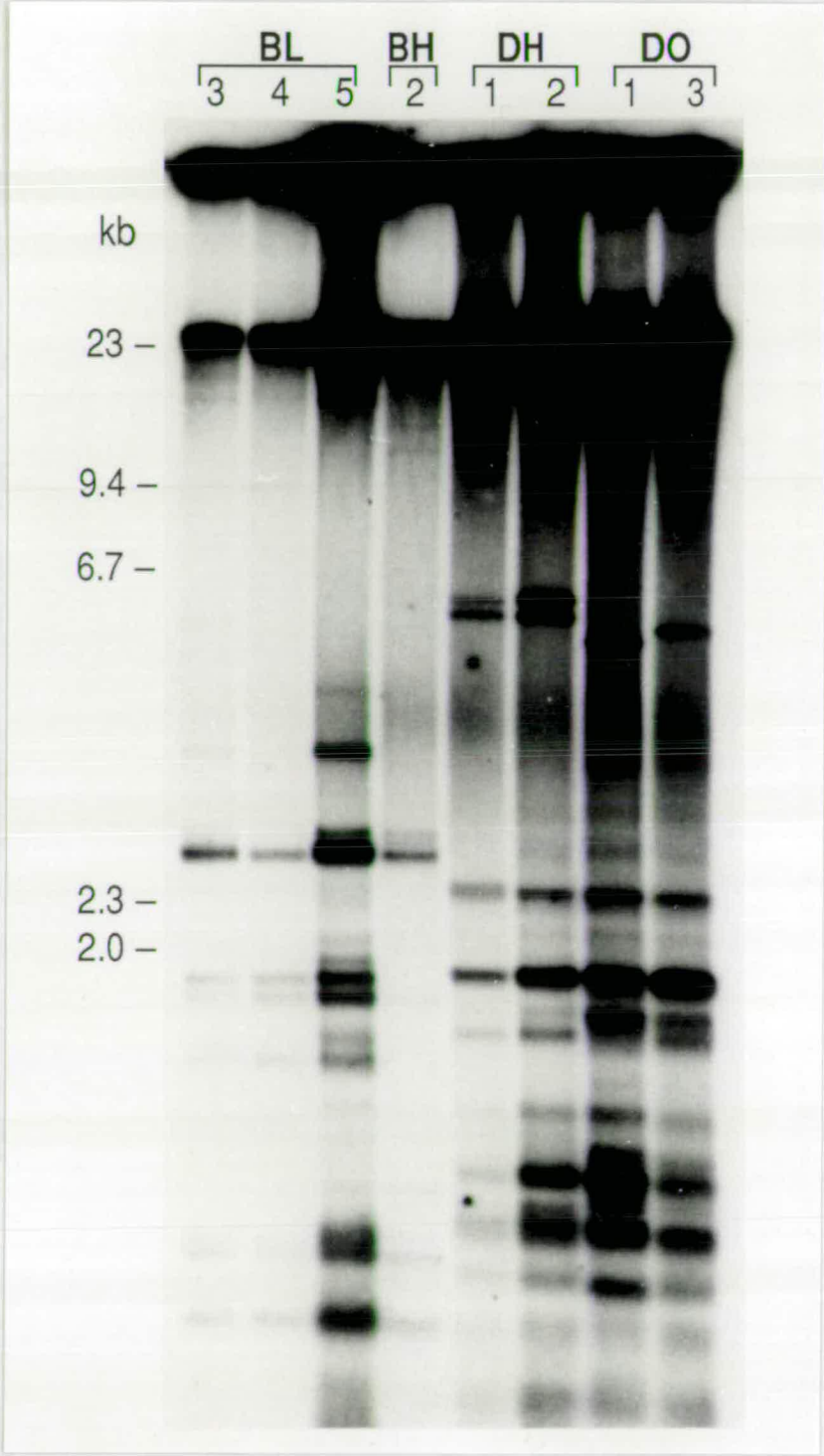
Figure 6.9 Comparison of $(TTAGGG)_n$ Containing Fragments Produced by Mnl I Digestion of Mouse Genomic DNA

Total genomic DNA from different individuals of the C57Bl/6 and DBA/2 inbred mouse strains was digested with *Mnl* I, the fragments separated by conventional gel electrophoresis, transferred to a nylon membrane and hybridised with the $(TTAGGG)_4$ oligonucleotide probe.

BL	=	C57Bl/6 mouse from a local colony
BH	=	C57Bl/6 mouse derived from the Hull colony
DH	=	DBA/2 mouse derived from the Hull colony
DO	=	DBA/2 mouse derived from the Oxford colony

The numbers above each track refer to each individual mouse.

Hind III digested λ DNA size markers are indicated (kb).


BL
BH
DH
DO
3 4 5
2
1 2
1 3
kb
23 –
9.4 –
6.7 –
2.3 –
2.0 –

banding pattern differences were not observed within the *Hae* III derived internal $(TTAGGG)_n$ containing loci. To further determine the variability of these *Mnl* I specific fragments, their inheritance into a second generation was investigated. As already shown, when the internal $(TTAGGG)_n$ containing loci were released from total genomic DNA of the F1 offspring of a DBA/2 x C57Bl/6 cross, both sets of parental bands were inherited into each individual. However, the same was not true of the telomere derived loci suggesting that the parents were heterozygous within this region of the genome (Chapter 5). If the restriction enzyme *Mnl* I was releasing different sized DNA fragments from individual telomeric loci, a reflection of their inheritance into some but not all of the F1 offspring might be expected. Hence, total genomic DNA from the parents and offspring of all three DBA/2 x C57Bl/6 pairs was digested with (a) *Hae* III and (b) *Mnl* I, the fragments separated by either pulsed-field or conventional gel electrophoresis, transferred to a nylon membrane and hybridised with the $(TTAGGG)_4$ oligonucleotide probe. In all cases the telomeric loci released by *Mnl* I were reduced in size compared to those produced with *Hae* III. For example, in one pair the telomeric bands produced by *Hae* III varied in size from 30-150kb compared to a size of <50kb when released by *Mnl* I, Figures 6.10a and b. This effect was not quite so pronounced for the other two pairs analysed but there was still a general ~30kb reduction in size compared to *Hae* III released telomeric loci. Analysis, by conventional gel electrophoresis of the same DNA's digested with *Hae* III and *Mnl* I is shown in Figure 6.11a and b. *Mnl* I specific $(TTAGGG)_4$ hybridising DNA fragments showed two forms of inheritance into the F1 offspring. Some were inherited into all of the offspring in common with the *Hae* III specific hybridising fragments suggesting that these loci were homozygous in the parental genomes. However, the other $(TTAGGG)_4$ hybridising bands were present in some but not all of the offspring to the extent that each individual had a different combination of $(TTAGGG)_n$ containing DNA fragments. This effect reflects the situation seen within the mouse telomeric loci. A similar result was obtained for the other two pairs analysed in this way, however, there were fewer heterozygous bands observed (data not shown).

6.6.3 Analysis of an Mnl I Specific New Variant

New variant sized telomeric $(TTAGGG)_4$ hybridising bands were observed at the telomeric loci when total genomic DNA from F1 mice + parents was digested with *Hae* III (As shown in Figures 5.2a, b and c). For the C57Bl/6 x DBA/2 cross shown in Figure 5.2a a new variant of 140kb was observed the F1 female 0003; the new variant sized band is indicated by an arrow in Figure 6.10a. The genomic DNA from the same individual contained an ~12kb new variant when cleaved with *Mnl* I and separated by conventional gel electrophoresis, as indicated in Figure 6.11b. It was possible that the origin of the new *Mnl* I derived ~12kb band may have coincided with the 140kb telomeric band. Hence, to determine the exact origin of the *Mnl* I

Figure 6.10 PFGE Analysis of the Telomeric $(TTAGGG)_n$ Containing Loci Following Digestion with (a)Hae III and (b)Mnl I

Total genomic DNA from a C57Bl/6 x DBA/2 pair and their F1 offspring was digested with (a) *Hae* III and (b) *Mnl* I, the fragments separated by PFGE, transferred to a nylon membrane and hybridised with the $(TTAGGG)_4$ oligonucleotide probe.

λ concatemer DNA size markers are indicated (kb).

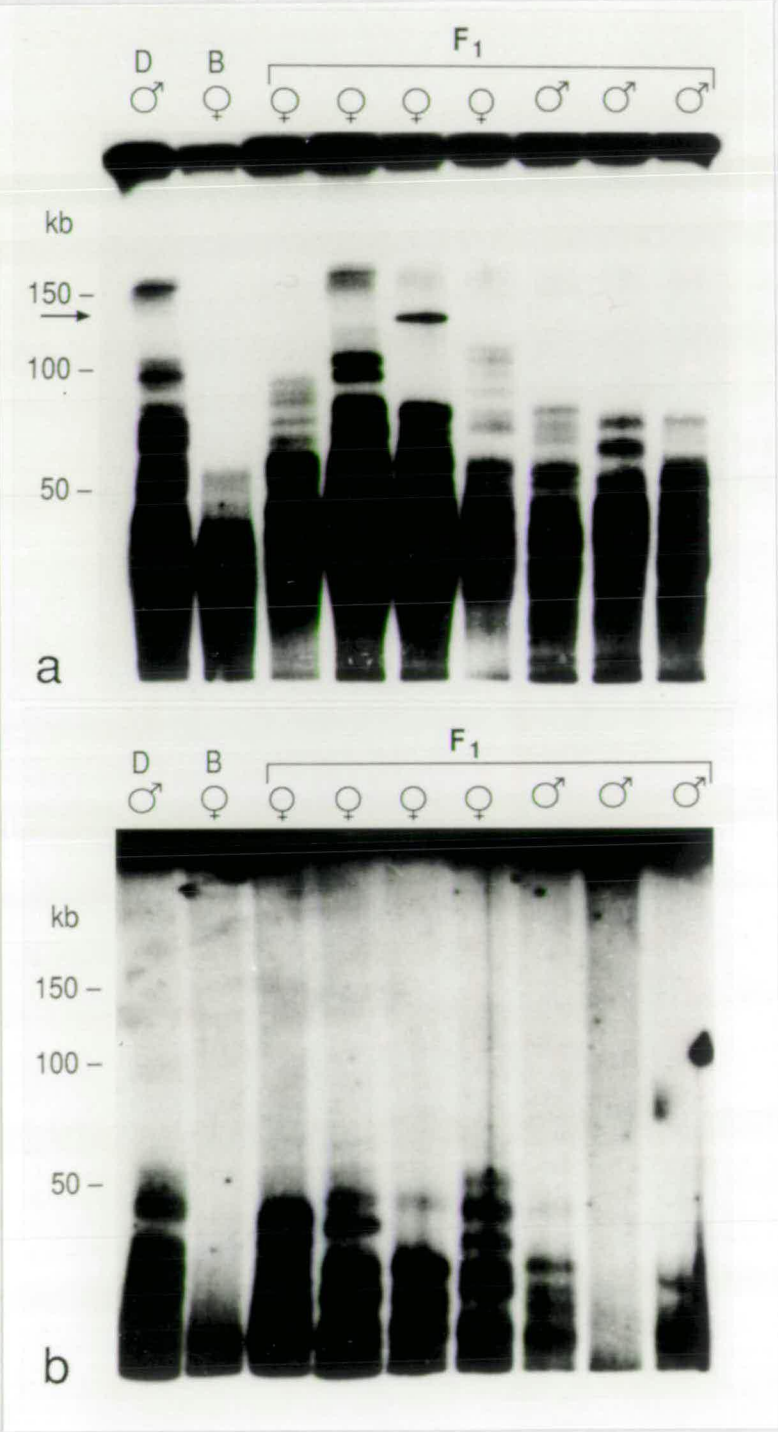
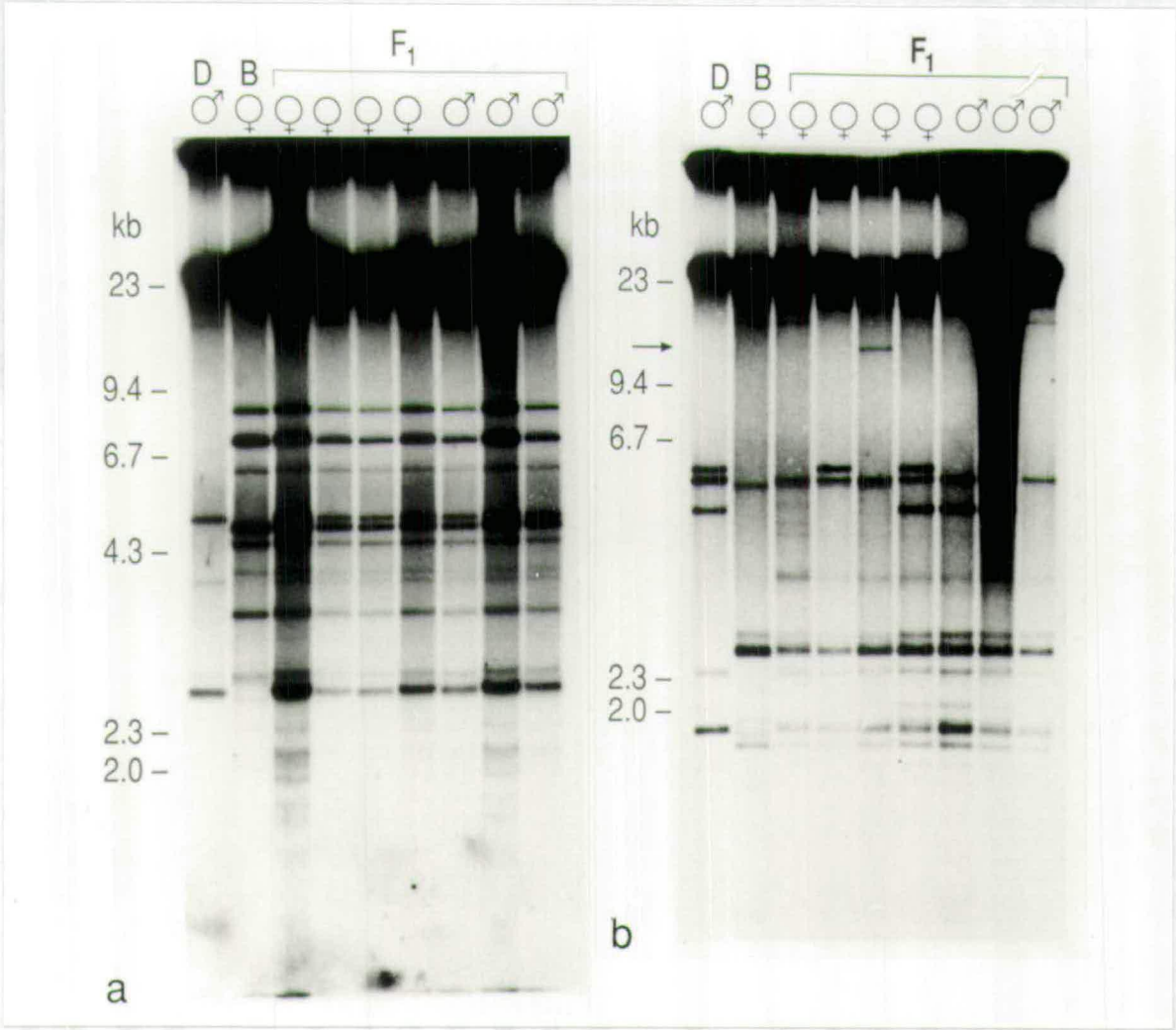

D
B
F1
♂
♀
♀
♀
♀
♀
♂
♂
♂
kb
150 –
100 –
50 –
a
D
B
F1
♂
♀
♀
♀
♀
♀
♂
♂
♂
kb
150 –
100 –
50 –
b

Figure 6.11 Conventional Gel Electrophoresis Analysis of $(TTAGGG)_n$ Containing DNA Fragments Produced by Digestion with (a) Hae III and (b) Mnl I

The same set of genomic DNA's observed in Figure 6.10(a) and (b) were used. Total genomic DNA was digested with (a) *Hae* III and (b) *Mnl* I, the fragments separated by conventional gel electrophoresis, transferred to a nylon membrane and hybridised with the $(TTAGGG)_4$ oligonucleotide probe.

Hind III digested λ DNA size markers are indicated (kb).


D B F1
♂ ♀ ♀ ♀ ♀ ♀ ♂ ♂ ♂
kb
23 –
9.4 –
6.7 –
4.3 –
2.3 –
2.0 –
a
D B F1
♂ ♀ ♀ ♀ ♀ ♀ ♂ ♂ ♂
kb
23 –
9.4 –
6.7 –
2.3 –
2.0 –
b

$(TTAGGG)_4$ hybridising fragments and if a connection existed between the new variants described, the smear of >23kb DNA, obtained by digestion with *Hae* III and separation by conventional gel electrophoresis, was isolated. This method of isolation would result in the removal of the internal $(TTAGGG)_n$ containing DNA fragments. The telomeric DNA was then digested with *Mnl* I, the fragments separated by conventional gel electrophoresis, transferred to a nylon membrane and hybridised with the $(TTAGGG)_4$ oligonucleotide probe. As shown in Figure 6.12a, several of the *Mnl* I specific bands were derived from the mouse telomere. The genomic DNA used in the experiment had been derived from the F1 female, 0003, containing the 140kb *Hae* III and ~12kb *Mnl* I new variants. The ~12kb fragment could just be seen in the *Mnl* I digested telomeric DNA. This suggests that the ~12kb variant was derived from the telomeric DNA.

To be sure of the exact origin of the 12kb DNA fragment, the *Hae* III 140kb telomeric band was isolated from a low melting point, pulsed-field gel and the plug redigested with *Mnl* I. The resultant fragments were separated by conventional gel electrophoresis, transferred to a nylon membrane and hybridised with the $(TTAGGG)_4$ oligonucleotide probe. As shown in Figure 6.12b, the only resolved band present was the 12kb variant. A >23kb smear of hybridisation corresponding to the remainder of the 140kb *Hae* III band was also observed. Its presence suggests that *Mnl* I cleavage sites were present at >23kb but <50 kb intervals within the *Hae* III derived band; the whole telomeric smear in this individual was <50kb when total genomic DNA was digested with *Mnl* I. Hence, it appeared that the 140kb telomeric locus contained the 12kb *Mnl* I band and both appeared as new variants when compared to the C57Bl/6 and DBA/2 parents.

To determine the stage in mouse development at which the new variants had been generated, their inheritance into the next generation was investigated. This would determine if the mutation had occurred at meiosis within the germ line in the DBA/2 and C57Bl/6 inbred mice or within the somatic cells of the F1 animal, *ie* a somatic mutation. Such an experiment would also help to determine the stability of the mutation event. Hence, the female F1, 0003, was mated with an F1 male and 10 F2 offspring were analysed for the presence of the *Hae* III specific 140kb and *Mnl* I specific ~12kb new variants. The DNA fragments were inherited into the same 7 out of 10 male and female offspring, Figures 6.13a and b). The pattern of inheritance suggested that the mutation event was stable and had occurred in the germline.

6.6.4 Determination of Other Forms of Divergence from the $(TTAGGG)_n$ Repeat within the Mouse Telomere

The results obtained from *Mnl* I analysis demonstrated the existance of enzyme cleavage sites within the mouse telomere. Hybridisation studies suggested that the $(TTGGGG)_n$ repeat, which by a mutation event at the second G deoxynucleotide gives rise to an *Mnl* I

Figure 6.12 Determination of the Origin of Mnl I Specific Variable Bands

(a) Mnl I specific bands derive from the mouse telomere

Total genomic DNA from the F1 female shown in lane 5 of Figures 10(a) and 11(a) was initially digested with *Hae* III and subjected to conventional gel electrophoresis in 1% LMP agarose gel. The >23kb telomeric smear at the top of the track was isolated as described in 2.9.6; resultant DNA can be seen in lane D. This was then redigested with *Mnl* I (lanes E1, E2, and E3), and subjected to conventional gel electrophoresis together with genomic DNA of the same individual digested with *Hae* III (lanes A1, A2), *Mnl* I (lanes B1 and B2) and *Hae* III + *Mnl* I (lanes C1 and C2). The separated fragments were transferred to a nylon filter and hybridized with the $(TTAGGG)_4$ oligonucleotide probe. The repeated lanes show different exposures of the DNA fragments produced by the stated combinations of restriction enzymes. The arrow indicates a new 12kb fragment produced by *Mnl* I.

Hind III digested λ DNA size markers are indicated (kb).

(b)The 12kb Mnl I band derives from the ~140kb telomeric new variant

Total genomic DNA from the same individual as in (a) was digested with *Hae* III and subjected to PFGE in 1% LMP agarose, together with controls to establish the position of the ~140kb band arrowed (lane 1). This was cut out as a block of gel + DNA and redigested with *Mnl* I as described in 2.9.6. The restriction fragments contained in the plug were separated by conventional gel electrophoresis (lane 4) together with genomic DNA of the same individual digested with *Hae* III (lane 2) and *Mnl* I (lane 3). The DNA was then transferred to a nylon filter and hybridised with the $(TTAGGG)_4$ oligonucleotide probe. The 12kb band, produced only by *Mnl* I and visualized by conventional gel electrophoresis, is indicated.

λ concatemer and *Hind* III digested DNA size markers are indicated (kb).

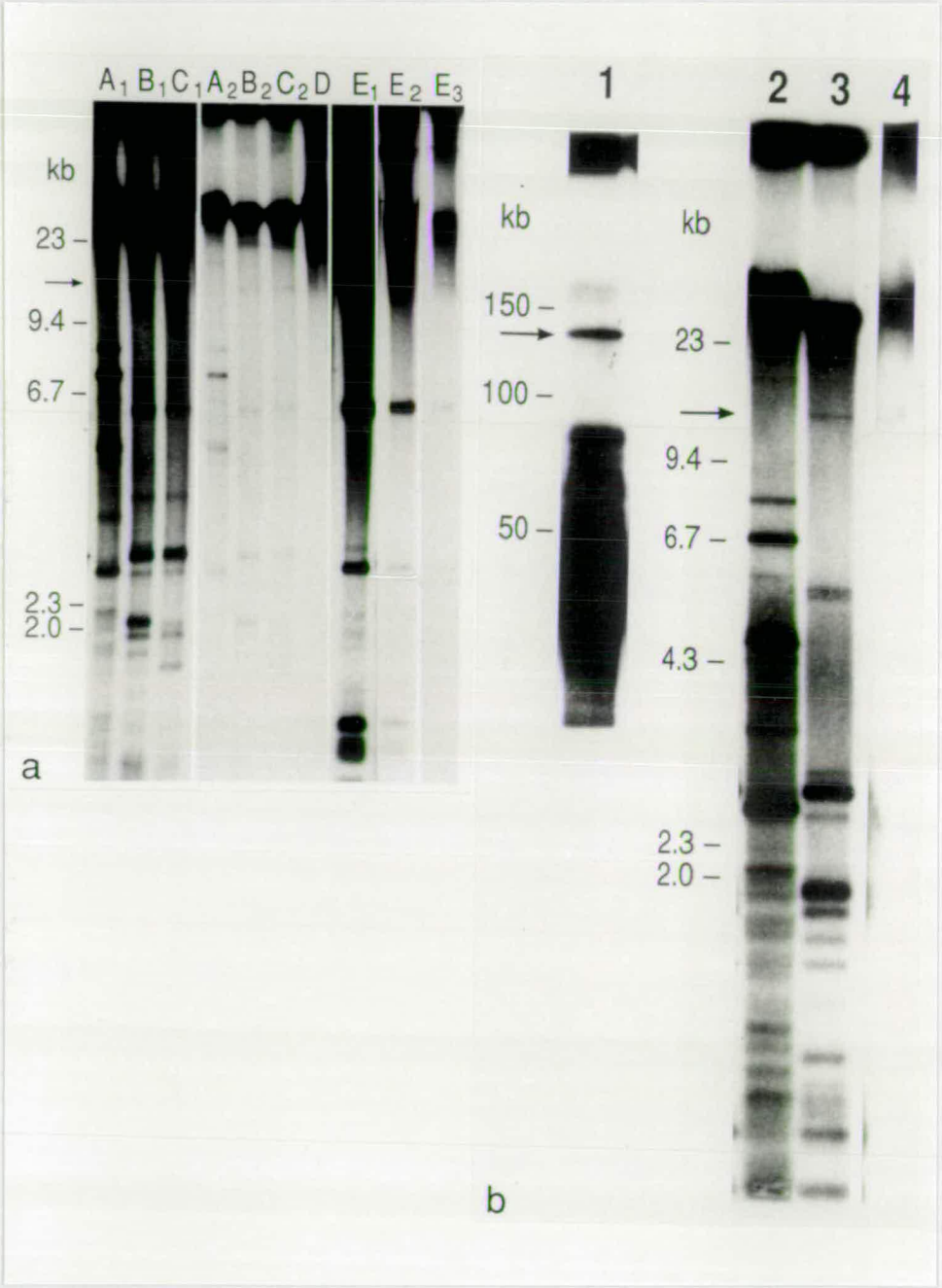

A1 B1 C1 A2 B2 C2 D E1 E2 E3
kb
23 –
9.4 –
6.7 –
2.3 –
2.0 –
a
1
kb
150 –
100 –
50 –
2 3 4
kb
23 –
9.4 –
6.7 –
4.3 –
2.3 –
2.0 –
b

Figure 6.13 Inheritance of the Telomeric $(TTAGGG)_n$ Containing ~140kb and 12kb New Variants into the F2 Generation

The female mouse containing the variants (Figures 10 and 11) was mated to another F1 mouse (from a separate C57Bl/6 x DBA/2 cross), resulting in the production of 10 F2 offspring.

(a) Total genomic DNA from the grandparents, F1 parents and F2 offspring was digested with *Mnl* I, the fragments separated by conventional gel electrophoresis, transferred to a nylon membrane and hybridised with the $(TTAGGG)_4$ oligonucleotide probe.

(b) Total genomic DNA from the F1 parents and F2 offspring was digested with *Hae* III, the fragments separated by PFGE, transferred to a nylon membrane and hybridised with the $(TTAGGG)_4$ oligonucleotide probe. The order of the F2 offspring is the same as in (a).

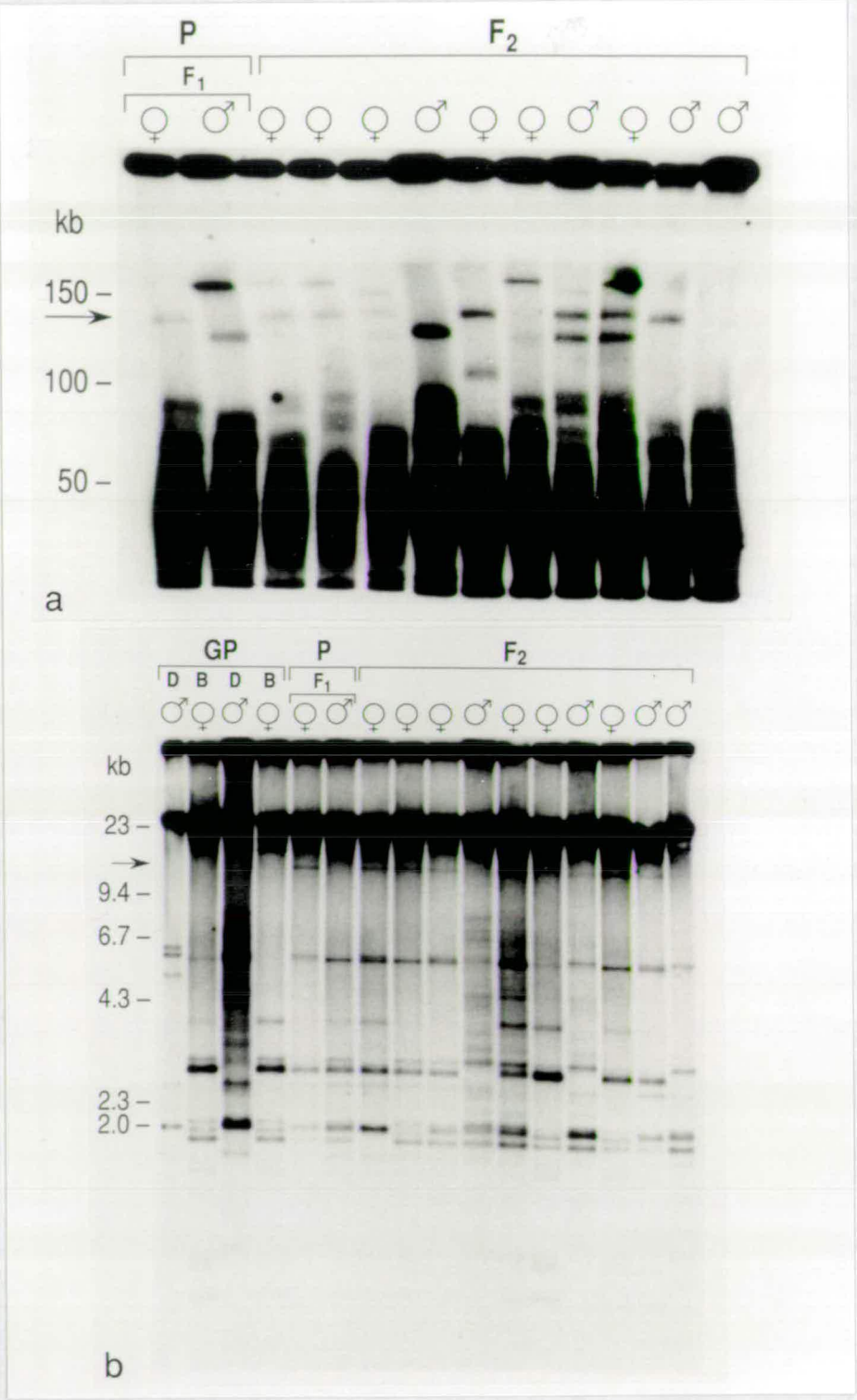
GP = grandparents

P = parents

D = DBA/2 inbred strain

B = C57Bl/6 inbred strain

λ concatemer and *Hind* III digested λ DNA size markers are indicated (kb).


