THE CONSTRUCTION OF VIRAL-BASED LINEAR VECTORS FOR MAMMALIAN CELLS

VASKER BHATTACHERJEE

PhD
UNIVERSITY OF EDINBURGH
1993

ABSTRACT

Recently, the DNA sequences from telomeres, the ends of chromosomes, were isolated from human cells. Previously in yeast and ciliate protozoa, cloned telomeric DNA fragments were shown to have telomeric function *in vivo* when reintroduced on linear vectors. The aim of this work was to test the replication and maintenance in immortalized human cells of a linear DNA vector which had been capped with human telomeric sequences.

The linear vector was based upon a pre-existing circular plasmid which contained replication sequences from Epstein-Barr virus, p220.2. p220.2 was modified by the insertion of two telomeric DNA fragments in opposing orientations separated by a selectable "stuffer" fragment (the kanamycin resistance marker). This construct which contained two potential internal telomeres was named p220.2TC. Excision of the stuffer fragment prior to transfection exposed the termini of the telomeric sequences to create the linear vector, pET which terminated in telomeric DNA sequences. A second linear molecule, pIT, was obtained from p220.2TC as a control by cleaving the plasmid in the ampicillin resistance marker leaving the telomeric DNA fragments in an internal position. pIT was designed as a negative control for telomere function.

p220.2, p220.2TC, pET and pIT were transfected into human Raji cells by electroporation. Southern blot analysis of the DNA from transfected cells showed that the vectors were not rearranged by this method.

Raji cells were selected as the host cell line because these cells constitutively express an EBV gene, *EBNA-1* (also present on p220.2) which is essential for replication of the plasmid. Initial experiments apparently showed that p220.2 successfully replicated in Raji cells. Southern blot analysis of both Hirt-extracted DNA and extracts of total genomic DNA from the transfected cells showed that p220.2 was retained in both the short and the long term. p220.2 also became sensitive to *Mbo*I indicating that the vector DNA had lost the bacterial methylation pattern through replication in the mammalian cell. The copy number of p220.2 was calculated to be 10-25/cell.

Subsequent experiments did not find evidence that p220.2TC replicates in Raji cells as an episome. The Southern blot analysis showed progressive loss of p220.2TC, pET and pIT from the cells without replication taking place. Additionally, p220.2TC could not be rescued in *E.coli*. In these same experiments, however, the control plasmid p220.2 failed to replicate as it did in the initial tests on the Raji cells. Despite many attempts using different experimental conditions, replication or retention of p220.2TC was not detected. Elongation of the telomeric sequences on pET was also not detected.

Towards the end of the project, plasmid rescue experiments were carried out on DNA samples extracted from the cell line which had been successfully transfected with p220.2. Although bacterially replicated plasmids could be retrieved from Hirt DNA samples taken as late as 53 days after electroporation, endogenously replicated plasmids which were resistant to *Dpn*I enzyme could not be detected. This may have been due to lack of sensitivity of the assay.

This work was originally undertaken in order to construct a stable linear vector for use in mammalian cells and, possibly, the testing and study of putative centromere sequences. Owing to the inconsistent replication of the original p220.2 vector, it may be concluded that a different approach should be considered in the study of the reintroduction of cloned telomeres and centromeric sequences. The transfection experiments are discussed along with some of the alternative lines of research in this areas.

ABBREVIATIONS

A	Adenine
abs	absorbance
Ap	Ampicillin
ARS	Autonomously replicating sequence
BAP	Bacterial alkaline phosphatase
bp	base pair
BSA	Bovine serum albumin
C	Cytosine
cpm	counts per minute
dam$^+$	DNA adenosine methylase positive
dATP	Deoxyadenosine triphosphate
dCTP	Deoxycytidine triphosphate
dGTP	Deoxyguanosine triphosphate
dTTP	Deoxythymidine triphosphate
dNTP	Deoxynucleotide triphosphate
DMSO	Dimethylsulphoxide
DNA	Deoxyribonucleic acid
ds	Double stranded
DTT	Dithiothreitol
EBNA-1	EBV nuclear antigen 1
EBV	Epstein Barr virus
EDTA	Ethylenediaminetetraacetic acid
FCS	Foetal Calf serum
G	Guanine
HmB	Hygromycin B
hph	hygromycin B phosphatase resistance gene
Kan	Kanamycin

kb	kilobase pairs
KCl	Potassium chloride
kDa	kiloDaltons
lHHR	λ DNA digested with *Hin*dIII plus l digested with *Eco*RI and *Hin*dIII
μF	microFarad
$MgCl_2$	Magnesium chloride
$MgSO_4$	Magnesium sulphate
min	minutes
NaCl	Sodium chloride
p	Short chromosome arm
PBS	Phosphate buffered saline
pET	Linear vector with exposed telomeric sequences (p220.2ET)
pIT	Linear vector with internal telomeric sequences (p220.2IT)
q	Long chromosome arm
RNA	Ribonucleic acid
rpm	revolutions per minute
SDS	Sodium dodecyl sulphate
ss	Single-stranded
SSC	Sodium saline citrate
T	Thymidine
TE	10mM Tris, 1mM EDTA pH 8.0
Tris	Tris (hydroxymethyl) aminomethan
UV	Ultraviolet
vol	Volume
YAC	Yeast artificial chromosome

ACKNOWLEDGEMENTS

I owe a debt of gratitude to members of all the MRC unit's sections for help I have received during my time here. My thanks to you all. Some of you deserve a special mention, however.

Many thanks to my section, Structural Studies (both past and present members) for all their help and encouragement, but the following, in particular: Howard for his endless patience, especially when waiting for me to hang up my Gilsons; Helen and Sylvia for their irrepressible cheerfulness (keep the faith, sisters!); Stewart for his technical help; Dave K for his friendship (despite everything!) and his thorough criticism of my thesis; Mary for almost always speaking to me in Scots instead of English and Elaine (Imelda) for reminding me that every silver lining comes with its own cloud.

This thesis could not have been brought to you without the considerable help of Sandy, Norman and Douglas in Photography. Many thanks as well to John Ireland along with other members of the "computer competent" who helped this Luddite operate the PCs and Sun workstations (text by Word for Windows 2, drawings by IslandDraw, Xenon2 Megablast by the Bitmap Brothers!). I am also grateful to Sheila for running such an efficient library service. I would like to acknowledge the financial support of the MRC Human Genome Mapping Program (as well as my numerous lodgers) during my studentship.

I am grateful to all the unsuspecting punters who sat through my seminar on the evolution of cichlid fish in the African Rift Valley lakes and actually enjoyed it.

On a personal note, I would like to thank the following people: first and foremost, tall Paul (the man who knows all) for "The Days That Used To Be"; Barbie, Isobel (I'll always know you as Isy, though!) and Kathy O for keeping my spirits up and my ego under restraint; and Ian for teaching me (though he may not have known it) to recognise my limits, but always to push to them. Many thanks as well to the Emaresea (pronounced MRC!) hockey team for keeping me in the squad despite being the lowest scoring centre-forward in their history.

Finally, I am grateful to all my flatmates, past and present, for their support and tolerance: Wendy (who took pity on a homeless vagabond); Paula; Robin (my absent landlord); Viv ("Red hair on black leather - my favourite colour scheme!"), Justin (who finally got his bookcase back); Maria (from Manchester) and latest, though probably not the last, Rebecca "Dancing Queen" Devon.

This thesis is dedicated to my parents, without whom I would not be here today.

I learnt a lot during the last three and a half years, only a fraction of which concerns science. Therefore, I feel that it is suitable to quote the following sentiments of a certain Mr Ian Kilminster, a rock musician.

"After an opposable thumb, the second greatest gift we've been given is a brain that has the ability to experience and learn. If you can't gather that into yourself then you're worth nothing, your life's wasted......" (Lemmy - Motorhead)

"Damn, I got hit, but I still complete the mission!"

THE CONSTRUCTION OF VIRAL-BASED LINEAR VECTORS FOR MAMMALIAN CELLS

CONTENTS

PUBLICATIONS

CHAPTER 1 INTRODUCTION

INTRODUCTION TO LINEAR VECTORS

The aim of this project was to construct a linear vector for mammalian cells to test the viability of reintroduced cloned telomeric sequences as functional telomeres.

The project involved the adaptation of an existing circular vector which was reportedly competent to replicate in human cells. Human telomeric sequences were inserted which were tested for function as telomeres once the construct was converted to a linear form. The vector was introduced in a variety of forms into human immortalized cells: circular, with and without internal telomeric sequences; and linear, with and without terminal telomeric sequences. The construction of these four vectors is described and their structure is illustrated in section 3.1. The fate of the DNA molecules was examined after transfection in order to determine 1) if the vectors were replicating, 2) if the vectors were retained within the cell in the long term, 3) if the structure was stable and 4) if the telomeres could prevent integration of the vector into the genome.