P
F2
F1
kb
150 –
100 –
50 –
a
GP
P
F2
D B D B
F1
kb
23 –
9.4 –
6.7 –
4.3 –
2.3 –
2.0 –
b

cleavage site, must have been dispersed throughout the telomere. This is in contrast to the organisation of telomeric repeats at the human telomere, where the $(TTGGGG)_n$ repeat is found in clusters (Allshire *et al.*, 1989, Brown *et al.*, 1990). In addition, the results obtained implied that there was divergence from the telomeric repeat sequence. The possibility of other such events generating alternative enzyme sites was investigated.

Total genomic DNA from a C57Bl/6 x DBA/2 cross and offspring was digested with the enzymes listed in Table 6.2, *ie Hph* I, *Mse* I and *Dde* I. The fragments were separated by conventional gel electrophoresis, transferred to a nylon membrane and hybridised with the $(TTAGGG)_4$ oligonucleotide probe. *Dde* I specific fragments demonstrated the same pattern of inheritance seen with *Hae* III specific fragments; each set of parental bands was inherited into the offspring, Figure 6.14a. A very similar result was obtained when the same set of DNA's was cleaved with *Hph* I, Figure 6.14b. Some variations were seen in that certain fragments were not inherited into all of the offspring, for example in the final F1 male in Figure 6.14b. However, the variations were not on the same scale as those seen with *Mnl* I. Cleavage of the same set of DNA samples with *Mse* I revealed many weakly hybridising bands ranging in size from ~8.5-20kb which demonstrated a pattern of inheritance similar to that seen for the *Mnl* I specific bands (Figure 6.14c). *Mse* I recognises the cleavage site TTAA; conversion of the first G to an A within the (TTAGGG) repeat would result in an *Mse* I site. Variation of the Mse I bands present in the F1 animals suggests that they were derived from the telomere.

From the results it appeared that firstly the *Hph* I restriction enzyme was not cleaving the DNA at the same site as *Mnl* I. If such cleavage had occurred the same pattern of bands would be seen with both enzymes. This suggests that the divergence event creating the *Mnl* I site might have occurred within randomly dispersed (TTGGGG) repeats. Alternatively, a second mutation within the TGAGGG repeat could have resulted in the loss of the *Hph* I site. Secondly, *Mse* I seemed to recognise cleavage sites within the telomeric loci, suggesting the presence of other mutation events within the telomeric repeats.

6.7 Summary

Using a variety of different restriction enzymes upon C57Bl/6, DBA/2 and a C57Bl/6 x DBA/2 derived F2 mouse total genomic DNA, several features of the internal and telomeric $(TTAGGG)_n$ containing loci were revealed. Firstly, the telomeric loci were found to be relatively devoid of most restriction enzyme sites suggesting that they were composed of simple repeats. *Mnl* I, *Hph* I and *Mse* I were able to cleave within these loci. All three enzymes have recognition sites that would be generated upon a single base change within the $(TTAGGG)_n$ repeat sequence, as outlined in Table 6.2. The absence of an *Hph* I site in the

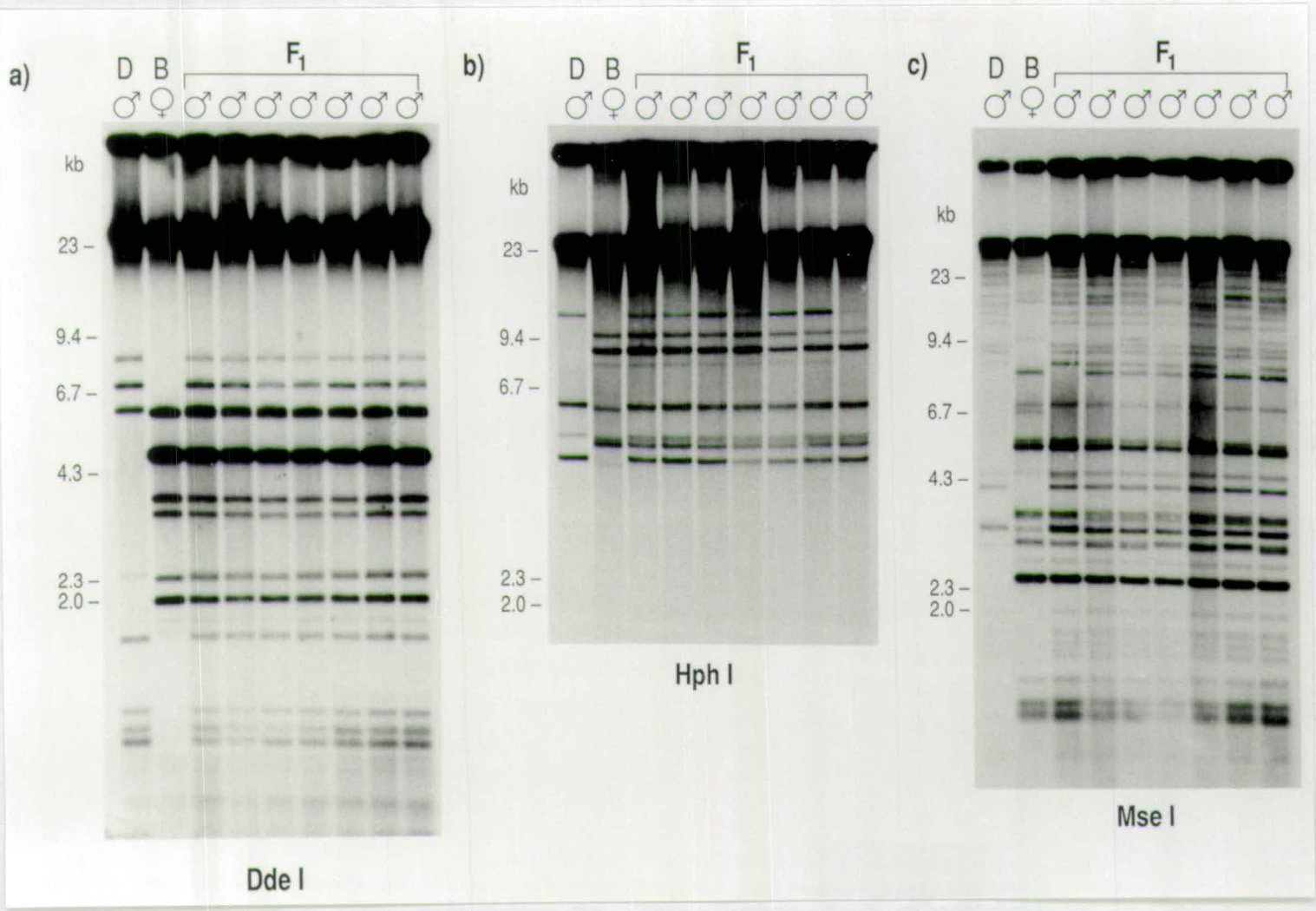
Figure 6.14 Determination of Additional Sequence Divergence within the Mouse Telomeric Repeat

Total genomic DNA from the parents and F1 offspring of a C57Bl/6 x DBA/2 mating was digested, in turn, with (a) *Dde* I, (b) *Hph* I and (c) *Mse* I, the fragments separated by conventional gel electrophoresis, transferred to a nylon membrane and hybridised with the $(TTAGGG)_4$ oligonucleotide probe.

D = DBA/2 inbred strain

B = C57Bl/6 inbred strain

Hind III digested λ DNA size markers are indicated (kb).


a)
D B F1
kb
23 –
9.4 –
6.7 –
4.3 –
2.3 –
2.0 –
Dde I
b)
D B F1
kb
23 –
9.4 –
6.7 –
2.3 –
2.0 –
Hph I
c)
D B F1
kb
23 –
9.4 –
6.7 –
4.3 –
2.3 –
2.0 –
Mse I

presence of an *Mnl* I site and *vice versa*, suggested that an additional repeat types were also present within the mouse telomere. Hybridisation analysis further showed that the mouse telomere was predominantly composed of $(TTAGGG)_n$ repeats interspersed with $(TTGGGG)_n$ repeats. Cloning and sequencing of the mouse telomere will determine the exact nature of the repeat.

Partial digestion analysis suggested that interstitial $(TTAGGG)_n$ containing loci might be composed of a single block of $(TTAGGG)_n$ together with an unrelated simple repeat. Isolation and sequencing of these loci will determine the nature of such an unrelated repeat.

Thirdly, hybridisation analysis of the mouse genomic DNA revealed that both the $(TTGGGG)_n$ and $(TGAGGG)_n$ repeats might be interspersed within the mouse genome in a similar manner to the *Alu* repeats within the human genome. *In situ* hybridisation analysis might help to determine if these repeats are dispersed throughout the mouse genome or if they are localised to specific areas like the telomere adjacent region or at the centromeres. In addition, there was very weak hybridisation of the $(TTGGGG)_4$ probe to the DNA fragments containing the $(TTAGGG)_n$ repeats. This suggests that the previously identified interstitial $(TTAGGG)_n$ containing loci also contain short stretches of this hexamer. Again, cloning and sequencing of the DNA from these loci would help to establish their exact nature.

Finally, analysis of the parents and offspring of a C57Bl/6 x DBA/2 mating revealed a 140kb new variant sized telomeric band in one of the female F1 animals. *Mnl* I digestion and conventional gel electrophoresis showed an ~12kb new variant in the same individual. It was further shown that the ~12kb band was derived from the 140kb new variant telomeric locus. Possible mechanisms for the generation of this new locus will be discussed in Chapter 8. In addition, the ~12kb, *Mnl* I new variant and the *Mse* I telomere derived bands hybridised weakly to the $(TTAGGG)_4$ oligonucleotide probe. This result might suggest that these DNA fragments contain a limited number of (TTAGGG) repeats together with some form of divergent telomere repeat that does not crosshybridise with the $(TTAGGG)_4$ probe.

Chapter 7

Mapping of Internal $(TTAGGG)_n$ Containing Loci within the Mouse Genome

7.1 Introduction

The C57Bl/6 x DBA/2 recombinant inbred (RI) mouse strains had previously been used to determine the genomic location of an interstitial $(TTGGGG)_n$ containing locus, which was mapped to the β2 microglobulin gene (Allshire *et al.*, 1989). Hence, the same method was used to attempt to map the mouse internal $(TTAGGG)_n$ containing loci. Two separate systems were used, firstly, the C57Bl/6 x DBA/2 RI mouse strains and secondly, the interspecific backcross progeny generated by mating (C57Bl/6 x *Mus spretus*) F1 females and C57Bl/6 males.

The recombinant inbred mouse strains were derived by the systematic inbreeding of two inbred, progenitor, strains, beginning at the F2 generation. An RI strain is generated after twenty such steps have been completed. Unlinked genes are randomised in the F2 generation and will be seen in either the parental or recombinant form in the RI strains. Linked genes will tend to be fixed in the original parental combination. Once each strain has been bred to homozygosity at each locus, they are typed with respect to the genetic differences found within the progenitor strains. Linkage analysis for a genetic locus or trait can be performed by finding its strain distribution pattern (SDP) in the RI strain series and comparing it to known SDP's for previously mapped loci within the same strains. It is also possible to estimate the genetic distance between two closely linked loci by determining the number of independent recombination events resulting in their separation (reviewed by Taylor, 1978).

The *Mus spretus* x C57Bl/6 interspecies backcrosses have been derived by first mating a female C57Bl/6 with a male *Mus spretus*, mouse. The resulting F1 females are fertile and the males sterile. Hence, the F1 females are crossed with C57Bl/6 males to give the backcross progeny. Unlinked genes would be randomised at meiosis by crossing-over between homologues from the two independent mouse species within the germ cells of the F1 females. Also, closely linked genes will tend to be fixed in the original parental configuration. Crossing back onto the C57Bl/6 genetic background results in all of the offspring containing the C57Bl/6 gene loci. Hence, only exchange events resulting in the randomisation of *Mus spretus* loci will be detected in the backcross progeny. Identification of linkage is conducted by determining which progeny contain two cosegregating fragments which are specific to the *Mus spretus* genome. Linkage with previously mapped loci and hence, positioning within the mouse genome, can be obtained by comparing the patterns of distribution within the backcross progeny. The extent of linkage and an estimate of the

genetic distance can be obtained by determining the recombination frequency between two such loci (reviewed by Avner *et al.*, 1988).

7.2 The Mapping of Internal $(TTAGGG)_n$ Containing Loci within the BxD Recombinant Inbred Strains of Mice

Total genomic DNA from one C57Bl/6 and one DBA/2 inbred mouse from the local and Hull colonies respectively, was digested with each of the five restriction enzymes *Hinf* I, *Hae* III, *Sau* 3A, *Alu* I and *Rsa* I. The resultant fragments were separated by conventional gel electrophoresis, transferred to a nylon membrane and hybridised with the $(TTAGGG)_4$ oligonucleotide probe, Figure 7.1. The hybridising DNA fragments produced by each enzyme showed distinct size differences when the C57Bl/6 and DBA/2 banding patterns were compared. Hence, it should be possible to determine the SDPs of internal $(TTAGGG)_n$ containing loci within the RI mouse strains.

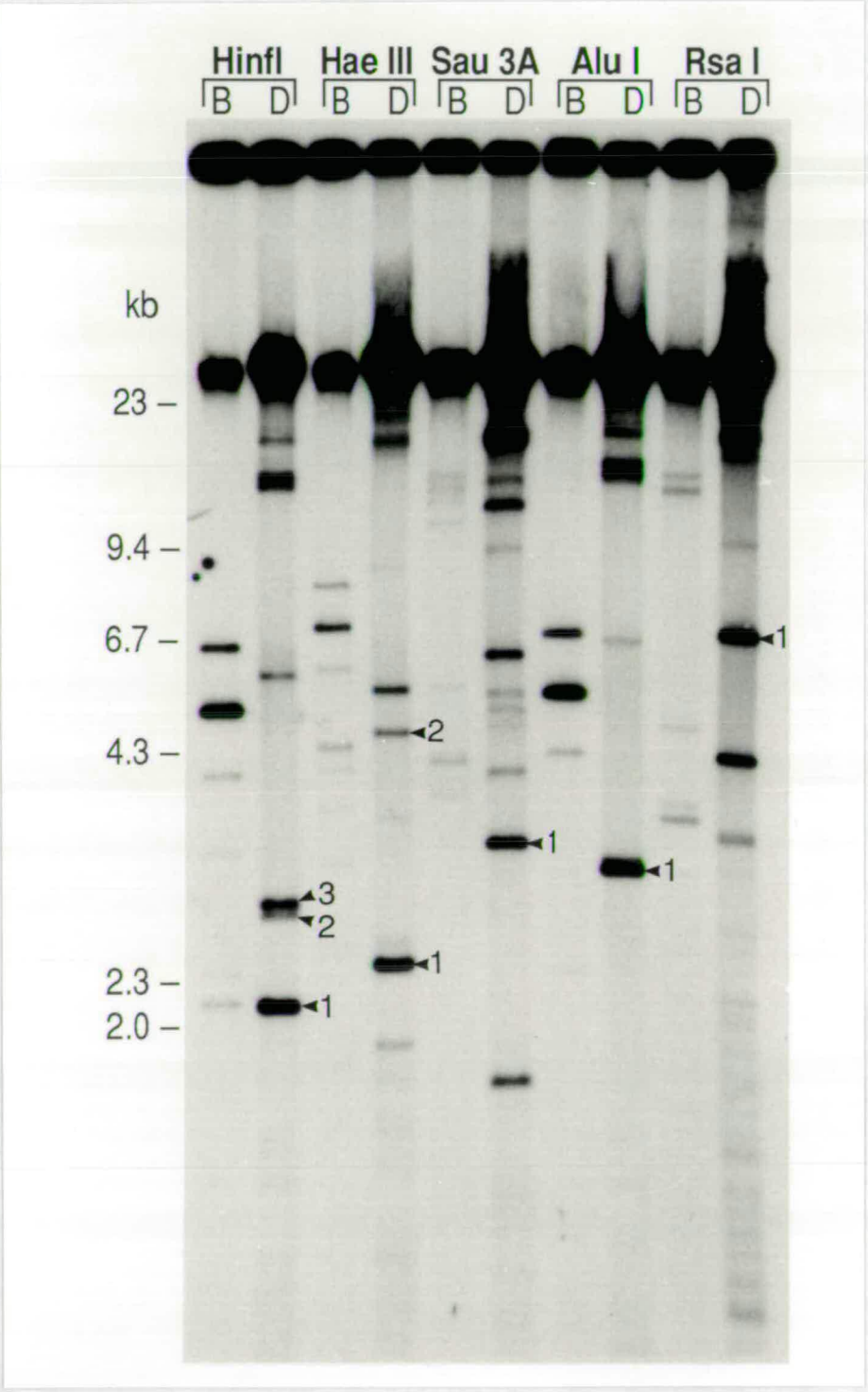
The breeding experiment conducted during the course of the project followed the inheritance of *Hae* III specific, C57Bl/6 (local colony) and DBA/2 (Hull colony) internal $(TTAGGG)_n$ containing loci through four generations. The results showed that there were two main loci specific to the C57Bl/6 genome and likewise in the DBA/2 inbred strain. The C57Bl/6 locus 1,2,3 was located upon the X chromosome and the DBA/2 locus,4, may be located within the X and Y chromosome pseudoautosomal region; the remaining loci were located upon autosomes. The C57Bl/6 and DBA/2 loci did not appear to be allelic. Hence, the SDP for each progenitor band was determined without comparison to an alternative fragment in the other progenitor strain. Each SDP was written such that the absence of the relevant band was scored as negative. The same principal was applied to mapping the hybridising fragments specific to each of the other four enzymes. It was shown previously that the bands produced by the different restriction enzymes were all derived from the same loci. Hence, total genomic DNA from the 26 RI strains together with the progenitor DNA's described in Figure 7.1 were digested with (a) *Hae* III, (b) *Hinf* I, (c) *Alu* I, (d) *Sau* 3A and (e) *Rsa* I. The fragments were separated by conventional gel electrophoresis, transferred to a nylon membrane and hybridised with the $(TTAGGG)_4$ oligonucleotide probe. The resultant SDP's obtained from the different enzymes were compared to establish the existance of any common patterns. These common bands are numbered 1-3 in Figure 7.1. The numbers, 1-3, of these bands are not intended to suggest a relationship with bands 1, 2 and 3 of the C57Bl/6 locus 1,2,3. The revised results were sent to B.A. Taylor at the Jackson Laboratories for linkage analysis. Three of the SDP's showed linkage to previously defined loci.

Figure 7.1 Comparison of $(TTAGGG)_n$ Containing Fragments Produced by Different Enzymes from the C57Bl/6 and DBA/2 Mouse Genomes

Total genomic DNA from one C57Bl/6 and DBA/2 inbred mouse from the local and Hull colonies respectively was digested with *Hinf* I, *Hae* III, *Sau* 3A, *Alu* I and *Rsa* I, the fragments separated by conventional gel electrophoresis, transferred to a nylon membrane and hybridised with the $(TTAGGG)_4$ oligonucleotide probe.

The enzyme used for digestion is indicated above each lane and D = DBA/2 and B = C57Bl/6. The numbered bands upon the photograph correspond to those DNA fragments whose inheritance was followed within the 26 RI mouse strains.

Hind III digested λ DNA size markers are indicated (kb).


HinfI
Hae III
Sau 3A
Alu I
Rsa I
B D B D B D B D B D
kb
23 –
9.4 –
6.7 –
4.3 –
2.3 –
2.0 –
1
2
3
2
1
1
2
1
1
1
1

(a) For the band designated 1 in Figure 7.1 and D1 in Figures 7.2a and b :

SDP :

RI strains	1	2	5	6	8	9	11	12	13	14	15	16	18
SDP	D	D	D	-	-	-	-	-	-	-	-	D	-
RI strains	19	20	21	22	23	24	25	27	28	29	30	31	32
SDP	-	-	D	D	D	D	-	-	-	-	D	-	D

Where D = a DNA fragment originating from the DBA/2 inbred strain.

- = absence of the DBA/2 specific band.

The same SDP was obtained with each of the five restriction enzymes used for mapping within the RI mouse strains, indicated in Figure 7.1. An example of the RI strains demonstrating the absence and presence of this band after *Hae* III digestion is shown in Figures 7.2a and b. An alternate C57Bl/6 pattern could be seen when the D1 band was absent from the *Hae* III digests of the RI strains, indicated as B1 in Figures 7.2 a and b. It was impossible to determine if this C57Bl/6 band was absent in the presence of D1. A C57Bl/6 allelic band could not be seen in the other restriction enzyme digests.

Comparison of the SDP for D1 with those of other previously defined loci showed linkage to the mouse cholecystokinin gene upon chromosome 9 (Freidman *et al.*, 1989). Three out of twenty-six of the RI strains showed recombination resulting in the separation of these two loci. From the previously defined map of chromosome 9 it appeared that the DBA/2 locus was located at the most centromere distal point; the arrangement of markers on chromosome 9 is shown in Figure 7.3 (derived from Freidman *et al.*, 1989).

(b) For the band designated 2 in Figure 7.1 and D2 in Figures 7.2 a and b.

SDP :

RI strain	1	2	5	6	8	9	11	12	13	14	15	16	18
SDP : *Hae* III	D	D	-	-	D	D	-	D	D	D	D	D	D
Hinf I	D	D	-	-	D	D	-	D	D	D	D	D	D

RI strain	19	20	21	22	23	24	25	27	28	29	30	31	32
SDP : *Hae* III	-	-	D	-	-	-	-	D	D	-	D	-	D
Hinf I	-	-	D	-	-	-	-	D	D	-	-	-	D

Where D = a DNA fragment originating from the DBA/2 inbred strain

- = absence of the DBA/2 specific band.

The *Hae* III derived band designated D2 showed linkage with one of A.Jeffreys' minisatellite family fragments. Three out of twenty-six recombination events resulting in the separation of these loci were detected. Unfortunately, the minisatellite fragment had not

Figure 7.2 SDP's of Two $(TTAGGG)_n$ Containing Loci within the RI Mouse Strains

Total genomic DNA from the DBA/2 and C57Bl/6 mice used in Figure 7.1 and from 26 RI mouse strains, numbered from 1-30, was digested with *Hae* III, the fragments separated by conventional gel electrophoresis, transferred to a nylon membrane and hybridised with the $(TTAGGG)_4$ oligonucleotide probe. The number of the inbred strain is indicated above each lane and D = DBA/2 and B = C57Bl/6. D1, D2 and B1 refer to DNA fragments that have been mapped to positions within the mouse genome.

Hind III digested λ DNA size markers are indicated (kb).

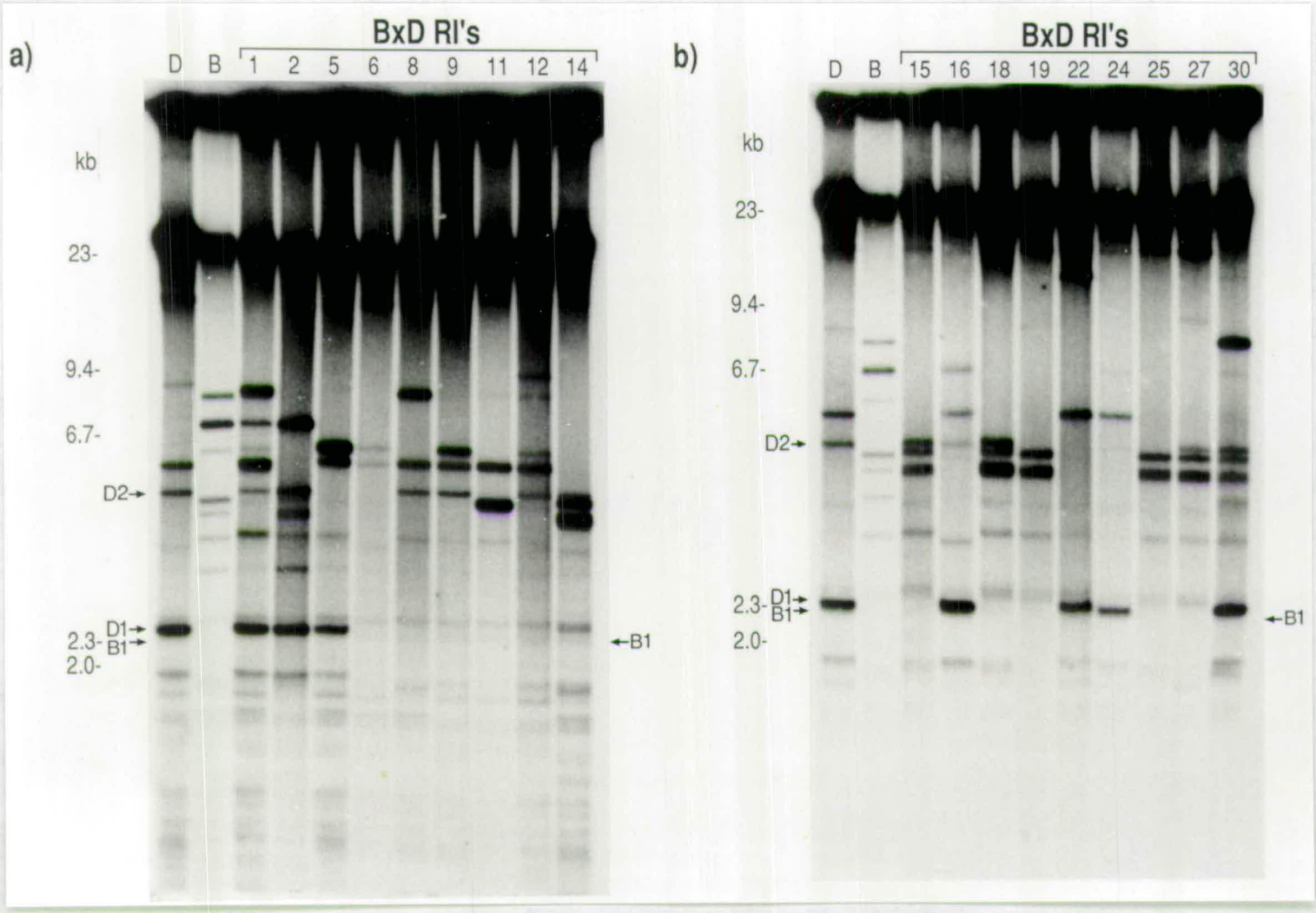

a)
BxD RI's
D B 1 2 5 6 8 9 11 12 14
kb
23-
9.4-
6.7-
D2→
2.3- D1→
B1→
2.0-
←B1
b)
BxD RI's
D B 15 16 18 19 22 24 25 27 30
kb
23-
9.4-
6.7-
D2→
2.3- D1→
B1→
2.0-
←B1

Figure 7.3 Map of the Mouse Chromosome 9 to show the Position of (TTAGGG)$_n$ Containing Locus D1

The map was derived from Friedman *et al.* (1989), which in turn was derived from *Mus spretus* backcross progeny typed for RFLP's with Cck.b5, transferrin and P450-1 (probes). The other markers were derived from the mouse gene map.

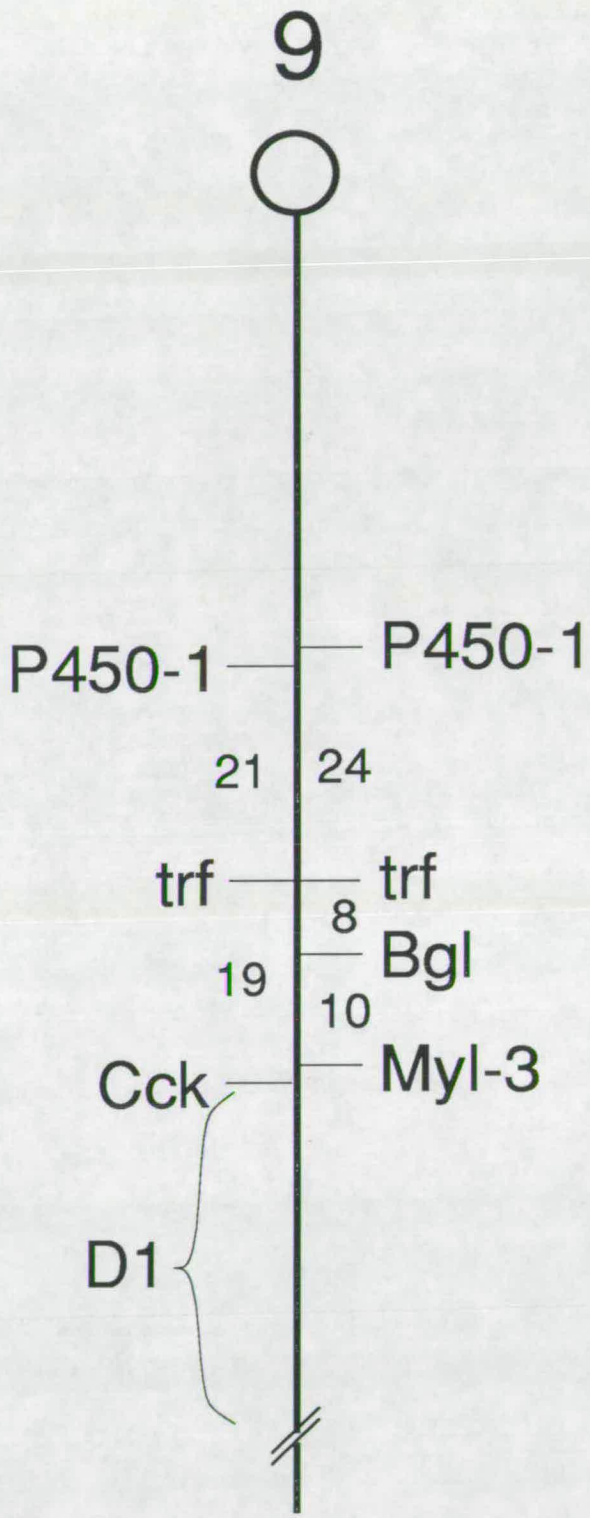
Myl-3 = myosin light chain

Bgl = β-galactosidase

The region bracketed and labelled as D1 refers to the general position of one of the DBA/2 interstitial (TTAGGG)$_n$ containing loci.

The circle just below the number 9 represents the centromere.

For the markers on the left hand side of the map: the information was derived from the *Mus spretus* backcross progeny typed for RFLP's with Cck.b5, transferrin and P450-1 (probes).
For the markers on the right hand side of the map: the information is as reported in the current mouse gene map.


9
P450-1
P450-1
21
24
trf
trf
8
Bgl
19
10
Cck
Myl-3
D1

been assigned to a mouse chromosome, hence position information for the $(TTAGGG)_n$ containing locus D2 was not obtained. The restriction enzyme *Hinf* I produced a fragment, which gave an SDP varying in only one strain from D2. The difference took the form of band absence from RI strain 30. The loss of this band resulted in the loss of linkage with the A.Jeffreys' minisatellite fragment upon likage analysis by B. A. Taylor. However, it is still likely that D2 and its *Hinf* I counterpart are related.

(c) For the band designated 3 in Figure 7.1

SDP :

RI strain	1	2	5	6	8	9	11	12	13	14	15	16	18
SDP	D	-	D	D	D	D	D	D	-	-	-	D	-

RI strain	19	20	21	22	23	24	25	27	28	29	30	31	32
SDP	-	D	-	D	D	D	-	-	-	-	-	D	D

Where D = a DNA fragment originating from the DBA/2 inbred strain.

\- = absence of the DBA/2 specific band.

The *Hinf* I derived band designated D3 showed linkage with a restriction fragment length polymorphism (RFLP) defined by Jake Lusis, HMG CoA- Synthetase (Hmgs). One out of the twenty-six RI strains showed loss of linkage between the two loci. Unfortunately, this gene had not been assigned to a chromosome. None of the other restriction enzymes used gave the same SDP seen for this *Hinf* I specific fragment.

The results of linkage analysis demonstrated that the DBA/2 internal $(TTAGGG)_n$ containing loci could be mapped within the mouse genome. It was unfortunate that two of the fragments showing linkage with previously defined loci could not be assigned to positions within the mouse genome. None of the SDP's obtained from the C57Bl/6 internal $(TTAGGG)_n$ containing loci showed linkage with the previously defined genetic markers. The inheritance patterns of the *Hae* III specific bands into several generations showed that the C57Bl/6 genome contained two main loci, one located on the X chromosome, the other autosomal and each composed of three hybridising fragments. Recombination events within the loci resulting in the separation of these fragments were observed, Chapter 5. In contrast, the DBA/2 $(TTAGGG)_n$ containing loci were composed of a single sized fragment. Hence, the absence of linkage may have been due to recombination within the loci causing disruption to the extent that authentic linkage with other markers could not be determined. Alternatively, the C57Bl/6 loci may be situated in a region of the mouse where no other markers have yet been localised.

The characteristic banding pattern identified for each restriction enzyme within the C57Bl/6 genome was extremely difficult to identify in the RI strains. For example, in Figure

7.2a and b, the *Hae* III specific locus 1,2,3 was rarely seen as a whole, more frequently the fragments were seen independently. The same was true of the 5,6,7 locus but to a lesser extent. New variant sized hybridising fragments were also seen and fragments with a stronger hybridisation intensity, Figure 7.2a and b. The latter result suggests the presence of additional copies of the (TTAGGG) repeat within these fragments perhaps in the form of extra blocks of repeat separated by unique sequence. The stated observations suggested that the C57Bl/6 derived loci were variable to the extent that mapping within the RI strains was not possible.

7.3 Mapping of Internal $(TTAGGG)_n$ Containing Loci within the Mus spretus x C57Bl/6 Interspecies Backcrosses

Total genomic DNA from the *Mus spretus* and C57Bl/6 progenitors supplied by N. Copeland and N. Jenkins was digested with each of the five restriction enzymes *Hae* III, *Hinf* I, *Alu* I, *Sau* 3A and *Rsa* I. The fragments were separated by conventional gel electrophoresis, transferred to a nylon membrane and hybridised with the $(TTAGGG)_4$ oligonucleotide probe. Comparison of the hybridising bands for each enzyme between the two progenitor strains revealed that each showed a different pattern of bands, Figure 7.4. Hence, any of the enzymes could be used to map within the backcross progeny.

Total genomic DNA from 98 of the (C57Bl/6J x *Mus spretus*)F1 female x C57Bl/6J male offspring together with the progenitors was digested with the restriction enzyme *Hae* III. The resultant fragments were separated by conventional gel electrophoresis, transferred to a nylon membrane and hybridised with the $(TTAGGG)_4$ oligonucleotide probe. Two groups of different sized fragments cosegregated in the backcross mice. The distribution patterns were sent to N. Copeland and N. Jenkins at the Frederick Cancer Research Facility for comparison with previously determined mouse genetic loci.

(a) For DNA fragments designated T1/Telo-1 indicated in Figure 7.5.

These bands showed close linkage with a number of chromosome 13 loci; the most closely linked being human interleukin in DA cells (*Hilda*) and cytotoxic T-cell lymphoid antigen-3 (*Ctla-3*). This locus was designated *Telo-1*. Eighty-two of the [(C57Bl/6J x *Mus spretus*) F1 x C57Bl/6J] mice were typed for *Telo-1*, *Ctla-3* and *Hilda*. Twelve recombinants were detected between *Hilda* and *Ctla-3*, while two were detected between *Ctla-3* and *Telo-1* (Figure 7.6a, supplied by N. Copeland and N. Jenkins). *Hilda* had previously been mapped proximal to *Ctla-3* (Justice *et al.*, 1990). Taken as a whole the linkage data indicated that *Telo-1* mapped distal to *Hilda* and *Ctla-3*, which in turn made *Telo-1* the most distal marker so far positioned

Figure 7.4 Comparison of $(TTAGGG)_n$ Containing Fragments Produced by Different Enzymes from the C57Bl/6 and Mus spretus Mouse Genomes

Total genomic from one C57Bl/6 and *Mus spretus* mouse was digested with *Hinf* I, *Hae* III, *Sau* 3A, *Alu* I and *Rsa* I, the fragments separated by conventional gel electrophoresis, transferred to a nylon membrane and hybridised with the $(TTAGGG)_4$ oligonucleotide probe. The lanes containing *Mus spretus* and C57Bl/6 DNA are indicated and, in addition, the enzyme used to digest the sample.

Hind III digested λ DNA size markers are indicated (kb).

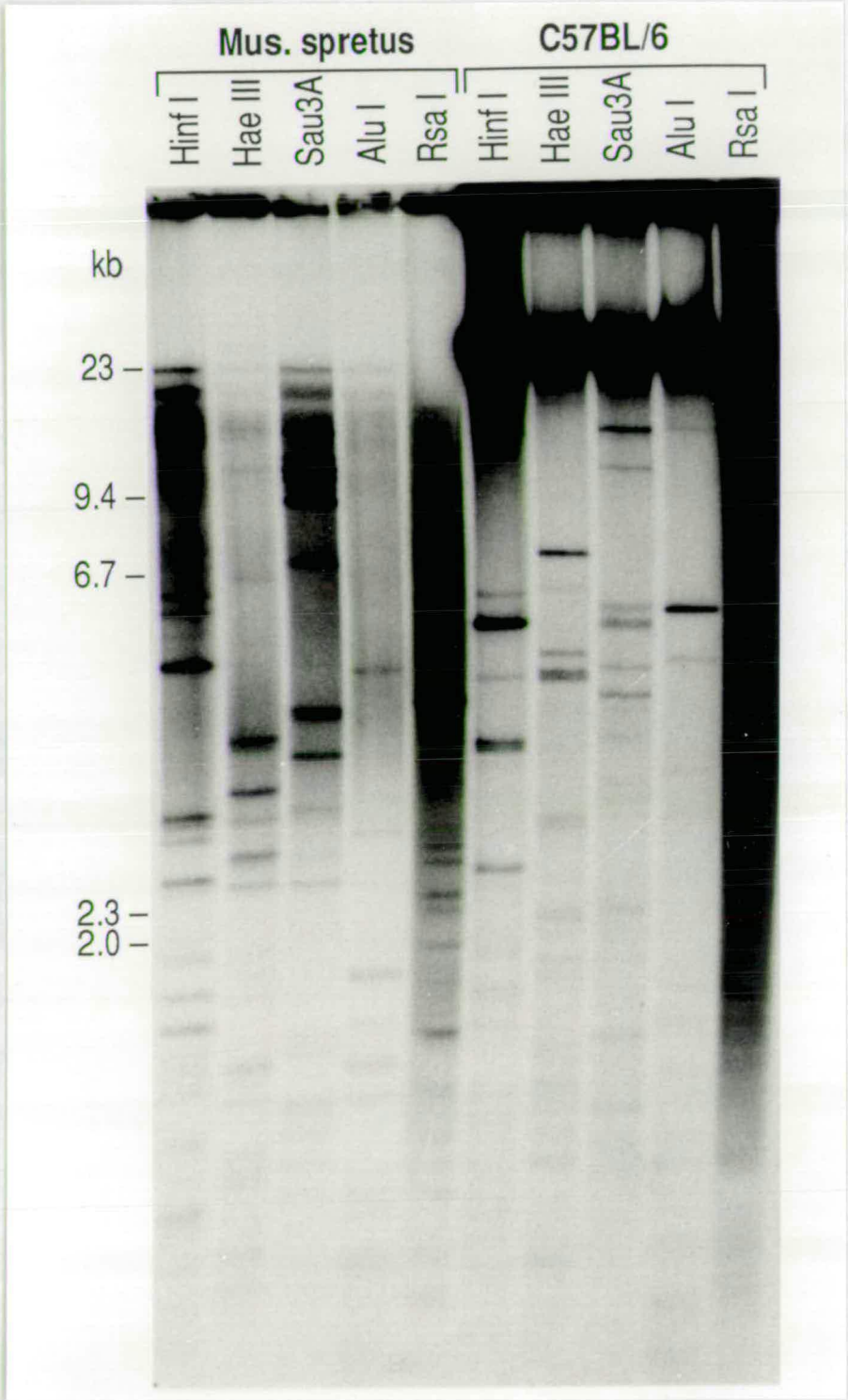

Mus. spretus
C57BL/6
Hinf I
Hae III
Sau3A
Alu I
Rsa I
Hinf I
Hae III
Sau3A
Alu I
Rsa I
kb
23
9.4
6.7
2.3
2.0

Figure 7.5 SDP's of Two $(TTAGGG)_n$ Containing Loci within the C57Bl/6 / Mus spretus Backcrosses

Total genomic DNA from the C57Bl/6 and *Mus spretus* mice used in Figure 7.4 and from 13 backcross animals, numbered 117-129, was digested with *Hae* III, the fragments separated by conventional gel electrophoresis, transferred to a nylon membrane and hybridised with the $(TTAGGG)_4$ oligonucleotide probe. The number of each backcross mouse is indicated above the relevent track and B = C57Bl/6 and S = *Mus spretus*.

T1 and T2 refer to two $(TTAGGG)_n$ containing loci that were mapped to positions within the *Mus spretus* genome.

Hind III digested λ DNA size markers are indicated (kb).

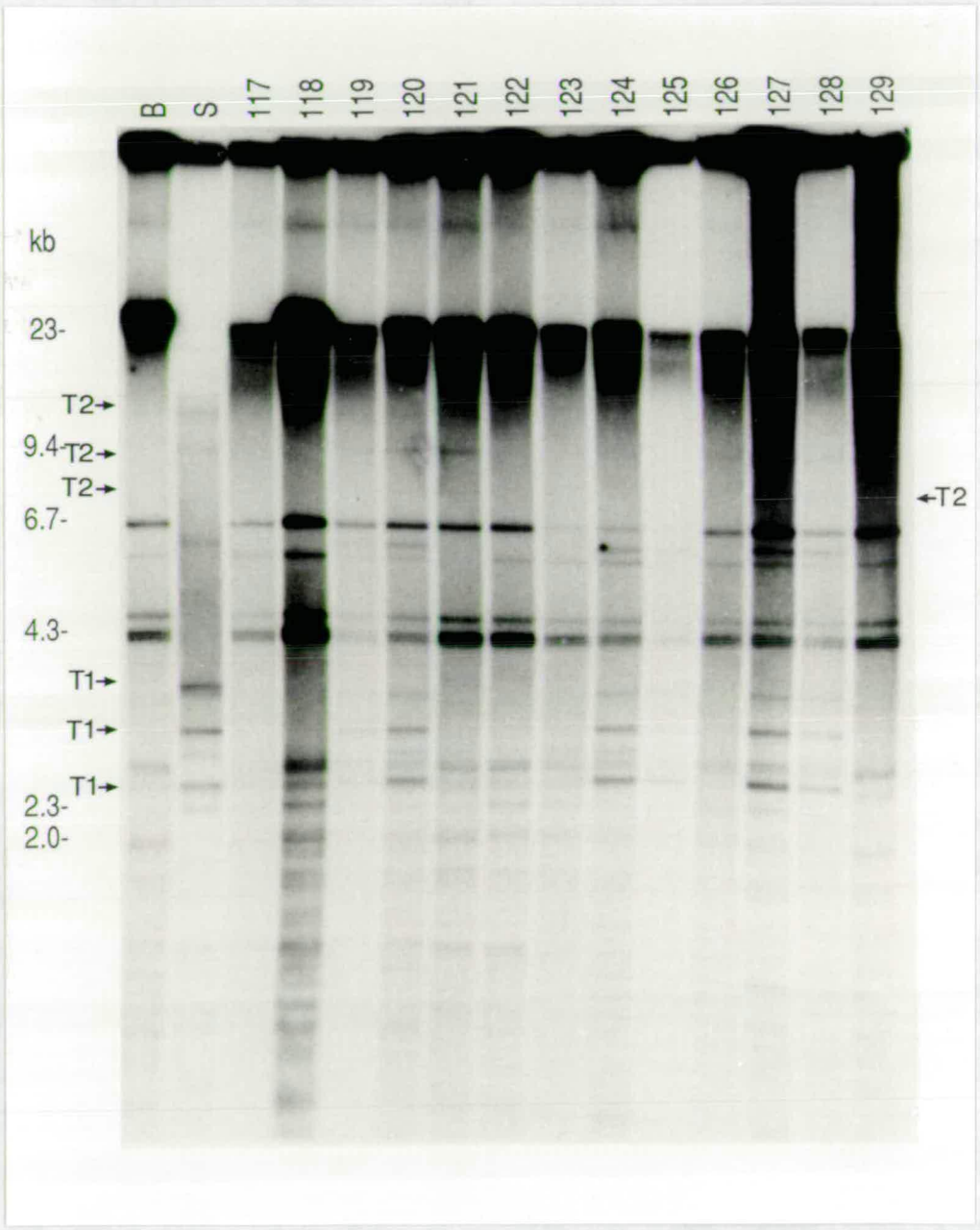
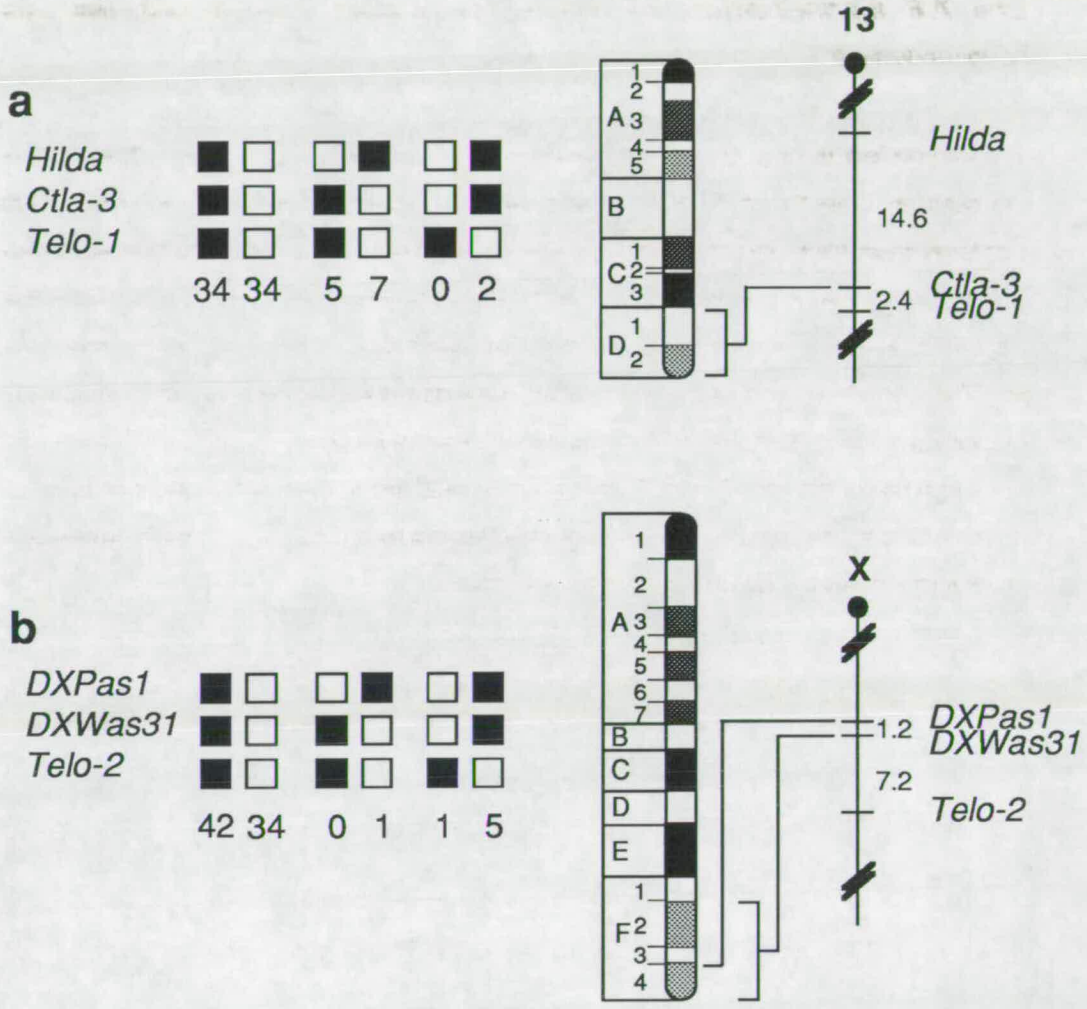

B
S
117
118
119
120
121
122
123
124
125
126
127
128
129
kb
23-
T2
9.4-
T2
T2
6.7-
4.3-
T1
T1
T1
2.3-
2.0-
T2

Figure 7.6 Localisation of Telo-1 (T1) and Telo-2 (T2) in Mouse Chromosomes

The recombination events that were observed between *Telo-1* and *Telo-2* and other loci used to position *Telo-1* and *Telo-2* on the genetic map are shown on the left of each panel. Each column represents the chromosome identified in backcross progeny that was inherited from the (C57Bl/6 x *Mus spretus*)F1 parent. The dark boxes represent the presence of a *Mus spretus* allele. The number of offspring inheriting each type of chromosome is shown at the bottom. The chromosomes in the middle of each panel represents the cytological map of chromosomes 13 and X, (Lyon and Kirby, 1990). Partial interspecific maps of chromosomes 13 and X indicating the location of *Telo-1* and *Telo-2*, are shown on the right of each panel. The interspecific map is not drawn to scale with respect to the cytological map. Where known, the location of loci on the cytological map is indicated.


a
Hilda
Ctla-3
Telo-1
34 34 5 7 0 2
13
A
B
C
D
Hilda
14.6
2.4
Ctla-3
Telo-1
b
DXPas1
DXWas31
Telo-2
42 34 0 1 1 5
X
A
B
C
D
E
F
1.2
DXPas1
DXWas31
7.2
Telo-2

upon chromosome 13. The map distance (in cM ± standard error) and the gene order of the three loci is :

centomere - *Hilda*- 14.6 ± 3.9 - *Ctla-3* 2.4 ± 1.7 - *Telo-1*(Figure 7.6a)

The linkage information was supplied by N. Copeland and N. Jenkins.

(b) For the DNA fragments designated T2/Telo-2 indicated in Figure 7.5.

These bands showed linkage with two anonymous DNA markers *DXPas1* and *DXWas31* (Amar *et al.*, 1988, Disteche *et al.*, 1989), both of which had previously been mapped to the X chromosome. Of the eighty-three [(C57Bl/6J x *Mus spretus*) F1 x C57Bl/6J] mice that were typed for all three loci, one recombinant was detected between *DXPas1* and *DXWas31* and six recombinants were detected between *DXWas31* and *Telo-2* (Figure 7.6b). Additional mapping studies, performed by N. Copeland and N. Jenkins, had demonstrated that *DXWas31* mapped distal to *DXPas1*. Hence, the resultant gene order and map distance was:

centomere - *DXPas1* - 1.2 ± - 1.2 - *DWWas31* - 7.2 ± 2.8 - *Telo-2* (Figure 7.6b)

Again, *Telo-2* is the most distal locus so far mapped on the X chromosome, Figure 7.6b. The linkage information was supplied by N. Copeland and N. Jenkins.

The previously defined loci *Hilda*, *Ctla-3*, *DXPas1* and *DXWas31* had been mapped by *in situ* hybridisation to mouse metaphase chromosomes. *Ctla-3* was positioned in band D of the distal region of chromosome 13, Figure 7.6a (Mattei *et al.*, 1987b). Since *Telo-1* maps 2.4cM distal to *Ctla-3*, the results predicted that *Telo-1* was also located in band D. Similarly, *DXWas31* had been localised to the distal region of the X chromosome, bands F2-F4 (Disteche *et al.*, 1989) while *DXPas1* had been localised to the interface of bands F3-F4, Figure 7.6b (Mattei *et al.*, 1987a, P. Avner, personal communication). From the linkage results, *Telo-2* mapped 7.2 and 8.4 cM distal to *DXWas31* and *DXPas1*, respectively, Hence, the *in situ* results strongly suggested that *Telo-2* also mapped to band F with the most likely location being band F4. Therefore, it appeared from both the genetic and cytological mapping data were consistant with a telomeric or subtelomeric chromosomal localisation of internal $(TTAGGG)_n$ containing loci within the *Mus spretus* mouse chromosomes. All of the linkage analysis, genetic maps and Figures 7.6a and b were supplied by N. Copeland and N. Jenkins.

7.4 Summary

From analysis of the RI and backcross strains of mice, one of the DBA/2 loci, D1 or 10, was mapped to the most centromere distal point upon chromosome 9, while in the backcross mice two loci were mapped to the most distal points upon the 13 and X chromosomes. The two main points to arise from mapping are that, firstly, the interstitial $(TTAGGG)_n$ containing

loci would appear to clustered within subtelomeric regions of the mouse chromosomes. Secondly, in both the C57Bl/6 and *Mus spretus* genomes a $(TTAGGG)_n$ locus was found upon the X chromosome. In addition, the *Mus spretus* locus appeared to located within the pseudoautosomal region. Hence, it is possible that the DBA/2 locus 4 is also present within this region and that a $(TTAGGG)_n$ containing locus is a conserved feature within the sex chromosomes of different mouse species. The possible role and origin of these loci within the subtelomeric regions of mouse chromosomes will be discussed in Chapter 8.

Chapter 8

Discussion

8.1 Summary of the Properties of Telomeric and Interstitial $(TTAGGG)_n$ Containing Loci

Characterisation of telomeric and interstitial $(TTAGGG)_n$ containing loci was carried out within the mouse genome. This involved the determination of the genomic location of these repeats, mapping within inbred mouse strains, restriction enzyme sequence analysis, assessment of the level of variability upon inheritance into subsequent generations and determination of the presence of other, related repeats within the mouse genome. From such analysis, several properties of telomere repeat containing loci were discovered.

8.1.1 Mouse Telomeres

The nonresolvable smear of $(TTAGGG)_n$ hybridising material previously observed within the DBA/2 inbred mouse genomic DNA (Allshire *et al.*, 1989) was shown, by PFGE, to be composed of discrete bands ranging in size from ~30-150kb. Upon analysis of other mouse strains, it was found that, with the exception of the *Mus spretus* and *Mus caroli* strains, the discrete bands were a common feature, with each strain having its own specific banding pattern, Figure 4.1b. In *Mus caroli*, the smear of hybridisation ranged from ~20-30kb. It was difficult to determine by PFGE analysis whether the hybridising material was composed of discrete bands, Figures 4.1a and 4.8a. The *Mus spretus* DNA lacked a nonresolvable smear of hybridisation; this was replaced by a smear of signal between 4.3-6.7kb (shown in Figure 4.1a).

Bal 31 and subsequent restriction enzyme, digestion, electrophoretic separation and hybridisation with the $(TTAGGG)_4$ oligonucleotide probe showed that the nonresolvable smear and discrete bands in the DBA/2 (Figure 4.2) and C57Bl/6 DNA (data not shown) were telomeric. In addition, the *Mus caroli* ~20-30kb region was also telomeric (Figures 4.4a). *In situ* hybridisation of the $(TTAGGG)_4$ oligonucleotide probe, by the PRINS technique, to mouse metaphase chromosomes provided further evidence that mouse telomeres were composed of *Trypanosoma*-like $(TTAGGG)_n$ telomere repeats (Figures 4.6a and b).

Analysis of the telomeric, $(TTAGGG)_4$ hybridising bands within the inbred mouse strains, DBA/2 and C57Bl/6, revealed that members of an inbred strain showed variability in the banding pattern observed (Figure 4.7a). This was unexpected as the inbred mouse strains have been bred to a stage where homologous chromosomes are homozygous along their entire length and exceptions to this rule are rare. Analysis of different tissues from inbred mice revealed that the variability was germ line rather than somatic in origin (Figure 4.9a).

Inheritance studies following the progression of the telomeric $(TTAGGG)_n$ containing loci into further generations of an initial DBA/2 x C57Bl/6 mating revealed that each was inherited as a distinct locus. In addition, the loci were inherited into some but not all of the offspring, suggesting that the inbred strains are heterozygous for these loci. The telomeric loci were also subject to a high rate of new variant generation (Chapter 5).

Restriction enzyme sequence analysis of the telomeric loci revealed that they were devoid of most enzyme sites, demonstrating that these regions are composed of simple repeats. Hybridisation analysis suggested that this repeat was predominantly $(TTAGGG)_n$. Restriction enzyme sites created by single base substitutions within the (TTAGGG) hexamers demonstrated that there was divergence from the repeat and that it was scattered at random through the telomeric DNA sequence. It was also possible that the *Tetrahymena*-like $(TTGGGG)_n$ telomere repeats were present scattered at random through the mouse telomere (Chapter 6).