A future aim of the work is the isolation of centromeres from mammalian cells by the insertion of putative centromeric sequences into a linear vector. In yeast, the presence of a linked centromere causes a plasmid to segregate 1:1 to daughter cells and to be maintained at a single copy (Clarke and Carbon, 1980). Although centromeres could also be isolated on circular plasmids, the linear vector approach might be easier to screen. The existing circular vectors which have been developed for mammalian cells would be expected to be quite stable whether or not they contained functional centromeres. However, on the basis of work carried out in yeast minichromosomes (Murray and Szostak, 1983) we can speculate that a linearized vector even when equipped with telomeres would not exhibit a high degree of stability unless a centromere was present. This increase in stability might greatly facilitate the screening for such a molecule. Another reason for using linear vectors concerns the probable size of a putative cloned centromere (Willard, 1990). Large fragments of DNA already cloned in yeast artificial chromosomes (YACs) could be suitable for the testing of centromeric function and it may be technically more difficult to convert a YAC into a circular form than to convert it into a linear vector.

Although plasmid-based vectors already exist for mammalian cells (Sambrook *et al.*, 1989), there are advantages to developing artificial chromosomes as vectors. 1) Most of the existing vectors are restricted to specific cell types especially if they utilise viral replication genes (see section 1.1). A linear vector/mini-chromosome might have a wide range of host cell types which might enable its use in gene therapy. 2) It is possible to clone groups of related genes which are positioned sequentially in the genome (such as the Hox developmental genes) on YACs. The reintroduction of the entire YAC as part of a mammalian linear vector may allow the study of the long range control of gene expression. 3) A stable linear vector might support long term expression of a gene/group of genes as opposed to a transient assay.

1.1 VECTORS FOR MAMMALIAN CELLS

Most DNA vectors which are used in mammalian cells have been designed primarily to produce high levels of transient expression of a specific protein in an immortalized cell line. Vectors which produce transient replication are either lost from the cells by dilution over successive cell divisions, or they replicate themselves to such a high copy number that the host cells can no longer support them and die. After this time, the only expression is obtained from cells in which the sequences have integrated into the host genome (Sambrook *et al.*, 1989).

To achieve this end, even prokaryotic plasmids carrying transcription units but with no form of eukaryotic replication origin have been used. Since prokaryotic origins of replication do not work in eukaryotic cells (Biamonti *et al.*, 1985)), bacterial plasmids must integrate into the host genome if long term maintenance and expression of the introduced gene is to occur. Expression of a gene after integration into the host genome can be obtained, but the level can be variable owing to position effects in the chromosome (Henikoff, 1990). If greater control of the expression is required, a eukaryotic replication origin should be included in the vector.

Attempts have been made to bypass these limitations by resorting to viral DNA replication sequences. Viral infections of mammalian cells are virtually the only situations where non-host origins of replication function in the higher eukaryotes (Diffley and Stillman, 1990). Somatic cell hybrids could be regarded as exceptions, since a chromosome from one mammalian type, often human, can exist inside the nucleus of another, for instance rodent. However, the foreign chromosome used in a somatic cell hybrid may be similar enough to the host to be considered as being the same (Weiss and Green, 1967). Plasmids which utilise viral sequences are usually bacterial shuttle vectors, containing a bacterial origin of replication and antibiotic resistance gene. The benefit of bacterial shuttle vectors is the ease with which plasmids can be manipulated and modified in *E.coli.* Large amounts of highly purified vector DNA, which is a prerequisite for transfecting mammalian cells, can also be obtained from bacteria by standard methods (Sambrook *et al.*, 1989).

A viral plasmid was chosen as the basis from which to construct a linear vector in this project. Since I wish to test reintroduced telomeres in the vector, it is essential that the vector will replicate extrachromosomally inside the mammalian cell. Although mammalian DNA sequences have been isolated which act as origins of replication, notably the dihydrofolate reductase locus, these only function as part of chromosomes and have not yet been shown to act as autonomously replicating sequence elements with the exception of the c-*myc* locus (see section 1.2). Viral replication sequences have, however, been cloned which promote extrachromosomal replication in mammalian cells.

1.1.1 SV40 Vectors

A common expression system which is in use involves simian virus 40 replication sequences. The SV40 genome replicates as a 5.2kb double-stranded circle of which the viral origin, the coding sequence for the large T-antigen and the T-antigen binding sites are required for vector replication. The expression of the SV40 large T-antigen is essential for activating the origin of replication (McKnight and Tijan, 1986). Although expression of cloned genes from SV40 promoters is obtained at low to moderate levels in a wide range of mammalian cells, the origin of replication only functions in simian cells. The highest levels of replication and expression are obtained from SV40 vectors when used in immortalized cell lines, such as COS cells (monkey *C*V-1 cells transformed with the *o*rigin of *S*V40), where the large T-antigen is already constitutively expressed from the host chromosomes (Gluzman, 1981; Mellon *et al.*, 1981). The SV40 vectors are mainly used to provide transient expression of a protein since a major drawback of using SV40 vectors is the short lifetime of the transfected cells. When the origin is working efficiently, unchecked replication of the transfected plasmids occurs. Cells which allow high levels of replication die at 70-90 hours post-transfection, presumably because they cannot tolerate the high levels of replicating DNA (Mellon *et al.*, 1981; DuBridge *et al.*, 1987).