8.1.2 Interstitial $(TTAGGG)_n$ Containing Loci

The discrete bands within the DBA/2 DNA hybridising with the $(TTAGGG)_4$ oligonucleotide probe (Allshire *et al.*, 1989) were shown to be present within the genomes of other mouse strains and species; each strain possessing its own specific pattern of bands (Figure 4.1a). These bands were shown to be insensitive to Bal 31 digestion and therefore, located at a site internal to the mouse telomeres (Section 4.3). *In situ* hybridisation of the $(TTAGGG)_4$ oligonucleotide probe to mouse metaphase chromosomes also suggested that these repeats were present at interstitial sites (Figure 4.6a and b).

Analysis of different members of the DBA/2 and C57Bl/6 inbred strains revealed no variation in the interstitial banding patterns between members of a colony. However, a low level of variation was detected when animals of the same inbred strain but from different colonies were compared (Figures 4.7b). Hence, it appeared that the inbred strains were homozygous for the interstitial $(TTAGGG)_n$ containing loci. Analysis of different tissues from the different mouse strains demonstrated that the interstitial loci were somatically stable (Section 4.6).

Inheritance studies following the DBA/2 and C57Bl/6 specific interstitial loci into subsequent generations after cross-strain mating revealed that each possessed two main loci. The DBA/2 loci were each composed of a single size of hybridising DNA fragment, while the C57Bl/6 loci were multicomponent in nature, each being composed of three different sized bands. New variant sized bands were generated during the course of the mating series, but at a much lower frequency than for the telomeric loci. Inheritance studies together with mapping in the BxD RI strains and *Mus spretus* / C57Bl/6 backcross mice revealed that in both the C57Bl/6 and *Mus spretus* strains one of the loci mapped to the X chromosome, and

in the DBA strain, the presence of a locus within the pseudoautosomal region was suspected (Chapters 5 and 7). In addition, mapping of these loci within the inbred strains revealed that they tended to be present within the subtelomeric regions of the mouse chromosomes (Chapter 7).

Restriction enzyme sequence analysis of the interstitial $(TTAGGG)_n$ containing loci showed that they were composed of simple repeats (Section 6.3). Partial enzyme digestion analysis suggested that the DBA/2 and C57Bl/6 specific bands may be composed of a single block of $(TTAGGG)_n$ repeats (Figure 6.5b and c). In addition, variations in hybridisiation signal intensity that were unrelated to size and a lack of frequently cutting restriction enzyme sites suggests that an unrelated simple repeat may also be present within these loci. Divergence from the $(TTAGGG)_n$ repeated sequence was detected using the restriction enzyme *Mnl* I (Section 6.3.3). However, as previously described, this divergence did not give rise to an *Hph* I site, suggesting that either, there were some *Tetrahymena*-like $(TTGGGG)_n$ repeats or, some other form of telomere-like hexamer repeat, within the interstitial loci. Hybridisation analysis using the $(TTGGGG)_4$ and $(TGAGGG)_4$ oligonucleotide probes showed that there may be some $(TTGGGG)_n$ repeat present within the interstitial telomere repeat containing loci and that both types of repeat maybe dispersed throughout the mouse genome in a similar manner to the *Alu* repeats found within the human genome (Section 6.5).

8.2 New Variant Generation at Telomeric and Interstitial $(TTAGGG)_n$ Containing Loci

Events resulting in the separation of bands from the C57Bl/6 loci, 1,2,3 and 5,6,7, were observed within the DBA/2 and C57Bl/6 breeding series. From such data a possible order for the bands making up each multicomponent locus and an estimate of the recombination frequency separating each band was calculated. Hence, for the 1,2,3 locus the map reads :

2 — 1 — 3

0.6% ~5% = Recombination frequency

and the locus 5,6,7 :

7 — 5 — 6

1.7% 0.6% = Recombination frequency

The new variant generation rate at telomeric and interstitial loci was calculated to be 20.9% and 0.51%, respectively, per gamete. Hence, new variant generation occurred at a rate ~41 times higher at telomeric compared to interstitial loci.

8.2.1 Possible Mechanisms of New Variant Generation at Telomeric Loci

When considering the possible mechanisms of new variant generation at telomeres, it is important to take into account the two proposed processes for the replication of these regions of the genome. Firstly, there is the recombination-mediated telomere replication model (Section 1.2.4(b)) which essentially suggests that after DNA replication, the 3' overhang, produced by primer excision from the end of the chromosome, strand invades and base pairs with a complementary region upon a sister chromatid or a homologous chromosome. Subsequent replication would proceed using the donor telomere as a template. After dissociation, the terminus would carry a single strand tail composed of new repeats that would serve as a template for primase and conventional DNA polymerase-mediated replication of the complementary strand. The second model involves telomere terminal transferase-mediated replication (Section 1.2.4(c)). Here an enzyme described as the telomere terminal transferase or Telomerase, adds telomere repeats to the ends of chromosomes in a non-templated manner. The loss of repeats from incomplete replication (Watson, 1972) is balanced with addition by the enzyme.

Both recombination and telomere terminal transferase, mediated replication could be applied to the generation of new variants at mouse telomeres. Hence, during recombination-mediated replication, invasion of the 3' single strand from one terminus into a second telomere acting as a template could result in an increase in size of the original locus. The extent of the size increase being dependent upon the size of the template telomere and the point of invasion and base pairing. Telomere terminal transferase mediated replication would predict the addition of a variable number of tandem repeats to the ends of different mouse chromosomes.

In order to determine which mechanism of telomere replication might play a role in the generation of new variant sized telomere derived bands, the two processes can be applied to the 140kb and ~12kb new variants described in Section 6.6. As indicated in Figure 6.10a, a 140kb, $(TTAGGG)_4$ hybridising, new variant sized band was observed in an F1 female of a DBA/2 x C57Bl/6 mating. This DNA fragment was also inherited from the F1 female into 7 out of 10 male and female F2 offspring (Figure 6.13b). In addition to the 140kb DNA band, an ~12kb new variant, produced by *Mnl* I, was also observed and shown to be derived from the 140kb DNA fragment and inherited into the F2 generation (Figures 6.11b, 6.12a and b, 6.13a). Hence, if the recombination mediated mechanism for telomere replication is considered for the generation of the 140kb and ~12kb new variants it would be predicted that the invading strand would contain an *Mnl* I enzyme site close to its terminus and that it would become aligned close to a second such site within the template strand. Subsequent replication would result in two *Mnl* I sites separated by ~12kb of telomere repeat and the 140kb new variant sized band.

If the second replication model is applied, it would be predicted that the 140kb new variant was generated by telomere repeat addition by a mouse 'telomerase' enzyme. As the parental telomeres that could be involved in the generation event were ~100kb or smaller (Figure 6.10a), ~40kb or greater of repeat would have to be added to one of the parental telomeres. The *Mnl* I sites could be generated by infidelity in the addition of deoxyribonucleotides to the 3' tail of the telomere.

The 140kb and ~12kb new variants were inherited into some but not all of the F2 offspring, as would be expected of a heterozygous locus. If either of the described models is the major mechanism of telomere replication in the mouse a further increase in size of the new variant sized telomere might be expected; this was not observed, Figure 6.13b. In addition, as shown in Figures 5.2a, b and c, and 5.4a and b, parental telomeric loci could be seen to be inherited into F1 and F2 offspring. Again, if the same mechanism giving rise to the 140kb new variant was involved in telomere replication, inheritance of the same sized loci would not be expected. If a mouse telomerase was responsible for the addition of ≥40kb of telomere repeat it might be predicted that similar size increases would be observed upon band inheritance into further generations; this was not the case (Figure 6.13b). Studies into telomere replication within *Tetrahymena* and *Trypanosoma* revealed that the telomeres increased in size by a few base pairs at each cell cycle and in *Tetrahymena* this was predicted to be catalysed by Telomerase (Bernards *et al.*, 1983, Larson *et al.*, 1987). It is possible that the same may be true of telomere replication at the ends of mouse chromosomes. If a mouse telomerase added only a small number of repeats per round of replication the overall size increase of the telomere would be so small as to make detection by PFGE very difficult. This would leave the recombination mechanism as the model for new variant generation. In addition, new variant generation at telomeres appears to be occuring during meiosis of gametogenesis as no somatic differences in telomere loci could be detected within mice (Section 4.6, Figure 4.9a) and the 140kb new variant was inherited into further generations. Hence, either the mouse telomerase is randomly overactive and extremely processive during gametogenesis, adding large arrays of telomeric repeat to pre-existing telomeres, or as would seem more likely, mouse telomeres demonstrate a high level of non-reciprocal, recombination-mediated new variant generation.

A second mechanism for new variant generation would predict that the 140kb new variant was generated from the ~150kb DBA/2 parental telomere DNA fragment by the loss of ~10kb of telomere repeat (Figure 6.10a). For the simultaneous generation of the ~12kb new variant an *Mnl* I site would be present ~22kb from the terminus of the DBA/2 ~150kb telomere. Digestion of genomic DNA from this individual and separation of the fragments by conventional gel electrophoresis would not allow the visualisation of a band of ~22kb. To determine whether the ~12kb new variant does represent the most terminal fragment of the

140kb new variant, Bal 31 digestion studies need to be performed. However, the fact that a discrete, non-smeared, ~12kb band is observed suggests that this is not a terminal fragment.

A third and final mechanism would predict the generation of new variants at telomeres by homologous recombination during meiosis of gametogenesis. Until mouse telomeres have been more extensively characterised and the existence of a mouse telomerase and the level of its catalytic activity determined, it is impossible to say which of the above mechanisms represents the method of new variant generation at mouse telomeres.

8.2.2 Possible Mechanisms of New Variant Generation at the Interstitial, $(TTAGGG)_n$ Containing Loci

Two new variant sized bands, NV2 and NVQ, were observed during the inheritance of the interstitial $(TTAGGG)_n$ containing loci into several generations of an initial C57Bl/6 x DBA/2 cross strain mating. NV2 is only slightly larger in size than band 2 of locus 1,2,3 but has a much lower hybridisation signal than 2 with the $(TTAGGG)_4$ oligonucleotide probe. NVQ is larger than band 1, so easily distinguishable and again, has a lower hybridisation signal than 1 and 2. Both bands show a similar level of hybridisation to that seen for band 3 of the C57Bl/6 derived locus 1,2,3 suggesting a much smaller amount of telomeric repeat within the new variants (Chapters 5 and 6).

The two new variants could have been produced by either recombination between homologous chromosomes or by unequal crossing over between sister chromatids, during gametogenesis. Both new variants were inherited into subsequent generations, pinpointing gametogenesis as their origin. Determination of the locus from which each new variant was generated and examination of the flanking markers will help to establish the mechanism of new variant generation. The unequal exchange of genetic material between sister chromatids would alter the amount of $(TTAGGG)_n$ repeat within the resultant loci but would not alter the flanking DNA as both chromatids are generated from the replication of the same DNA molecule. However, homlogous recombination would result in the exchange of DNA between non-identical chromosomes which may be reflected in differences in the DNA flanking the new variant loci.

Mapping of the interstitial loci within the BxD RI strains also revealed deviations from the expected parental banding patterns. New variant sized bands were detected in some of the inbred strains; being identified as new variants by their presence in only one of the RI strains, Figure 7.2. A more common occurrence was the presence of a band that hybridised more intensely than the same DNA fragment within the derivative parental strain, Figure 7.2. For example, in the BxD RI strains 1 and 8, the DNA fragment corresponding to band 1 of the C57Bl/6 locus 1,2,3, hybridised more intensely with the $(TTAGGG)_4$ oligonucleotide probe than in the C57Bl/6 parental DNA. As discussed above, the new variant sized bands could

have been generated by either crossing over between homologous chromosomes or unequal crossing over between sister chromatids. The increased hybridisation signal intensity seen with some of the parental bands in the RI strains suggests the presence of additional $(TTAGGG)_n$ repeat. Such expansion of the locus could arise by unequal crossing over between either sister chromatids. The tandemly repeated minisatellites identified by Jeffreys *et al.* (1985b) are postulated to generate new variants by an increase in the number of tandem repeats by sister chromatid exchange (Jeffreys *et al.*, 1985b). However, unlike the minisatellites, the parental bands showing an increase in hybridisation intensity, remain the original size; a size increase would be expected to accompany an expansion of $(TTAGGG)_n$ repeats. Hence, these DNA fragments may be derived from the smaller C57Bl/6 and DBA/2 bands, an increase in the number of repeats leading to a size increase equal to that of, for example, band 1 of the C57Bl/6 locus 1,2,3. Only analysis of the flanking DNA sequences of the new variants and more intensely hybridising DNA fragments will determine the exact mechanism of new variant generation and repeat expansion repectively, in the C57Bl/6 and DBA/2 derived $(TTAGGG)_n$ containing loci.

8.2.3 Comparison of the Levels of Variability at Telomeric Compared to Interstitial $(TTAGGG)_n$ Containing Loci

The new variant generation rate was shown to be markedly higher at telomeric compared to interstitial $(TTAGGG)_n$ containing loci. One possible reason for this vast difference might result from the different locations of the two types of locus within the mouse genome. As discussed in Section 1.2.1, telomeres are thought to take part in telomere-telomere associations which may play a role in the process of chromosome pairing prior to meiosis by bringing all of the chromosomes into close proximity within the nucleus. Such associations could promote recombination events and so new variant generation at telomeric loci. The same level of association would not be expected to occur at the interstitial loci, where recombination events within the loci would expected to occur purely by chance crossing over events between either sister chromatids or homologous chromosomes.

An alternative explanation would suggest that either telomerase- or recombination-mediated replication of mouse telomeres results in some large size changes at mouse telomeres, resulting in a high level of new variant generation. The interstitial loci would be replicated by conventional DNA replication and so not subject to the same level of new variant generation. Determination of the mechanism of replication at mouse telomeres is required before any real conclusions about the difference in new variant generation at the two different loci can be formed.

8.3 Sequence Analysis of Telomeric and Interstitial $(TTAGGG)_n$ Containing Loci

Using a selection of different restriction enzymes it was shown that both the telomeric and interstitial $(TTAGGG)_n$ containing loci were composed of simple repeats, Sections 6.2 and 6.3. Comparison of the hybridisation signal intensity produced by the $(TTAGGG)_4$ oligonucleotide probe between mouse and human telomeres suggested that mouse telomeres were composed predominantly of $(TTAGGG)_n$ repeats (Figure 6.8). While restriction enzyme analysis, using enzymes that recognise single base changes within the telomere repeat, demonstrated that other forms of simple repeat were present, dispersed at random through the mouse telomere. Further evidence for such sequence organisation was revealed upon hybridisation of the *Tetrahymena*-like telomere repeat $(TTGGGG)_4$, and the divergent $(TGAGGG)_4$, oligonucleotide probes to mouse genomic DNA. In addition, it appeared that the $(TTGGGG)_n$ repeat was also dispersed within the interstitial loci (Section 6.5). Hybridisation of both oligonucleotides to a conventional gel separation of mouse genomic DNA fragments gave a similar pattern of hybridisation to that observed for the *Alu* repeats within the human genome (Figure 6.7). This may suggest that both types of hexamer repeat are dispersed throughout the mouse genome.

Human telomeres have been shown to be composed of $(TTAGGG)_n$ repeats at the most terminal region of the chromosomes, with clusters of (TTGGGG) and (TGAGGG) repeats forming the border with the subtelomeric repeated DNA (Allshire *et al.*, 1989, Brown *et al.*, 1990). As discussed above, no such clustering of repeats was seen within the mouse telomere. It has been proposed that the human (TTGGGG)- and (TGAGGG)- like repeats arose by the acquisition of mutations in copies of the TTAGGG repeat to give variant forms which subsequently became amplified and transferred to all telomeres by recombinogenic interactions. If this does prove to be the case then it might be expected that the heterogenity would remain at a proximal location, since this region of the telomeric repeat would be less likely to be exposed and therefore, corrected by a putative human telomerase. In addition, the possible presence of such clusters of repeat at all human telomeres suggests that they may have some role in telomere function, maybe in providing binding sites for proteins involved in a putative human telomere-protein complex (Allshire *et al.*, 1989).

Clustering of TTGGGG and TGAGGG repeats was not observed in the mouse telomere, instead it appeared that there were very few variant repeats and that these were dispersed throughout the repeat array. From the above hypothesis for human telomeres, this would suggest that the whole of the mouse telomere was accessible to the actions of a putative mouse telomerase. However, the faithful inheritance of the 140kb and ~12kb new variants into a second generation suggested that if a telomerase enzyme was involved in mouse

telomere replication, it would only act upon the most terminal few kb of repeat. Hence, the presence of random divergent repeats within the mouse telomere suggests that point mutations are altering the TTAGGG repeats but that these events are not being expanded to generate tandem arrays of either TTGGGG or TGAGGG or a related repeat. Such conservation may indicate that, in contrast to human telomeres, the whole of the $(TTAGGG)_n$ repeat array making up the mouse telomere is involved in its structure and function within the nucleus. The presence of base substitutions would not interfere with this function, but the need to retain almost pure $(TTAGGG)_n$ repeats would prevent expansion of the mutated units.

Analysis of the $(TTAGGG)_n$ repeats using restriction enzymes that recognised sites resulting from single base changes within the repeat unit demonstrated that an *Mnl* I site was created in the absence of an *Hph* I site. This suggested that the *Mnl* I site had arisen from a single base change within a *Tetrahymena*-like (TTGGGG) or related repeat unit, Section 6.2.2. As has already been discussed, hybridisation analysis revealed that it was unlikely that the TTGGGG repeats were present in clusters and it seems unlikely that single repeat units occuring at random within the mouse telomere could become specifically mutated. It is possible that the TTGGGG repeats, although present within the repeat array, are not involved in its function, while the surrounding TTAGGG repeats are all involved in telomere function and so highly conserved. Hence, these TTGGGG repeats would be more susceptible to mutation.

An alternative explanation would predict that the *Mnl* I site could have been created within a block of repeats that differ from those examined in this study and that do not crosshybridise with any of the probes used. Hence, the mouse telomere could be composed of arrays of $(TTAGGG)_n$ repeats interspersed with an, as yet, undefined simple repeat. This possibility is reinforced by the results obtained from digestion of mouse genomic DNA with the restriction enzyme *Mse* I. *Mse* I cleavage of mouse genomic DNA, from a C57Bl/6 x DBA/2 cross plus the F1 offspring, followed by conventional gel electrophoresis and hybridisation with the $(TTAGGG)_4$ oligonucleotide probe revealed many weakly hybridising bands ranging in size from ~8.5-20kb which were inherited into some but not all of the offspring (Figure 6.14c). This pattern of inheritance being similar to that seen for the *Mnl* I specific bands (Figure 6.11b). Hence, it was possible that *Mse* I was cutting within the mouse telomere. The low level of hybridisation of the $(TTAGGG)_4$ oligonucleotide probe to the resultant fragments might suggest the presence of a simple repeat other than the $(TTAGGG)_n$ hexamer. A similar, low level of hybridisation was also observed for the ~12kb, *Mnl* I new variant (discussed in Section 6.6, Figure 6.11b) again suggesting the presence of an undefined, simple repeat. It is impossible to predict from the results presented in this study how such a simple repeat might be organised within the mouse telomere. However, Bal 31 digestion of *Mus caroli* total genomic DNA, followed by PFGE and hybridisation with the

$(TTAGGG)_4$ oligonucleotide probe, suggested that the most internal ~6-8kb of the telomere was composed of an undefined simple repeat (Section 4.3, Figures 4.4a and b). Hence, the mouse telomere, like the human telomere, could be composed of an array of $(TTAGGG)_n$ repeats within the most terminal region and an internal segment of an alternative repeat; the *Mnl* I and *Mse* I restriction enzyme analysis and hybridisation with the $(TTAGGG)_n$ would predict that this region would contain $(TTAGGG)_n$ repeats interspersed with variations of this plus an unrelated simple repeat.

Both of the above, described models for the sequence composition and organisation of mouse telomeres are highly speculative. The cloning and sequencing of these regions of the mouse genome will determine which of the models is correct and the similarites and differences between mouse and human telomeres.

The interstitial loci were also shown to contain *Mnl* I sites in the absence of *Hph* I sites (Section 6.3.3, Figure 6.4). Hybridisation analysis suggested that small amounts of the TTGGGG and TGAGGG repeats might exist within these loci. Hence, it seems possible that the interstitial telomere repeat containing loci also contain divergent, interspersed repeats. Again only isolation and sequencing of these loci will allow their true nature to be determined. Hybridisation analysis of mouse genomic DNA also revealed the possibility that TTGGGG and TGAGGG repeats are dispersed throughout the mouse genome in as similar manner to the human *Alu* repeats. Further investigations into the presence of these repeats within the mouse genome are required before this can be confirmed.

8.4 Why Do Mouse Chromosomes Possess Such Long Telomeres?

The telomeres of the mouse species studied fall within the 20-150kb size range, with the exception of the *Mus spretus* mouse strain, where the telomeric repeat ranges from 4.3-6.7kb (Figures 4.1a and b). With the exception of the *Mus Caroli* and *Mus spretus* telomeres, this is ~10x larger than the telomeres of human chromosomes, and larger again than the telomeres isolated from other plant and animal species (Table 1.1). Long telomere repeat arrays have also been observed within tomato and melon plants (Broun, *et al.*, 1992). Analysis of several individuals from different mouse strains also revealed a tendancy towards the same size range of telomeres within the strains. For example, in the DBA/2 strain, the telomeres ranged from 30-150kb in all of the animals studied, while in the C57Bl/6 strain, the telomere size tended to be within the smaller size range of ~30-100kb (Figure 4.7). In the *Mus Caroli* mice all of the telomeres were within the 20-30kb size range (Figure 4.8a). Hence, not only are mouse telomeres larger than for most other species examined so far, but there appears to be strain and species specific factors controlling their size.

8.4.1 Why are Mouse Telomeres so Large ?

A number of possible functions could be predicted to explain the size difference between mouse telomeres and those of other species. Firstly, the long array of telomere repeat could be involved in maintaining the mouse chromosome. All of the mouse strains involved in this study have acrocentric chromosomes. Hence, the large region of telomeric repeat could protect the telocentric centromere from degradation and loss of function and so ensure maintenance of the whole chromosome. However, this hypothesis seems unlikely as the *Mus spretus* mouse strain has telomeres ranging in size from ~4.3-6.7kb. Secondly, the large repeat array may be essential for the generation of a telomere/protein complex that is known to exist at the termini of *Oxytricha*, *Tetrahymena* and *Dictyostelium* macronuclear, DNA molecules (Blackburn and Chiou, 1981, Edwards and Firtel, 1984, Gottschling and Cech, 1984, Budarf and Blackburn, 1986). However, as before, the *Mus spretus* mouse strain has much smaller telomeres suggesting that the hypothesis is incorrect. Finally, there may be no real explanation for the length of the mouse telomere repeat array other than it being a strain and species specific feature. Similar examples are present within the hypotrichous ciliates, *Oxytricha nova*, *Oxytricha spp.*, *Oxytricha fallax* and *Stylonicha pustulata*. Here the telomere repeat is very short and defined, comprising 36bp of (G_4T_4) and 20bp of (C_4A_4) (Table 1.1). This is in contrast to the holotrichous cilates where the telomere repeat ranges from 0.12-0.42kb (Table 1.1). It is as yet unknown why such a difference exists between two groups of ciliates.

8.4.2 Why are there Strain Specific Size Differences in Mouse Telomeres ?

As discussed above and in Section 4.2, the telomeres of different mouse strains varied when compared between strains but were of a similar size within a group. In addition, most mouse groups examined had telomeres ranging from 30-150kb, with exceptions for *Mus Caroli*, telomere size from ~20-30kb, and *Mus spretus*, telomere size from ~4.3-6.7kb. The tendency for the telomeres within a mouse strain to fall within a certain size range can be explained by considering the high level of new variant generation at mouse telomeres. It is possible that new variant length telomeres are generated by non-reciprocal crossing over at telomeres (Section 8.2.1). Hence, the 3' single strand tail of one terminus could strand invade into a second telomere which would then serve as a template for elongation of the invading strand. This, in addition to recombination between the telomeres of homologous chromosomes at meiosis could maintain and generate a strain and species specific range of telomeres. The generation of uniform telomere size may also have been taking place upon the inheritance of telomeric loci into the F2 offspring of an initial C57Bl/6 x DBA/2 cross strain mating. Further evidence for this can be drawn from the fact that it was extremely difficult to

score new variant telomeres within the F2 offspring, as these possible loci varied only slightly in size from those present within the F1 parents (Section 5.3.3, Figures 5.4a and b).

An alternative explanation for the telomere size similarities within a mouse strain would predict that strain specific factors exist that limit the size of the telomere repeat. It has been shown that there are variations in the overall size of the telomere repeat array between different strains of yeast, but within a strain all of the telomeres fall within the same size range. In addition, mating studies have revealed that yeast telomere length is under the control of more than one gene and that these genes are codominant; the telomeres of the diploid offspring of a mating between two strains with distinctly different size ranges of telomeres demonstrate a size intermediate between that of the initial crossed strains (Walmsley and Petes, 1985). Additional evidence for genetic control of telomere length is provided by yeast mutants that have either shorter, *TEL1* and *TEL2*, or longer, *CDC17*, telomeres than the strain from which they were derived (Lustig and Petes, 1986, Carson and Hartwell, 1985). Both *TEL1* and *TEL2* are thought to represent some form of defect in the telomere elongation reaction. Such defects would predict that the balance between telomere elongation, by nontemplated repeat addition, and shortening, by primer excision after DNA replication, would be disturbed resulting in a gradual shortening of the telomeres (Lustig and Petes, 1986). The *CDC17* gene product has been identified as a DNA polymerase, however, its role in aberrant telomere elongation has not yet been defined (cited by Lucchini *et al.*, 1990). Evidence for a telomere size limiting factor in yeast has also been obtained from experiments which introduced additional ($C_{1-3}A$) telomere repeats into yeast. The increase of repeats led to an overall increase in the length of the chromosomal telomeres and it was suggested that the extra telomere repeats had caused a dilution of a size limiting factor (Runge and Zakian, 1989). Hence, it is possible that the mouse, strain specific, telomere size differences are under the control of specific genes.

At present it is impossible to predict which of the above explanations for strain specific sized telomeres within the mouse is correct. A line of investigation that might help to solve this question would involve cross strain matings bewteen mice with distinctly different telomere sizes. For example, if the *Mus spretus* and C57Bl/6 mouse strains were mated and the telomeres of the offspring analysed two possible results would be obtained. If recombination was the mechanism producing strain and species specific telomeres, then the offspring would be expected to have telomeres within a similar size range. However, if genetic factors are controlling telomere size than these would be expected to segregate into the offspring. The heterozygotes of such a cross would either show both sizes of telomere, if the factors are codominant, or one or other size class if one of the factors was dominant.

8.5 What Function do Telomere Repeats have within Mouse Chromosomes ?

Bal 31 digestion of total genomic DNA, *in situ* hybridisation and mapping in inbred strains suggested that the C57Bl/6 and DBA/2 loci, 1,2,3, 4, 5,6,7, 10, exist at interstitial sites within mouse chromosomes (Chapters 4, 5 and 7). In addition, mapping within the BxD RI and *Mus spretus* x C57Bl/6 backcross mice placed three different loci at the most terminal point on chromosomes 9, 13 and X. Internalised telomere repeats have also been found within the telomere associated sequences of yeast and *Plasmodium* chromosomes (Pace *et al.*, 1987, Walmsley *et al.*, 1984) and have been shown, by *in situ* hybridisation, to cluster at the subtelomeric regions of human chromosomes (Wells *et al.*, 1990).

The function of internalised telomere repeats has not yet been established. In *Plasmodium*, their presense has been implicated in the generation of chromosome length polymorphisms which result from the gain or loss of the telomere-associated sequences within which they are embedded. These length polymorphisms are thought to arise via mitotic recombination between a telomere and the internalised, telomere repeats (Pace *et al.*, 1990). As yet no possible function or reason has been described to explain the presence of internalised telomere repeats within human and yeast chromosomes. However, such loci can be postulated to have several, contrasting explanations for their presence within the mouse genome. Firstly, these proterminal loci could be acting as a buffer against chromosome breakage. This model would predict that the generation of an unstable, and susceptible to fusion, breakpoint by the loss of a telomere from a chromosome could be prevented by the presence of internalised telomere repeats which could form a new, functional telomere. However, it is unlikely that such breakpoints would always occur in the vicinity of these loci. In addition, recent studies of the human telomerase enzyme have revealed that very little homology of a broken chromosome end to its RNA primer is required for the telomerase mediated addition of telomere repeats and so healing of the chromosome (Morin, 1991).

Secondly, interstitial telomere repeat containing loci could represent the remains of a telomere-to-telomere, chromosome fusion event that was involved in the generation of the mouse karyotype. An example of such an event is provided by the human chromosome 2, which was generated from the fusion of two ape chromosomes Yunis and Prakash, 1982). *In situ* hybridisation of the *Tetrahymena*-like telomere repeat probe to human metaphase chromosomes demonstrated that these repeats exist within the region of fusion, 2q11-2q14 (Allshire *et al.*, 1988). In addition, this region has been cloned and shown to contain two inverted arrays of telomere repeats and flanking sequences characteristic of telomere-associated sequences (IJdo *et al.*, 1991). If this is also true of mouse internalised telomere repeats common locations for these loci would be expected to be present within different

mouse strains. Such an analysis has yet to be performed. Thirdly, they could represent regions of the genome that are susceptible to breakage as a result of telomere resolution (Section1.3.2). Until the exact locations of these loci within the mouse genome have been established this function cannot be explored. However, the genomic locations of three of these loci have been mapped within the BxD RI and *Mus spretus* x C57Bl/6 backcross mouse strains and no connection with 'fragile' loci has been made (Chapter 7). Finally, there may be no real reason for their presence within the protermini of some mouse chromosomes. They may simply represent repetitive DNA that has arisen from an initial translocation of DNA from a telomere to an internal site. However, until more is known about these loci it is impossible to predict what their function might be within the mouse genome.

8.6 Mouse Telomeres Represent Distinct Genetic Loci : How Might They be Useful in the Study of Recombination at Telomeres ?

The identification of distinct, high molecular weight, DNA fragments corresponding to mouse telomeres and the demonstration of their faithful inheritance suggested that mouse telomeres represent distinct genetic loci (Chapters 4 and 5). With the exception of the hypotrichous ciliates (Klobutcher *et al.*, 1981, Pluta *et al.*, 1982, Dawson and Herrick, 1982), tomato and melon plants (Broun *et al.*, 1992), all other organisms examined showed a variable number of repeats at their telomeres which produced a smear of hybridising DNA fragments upon visualisation with the relevant telomere probe (for example, Blackburn and Gall, 1978, Allshire *et al.*, 1988). In mice and tomato and melon plants the lack of obvious differences in the number of repeats at telomeres within different cells of the same organism probably relates to the overall size of these loci. For example, if a 150kb telomere from one chromosome varied by only ~1kb between different cells within a mouse, this size difference would be extremely difficult to visualise using PFGE which separates fragments within the size range of ~20-150kb.

The inheritance of mouse telomeres as distinct genetic loci has allowed their behaviour upon passage into several mouse generations to be studied (Chapter 5) and it has been shown that there is a high level of new variant generation occuring within these regions of the mouse genome (Chapter 5). As discussed in Section 8.2.1, it is possible that new variants are generated by some form of recombination based mechanism at meiosis of gametogenesis. The concept of recombination at telomeres is not new and has been demonstrated within the yeast *S. cerevisiae* using linear plasmids containing telomere repeats from *Tetrahymena* and *Oxytricha*. (Section 1.2.4(b)). However, these experiments do not strictly portray recombination events that might be occuring naturally at telomeres.

Hence, investigation of these events within mouse telomeres will not only help to define the mechanism of new variant generation, but may also provide some insight into the possibilty of these events taking place within other organisms.

From the initial observations it can be said that mouse telomeres represent distinct genetic loci that can be faithfully inherited and that demonstrate a high level of new variant generation. However, a mechanism for new variant generation could only be speculated upon as the telomere giving rise to a new variant could not be examined. However, the isolation of individual mouse telomeres and, more importantly, the unique DNA sequences that border such loci would allow mouse telomeres to be followed into further generations. This, in turn, would provide an insight into the level of new variant generation at individual telomeres and the mechanism by which such loci could have been created. If recombination does form the basis of new variant generation then two possible products can be predicted which depend upon either non-reciprocal or reciprocal crossing at telomeres. Hence, with non-reciprocal recombination, where strand invasion of a donor telomere is predicted to occur followed by replication of the invading strand, an increase in size of the locus would be seen (As discussed in Section 8.2.1). Alternatively, if both gains and losses were observed upon the generation of new variants, it would be predicted that telomeric repeats were being exchanged between the homologous chromosomes involved in the recombination event.

8.7 Interstitial $(TTAGGG)_n$ Containing Loci as Markers for Genome Mapping

From mapping of interstitial $(TTAGGG)_n$ containing loci within the BxD RI and *Mus spretus* x C57Bl/6 backcross, mouse strains, three loci were mapped to the most terminal points of chromosomes 9, 13 and X. In addition, it is possible that other $(TTAGGG)_n$ containing loci are proterminal, for example locus D2 (Section 7.2, Figures 7.2a and b). However, this cannot be confirmed until other markers, that are linked to these loci, have been placed upon the mouse genetic map, for example the A. J. Jeffreys minisatellite linked to D2 (Section 7.2). Irrespective of a possible function for telomere repeats within the mouse genome, such loci would be valuable for gene mapping as their proterminal location could provide an end to the genetic map for these three chromosomes. Hence, genes and other genetic markers could be positioned along the chromosome relative to this physical end. Their value to genome mapping would be enhanced if unique, locus-specific probes could be obtained from these loci. This would allow the generation of linkage maps for loci lying close to mouse telomeres. Although it proved extremely difficult to isolate interstitial telomere repeat containing loci from both the mouse and human genomes, new approaches, that will be discussed in Section

8.9, to the problem of cloning these loci may yield a set of extremely useful markers for genome mapping.

8.8 In What Ways do the Interstitial, $(TTAGGG)_n$ Containing Loci Resemble Minisatellite DNA ?

In Chapter 1, it was speculated that while internalised telomere repeats might retain the properties of chromosome termini resulting in a region of the mouse genome that would be susceptible to breakage, an alternative theory would predict that they take on the properties of minisatellite DNA. Hence, from the findings of this study, firstly, it would appear that both minisatellites and internalised telomere repeat containing loci tend to cluster within the proternimal regions of chromosomes (Royle, *et al.*, 1988, Wells *et al.*, 1989). As yet there are no clues to indicate the reason for the unusual distribution of these minisatellites. In addition, while minisatellites provide an individual specific DNA fingerprint, internalised telomere repeat loci give strain specific fingerprints within the mouse. While minisatellite DNA fingerprints within humans are used in forensic science, paternity testing and general individual identification, internal telomere loci may prove useful in the identification of different mouse strains. Secondly, both the minisatellite and internal telomere loci are devoid of restriction enzyme sites suggesting that both are composed of simple repetitive DNA. Thirdly, minisatellites have been shown to be composed of tandem copies of a repeat which contains a known core sequence. In addition, hybridisation studies have shown that each minisatellite DNA fragment is composed almost entirely of these repeat units (Jeffreys *et al.*, 1985b). The same cannot be said for the internalised telomere repeat loci. Partial restriction enzyme and hybridisation analysis of the DNA fragments making up these loci suggest that they may be composed of a block of $(TTAGGG)_n$ repeats, the number of which varies between different loci, embedded within an unrelated form of simple repeat (Chapter 6). However, until the interstitial telomere repeat containing loci have been isolated and charaterised the nature of a possible additional simple repeat remains unknown. Fourthly, both minisatellites and interstitial telomere repeat containing loci have been shown to be stably inherited into further generations and new variants appear to be generated within the germ line rather than somatic cells. Finally, minisatellite loci demonstrate a high level of heterozygosity and new variant generation. In addition, minisatellites may be hotspots for recombination (Jeffreys *et al.*, 1985b, 1986 and 1988, Wong *et al.*, 1986 and 1987). From the results obtained for the interstitial telomere repeat containing loci, it is impossible to speculate whether these loci have similar properties. Isolation, sequencing and mapping of the interstitial telomere repeat containing loci is required before their levels of new variant generation and heterozygosity can be determined.

8.9 Future Direction

From the work carried out within this research project, it became apparent that several areas either required a new technical approach or that additional studies could be performed that would both clarify and increase the information known about telomeric and interstitial telomere repeat containing loci. Hence, below are presented some possible areas for future research into the telomeric and interstitial $(TTAGGG)_n$ containing loci.

8.9.1 Isolation of Telomeric and Interstitial $(TTAGGG)_n$ Containing Loci from the Mouse Genome

As indicated throughout the discussion, the isolation, sequencing and mapping of both types of loci could answer many questions concerning the properties of these loci within the mouse genome.

8.9.1(a) Isolation of telomeric loci

The isolation and sequencing of mouse telomeres may provide an insight into their organisation and demonstrate how they differ from human telomeres. In addition, the isolation of the unique sequence that flanks a single mouse telomere may help to establish the mechanism of new variant generation at mouse telomeres.

A procedure for cloning mouse telomeres together with unique flanking sequences would have to take into account the fact that mouse telomeres are under-represented within the mouse genome, represent a stretch of highly repetitive DNA and contain a non-clonable end. Hence, mouse telomeres could be cloned using the strategy for the cloning of human telomeres in yeast. Here genomic DNA, enriched for human telomeres, was ligated to a linearised vector containing a *Tetrahymena* telomere at one end, a yeast centromere, a yeast selectable marker and a yeast ARS and the ligation reaction transformed into yeast. Only those ligation products containing a functional human telomere were maintained within yeast. (Brown, 1989, Cross *et al.*, 1989).

If a similar approach was adopted for the cloning of mouse telomeres, terminal DNA fragments could be selected for by isolation of the >23kb, non-resolvable region of DNA from a conventional gel (Figure 4.1a). Alternatively, individual mouse telomeres could be isolated from pulsed-field gels, in the same manner as described for the isolation of the 140kb new variant (Chapter 6). The initial digestion of the genomic DNA to produce the telomeric fragments would require a restriction enzyme that did not cut too close to the telomeric repeats; in cloning mouse telomeres it would be important to obtain unique flanking DNA.

Hence, rare cutting enzymes such as *Not* I or *Xho* I could be used to obtain the telomeric fragments.

8.9.1(b) Isolation of interstitial $(TTAGGG)_n$ containing loci

As described in Chapter 3, cloning of the interstitial $(TTGGGG)_n$ and $(TTAGGG)_n$ loci from the human and mouse genomes, respectively, proved to be extremely difficult. This may have been due to the highly recombinogenic nature of tandem arrays of a simple hexamer repeat. In addition, it was also possible that the presence of these repeats caused rearrangements within the flanking DNA, hence preventing mapping of isolated loci within either the human or mouse genomes. As the most important element of the locus is the unique, flanking DNA, an alternative approach could be adopted which would not require the propagation of telomeric repeats within bacteria. Such a strategy would involve the use of the polymerase chain reaction (PCR) to specifically amplify the DNA flanking the telomere repeats. Hence, the first step in this strategy would involve digestion of total genomic mouse DNA with restriction enzymes that would leave at least 1kb of the DNA that flanks the telomere repeat containing loci. Six bp cutting enzymes, like *Bam* HI or *Eco* RI, could be used although the length of flanking DNA remaining after digestion might be to large for the subsequent PCR reactions. Alternatively, partial digestion of the genomic DNA with a frequently cutting enzyme, like *Hae* III, may provide a suitable length of flanking DNA. The second step would involve the ligation of linkers of a known sequence to the digested DNA followed by primer extension using primers, with biotin attached at the 5'end, complementary to the linkers. After extension, the single stranded product would be isolated by adhesion of the biotin molecule to streptavidin beads. Thirdly, the single stranded DNA product of the primer extension reaction would be amplified by PCR between a primer complementary to the linker and one complementary to $(TTAGGG)_n$ repeats. The double strand products could then be cloned into plasmid vectors and their ability to identify specific interstitial $(TTAGGG)_n$ containing loci within the mouse genome determined. For this last step to be performed the $(TTAGGG)_n$ repeats within the clones would have to be removed; this could be achieved by digestion with frequently cutting restriction enzymes.

8.9.2 Identification of Factors Involved in Strain Specific Telomere Size Determination

From the analysis of telomeres from different strains of mice it became apparent that there was a strain specific telomere size range. For example, all of the *Mus caroli* mice analysed had telomeres within the ~20-30kb size range, while the DBA/2 mice had telomeres ranging in size from ~30-150kb. It would be interesting to investigate the possibility that there might be a genetic factor involved in determination of telomere size that differed between different

mouse strains. This might take the form of a telomere binding protein or subtle differences in the activity of a putative mouse telomere terminal transferase. Alternatively, the strain differences may result from an homogenisation of telomere size, by a recombination-based mechanism, within each separate strain. It might be possible to resolve this question by following the inheritance of the telomeres, from strains with very different telomere size ranges, into further generations. Hence, upon cross breeding, if the size differences are due to a genetic factor, segregation of the two types of telomeres would be expected. However, if recombination-based homogenisation is taking place at mouse telomeres then a general mixing of telomere size might be expected resulting in a new average telomere length.

The DBA/2 and C57Bl/6 mice used for inheritance studies in this project (Chapter 5) have telomeres ranging in size from ~30-150kb and ~30-100kb respectively. Upon crossbreeding and analysis of the F2 offspring, most animals possessed telomeres ranging in size from ~30-150kb. This result could predict that telomere homogenisation was taking place between the two strains. However, the result is ambiguous as the two strains possess telomeres within very similar size ranges. It would be far more interesting and informative to follow telomere inheritance from an initial C57Bl/6 x *Mus spretus* cross strain mating. Here it should be far easier to follow the 4.3-6.7kb *Mus spretus* telomeres within the ~30-100kb C57Bl/6 background. Such cross strain matings have already been performed in producing the *Mus spretus* / C57Bl/6 backcross mice from Copeland and Jenkins (Chapter 7). During the course of this project the DNA samples from these crosses were used for linkage analysis which required DNA fragment separation by conventional gel electrophoresis (Figure 7.5). Hence, it was very difficult to determine from the results obtained if telomere size was affected by crossing the strains. In addition, the backcross mice have been generated by mating the F1 progeny of the initial C57Bl/6 x *Mus spretus* cross with C57Bl/6 inbred mice. This places the recombination events from the first mating onto a C57Bl/6 background. This may bias telomere size towards that of the C57Bl/6 mouse strain. Hence, it would be more interesting to analyse the F1 progeny of the initial C57Bl/6 x *Mus spretus* mating for the presence of telomere size regulating factors.

8.9.3 Identification of the Putative Mouse Telomerase

Identification of a mouse telomere terminal transferase will allow its role, if any, in the generation of new variants at mouse telomeres to be determined. In addition, this would also allow the mechanism by which mouse telomeres are replicated to be identified and the level of catalytic activity of this enzyme to be determined. Once a mouse telomere terminal transferase has been cloned it would be interesting to look at the effects of disabling this enzyme within both mouse cells and transgenic mice. It is possible that a similar phenotype to that seen with the yeast mutant gene, *EST1* (Lundblad and Szostak, 1989), might be

observed. The gradual loss of repeats from the mouse telomere may predispose the animal to telomere fusion and, if there is a real correlation between telomere shortening and some forms of cancer (Hastie *et al.*, 1990), and the occurrence of tumours.

8.10 Summary

Trypanosoma-like $(TTAGGG)_n$ telomere repeats are present at the telomeres and at interstitial sites within the mouse genome. Within both of these regions of the mouse genome the $(TTAGGG)_n$ repeats are present within distinct genetic loci that are stably inherited into subsequent generations. New variant generation occurs at both the telomeric and interstitial loci, takes place at a significantly higher frequency at the telomeric compared to interstitial loci and occurs during gametogenesis. It is possible that the higher rate of new variant generation at mouse telomeres compared to interstitial sites relates to their positions within mouse chromosomes. Restriction enzyme and hybridisation sequence analysis of the loci demonstrated that both were composed of simple repeats. It is possible that the interstitial loci contain a single block of $(TTAGGG)_n$ repeats which vary in number between different loci and an additional form of simple repeat. The telomeric loci appeared to be predominantly composed of $(TTAGGG)_n$ repeats, interspersed with related, divergent telomere-like repeats. In addition, there may also be some other form of simple repeat within mouse telomeres, however, its organisation within the telomere repeat array and its sequence composition could not be determined from the present study. Mapping of the interstitial $(TTAGGG)_n$ containing loci within the BxD RI, and *Mus spretus* / C57Bl/6 backcross, mice demonstrated their presence within the protermini of chromosomes 9, 13 and X. It remains to be determined whether this distribution is functionally significant. However, such loci would provide an end to the genetic map for these chromosomes and hence, would be extremely useful for genome mapping.

From the results obtained in this study the next step in the charaterisation of telomeric and interstitial $(TTAGGG)_n$ containing loci would involve their isolation from the mouse genome. This would not only allow their sequence organisation to be determined but would also help to reveal the mechanism of new variant generation at both types of loci. In addition, from such analyses it should be possible to determine how the $(TTAGGG)_n$ containing terminal repeat arrays function as telomeres, and whether internalised telomere repeats have a function, within the mouse genome.

References

References

Agard, D. A. and Sedat, J. W. (1983). Three-dimensional architecture of a polytene nucleus. *Nature* **302**, 676-681.

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

Allshire, R. C., Dempster, M. and Hastie, N. D. (1989). Human telomeres contain at least three types of G-rich repeat distributed non-randomly. *Nucl. Acids Res.* **17**, 4611-4627.

Allshire, R. C., Gosden, J. R., Cross, S. H., Cranston, G., Rout, D., Sugawara, N., Szostak, J. W., Fantes, P. A. and Hastie, N. D. (1988). Telomeric repeat from *T. thermophilia* cross hybridizes with human telomeres. *Nature* **332**, 656-659.

Amar, L. C., Arnaud, D., Cambrou, J., Guenet, J. L. and Avner, P. R. (1985). Mapping of the mouse X chromosome using random genomic probes and an interspecific mouse cross. *EMBO J*. **4**, 3695-3700.

Armour, J. A. L., Patel, I., Thein, S. L., Fey, M. F. and Jeffreys, A. J. (1989). Analysis of somatic mutations at human minisatellite loci in tumors and cell lines. *Genomics* **4**, 328-334.

Ashley, T. and Wagenaar, E. B. (1974). Telomeric associations of gametic and somatic chromosomes in diploid and autotetraploid *Ornithogalum virens*. *Can. J. Genet. Cytol.* **16**, 61-76.

Avner, P., Amar, L., Dandolo, L. and Guenet, J. L. (1988). Genetic analysis of the mouse using interspecific crosses. *TIG* **4**, 18-23.

Baroin, A., Prat, A. and Caron, F. (1987). Telomeric site position heterogeneity in macronuclear DNA of *Paramecium primaureila*. *Nucl. Acids Res.* **15**, 1717-1728.

Bateman, A. J. (1975). Simplification of the palindromic telomere theory. *Nature* **253**, 379.

Bedbrook, J. R., Jones, J., O'Dell, M., Thompson, R. D. and Flavell, R. B. (1980). A molecular description of telomeric heterochromatin in *Secale* species. *Cell* **19**, 545-560.

Beissmann, H., Carter, S. B. and Mason, J. M. (1990a). Chromosome ends in *Drosophila* without telomeric DNA sequences. *Proc. Natl. Acad. Sci. USA* **87**, 1758-1761.

Beissmann, H., Mason, J. M., Ferry, K., d'Hulst, M., Valgeirsdottir, K., Traverse, K. L. and Pardue, M-L. (1990b). Addition of telomere-associated HeT DNA sequences "heals" broken chromosome ends in *Drosophila*. *Cell* **61**, 663-673.

Bell, G. I., Selby, M. J. and Rutter, W. J. (1982). The highly polymorphic region near the human insulin gene is composed of simple tandemly repeating sequences. *Nature* **295**, 31-35.

Berger, R., Bloomfield, C. D. and Sutherland, G. R. (1985). Report of the committee on chromosome rearrangements in neoplasia and on fragile sites. *Cytogenet. Cell Genet.* **40**, 490-535.

Bergold, P. J., Campbell, G. R., Littau, V. C. and Johnson, E. M. (1983). Sequence and hairpin structure of an inverted series at termini of the *Physarum* extrachromosomal rDNA molecule. *Cell* **32**, 1287-1299.

Berman, J., Tachibana, C. Y. and Tye, B-K. (1986). Identification of a telomere-binding activity from yeast. *Proc. Natl. Acad. Sci. USA* **83**, 3713-3717.

Bernards, A., Michels, P. A. M., Linke, C. R. and Borst, P. (1983). Growth of chromosome ends in multiplying Trypanosomes. *Nature* **303**, 592-597.

Birnboim, H. C. and Doly, J. (1979). A rapid alkaline extraction method for screening recombinant plasmid DNA. *Nucl. Acids Res.* **7**, 1513-1516.

Blackburn, E. H. (1984). Telomeres: Do the ends justify the means ? *Cell* **37**, 7-8.

Blackburn, E. H., Budarf, M. L., Challoner, P. B., Cherry, J. M., Howard, E. A., Katzen, A. L., Pan, W.-C. and Ryan, T. (1982). DNA termini in ciliate macronuclei. *Cold Spring Harbor Symp. Quant. Biol.* **47**, 1195-1207.

Blackburn, E. H. and Challoner, P. B. (1984). Identification of a telomeric DNA sequence in *Trypanosoma brucei*. *Cell* **36**, 447-457.

Blackburn, E. H. and Chiou, S-S. (1981). Non-nucleosomal packaging of a tandemly repeated DNA sequence at termini of extrachromosomal DNA coding for rRNA in *Tetrahymena*. *Proc. Natl. Acad. Sci. USA* **78**, 2263-2267.

Blackburn, E. H. and Gall, J. G. (1978). A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in *Tetrahymena*. *J. Mol. Biol.* **120**, 33-53.

Blackburn, E. H. and Karrer, K. M. (1986). Genomic organisation in ciliated protozoans. *Ann. Rev. Genet.* **20**, 501-521.

Blackburn, E. H. and Szostak, J. W. (1984). The molecular structure of centromeres and telomeres. *Ann. Rev. Biochem.* **53**, 163-194.

Boliviar, F., Roderiguez, P. J., Greene, P. J., Betlach, M. C. and Boyer, H. W. (1977). Construction and characterisation of new cloning vehicles. I. Ampicillin resistance derivatives of the plasmid pMB9. *Gene* **2**, 75-93.

Borst, P., Bernards, A., Van der Ploeg, L. H. T., Michals, P. A. M., Lui, A. Y. C., de Lange, T., Sloof, P., Schwartz, D. C. and Cantor, C. R. (1983). The role of mini-chromosomes and gene translocation in the expression and evolution of VSG genes. In Gene Expression, UCLA Symp. on Molecular and Cellular Biology, New Series, 8, D. Hammer and M. Rosenberg, eds. (New York: Liss), pp413-435.

Boswell, R. E., Klobutcher, L. A. and Prescott, D. M. (1982). Inverted terminal repeats are added to genes during macronuclear development in *Oxytricha nova*. *Proc. Natl. Acad. Sci. USA* **79**, 3255-3259.

Brown, W. R. A. (1989). Molecular cloning of human telomeres in yeast. *Nature* **338**, 774-776.

Brown, W. R. A., MacKinnon, P. J., Villasante, A., Spurr, N., Buckle, V. J. and Dobson, M. J. (1990). Structure and polymorphism of human telomere-associated DNA. *Cell* **63**, 119-132.

Broun, P., Ganal, M. W. and Tanksley, S. D. (1992). Telomeric arrays display high levels of heritable polymorphism among closely related plant species. *Proc. Natl. Acad. Sci. USA* **89**, 1354-1357.

Budarf, M. L. and Blackburn, E. H. (1986). Chromatin structure of the telomeric region and 3'-nontranscribed spacer of *Tetrahymena* ribosomal RNA genes. *J. Biol. Chem.* **261**, 363-369.

Burgoyne, P.S. (1982). Genetic homology and crossing over in the X and Y chromosomes of mammals. *Hum. Genet.* **61**, 85-90.

Burmeister, M. and Lehrach, H. (1989). Isolation of large DNA fragments from agarose gels using agarase. *Trends Genet.* **5**, 41.

Capon, D. J., Chen, E. Y., Levinson, A. D., Seeburg, P. H. and Goeddel, D. V. (1983). Complete nucleotide sequences of the T24 human bladder carcinoma oncogene and its normal homologue. *Nature* **302**, 33-37.

Carle, G. F. and Olson, M. V. (1985). An electrophoretic karyotype for yeast. *Proc. Natl. Acad. Sci. USA* **82**, 3756-3760.

Carson, M. J. and Hartwell, L. (1985). *CDC17*: An essential gene that prevents telomere elongation in yeast. *Cell* **42**, 249-257.

Cavalier-Smith, T. (1974). Palindromic sequences and replication of eukaryote chromosome ends. *Nature* **250**, 467-470.

Chan, C. S. M and Tye, B-K. (1980). Autonomously replicating sequences in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* **77**, 6239-6333.

Chan, C. S. M. and Tye, B-K. (1983a). A family of *Saccharomyces cerevisiae* repetitive autonomously replicating sequences that have very similar genomic environments. *J. Mol. Biol.* **168**, 505-523.

Chan, C. S. M. and Tye, B-K. (1983b). Organisation of DNA sequences and replication origins at yeast telomeres. *Cell* **33**, 563-573.

Chandley, A. C. and Mitchell, A. R. (1988). Hypervariable minisatellite regions are sites for crossing-over at meiosis in man. *Cytogenet. Cell Genet.* **48**, 152-155.

Cheng, J-F., Smith, C. L. and Cantor, C. R. (1989). Isolation and characterisation of a human telomere. *Nucl. Acids Res.* **17**, 6109-6127.

Cheung, M. K., Drivas, D. T., Littau, V. C. and Johnson, E. M. (1981). Protein tightly bound near the termini of the *Physarum* extrachromosomal rDNA palindrome. *J. Cell Biol.* **91**, 309-314.

Chung, H-M. M., Shea, C., Fields, S., Taub, R. N. and Van der Ploeg, L. H. T. (1990). Architectural organisation in the interphase nucleus of the protozoan *Trypanosoma brucei*: location of telomeres and mini-chromosomes. *EMBO J.* **9**, 2611-2619.

Church, G. M. and Gilbert, W. (1984). Genomic sequencing. *Proc. Natl. Acad. Sci. USA.* **81**, 1991-1995.

Comings, D. E. and Okada, T. A. (1972). Architecture of meiotic cells and mechanisms of chromosome pairing. *Adv. Cell Mol. Biol.* **2**, 309-384.

Conrad, M. N., Wright, J. H., Wolf, A. J. and Zakian, V. A. (1990). RAP1 protein interacts with yeast telomeres *in vivo*: Overproduction alters telomere structure and decreases chromosome stability. *Cell* **63**, 739-750.

Cooke, H. J. and Smith, B. A. (1986). Variability at the telomeres of the human X/Y pseudoautosomal region. *Cold Spring Harbor Symp. Quant. Biol.* **51**, 213-219.

Corcoran, L. M., Thompson, J. K., Walliker, D. and Kemp, D. J. (1988). Homologous recombination within subtelomeric repeat sequences generates chromosome size polymorphisms in *P. falciparum*. *Cell* **53**, 807-813.

Cross, S. H., Allshire, R. C., McKay, S. J., McGill, N. I. and Cooke, H. J. (1989). Cloning of human telomeres by complementation in yeast. *Nature* **338**, 771-774.

Cross, S., Lindsey, J., Fantes, J., McKay, S., McGill, N. and Cooke, H. (1990). The structure of a subterminal repeated sequence present on many human chromosomes. *Nucl. Acids Res.* **18**, 6649-6657.

Dancis, B. M. and Holmquist, G. P. (1979). Telomere replication and fusion in eukaryotes. *J. Theor. Biol.* **78**, 211-224.

Dawson, D. and Herrick, G. (1982). Micronuclear DNA sequences of *Oxytricha fallax* homologous to the macronuclear inverted terminal repeat. *Nuc. Acids Res.* **10**, 2911-2924.

Disteche, C. M., McConnell, G. K., Grant, S. G., Stephenson, D. A., Chapman, V. M., Gandy, S. and Adler, D. A. (1989). Comparison of the physical and recombination maps of the mouse X chromosome. *Genomics* **5**, 177-184.

Dore, E., Pace, T., Ponzi, M., Picci, L. and Frontali, C. (1990). Organisation of subtelomeric repeats in *Plasmodium berghei*. *Mol. Cell. Biol.* **10**, 2423-2427.

Drets, M. E. and Therman, E. (1983). Human telomeric 6;19 translocation chromosome with a tendency to break at the fusion point. *Chromosoma (Berl)* **88**, 139-144.

Dunn, B., Szauter, P., Pardue, M. L. and Szostak, J. W. (1984). Transfer of yeast telomeres to linear plasmids by recombination. *Cell* **39**, 191-201.

Dutrillaux, B., Croquette, M. F., Viegas-Pequignot, E., Aurias, A., Coget, J., Couturier, J. and Lejeune, J. (1978). Human somatic chromosome chains and rings. *Cytogenet. Cell Genet.* **20**, 70-77.

Edwards, C. A. and Firtel, R. A. (1984). Site-specific phasing in the chromatin of the rDNA in *Dictyostelium discoideum*. *J. Mol. Biol.* **180**, 73-90.

Emery, H. S. and Weiner, A. M. (1981). An irregular satellite sequence is found at the termini of the linear extrachromosomal rDNA in *Dictyostelium discoideum*. *Cell* **26**, 411-419.

Farr, C., Fantes, J., Goodfellow, P. and Cooke, H. (1991). Functional reintroduction of human telomeres into mammalian cells. *Proc. Natl. Acad. Sci. USA* **88**, 7006-7010.

Feinberg, A. P. and Vogelstein, B. (1983). A technique for radiolabelling DNA restriction endonuclease fragments to a high specific activity. *Anal. Biochem.* **132**, 6-13.

Feinberg, A. P. and Vogelstein, B. (1984). Addenum: a technique for radiolabelling DNA restriction endonuclease fragments to a high specific activity. *Anal. Biochem.* **137**, 266-267.

Fitzgerald, P. H. and Morris, C. M. (1984). Telomeric association of chromosomes in B-cell lymphoid leukaemia. *Hum. Genet.* **67**, 385-390.

Foe, V. E. and Alberts, B. M. (1985). Reversible chromosome condensation induced in *Drosophila* embryos by anoxia: Visualization of interphase nuclear organisation. *J. Cell Biol.* **100**, 1623-1636.

Formosa, T. and Alberts, B. M. (1986). DNA synthesis dependent on genetic recombination: Characterisation of a reaction catalysed by purified bacteriophage T4 proteins. *Cell* **47**, 793-806.

Forney, J. D. and Blackburn, E. H. (1988). Developmentally controlled telomere addition in wild-type and mutant *Paramecia*. *Mol. Cell. Biol.* **8**, 251-258.

Forney, J., Henderson, E. R. and Blackburn, E. H. (1987). Identification of the telomeric sequence of the acellular slime molds *Didymium iridis* and *Physarum polycephalum*. *Nucl. Acids Res.* **15**, 9143-9152.

Friedman, J. M., Schneider, B. S., Barton, D. E. and Francke, U. (1989). Level of expression and chromosome mapping of the mouse Cholecystokinin gene: Implications for murine models of genetic obesity. *Genomics* **5**, 463-469.

Gall, J. G. (1974). Free ribosomal RNA genes in the macronucleus of *Tetrahymena*. *Proc. Natl. Acad. Sci. USA* **71**, 3078-3081.

Gasser, S. M. and Laemmli, U. K. (1987). A glimpse at chromosomal order. *TIG* **3**, 16-22.

Gehring, W. J., Klemenz, R., Weber, U. and Kloter, U. (1984). Functional analysis of the (*white*$^+$) gene of *Drosophila* by P-factor-mediated transformation. *EMBO J.* **3**, 2077-2085.

Gill, P., Jeffreys, A. J. and Werrett, D. J. (1985). Forensic applications of DNA 'fingerprints'. *Nature* **318**, 577-579.

Goodbourn, S. E. Y., Higgs, D. R., Clegg, J. B. and Weatherall, D. J. (1983). Molecular basis of length polymorphisms in the human ζ-globin gene complex. *Proc. Natl. Acad. Sci. USA* **80**, 5022-5026.

Gosden, J., Hanratty, D., Starling, J., Fantes, J., Mitchell, A. and Porteous, D. (1991). Oligonucleotide-primed *in situ* DNA synthesis (PRINS): a method for chromosome mapping, banding, and investigation of sequence organisation. *Cytogenet. Cell Genet.* **57**, 100-104.

Gottschling, D. E., Aparicio, O. M., Billington, B. L. and Zakian, V. A. (1990). Position effect at *S. cerevisiae* telomeres: Reversible repression of Pol II transcription. *Cell* **63**, 751-762.

Gottschling, D. E. and Cech, T. R. (1984). Chromatin structure of the molecular ends of *Oxytricha* macronuclear DNA: Phased nucleosomes and a telomeric complex. *Cell* **38**, 501-510.

Gottschling, D. E. and Zakian, V. A. (1986). Telomere proteins: Specific recognition and protection of the natural termini of *Oxytricha* macronuclear DNA. *Cell* **47**, 195-205.

Greider, C. W. and Blackburn, E. H. (1985). Identification of a specific telomere terminal transferase activity in *Tetrahymena* extracts. *Cell* **43**, 405-413.

Greider, C. W. and Blackburn, E. H. (1987). The telomere terminal transferase of *Tetrahymena* is a ribonucleoprotein enzyme with two kinds of primer specificity. *Cell* **51**, 887-898.

Greider, C. W. and Blackburn, E. H. (1989). A telomeric sequence in the RNA of *Tetrahymena* telomerase required for telomere repeat synthesis. *Nature* **337**, 331-337.

Grunstein, M. and Hogness, D. S. (1975). Colony hybridisation: A new method for the isolation of cloned cDNAs that contain a specific gene. *Proc. Natl. Acad. Sci. USA* **72**, 3961-3965.

Haber, J. E. and Thorburn, P. C. (1984). Healing of broken linear dicentric chromosomes in yeast. *Genetics* **106**, 207-226.

Hanahan, D. (1983). Studies on transformation of *Escherischa coli* with plasmids. *J. Mol. Biol.* **166**, 577-580.

Harrington, L. A. and Greider, C. W. (1991). Telomerase primer specificity and chromosome healing. *Nature* **353**, 451-454.

Harley, C. B., Futcher, A. B. and Greider, C. W. (1990). Telomeres shorten during ageing of human fibroblasts. *Nature* **345**, 458-460.

Hartwell, L. H. (1973). Three additional genes required for deoxyribonucleotide acid synthesis in *Saccharomyces cerevisiae*. *J. Bacteriol.* **115**, 966-974.

Hartwell, L. H. and Smith, D. (1985). Altered fidelity of chromosome metabolism in cell cycle mutants in *S. cerevisiae*. *Genetics* **110**, 381-395.

Hastie, N. D. and Allshire, R. C. (1989). Human telomeres: fusion and interstitial sites. *TIG* **5**, 326-331.

Hastie, N. D., Dempster, M., Dunlop, M. G., Thompson, A. M., Green, D. K. and Allshire, R. C. (1990). Telomere reduction in human colorectal carcinoma and ageing. *Nature* **346**, 866-868.

Henderson, E. R. and Blackburn, E. H. (1989). An overhanging 3' terminus is a conserved feature of telomeres. *Mol. Cell. Biol.* **9**, 345-348.

Henderson, E., Hardin, C. C., Walk, S. K., Tinoco, I. Jr. and Blackburn, E. H. (1987). Telomeric DNA oligonucleotides form novel intramolecular structures containing Guanine.Guanine base pairs. *Cell* **51**, 899-908.

Henderson, E. R., Moore, M. and Malcolm, B. A. (1990). Telomere G-strand structure and function analysed by chemical protection, base analogue substitution, and utilisation by telomerase *in vitro*. *Biochemistry* **29**, 732-737.

Heumann, J. M. (1976). A model for the replication of the ends of linear chromosomes. *Nuc. Acids Res.* **3**, 3167-3171.

Hicke, B. J., Celander, D. W., MacDonald, G. H., Price, C. M. and Cech, T. R. (1990). Two versions of the gene encoding the 41-kilodalton subunit of the telomere binding protein of *Oxytricha nova*. *Proc. Natl. Acad. Sci. USA* **87**, 1481-1485.

Higgs, D. R., Goodbourn, S. E. Y., Wainscoat, J. S., Clegg, J. B. and Weatherall, D. J. (1981). Highly variable regions of DNA flank the human α globin genes. *Nucl. Acids Res.* **9**, 4213-4224.

Hill, R. E., Shaw, P. H., Barth, R. K. and Hastie, N. D. (1985). A genetic locus closely linked to a protease inhibitor gene complex controls the level of multiple RNA transcripts. *Mol. Cell Biol.* **5**, 2114-2122.

Horowitz, H., Thorburn, P. and Haber, J. E. (1984). Rearrangements of highly polymorphic regions near telomeres of *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **4**, 2509-2517.

Hunt, P. A. and Burgoyne, P. S. (1987). A tissue-specific fragile site associated with the sex reversed (*Sxr*) mutation in the mouse. *Chromosoma (Berl)* **96**, 67-71.

IJdo, W., Baldini, A., Ward, D. C., Reeders, S. T. and White, R. A. (1991). Origin of human chromosome 2: An ancestral telomere-telomere fusion. Proc. Natl. Acad. Sci. USA 88, 9051-9055.

Jarman, A. P., Nicholls, R. D., Weatherall, D. J., Clegg, J. B. and Higgs, D. R. (1986). Molecular characterisation of a hypervariable region downstream of the human α-globin gene cluster. *EMBO J.* **5**, 1857-1863.

Jeffreys, A. J., Neumann, R. and Wilson, V. (1990). Repeat unit sequence variation in minisatellites: A novel source of DNA polymorphism for studying variation and mutation by single molecule analysis. *Cell* **60**, 473-485.

Jeffreys, A. J., Royle, N. J., Wilson, V. and Wong, Z. (1988). Spontaneous mutation rates to new length alleles at tandem-repetitive hypervariable loci in human DNA. *Nature* **332**, 278-281.

Jeffreys, A. J., Wilson, V., Kelly, R., Taylor, B. A. and Bulfield, G. (1987). Mouse DNA 'fingerprints': analysis of chromosome localisation and germ-line stability of hypervariable loci in recombinant inbred strains. *Nucl. Acids Res.* **15**, 2823-2836.

Jeffreys, A. J., Brookfield, J. F. Y. and Semeonoff, R. (1985a). Positive identification of an immigration test-case using human DNA fingerprints. *Nature* **317**, 818-819.

Jeffreys, A. J. and Morton, D. B. (1987). DNA fingerprints of dogs and cats. *Animal Genet.* **18**, 1-15.

Jeffreys, A. J., Wilson, V. and Thein, S. L. (1985b) Hypervariable 'minisatellite' regions in human DNA. *Nature* **314**, 67-73.

Jeffreys, A. J., Wilson, V. and Thein, S. L. (1985c). Individual-specific 'fingerprints' of human DNA. *Nature* **316**, 76-79.

Jeffreys, A. J., Wilson, V., Thein, S. L., Weatherall, D. J. and Ponder, B. A. J. (1986). DNA 'fingerprints' and segregation analysis of multiple markers in human pedigrees. *Am. J. Hum. Genet.* **39**, 11-24.

Johnson, E. M. (1980). A family of inverted repeat sequences and specific single-strand gaps at the termini of the *Physarum* rDNA palindrome. *Cell* **22**, 875-886.

Joseph, A., Mitchell, A. R. and Miller, O. J. (1989). The organisation of the mouse satellite DNA at centromeres. *Ex. Cell Res.* **183**, 494-500.

Justice, M. J., Silan, C. M., Ceci, J. D., Buchberg, A. M., Copeland, N. G. and Jenkins, N. A. (1990). A molecular genetic linkage map of mouse chromosome 13 anchored by the beige (*bg*) and satin (*sa*) loci. *Genomics* **6**, 341-351.

Katzen, A. L., Cann, G. M. and Blackburn, E. H. (1981). Sequence-specific fragmentation of macronuclear DNA in a holotrichous ciliate. *Cell* **24**, 313-320.

Kelly, R., Bulfield, G., Collick, A., Gibbs, M. and Jeffreys, A. J. (1989). Characterisation of a highly unstable mouse minisatellite locus: Evidence for somatic mutation during early development. *Genomics* **5**, 844-856.

Klobutcher, L. A., Jahn, C. L. and Prescott, D. M. (1984). Internal sequences are eliminated from genes during macronuclear development in the ciliated protozoan *Oxytricha nova*. *Cell* **36**, 1045-1055.

Klobutcher, L. A., Swanton, M. T., Donini, P. and Prescott, D. M. (1981). All gene-sized DNA molecules in four species of hypotrichs have the same terminal sequence and an unusual 3' terminus. *Proc. Natl. Acad. Sci. USA* **78**, 3015-3019.

Kornberg, R. D. (1977). Structure of chromatin. *Ann. Rev. Biochem.* **46**, 931-954.

Kovacs, G., Muller-Brechlin, R. and Szucs, S. (1987). Telomeric associations in two human renal tumours. *Cancer Genet. Cytogenet.* **28**, 363-366.

de Lange, T., Shiue, L., Myers, R. M., Cox, D. R., Naylor, S. L., Killery, A. M. and Varmus, H. E. (1990). Structure and variability of human chromosome ends. *Mol. Cell. Biol.* **10**, 518-527.

Larson, D. D., Spangler, E. A. and Blackburn, E. H. (1987). Dynamics of telomere length variability in *Tetrahymena thermophilia*. *Cell* **50**, 477-483.

Lawn, R. M., Heumann, J. M., Herrick, G. and Prescott, D. M. (1978). The gene size DNA molecules in *Oxytricha*. *Cold Spring Harbor Symp. Quant. Biol.* **42**, 483-492.

Leutweiler, L. S., Hough-Evans, B. R. and Meyeroweitz, E. M. (1984). The DNA of *Arabidopsis thaliana*. *Mol. Gen. Genet.* **194**, 15-23.

Levis, R. W. (1989). Viable deletions of a telomere from a *Drosophila* chromosome. *Cell* **58**, 791-801.

Levis, R., Hazelrigg, T. and Rubin, G. M. (1985). Effects of genomic position in the expression of transduced copies of the White gene of *Drosophila*. *Science* **229**, 558-561.

Lewin, B. (1985). Genes II. Published by John Wiley and Sons, Inc.

Lipps, H. J. (1980). *In vitro* aggregation of the gene-sized DNA molecules of the ciliate *Stylonychia mytilus*. *Proc. Natl. Acad. Sci. USA* **77**, 4104-4107.

Lipps, H. J., Gruissem, W. and Prescott, D. M. (1982). Higher order DNA structure in macronuclear chromatin of the hypotrichous ciliate *Oxytricha nova*. *Proc. Natl. Acad. Sci. USA* **79**, 2495-2499.

Longtine, M. S., Wilson, N. M., Petracek, M. E. and Berman, J. (1989). A yeast telomere binding activity binds to two related telomere sequence motifs and is indistinguishable from RAP1. *Curr. Genet.* **16**, 225-239.

Lucchini, G., Falconi, M. M., Pizzagalli, A., Aguilera, A., Klein, L. and Plevani P. (1990). Nucleotide sequence and characterisation of temperature-sensitive *pol1* mutants of *Saccharomyces cerevisiae*. *Gene* **90**, 99-104.

Lundblad, V. and Blackburn, E. H. (1990). RNA-dependent polymerase motifs in EST1: Tentative identification of a protein component of an essential yeast telomerase. *Cell* **60**, 529-530.

Lundblad, V. and Szostak, J. W. (1989). A mutant with a defect in telomere elongation leads to senescence in yeast. *Cell* **57**, 633-643.

Lustig, A. L. and Petes, T. D. (1986). Identification of yeast mutants with altered telomere structure. *Proc. Natl. Acad. Sci. USA* **83**, 1398-1402.

Lyon, M. F. and Kirby, M. C. (1990). Mouse chromosome atlas. *Mouse Genome* **87**, 28-50.

Maguire, M. P. (1984). The mechanism of meiotic homologue pairing. *J. theor. Biol.* **106**, 605-615.

Mandahl, N., Heim, S., Kristoffersson, U., Mitelman, F., Rooser, B., Rydholm, A. and Willen, H. (1985). Telomeric associations in a malignant fibrous histiocytoma. *Hum. Genet.* **71**, 321-324.

Mann, C. and Davis, R. W. (1983). Instability of dicentric plasmids in yeast. *Proc. Natl. Acad. Sci. USA* **80**, 228-232.

de Martinville, B., Wyman, A. R., White, R. and Francke, U. (1982). Assignment of first restriction fragment length polymorphism (RFLP) locus (D14S1) to a region of human chromosome 14. *Am. J. Hum. Genet.* **34**, 216-226.

Mathog, D., Hochstrasser, M., Gruenbaum, Y., Saumweber, H. and Sedat, J. (1984). Characteristic folding pattern of polytene chromosomes in *Drosophila* salivary gland nuclei. *Nature* **308**, 414-421.

Mattei, M. G., Amar, L., Arnaud, D., Dandolot, L., Mattei, J. F., Guenet, J. L., Avner, P. R. and Giraud, F. (1987a). Localisation of eleven unique sequences on mouse C chromosome by *in situ* hybridisation. *Cytogenet. Cell Genet.* **46**, 657.

Mattei, M. G., Harper, K., Brunet, J. F., Denizot, F., Mattei, J. F., Golstein, P. and Giraud, F. (1987b). *In situ* mapping of four cytotoxic T-lymphocyte associated proteins (CTLA) to the mouse genome. *Cytogenet. Cell Genet.* **46**, 657-658.

Maxam, A. M. and Gilbert, W. (1980). Sequencing end-labelled DNA with base-specific chemical cleavages. *Methods in Enzymology* **65**, 449-580.

McClintock, B. (1940). The stability of broken ends of chromosomes in *Zea mays*. *Genetics* **26**, 234-282.

McClintock, B. (1942). The fusion of broken chromosomes following nuclear fusion. *Proc. Natl Acad. Sci. USA* **28**, 458-463.

McCusker, J. H. and Haber, J. E. (1981). Evidence of chromosomal breaks near the mating-type locus of *Saccharomyces cerevisiae* that accompany *MATα x MATα* matings. *Genetics* **99**, 383-403.

Messing, J. (1983). New M13 vectors for cloning. *Met. Enzymol.* **101**, 20-78.

Meyne, J., Baker, R. J., Hobart, H. H., Hsu, T. C., Ryder, O. A., Ward, O. G., Wiley, J. E., Wurster-Hill, D. H., Yates, T. L. and Moyzis, R. K. (1990). Distribution of non-telomeric sites of the $(TTAGGG)_n$ telomeric sequence in vertebrate chromosomes. *Chromosoma (Berl)* **99**, 3-10.

Morin, G. B. (1989). The human telomere terminal transferase enzyme is a ribonucleoprotein that synthesizes TTAGGG repeats. *Cell* **59**, 521-529.

Morin, G. B. (1991). Recognition of a chromosome truncation site associated with α-thalassaemia by human telomerase. *Nature* **353**, 454-456.

Morin, G. B. and Cech, T. R. (1988). Mitochondrial telomeres: Surprising diversity of repeated telomeric DNA sequences among six species of *Tetrahymena*. *Cell* **52**, 367-374.

Moyzis, R. K., Buckingham, J. M., Cram, L. S., Dani, M., Deaven, L. L., Jones, M. D., Meyne, J., Ratliff, R. L. and Wu, J-R. (1988). A highly conserved repetitive DNA sequence, $(TTAGGG)_n$, present at the telomeres of human chromosomes. *Proc. Natl. Acad. Sci. USA* **85**, 6622-6626.

Müller, H. J. (1938). The remaking of chromosomes. *The Collecting Net* **13**, 181-195.

Muller-Hill, B., Crapo, L. and Gilbert, W. (1968). Mutants that make more Lac repressor. *Proc. Natl. Acad. Sci. USA* **59**, 1259-1264.

Müller, H. J. and Herskowitz, I. R. (1954). Concerning the healing of chromosome ends produced by breakage in *Drosophila melanogaster*. *Am. Nat.* **88**,177-208.

Müller, F., Wickey, C., Spicher, A. and Tobler, H. (1991). New telomere formation after developmentally regulated chromosomal breakage during the process of chromatin diminution in *Ascaris lumbricoides*. *Cell* **67**, 815-822.

Murray, A. W., Claus, T. E. and Szostak, J. W. (1988). Characterisation of two telomeric DNA processing reactions in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **8**, 4642-4650.

Murray, A. W. and Szostak, J. W. (1983). Construction of artificial chromosomes in yeast. *Nature* **305**, 189-193.

Nakamura, Y., Carlson, M., Krapcho, K., Kanamori, M. and White, R. (1988). New approach for isolation of VNTR markers. *Am. J. Hum. Genet.* **43**, 854-859.

Nakamura, Y., Leppert, M., O'Connell, P., Wolff, R., Holm, T., Culver, M., Martin, C., Fujimoto, E., Hoff, M., Kumlin, E. and White, R. (1987). Variable number of tandem repeat (VNTR) markers for human gene mapping. *Science* **235**, 1616-1622.

Norrander, J. T., Kempe, T. and Messing, J. (1983). Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. *Gene* **26**, 101-106.

Oka, Y. and Thomas, C. A. Jr. (1987). The cohering telomeres of *Oxytricha*. *Nuc. Acids Res.* **15**, 8877-8898.

Olins, A. L., Carlson, R. D., Wright, E. B. and Olins, D. E. (1976). Chromatin ν-bodies; isolation, sub-portionation and physical characterisation. *Nucl. Acids Res.* **3**, 3271-3291.

Olszewska, E. and Jones, K. (1988). Vacuum blotting enhances nucleic acid transfer. *TIG* **4**, 92-94.

Pace, T., Ponzi, M., Dore, E. and Frontali, C. (1987). Telomeric motifs are present in a highly repetitive element in the *Plasmodium berghei* genome. *Mol. Biochem. Parasit.* **24**, 193-202.

Pace, T., Ponzi, M., Dore, E., Janse, C., Mons, B. and Frontali, C. (1990). Long insertions within telomeres contribute to chromosome size polymorphisms in *Plasmodium berghei*. *Mol. Cell. Biol.* **10**, 6759-6764.

Pathak, S., Wang, Z., Dhaliwal, M.K. and Sacks, P.C. (1988). Telomeric association: another characteristic of cancer chromosomes ? *Cytogenet. Cell Genet.* **47**, 227-229.

Pays, E. and Steinert, M. (1988). Control of antigen gene expression in African trypanosomes. *Ann. Rev. Genet.* **22**, 107-126.

Pietras, D. F., Bennett, K. L., Siracusa, L. D., Woodworth-Gutai, M., Chapman, V. M., Gross, K. W., Kane-Haas, C. and Hastie, N. D. (1983). Construction of a small *Mus musculus* repetitive DNA library: identification of a new satellite sequence in *Mus musculus*. *Nucl. Acids Res.* **11**, 6965-6983.

Pluta, A. F., Kaine, B. P. and Spear, B. B. (1982). The terminal organisation of macronuclear DNA in *Oxytricha fallax*. *Nucl. Acids Res.* **10**, 8145-8154.

Pluta, A. F. and Zakian, V. A. (1989). Recombination occurs during telomere formation in yeast. *Nature* **337**, 429-433.

Pologe, L. G. and Ravetch, J. V. (1988). Large deletions result from breakage and healing of *P. falciparum* chromosomes. *Cell* **55**, 869-874.

Ponzi, M., Pace, T., Dore, E. and Frontali, C. (1985). Identification of a telomeric DNA sequence in *Plasmodium berghei*. *EMBO J.* **4**, 2991-2996.

Preer, J. R. and Preer, L. B. (1979). The size of macronuclear DNA and its relationship to models for maintaining genic balance. *J. Protozool.* **26**, 14-18.

Price, C. M. and Cech, T. R. (1987). Telomeric DNA-protein interactions of *Oxytricha* macronuclear DNA. *Genes Dev.* **1**, 783-793.

Pritchard, C. and Goodfellow, P. N. (1985). The pseudoautosomal region and telomeres: the beginning of the end ? *TIG* **1**, 289.

Proudfoot, N. J., Gil, A. and Maniatis, T. (1982). The structure of the human zeta-globin gene and a closely linked, nearly identical pseudogene. *Cell* **31**, 553-563.

Raghuraman, M. K. and Cech, T. R. (1989). Assembly and self-association of *Oxytricha* telomeric nucleoprotein complexes. *Cell* **59**, 719-728.

Raghuraman, M. K., Dunn, C. J., Hicke, B. J. and Cech, T. R. (1989). *Oxytricha* telomeric nucleoprotein complexes reconstituted with synthetic DNA. *Nucl. Acids Res.* **17**, 4235-4253.

Reed, K. C. and Mann, D. A. (1985). Rapid transfer of DNA from agarose gels to nylon membranes. *Nucl. Acids Res.* **13**, 7207-7221.

Reeders, S. T., Breuning, M. H., Davies, K. E., Nicholls, R. D., Jarman, A. P., Higgs, D. R., Pearson, P. L. and Weatherall, D. J. (1985). A highly polymorphic DNA marker linked to adult polycystic kidney disease on chromosome 16. *Nature* **317**, 542-544.

Richards, E. J. and Ausubel, F. M. (1988). Isolation of a higher eukaryotic telomere from *Arabidopsis thaliana*. *Cell* **53**, 127-136.

Riethman, H. C., Moyzis, R. K., Meyne, J., Burke, D. T. and Olsen, M. V. (1989). Cloning human telomeric DNA fragments into *Saccharomyces cerevisiae* using a yeast-artificial-chromosome vector. *Proc. Natl. Acad. Sci. USA* **86**, 6240-6244.

Roth, M. and Prescott, D. M. (1985). DNA intermediates and telomere addition during genome reorganisation in *Euplotes crassus*. *Cell* **41**, 411-417.

Rouyer, F., de la Chapelle, A., Andersson, M. and Weissenbach, J. (1990). An interspersed repeated sequence specific for human subtelomeric regions. *EMBO J.* **9**, 505-514.

Rouyer, F., Simmler, M-C., Johnsson, C., Vergnaud, G., Cooke, H. J. and Weissenbach, J. (1986). A gradient of sex linkage in the pseudoautosomal region of the human sex chromosomes. *Nature* **319**, 291-295.

Royle, N. J., Clarkson, R. E., Wong, Z. and Jeffreys, A. J. (1988). Clustering of hypervariable minisatellites in the proterminal regions of human autosomes. *Genomics* **3**, 352-360.

Rubin, G. M. (1977). Isolation of a telomeric DNA sequence from *Drosophila melanogaster*. *Cold Spring Harbor Symp. Quant. Biol.* **42**, 1041-1046.

Runge, K. W. and Zakian, V. A. (1989). Introduction of extra telomeric DNA sequences into *Saccharomyces cerevisiae* results in telomere elongation. *Mol. Cell. Biol.* **9**, 1488-1497.

Saiga, H. and Edström, J.-E. (1985). Long tandem arrays of complex repeat units in *Chironomus* telomeres. *EMBO J.* **4**, 799-804.

Saltman, D., Ross, F. M., Fantes, J. A., Allshire, R. C., Turner, G. E. and Evans, H. J. (1989). Telomeric associations in a lymphoblastoid cell line from a patient with B-cell follicular lymphoma. *Cytogenet. Cell Genet.* **50**, 230-233.

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989). *Molecular cloning. A laboratory manual*, 2nd Edn. Cold Spring Harbour Laboratory Press.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977). DNA sequencing with chain termination inhibitors. *Proc. Natl.. Acad. Sci. USA.* **74**, 5463-5467.

Sealy, P. and Southern, A. (1982). *Gel Electrophoresis*,2nd Edn, vol. 1 Reed 1990: Oxford University Press. (Chapter 2, pp51-99).

Sen, D. and Gilbert, W. (1988). Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature* **334**, 364-366.

Shampay, J., Szostak, J. W. and Blackburn, E. H. (1984). DNA sequences of telomeres maintained in yeast. *Nature* **310**, 154-157.

Sherwood, S. W., Rush, D., Ellsworth, J. L. and Schmicke, R. T. (1988). Defining cellular senescence in IMR-90 cells: A flow cytometric analysis. *Proc. Natl. Acad. Sci. USA* **85**, 9086-9090.

Shippen-Lentz, D. and Blackburn, E.H. (1989). Telomere terminal transferase activity from *Euplotes crassus* adds large numbers of TTTTGGGG repeats onto telomeric primers. *Mol. Cell. Biol.* **9**, 2761-2764.

Shippen-Lentz, D. and Blackburn, E. H. (1990). Functional evidence for an RNA template in telomerase. *Science* **247**, 546-552.

Shore, D. and Nasmyth, K. (1987). Purification and cloning of a DNA binding protein from yeast that binds to both silencer and activator elements. *Cell* **51**, 721-732.

Simmler, M. C., Johnsson, C., Petit, C., Rouyer, F., Vergnaud, G. and Weissenbach, J. (1987). Two highly polymorphic minisatellites from the pseudoautosomal region of the human sex chromosomes. *EMBO J.* **6**, 963-969.

Simmler, M-C., Rouyer, F., Vergnaud, G., Nystrom-Lahti, M., Ngo, K. Y., de la Chapelle, A. and Weissenbach, J. (1985). Pseudoautosomal DNA sequences in the pairing region of the human sex chromosomes. *Nature* **317**, 692-697.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98**, 503-517.

Stryer, L., (1981). Biochemistry. Published by W. H. Freeman and Company, San Francisco.

Sugawara, N. and Szostak, J. W. (1986). Telomeres of *Saccharomyces pombe. 13th Int. Conf. on Yeast Genetics and Molecular Biology*, S373.

Sundquist, W. I. and Klug, A. (1989). Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. *Nature* **342**, 825-829.

Szostak, J. W. (1983). Replication and resolution of telomeres in yeast. *Cold Spring Harbor Symp. Quant. Biol.* **47**, 1187-1194.

Szostak, J. W. and Blackburn, E. H. (1982). Cloning yeast telomeres on linear plasmid vectors. *Cell* **29**, 245-255.

Tabor, S. and Richardson, C. C. (1987). DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. *Proc. Natl. Acad. Sci. USA.* **84**, 4767-4771.

Taylor, B. A. (1978). Recombinant inbred strains: Use in gene mapping. *Origins of inbred mice*, 423-438. Published by Academic Press.

Thein, S. L., Jeffreys, A. J., Gooi, H. C., Cotter, F., Flint, J., O'Conner, N. T. J., Weatherall, D. J. and Wainscoat, J. S. (1987). Detection of somatic changes in human cancer DNA by DNA fingerprint analysis. *Br. J. Cancer* **55**, 353-356.

Traverse, K. L. and Pardue, M. L. (1989). Studies of He-T DNA sequences in the pericentric regions of *Drosophila* chromosomes. *Chromosoma (Berl)* **97**, 261-271.

Van der Ploeg, L. H. T., Liu, A. Y. C. and Borst, P. (1984). Structure of the growing telomeres of *Trypanosomes. Cell* **36**, 459-468.

Verkerk, A. J. M. H., Pieretti, M., Sutcliffe, J. S., Fu, Y-H., Kuhl, D. P. A., Pizzuti, A., Reiner, O., Richards, S., Victoria, M. F., Zhang, F., Eussen, B. E., van Ommen, G-J. B., Blonden, L. A. J., Riggens, G. J., Chastin, J. L., Kunst, C. B., Galjaard, H., Caskey, C. T., Nelson, D. L., Oostra, B. A. and Warren, S. T. (1991). Identification of a gene (*FMR-1*) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in Fragile X Syndrome. *Cell* **65**, 905-914.

Vieira, J. and Messing, J. (1982). The pUC plasmids, an M13mp7 derived system for the insertion mutagenesis and sequencing with synthetic universal primers. *Gene* **19**, 269-276.

Walmsley, R. W., Chan, C. S. M., Tye, B-K. and Petes, T. D. (1984). Unusual DNA sequences associated with the ends of yeast chromosomes. *Nature* **310**, 157-160.

Walmsley, R. M. and Petes, T. D. (1985). Genetic control of chromosome length in yeast. *Proc. Natl. Acad. Sci. USA* **82**, 506-510.

Wang, S-S., Pluta, A. F. and Zakian, V. A. (1989). DNA sequence analysis of newly formed telomeres in yeast. *Mechanisms of Chromosome Distribution and Aneuploidy*, Ed A. R. Liss, Inc., 81-89.

Wang, S-S. and Zakian, V. A. (1990). Telomere-telomere recombination provides an express pathway for telomere acquistion. *Nature* **345**, 456-458.

Watson, J. D. (1972). Origin of concatameric T7 DNA. *Nature* **239**, 197-201.

Weiss, B., Jacquemin-Sablon, A., Live,T. R., Fareed, G. C. and Richardson, C. C. (1968). Enzymatic breakage and joining of deoxyribonucleic acid. VI. Further purification and properties of polynucleotide ligase from *Escherischia coli* infected with bacteriophage T4. *J.Biol. Chem.* **243**, 4543-4555.

Weller, P., Jeffreys, A. J., Wilson, V. and Blanchetot, A. (1984). Organisation of the human myoglobin gene. *EMBO J.* **3**, 439-446.

Wells, R. A., Germino, G. G., Krishna, S., Buckle, V. J. and Reeders, S. T. (1990). Telomere-related sequences at interstitial sites in the human genome. *Genomics* **8**, 699-704.

Wells, R. A., Green, P. and Reeders, S. T. (1989). Simultaneous genetic mapping of multiple human minisatellite sequences using DNA fingerprinting. *Genomics* **5**, 761-772.

Wilimzig, M. (1985). LiCl boiled method for plasmid mini preps. *TIG* **1**, 158.

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

Wilkie, A. O. M., Lamb, J., Harris, P. C., Finney, R. D. and Higgs, D. R. (1990). A truncated human chromosome 16 associated with α thalassaemia is stabilized by addition of telomeric repeat $(TTAGGG)_n$. *Nature* **346**, 868-871.

Williamson, J. R., Raghuraman, M. K. and Cech, T. R. (1989). Monovalent cation-induced structure of telomeric DNA: The G-quartet model. *Cell* **59**, 871-880.

Wolff, R. K., Nakamura, Y. and White, R. (1988). Molecular charaterisation of a spontaneously generated new allele at a VNTR locus: No exchange of flanking DNA sequence. *Genomics* **3**, 347-351.

Wolff, R. K., Plaetke, R., Jeffreys, A. J. and White, R. (1989). Unequal crossing over between homologous chromosomes is not the major mechanism involved in the generation of new alleles at VNTR loci. *Genomics* **5**, 382-384.

Wong, Z., Wilson, V., Jeffreys, A. J. and Thein, S. L. (1986). Cloning a selected fragment from a human DNA 'fingerprint': isolation of an extremely polymorphic minisatellite. *Nucl. Acids Res.* **14**, 4605-4616.

Wong, Z., Wilson, V., Patel, I., Povey, S. and Jeffreys, A. J. (1987). Characterisation of a panel
of highly variable minisatellites cloned from human DNA. *Ann. Hum. Genet.* **51**, 269-288.

Wyman, A. and White, R. (1980). A highly polymorphic locus in human DNA. *Proc. Natl. Acad. Sci. USA* **77**, 6754-6758.

Yanisch-Peron, C., Vieira, J. and Messing, J. (1985). Improved M13 Phage cloning vectors and host strains : nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**, 103-119.

Yao, M-C., Blackburn, E. H. and Gall, J. G. (1978). Amplification of the rRNA genes in *Tetrahymena. Cold Spring Harbor Symp. Quant. Biol.* **43**, 1293-1296.

Yao, M. -C. and Gall, J. G. (1977). A single integrated gene for ribosomal RNA in a eukaryote, *Tetrahymena pyriformis. Cell* **12**, 121-132.

Yao, M-C., Yao, C-H. and Monks, B. (1990). The controlling sequence for site-specific chromosome breakage in *Tetrahymena. Cell* **63**, 763-772.

Yerganian, G., Shein, H. M. and Enders, J. F. (1962). Chromosomal disturbances observed in human fetal renal cells transformed *in vitro* by Simian Virus 40 and carried in culture. *Cytogenetics* **1**, 314-324.

Young, B. S., Pession, A., Traverse, K. L., French, C. and Pardue, M. L. (1983). Telomere regions in *Drosophila* share complex DNA sequences with pericentric heterochromatin. *Cell* **34**, 85-94.

Yu, G-L. and Blackburn, E. H. (1991). Developmentally programmed healing of chromosomes by telomerase in *Tetrahymena. Cell* **67**, 823-832.

Yu, G-L., Bradley, J. D., Attardi, L. D. and Blackburn, E. H. (1990). *In vivo* alteration of telomere sequences and senescence caused by mutated *Tetrahymena* telomerase RNAs. *Nature* **344**, 126-132.

Yunis, J. J. and Prakash, O. (1982). The origin of man: A chromosomal pictorial legacy. *Science* **215**, 1525-1530.

Zakian, V. A. (1989). Structure and function of telomeres. *Annu. Rev. Genet.* **23**, 579-604.

Zakian, V. A. and Blanton, H. M. (1988). Distribution of telomere-associated sequences on natural chromosomes in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **8**, 2257-2260.

Zahler, A. M. and Prescott, D. M. (1988). Telomere terminal transferase activity in the hypotrichous ciliate *Oxytricha nova* and a model for replication of the ends of linear DNA molecules. *Nucl. Acids. Res.* **16**, 6953-6972.

Zahler, A. M., Williamson, J. R., Cech, T. R. and Prescott, D. M. (1991). Inhibition of telomerase by G-quartet DNA structures. *Nature* **350**, 718-720.

Publications

Extensive telomere repeat arrays in mouse are hypervariable

J.A.Starling, J.Maule, N.D.Hastie and R.C.Allshire[1]
MRC Human Genetics Unit, Western General Hospital, Crewe Road, Edinburgh EH4 2XU and
[1]Cold Spring Harbor Laboratory, PO Box 100, Cold Spring Harbor, NY 11724, USA

Received September 10, 1990; Revised and Accepted October 17, 1990

ABSTRACT

In this study we have analysed mouse telomeres by Pulsed Field Gel Electrophoresis (PFGE). A number of specific restriction fragments hybridising to a $(TTAGGG)_4$ probe in the size range 50–150kb can be detected. These fragments are devoid of sites for most restriction enzymes suggesting that they comprise simple repeats; we argue that most of these are likely to be $(TTAGGG)_n$. Each discrete fragment corresponds to the telomere of an individual chromosome and segregates as a Mendelian character. However, new size variants are being generated in the germ line at very high rates such that inbred mice are heterozygous at all telomeres analysable. In addition we show that specific small (~4–12kb) fragments can be cleaved within some terminal arrays by the restriction enzyme MnlI which recognises 5′(N_7)GAGG3′. Like the complete telomere-repeat arrays (TRA's) these fragments form new variants at high rates and possibly by the same process. We speculate on the mechanisms that may be involved.

INTRODUCTION

The ends of mammalian chromosomes, like those of simple eukaryotes, terminate with stretches of G-rich repeats (1, 2, 3,). The special structures adopted by these repeats may render the termini (telomeres) resistant to degradation and fusion (4, 5, 6, 7). In ciliates and humans, at least, these repeat arrays are added to the ends of chromosomes de novo by a ribonucleoprotein enzyme, telomerase (8, 9, 10, 11, 12, 13, 14). In *S.cerevisiae* TRA's at the ends of linear plasmids participate in intermolecular exchanges, converting each other at high frequency, but it is not known whether this occurs between natural chromosome ends (15, 16, 17).

The termini of human chromosomes are composed of 10–15kb arrays which are mainly of the form TTAGGG (18, 19, 20, 21, 22); at the proximal end of the arrays are clusters of degenerate repeats such as $(TTGGGG)_n$ or $(TGAGGG)_n$ (3, 23). On conventional agarose gels the human TRA's can only be revealed as heterogeneous smears in Southern analysis using a labelled $(TTAGGG)_4$ probe. This is, presumably, because 46 different chromosome ends are being detected simultaneously and any particular end is likely to vary from cell to cell. Thus it has not been possible to follow the genetic behaviour of individual telomeres.

In a previous study, we showed that mouse TRA's which also hybridised to $(TTAGGG)_4$ were very much larger than those of human, failing to enter conventional agarose gels (23). Here we show that it is possible using PFGE to characterise the terminal fragments of specific mouse chromosomes and to follow their inheritance.

MATERIALS AND METHODS

Total genomic mouse DNA

(a)Strains used

Total spleen and kidney DNA's extracted from *Mus musculus, Mus domesticus, Apodemus sylvaticus* and the inbred Mus musculus strain AKR/J mice were a gift from R Hill. *Mus spretus* DNA was a gift from N Copeland and N Jenkins. The *Mus caroli* mouse was a gift from J Ansell. The *Mus musculus, Mus domesticus* and *Apodemus sylvaticus* mice were trapped wild in Greece, Denmark and locally respectively. The *Mus spretus* and *Mus caroli* mice were trapped and then held in the laboratory for a period of time. The inbred strains used were DBA/2, C57BL/6 and AKR/J. DBA/2 mice were obtained from colonies in Hull and Oxford. C57BL/6 mice were obtained from both local and Hull colonies. Cross strain matings were set up between Hull DBA/2 and local C57BL/6 mice. The Fl progeny were then brother/sister mated to obtain the F2 generation.

(b)DNA isolation

Two different methods were used:

(1)Nuclei were isolated and DNA recovered as described by Hill *et al.* (24) from mouse spleen, liver and kidney. This method was used for DBA/2 and C57BL/6 DNA used in Figure 4.

(2)The spleen, liver or kidney from a freshly killed mouse was dropped into buffer (100 mM NaCl, 50 mM Tris, 10 mM EDTA, pH8.0) and processed for 10 seconds with a Silversun homogenizer. Next, proteinase K to 100 μg/ml and SDS to 0.5% were added and the whole incubated for 16 hours at 37°C. Next day, the preparations were extracted with phenol (1 volume (vol)), phenol (0.5 vol)/chloroform (0.5 vol), and chloroform (1 vol). The DNA was precipitated with 0.3 vol of 7.5M NH_4OAc and 2.5 vols of ethanol and recovered by spooling into TE (1 mM EDTA, 10 mM Tris, pH8.0). This method was used for the *Mus*

caroli mouse and those mice involved in the cross strain mating experiment.

Southern blotting

Conventional gel electrophoresis

Restriction enzyme digests were performed according to the manufacturer's recommended conditions. Restricted DNA was separated in a 20 cm by 20 cm, 0.8% agarose gel at 50 volts in TAE buffer (0.8 M Tris, 0.4 M NaOAc, 0.02 M EDTA pH8.2) for 16 hours. The separated DNA fragments were then transferred to Hybond-N (Amersham) using a Vacugene vacuum transfer apparatus (LKB).

Pulse field gel electrophoresis (PFGE)

Restriction enzyme digests were performed as above, with extra care taken to ensure minimal shearing of the liquid DNA. Restricted DNA was separated using a Chef DR II Biorad pulse field system in a 1% agarose gel, run in 0.5 ×TAE at 6 volts/cm, using a 5 second pulse time at 10°C for 23 hours. Transfer to Hybond-N was as described above.

Oligomer hybridizations

The $(TTAGGG)_4$ synthetic oligonucleotide was manufactured on an Applied Biosystems 381A DNA Synthesizer by D Chambers. For a standard reaction, 25 ng of the oligonucleotide was labelled in the presence of 5O uCi of gamma ^{32}P ATP (Amersham, 5000 Ci/ml) using 10 units of Polynucleotide Kinase (BCL) in the recommended buffer. Hybridizations were performed in 5×SSC, 5 ×Denharts, 0.1% SDS and 0.1% NaPPi at 48°C for 16 hours. The filters were washed four times for 15 minutes in 4×SSC, 0.1% SDS and 0.1% NaPPi at 56°C.

Random prime hybridizations

The 196 mouse major satellite double strand DNA probe was a gift from A Mitchell originally isolated by Pietras *et al.* (25). It was radiolabelled with alpha^{32}p CTP (Amersham; 800 Ci/mMol) to a specific activity greater than 10^8 cpm using a random priming kit (BCL). Hybridizations were performed under the conditions of Church and Gilbert at 68°C (26). Filters were washed four times for 15 minutes in 2×SSC, 0.1% SDS and 0.1% NaPPi at 68°C.

Bal31 time course

The reaction mix contained DBA/2 total genomic DNA at a concentration of 200 μg/ml which was digested with 40 units/mi of Bal31 (BCL) in the manufacturers recommended buffer at 30°C. Initially, the mix, without Bal31, was preincubated at 30°C for 10 minutes, one aliquot removed and incubated in the absence of Bal31 for the duration of the time course (timepoint = 0). Subsequently, the Bal31 was added and aliquots were removed at the required timepoints. The reaction was stopped by the addition of 0.1 vols of 200 mM EGTA pH8.0. Each aliquot was then extracted once each with phenol (1 vol), phenol (0.5 vol) /chloroform (0.5 vol), and chloroform (1 vol). The DNA was recovered by precipitation with 0.3 vol of NH_4OAc and 2.5 vols of ethanol and spooled into TE pH8.0. This was then digested with HaeIII. For the complete timecourse 390 μg of total genomic DBA DNA was used; each aliquot contained 30 μg of DNA.

Determination of origin of MnlI specific DNA fragments

(a) Analysis of total telomeric smear

Total mouse genomic DNA of the required individual was digested with the restriction enzyme HaeIII (BCL). The sample was separated by conventional gel electrophoresis in 1% low melting point (LW), agarose (BRL). DNA in the size range > 23 kb was purified as follows; the gel slice was melted to 68°C in an equal volume of 100 mM NaCl, 10 mM Tris pH8.0, ImM EDTA for 30 minutes. The molten gel was cooled to 37°C and digested for 16 hours with 5 units/ml of agarose (Calbiochem). The DNA was then extracted once each with phenol (1 vol), phenol (0.5 vol)/chloroform (0.5 vol), and chloroform (1 vol) and recovered by precipitation with 0.3 vols of NH_4OAc and 2.5 vols of ethanol. The DNA obtained was then digested with MnlI (New England Biolabs) and the fragments separated by conventional gel electrophoresis.

(b) Analysis of a specific MnlI DNA fragment

Total genomic DNA from the required individual was digested with HaeIII and the fragments separated by PFGE (as above) into 1% LMP agarose. The position of the required band was determined and separated from the rest of the gel. The plug was redigested with 10−20 units of MnlI in an equal volume of the manufacturers recommended buffer together with 1 mM spermidine and 0.1% Triton×100 for 16 hours at 37°C. The plug was washed in ice cold TE PH8.0, the digest stopped with 10 mM EDTA and then subjected to conventional gel electrophoresis as previously described.

RESULTS

A telomere probe recognises discrete high molecular weight fragments in mouse DNA.

DNA from different mouse species and inbred strains was digested with the 4bp-cutting restriction enzyme HaeIII, which should not cut within the TRA, and electrophoresed on a conventional 0.8% agarose gel. Following Southern transfer, hybridization with a $(TTAGGG)_4$ probe showed that, in all cases, with the exception of *Mus caroli* and *Mus spretus*, the TRA's ran as an unresolved band > 23 kb (Fig 1). The telomeres from *Mus caroli* and *Mus spretus* were very much smaller, ranging in size from 20−30 kb and 5−20 kb respectively.

To determine the size of the TRA's in *Mus musculus*, DNA from the inbred strains C57BL/6 and DBA/2 was digested with a series of 4bp-cutting restriction enzymes and subject to PFGE. Hybridisation to the $(TTAGGG)_4$ probe revealed a series of discrete bands in DBA/2 mice ranging in size from 50−150 kb (Fig 2). The majority of the probe hybridized to a complex mixture of poorly resolved fragments between 30−50 kb. An identical pattern of fragments was observed with several different 4bp-cutting restriction enzymes (data not shown); this is to be expected if the telomeres are composed mainly of TTAGGG stretches which contain no sites for any of these enzymes. Fewer bands were observed in C57BL/6 animals in the size range > 50 kb.

The high molecular weight fragments are terminal and behave as highly variant Mendelian loci

It seemed highly likely that these very long fragments hybridizing to $(TTAGGG)_4$ were terminal fragments; also the discrete nature of each band suggested to us that it derived from an individual chromosome end. To confirm these suspicions, we carried out both biochemical and genetic analysis.

If these long fragments derive from individual chromosome ends they might be expected to segregate as single genetic loci in families. Thus we followed the segregation of these TRA bands

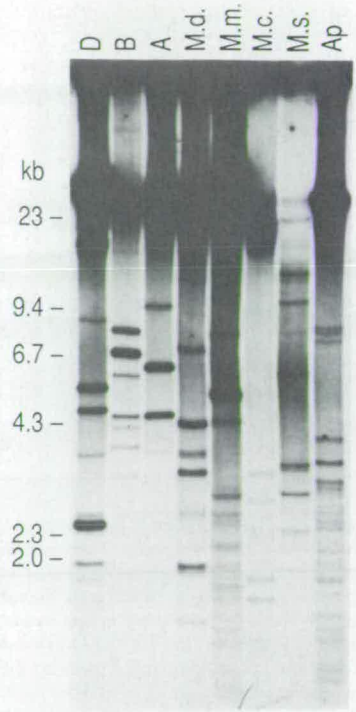


Figure 1. Analysis of TRAs in different mice strains and species by conventional gel electrophoresis. 10 μg samples of DBA/6 (D), C57BL/6 (B), Akr/J (A), *Mus domesticus* (M.d.), *Mus musculus* (M.m.), *Mus caroli* (M.c.), *Mus spretus* (M.s.), and *Apodemus sylvaticus* (Ap) DNA were digested with HaeIII, electrophoresed through a 0.8% agarose gel and transfered to a nylon filter. The filter was then hybridised with $(TTAGGG)_4$ as described in the materials and methods. HindIII digest of lambda DNA is indicated as size marker (kb).

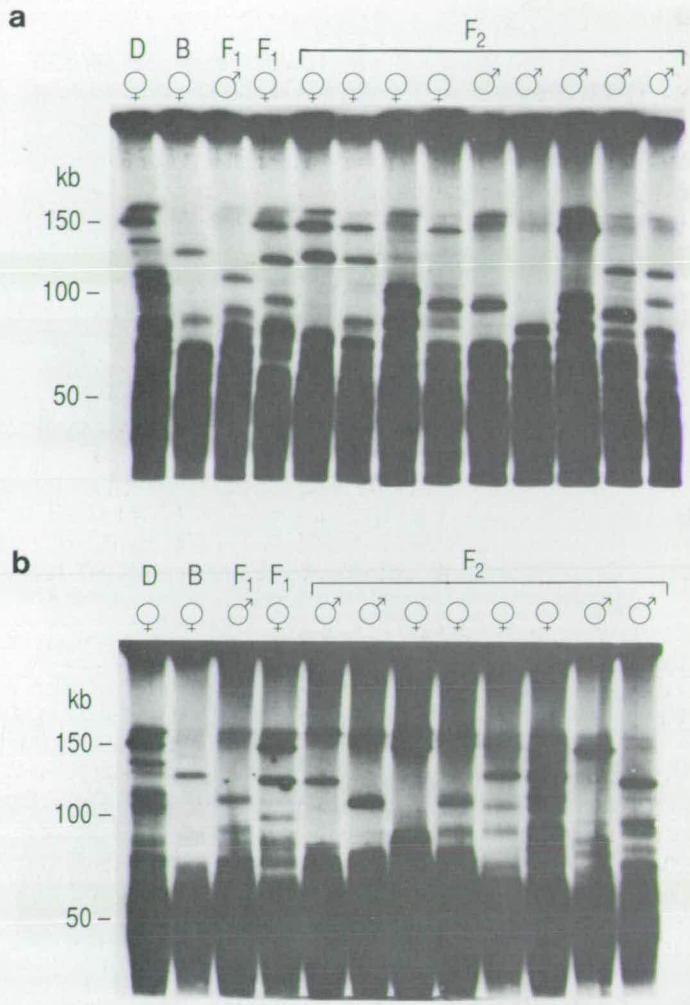


Figure 2. Analysis by PFGE of the inheritance of mouse TRAs. (a) and (b) Cross strain mating between a DBA/2 mouse (D) and a C57BL/6 (B) mouse. From the resultant F1 progeny a brother ≈ × sister pair was set up to generate the F2 offspring. (a) and (b) show the same initial DBA/2 and C57BL/6 pair and the same F1 brother × sister pair but 2 sets of F2 offspring respectively. 10 μg of total genomic spleen DNA from each mouse was digested with HaeIII. The restricted fragments separated by PFGE, transferred to a nylon filter and hybridized with $(TTAGGG)_4$ as described in the materials and methods. Lambda concatemer size markers are indicated (kb).

in several generation crosses between C57BL/6 and DBA/2 inbred strains. As shown in Figure 2 (a and b) the sibling F1's were not identical, inheriting a different combination of bands from their parents. Not one of the parental fragments was found in all the F1's (only four shown). Each F1 received about half of the DBA parent's bands. The most plausible explanation for these observations is that the 2 allelic terminal restriction fragments for any particular chromosome differ in size even in inbred mice. For this to be the case new length variants would have to be generated in the germ line at high frequency. It is possible to see new size bands in the F2 generation upon careful examination of the data but these are very close in size to bands inherited from the parents. The analysis is more difficult by PFGE than for small fragments on conventional gels. A more clear-cut case of a new variant observed in the F1 generation is shown below. It was quite easy to follow the stable segregation of most bands through several generations to approximately 50% of offspring as expected for a single genetic locus (Fig 2a and b). All possible combinations of DBA/2 fragments were found in the F1 generation so none of these bands is allelic with another. We must assume that the alleles of these termini are within the 30–50 kb range.

To test the terminal nature of the fragments the DNA was treated with the exonuclease Bal31 for various lengths of time prior to digestion with the restriction enzyme. As shown in Figure 3 these very long fragments become progressively smaller with increasing duration of Bal31 digestion. The discrete high molecular weight fragments had reduced in size by 30–40 kb by the final time point. The same can be said for the bulk of the telomeres represented by the 30–50 kb smear, also there is a considerable reduction in signal strength within the smear by the final time point. The rate of reduction in the length of the TRA's is approximately 30 kb in 90 minutes using 1 unit of Bal31 per 5 μgs of DNA. This compares favourably with our previous experiments showing a rate of loss of 15 kb in 90 minutes for human telomeres using a lower concentration of Bal31 (1 unit per 13.3 μgs of DNA)(18). The progressive decrease in size of these larger fragments with time and with the expected kinetics argues against non-specific degradation during the Bal31 digestion. There is no ideal control to test for degradation over this size range. However, the smear hybridising to a mouse satellite probe and extending up to 70 kb was not reduced in size and intensity during the course of the Bal31 digestion (Fig 3b).

Given the rate of acquisition of new variants and heterozygosity of these fragments we should expect that different colonies of the same inbred strain and even members of the same colony would show different patterns; this was indeed found to be the

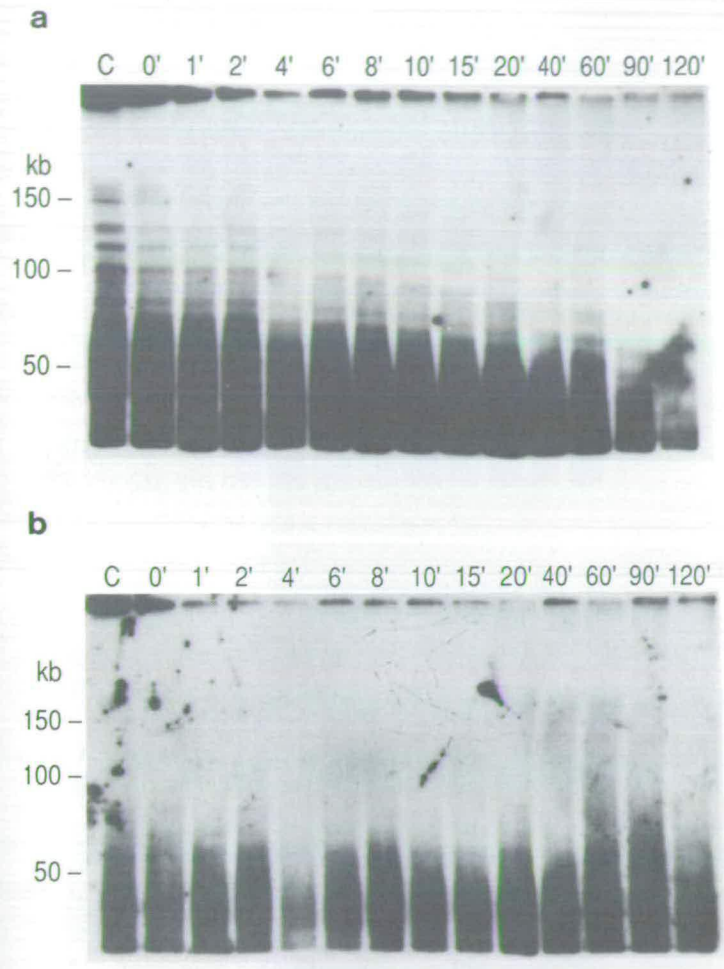


Figure 3. Bal31 sensitivity of mouse TRAs. The Bal31 time course was performed on total genomic spleen DNA from a DBA/2 mouse as described in the materials and methods. 10 μg of DNA from each time point was digested with HaeIII. The restricted fragments were separated by PFGE transferred to a nylon filter and hybridized with (a) $(TTAGGG)_4$ and (b) the mouse 196 major satellite as described in the materials and methods. Between the two hybridizations the filter was boiled for 10 minutes in distilled water. Track C refers to the starting DBA/2 DNA digested with HaeIII. 0′, 1′ etc refer to the time of exposure to Bal31. Lambda concatemer DNA size markers are indicated (kb).

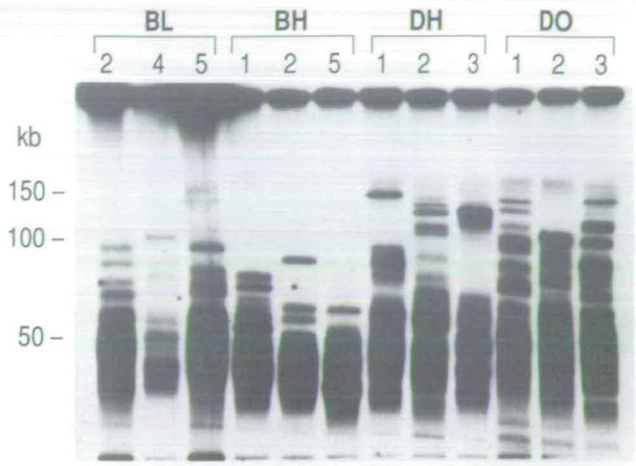


Figure 4. PFGE comparing TRAs of different individuals of C57BL/6 and DBA/2 mouse strains from different inbred colonies. Total genomic DNA from each individual was digested with HaeIII, subjected to PFGE, transferred to a nylon filter and hybridized with $(TTAGGG)_4$ as described in the materials and methods. B refers to the C57BL/6 and D to the DBA/2 mouse strains. L refers to a local inbred colony of mice, H to colonies in Hull and O to a colony in Oxford. The numbers above each lane refer to the number given to each individual mouse received at different times and bear no relation to other mice of the same number from different colonies. Lambda concatemer DNA size markers are indicated (kb).

case as shown in Figure 4 but it was also clear that the mice in the 2 different DBA colonies (one maintained in Hull, the other in Oxford) were more similar to each other in banding pattern than they were to any of the C57BL/6 mice, whose fragments were smaller.

MnlI releases small highly variable fragments from some telomeres

No restriction enzymes will cleave tandem arrays of TTAGGG. However, the type IIS restriction enzyme MnlI, which recognises the nonpallindromic sequence $5'(N)_7GAGG3'$, will cut within the TRA where there is a single T-G transition, caused either by occasional misincorporation by the telomerase or by mutation, resulting in a TGAGGG repeat. We demonstrated previously that a cluster of MnlI sites resides within the proximal end of human TRA's (23). Figure 5 shows PFGE analysis of DBA/2,C57BL/6 and F1's DNA cut with HaeIII or MnlI and probed with $(TTAGGG)_4$. The very large fragments of 50 kb and greater seen with HaeIII (and other 4bp-cutting restriction enzymes tested) were not observed upon MnlI digestion. Also, there has been a shift in the size of the smear to a lower molecular weight range. The most likely exphination is that MnlI is cutting at least once within most telomeres. If this were the case we might expect that small MnlI fragments released from some of the TRA's could now be resolved on conventional agarose gels. This was found to be the case. As shown in Figure 6 fragments ranging in size from 1−12 kb were observed when DNA was cut with MnlI and probed with $(TTAGGG)_4$. All the bands were variant between DBA/2 and C57BL/6 animals. The cluster of bands between 1 and 3 kb long behaved like most other genetic loci in the F1 generation; that is, all the bands of both parents were represented in all the F1's. However some bands from 4 kb and upwards are behaving quite differently. These bands were found in only a proportion of F1's, indicating that like the entire TRA'S, they are heterozygous even in inbred mice and therefore segregate in the F1 generation. We have observed bands of 1−10 kb hybridizing to the $(TTAGGG)_4$ probe with a variety of 4bp-cutting restriction enzymes, including HaeIII, as shown in Figure 6a. These HaeIII banding patterns differed between inbred strains but not between individuals within a strain (data not shown). Also, unlike the MnlI fragments of 4 kb and upwards (Fig. 6b), all of the bands from both parents were detected in all F1 progeny suggesting that the inbred mice are homozygous at these loci (Fig 6a).

To explain the degree of heterozygosity of the MnlI fragments >4 kb in size new variants must be generated at high frequency. In support of this expectation a new variant of ~12 kb can be seen in one of the F1's (Fig 6b). The signal strength for this new fragment is less than expected in relation to the smaller fragments. This could either mean that the band is not represented in all spleen cells, and hence is a somatic variant; or that it does not comprise TTAGGG alone or that the repeats are degenerate. Provocatively, the mouse with this ~12 kb variant was the F1 in which a new TRA length variant of 140 kb was seen, raising the possibility that they have arisen on the same chromosome and by the same process (Fig 5a). Therefore, we next tested

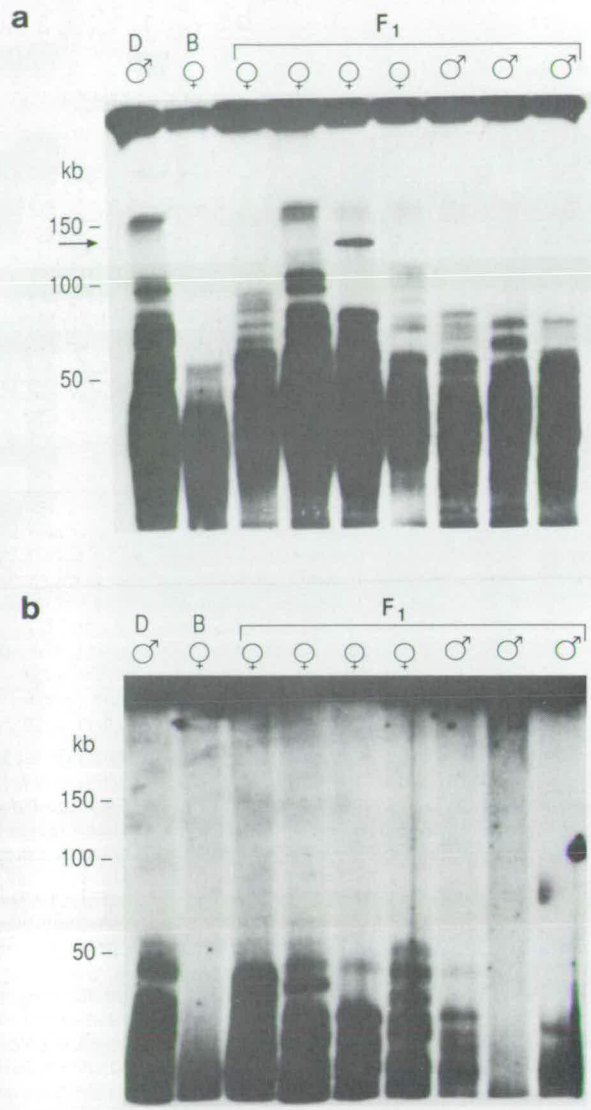


Figure 5. PFGE analysis of TRAs following digestion with (a) HaeIII and (b) MnlI. The C57BL/6 and DBA/2 individuals were different from those used previously. Total genomic spleen DNA of the above and their F1 offspring were digested with (a) HaeIII and (b) MnlI. The restricted fragments were separated by PFGE, transferred to a nylon filter and hybridized as described in the materials and methods. Lambda concatemer DNA size markers are indicated (kb).

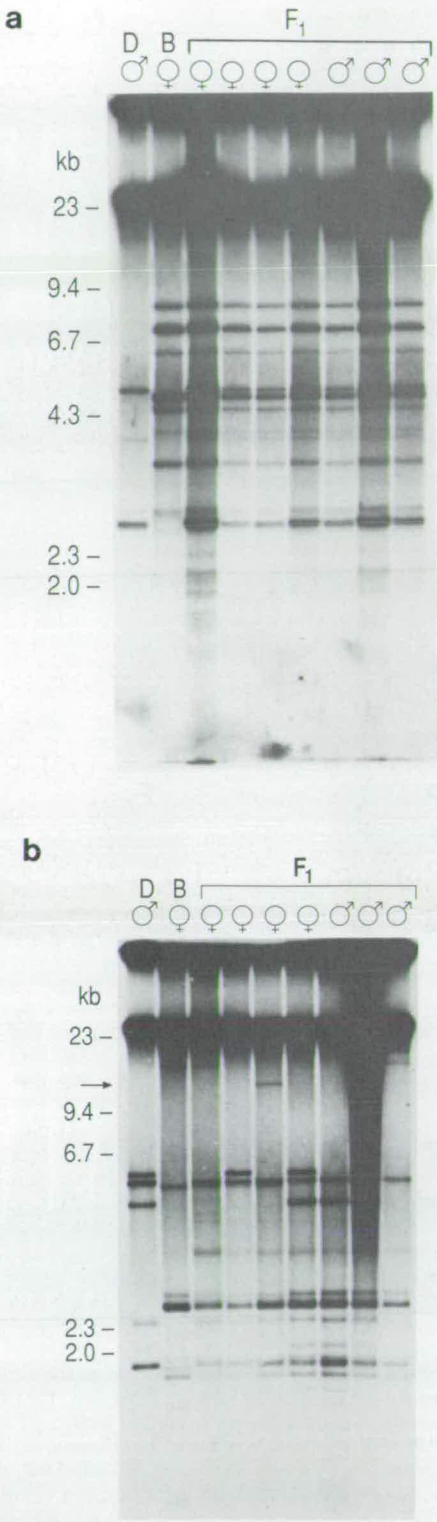
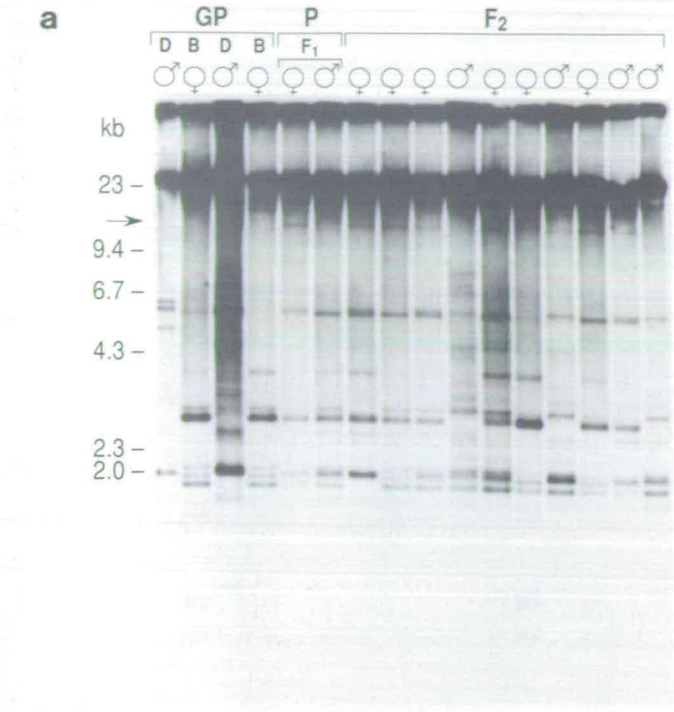


Figure 6. Comparison by conventional gel electrophoresis of TRAs following digestion with (a) HaeIII and (b) MnlI. The same set of genomic DNA's observed in Figure 5 (a) and (b) were used here. The DNA was digested with (a) HaeIII and (b) MnlI. The restricted fragments were separated by conventional gel electrophoresis, transferred to a nylon filter and hybridized with $(T7AGGG)_4$ as described in the materials and methods. HindIII digested Lambda DNA size markers are indicated (kb).

whether the 140 kb TRA length variant and the 12 kb MnlI variant could be co-transmitted through the germ line to the next generation. As shown in Figures 7a and b both these variants were indeed co-transmitted to 7 out of 10 F2 mice. It is not clear whether the signal strength with the 12 kb variants is lower than expected relative to other fragments. Sequencing will be necessary to determine whether this fragment diverges from the $(TTAGGG)_n$ pattern.

The very high rate of variation of the MnlI fragments (>4kb) is likely to be due to their location within the TRA'S. To prove that these fragments do arise from the telomere array, DNA from the F1 showing the variant was first cleaved with HaeIII and then run on a conventional agarose gel (Fig 8a). DNA >23 kb was recovered from the gel and then digested with MnlI. This produced several TTAGGG-hybridising bands which corresponded in size to heterozygous bands revealed by MnlI digestion of unfractionated DNA. Also, the new variant band of ~12 kb was released from the TRA by MnlI digestion but it is obscured by the smear. However, the less variant bands which were homozygous in the parental inbred strains were not



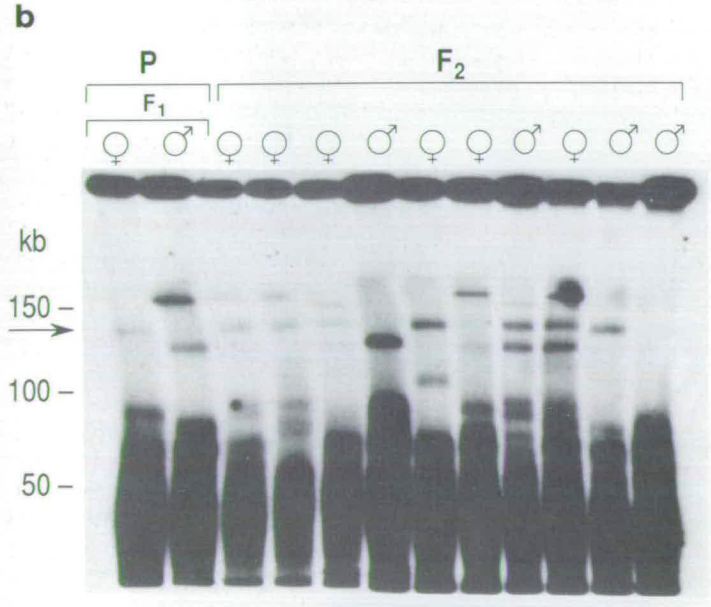


Figure 7. The newly acquired 140 kb TRA length variant and 12 kb MnlI variant are co-transmitted through the germline to the F2 generation. The female F1 mouse showing the variants (Fig 5 and 6) was crossed to another F1 mouse (from a separate C57BL/6×DBA/2 cross), generating 10 offspring. (a) 5 μgs of DNA from the grandparents, the F1's and F2's was digested with MnlI, electrophoresed on conventional agarose gels, blotted and hybridised to a $(TTAGGG)_4$ probe. (b) 10 μg of DNA from the parents an the F2 mice was digested with HaeIII and subjected to PFGE, blotted and hybridised to a $(TTAGGG)_4$ probe. The order of F2 progeny is the same as in (a). GP = grandparents, P = parents.

released from the TRA. To determine if the variant ~12 kb fragment does indeed arise from the variant 140 kb TRA fragment (Fig. 5a) the latter was purified by PFGE and then cleaved with MnlI. As shown in Figure 8b, the only specific fragment released was one of ~12 kb, corresponding to the variant fragment. The rest of the signal was in the unresolved region of >23 kb.

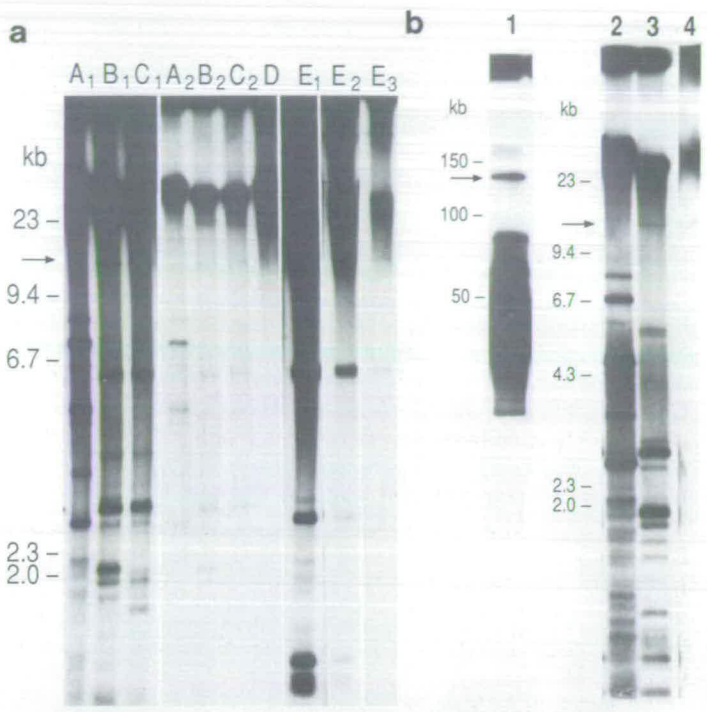


Figure 8. Determination of the origin of MnlI specific variable bands. (a) MnlI specific bands derive from the TRA smear. Total genomic spleen DNA from the F1 female shown in lane 5 of Figures 5a and 6a was initially digested with HaeIII and subjected to conventional gel electrophoresis in 1% LMP agarose gel. The >23 kb telomeric smear at the top of the track was isolated as described in the materials and methods; resultant DNA can be seen in lane D. This was then redigested with MnlI (lanes E1, E2, and E3), and subjected to conventional gel electrophoresis together with genomic DNA of the same individual digested with HaeIII (lanes A1, A2), MnlI (lanes B1 and B2) and HaeIII + MnlI (lanes C1 and C2). The separated fragments were transferred to a nylon filter and hybridized with $(TTAGGG)_4$ as described previously. The repeated lanes show different exposures of the DNA fragments produced by the stated combinations of restriction enzymes. The arrow indicates a new ~12 kb fragment produced by MnlI (see text for further details). HindIII digested lambda DNA size markers are indicated (kb). (b)The ~12 kb MnlI band derives from the 140 kb TRA length variant. Total genomic spleen DNA from the same individual as in (a) was digested with HaeIII and subjected to PFGE in 1% LMP agarose, together with controls to establish the position of the ~140 kb band arrowed (lane 1). This was cut out as a block of gel + DNA and redigested with MnlI as described in the materials and methods. The restriction fragments contained in the plug were separated by conventional gel electrophoresis (lane 4) together with genomic DNA of the same individual digested with HaeIII (lane 2) and MnlI (lane 3). The DNA was then transferred to a nylon filter and hybridized with $(TTAGGG)_4$ as described previously. The ~12 kb band only produced by MnlI and visualized by conventional gel electrophoresis is indicated (see text for further details). Lambda concatemer and HindIII digest size markers are indicated (kb).

DISCUSSION

We have shown that mouse TRA's are extremely long, that individual telomeres can be followed genetically and that new length variants arise at high frequency. They range in size from 30–150 kb, the upper limit being several fold longer than the largest human telomere and 100 fold longer than the TRA's of simple eukaryotes such as *Tetrahymena*. Do the mouse terminal restriction fragments comprise solely TTAGGG repeats or slight variations thereof or could a substantial part of these fragments consist of completely unrelated repeats which also lack sites for most restriction enzymes? A clear understanding will only come from cloning and sequencing of the terminal arrays, which is made difficult by their extreme length. Our data suggests that

mouse telomeres are composed mostly of $(TTAGGG)_n$ repeats from the following observations. Firstly, several lines of evidence support the idea that human telomere repeat arrays include 5–10 kb of monotonous TTAGGG repeats at their distal ends (18, 19, 20, 22). The average size of mouse telomeres is of the order of 40–50 kb, so if the mouse arrays comprise entirely TTAGGG motifs we should expect to see a 5–10 fold increased hybridisation signal for mouse DNA compared to the same amount of human DNA. Our unpublished experiments (JS) suggest that there is an approximately 10 fold increased hybridisation signal using mouse DNA relative to human DNA. Secondly, in the Bal31 experiments shown in Figure 3, the high molecular weight specific bands continue to hybridise to $(TTAGGG)_4$ when 30–40 kb of sequence has been removed; and there is very little reduction in the signal intensity. For the bulk of DNA in the 30–50 kb range the reduction in intensity of hybridisation signal falls off more or less in proportion to the reduction in average size of the fragments. Finally, preliminary Bal31 experiments have shown that the specific variable MnlI fragments which hybridise to $TTAGGG_4$ are not located in the distal 20–30 kb of the TRAs (data not shown). Taken together this data suggests that TTAGGG is found throughout the length of the TRA.

We and others have obtained evidence suggesting that the proximal end of the human TRAs comprise 1–2 kb of degenerate repeats such as $(TTGGGG)_n$ and $(TGAGGG)_n$ (23, 27). Given the length of the mouse TRAs we might expect that they would also have such degenerate arrays and possibly in greater amounts. However, we have not been able to detect hybridisation of radiolabelled $(TTGGGG)_4$ or $(TGAGGG)_4$ probes to mouse TRAs under conditions where they hybridise specifically to human TRAs, (data not shown). Sites for the restriction enzyme MnlI which only requires a single base change within TTAGGG appear to occur every 5–30 kb in the mouse TRA's suggesting little divergence from the consensus repeat. It appears that mouse telomeres are retaining sequence fidelity over much longer regions than those of human. We have no idea why this is the case but it may relate to the frequency of generation of new variants which could maintain these repeats.

It is not clear why mice should have such large telomeres; it does not seem to be essential as two species, *Mus caroli* and *Mus spretus* have TRA's similar in size to those found in humans. It is interesting that all the telomeres of *Mus spretus* appear to be much smaller than those of *Mus musculus*. Is this due to a difference in some aspect of the telomerase apparatus between the two species? It may be possible to address this question through the analyses of crosses between the two species.

Recently it has been shown that there is reduction in TRA length with passage number of human fibroblasts *in vitro* and that cells in a senescent population may lack telomeres at some ends altogether (28). Thus *in vitro*, telomere loss may play a role in senescence, a scenario for which there is evidence in *S. cerevisae* and *Tetrahymena* (29, 30). Some of the mice we have been studying are old in mouse terms, one and a half years, yet they still have TRA's greater than 30 kb in all tissues studied (data not shown). In humans, telomeres shorten with age at a rate of 100 bp per year (31), hence, it is conceivable that the same is happening in the mouse, but the removal of a few 100 bps of terminal DNA during its lifetime would not be detectable.

The most striking aspect of mouse telomeres is their variability. We observed new variants at a minimum rate of 2 in 10 mice in our breeding experiments. If we score this as 2 out of 4 bands clearly analysable in our experiments and assume the rate is the same for all 40 ends, we come to a rate of 1 new variant being generated per end per 20 mice. A similar rate was observed for the specific MnlI fragments contained within the TRA. This is similar to the frequency for the most hypervariable minisatellite described (32). What is the mechanism behind this variation? We must take into account the fact that although the band patterns are different in mice within the same colony and between different DBA colonies, the patterns of the different DBA mice are much more similar to each other than they are to C57BL/6 animals which consistently have fewer bands in the higher molecular weight range (refer to Figure 4). Also new variants are close in size to prexisting bands. This suggests to us that new variants are close in size to the fragments from which they arise; although the new variant of 140 kb shown in Figure 5 must have arisen from a fragment of l00 kb or less suggesting that a 40 kb addition can take place.

Several alternative mechanisms for generating this variation are conceivable. The first could be a recombination/conversion process between sister chromatids or between chromosomes as has been shown for TRA's on plasmids in *S. cerevisae* (15, 16, 17). The very long stretches of TTAGGG may be prone to recombination as proposed by Sen and Gilbert (5). *Tetrahymena* telomeres have nicks and gaps (Reviewed in 1). If such features were found in mouse telomeres it is conceivable that they could promote strand invasion to initiate recombination. A second mechanism could be slippage during conventional replication of the TRA's. A third could be telomere reduction followed by de novo addition by telomerase or vice versa; such expansion and contraction events have been observed for telomeres in *Trypanosoma* and *Tetrahymena* (33, 34). We cannot distinguish between these possibilities. It is interesting though that variants of the MnlI fragments of 4 kb and above within the TRA are arising at about the same frequency as the whole TRA length variants. Elsewhere we will describe how TTAGGG stretches of similar length, but outside and proximal to the TRA generate variants at a frequency of less than 1 in 200 mice. These are typified by the HaeIII bands shown in Figure 6a. It is tempting to speculate that the process which generates variants in the whole TRA simultaneously generates the variants of the MnlI fragments. In fact, we have shown that one MnlI variant of ~12 kb arises within and simultaneously with a 140 kb TRA length variant, supporting this notion. There is only a 1 in 20 probability that both these should arise together independently. In fact the chance of any one of the MnlI fragments of 5–10 kb generating a new variant is close to that for the TRA's, supporting the idea that they arise by the same process. Thus, a process which generates a change in the length of a whole TRA may simultaneously alter the size of an MnlI fragment which makes up only 10% of the TRA length. Again, this could arise through recombination, de novo addition of repeats or a combination of both. Additional experiments will be required to distinguish between these possibilities.

ACKNOWLEDGEMENTS

We thank Professor H.J. Evans for his encouragement throughout the course of this work; D. Stewart, S. Bruce and N. Davidson for high quality photography; E. Kay, L. Campbell and A. Wardrope for excellent secretarial work under difficult conditions. Also we thank D. Chambers for synthesising the oligonucleotide

and the staff of the animal unit for their help. J. Starling is supported by an MRC studentship.

REFERENCES

1. Blackburn,E.H. and Szostak,J.W. (1984) *Annu. Rev. Biochem.* **53**, 163–194.
2. Zakian,V.A. (1989) *Annu. Rev. Genet.* **23**, 579–604.
3. Hastie,N.D. and Allshire,R.C. (1989) *Trends in Genetics* **5**, 326–331.
4. Henderson,E., Hardin,C.C., Walk,S.K., Tinoco,Jr.,I. and Blackburn,E.H. (1987) *Cell* **51**, 899–908.
5. Sen,D. and Gilbert,W. (1988) *Nature* **334**, 364–366.
6. Williamson,J.R., Raghuraman,M.K. and Cech,T.R. (1989) *Cell* **59**, 871–880.
7. Sundquist,W.I. and Klug,A. (1989) *Nature* **342**, 825–829.
8. Greider,C.W. and Blackburn,E.H. (1985) *Cell* **43**, 405–413.
9. Greider,C.W. and Blackburn,E.H. (1987) *Cell* **51**, 887–898.
10. Greider,C.W. and Blackburn,E.H. (1989) *Nature* **337**, 331–337.
11. Shippen-Lenz,D. and Blackburn,E.H. (1989) *Mol. Cell. Biol.* **9**, 2761–2764.
12. Shippen-Lenz,D. and Blackburn,E.H. (1990) *Science* **247**, 546–552.
13. Zahler,A.M. and Prescott,D.M. (1988) *Nucl. Acids Res.* **16**, 6953–6972.
14. Morin,G.B. (1989) *Cell* **59**, 521–529.
15. Dunn,B., Szauter,P., Pardue,M.L. and Szostak,J.W. (1984) *Cell* **39**, 191–201.
16. Pluta,A.F. and Zakian,V.A. (1989) *Nature* **337**, 429–433.
17. Wong,S. and Zakian,V.A. (1990) *Nature* **345**, 456–458.
18. Allshire,R.C., Gosden,J.R., Cross,S.H., Cranston,G., Rout,D., Sugawara,N., Szostak,J.W., Fantes,P.A. and Hastie, N.D. (1988) *Nature* **332**, 656–659.
19. Moyzis,R.K., Buckingham,J.M., Cram,L.S., Dani,M., Deaven,L.L., Jones,M.D., Meyne,J., Ratlier,R.L. and Wu,J.-R. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 6622–6626.
20. De Lange,T., Shine,L., Myers,R.M., Cox,D.R., Naylor, S.L., Killery,A.M. and Vannus,H.E. (1990) *Mol. Cell. Biol.* **10**, 518–527.
21. Cross,S.H., Allshire,R.C., McKay,S., McGill,N. and Cooke,H.J. (1989) *Nature* **338**, 771–774.
22. Brown,W.R.A. (1989) *Nature* **338**, 774–776.
23. Allshire,R.C., Dempster,M. and Hastie,N.D. (1989) *Nucl. Acids Res.* **17**, 4611–4627.
24. Hill,R.E., Shaw,P.H., Barth,R.K. and Hastie,N.D. (1985) *Mol. Cell. Biol.* **5**, 2114–2122.
25. Pietras,D.F., Bennett,K.L., Siracusa,L.D., Woodworth-Gutai,M., Chapman,V.M., Gross,K.W., Kane-Haas,C. and Hastie, N.D. (1983) *Nucl. Acids Res.* **11**, 6965–6983.
26. Church,G.M. and Gilbert,W. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 1991–1995.
27. Brown,W.R.A., MacKinnon,P.J., Villasante,A., Spurr,N., Buckle,V.J. and Dobson,M.J. (1990) *Cell* **63**,119–132.
28. Harley,C.B., Futcher,A.B. and Greider,C.W. (1990) *Nature* **345**, 458–460.
29. Lundblad,V. and Szostak,J.W. (1989) *Cell* **57**, 633–643.
30. Yu,G.L., Bradley,J.D., Attardi,L.D. and Blackburn,E.H. (1990) *Nature* **344**, 126–132.
31. Hastie,N.D., Dempster,M., Dunlop,M.G., Thompson,A.M., Green,D.K. and Allshire,R.C. (1990) *Nature* **346**, 866–868.
32. Kelly,R., Bulfield,G., Collick,A., Gibbs,M. and Jeffreys,A.J. (1989) *Genomics* **5**, 844–856.
33. Bernards,A., Michels,P.A.M., Lincke,C.R. and Borst,P. (1983) *Nature* **303**, 592–597.
34. Larson,D.D., Sprangler,E.A. and Blackburn,E.H. (1987) *Cell* **50**, 477–483.