1.1.2 Epstein-Barr Virus Vectors

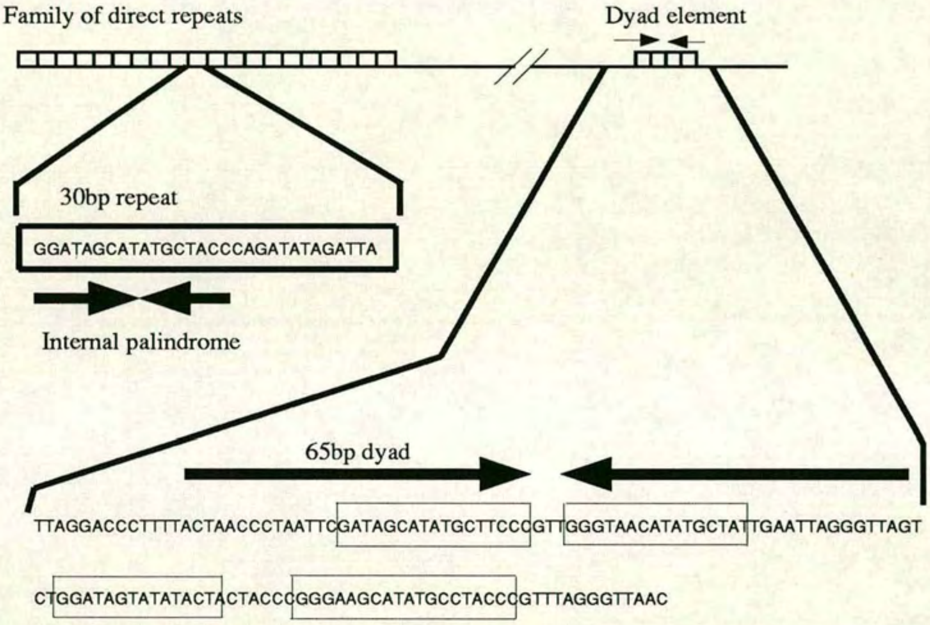
Other viral-based vectors have been developed which can replicate and express genes at a moderate level over long term periods instead of transiently. Prime examples are the vectors based upon the replication sequences of Epstein-Barr virus (a member of the herpesvirus family).

EBV infects human B-lymphocytes and can transform them into immortal blast cells. The EBV genome is a DNA 172kb circle which replicates extrachromosomally once per cell cycle and is maintained in multiple copies in transformed cells as well as integrated into the host genome (Lindahl *et al.*, 1976). (During mitosis, the extrachromosomal copies are not evenly segregated between the daughter cells leading to multiple copies in certain cells.) The essential sequences from the EBV genome which are required for replication have been isolated and cloned into plasmids. These sequences are analogous to those of SV40 in that a viral antigenic protein, EBNA-1, is required in addition to the origin of replication, *oriP*. EBNA-1 may activate both viral DNA transcription and replication by binding to specific sites in the EBV genome in an analogous way to SV40 large T antigen (Yates *et al.*, 1984 and 1985).

The *oriP* consists of two parts, a family of direct repeats and a region of dyad symmetry (Reisman *et al.*, 1985) (see figure 1.1). The family of repeats consists of 17 imperfect copies of a 30bp sequence with an internal palindrome plus another three truncated copies of the repeat. At about 1kb from the direct repeats lies the region of dyad symmetry. Four partial copies of the repeat are present in the dyad which is a 65bp long palindrome, part of which can form a 31bp stem and 3bp loop structure (Gahn and Schildkraut, 1989). The replication origins of the viruses SV40 and

Figure 1.1

Structure of the EBV origin of replication oriP



the dyad symmetry element contains four copies of the core repeat sequence from the family of repeats (boxed) cytomegalovirus also have sequences containing repeats and palindromes, so this may be a conserved feature through evolution (Tijan, 1978; Masse *et al.*, 1992).

The *EBNA-1* sequence codes for the only viral replication protein which is expressed during the lytic phase of the infection. The protein binds to each copy of the 30bp repeat in *oriP* (Rawlins *et al.*, 1985). Although it is a necessary factor for nuclear retention of the EBV genome and EBV-based plasmids (Yates *et al.*, 1985), the biological interaction of EBNA-1 with the human cell is unknown (Krysan *et al.*, 1989). Possibly, the protein binds to the nuclear membrane or to the nuclear matrix (Jankelevich *et al.*, 1992).

Unlike SV40 vectors, EBV-based vectors are not normally used as transient expression systems (Teshigarawa and Katsura, 1992). Far higher levels of expression are obtained using SV40 vectors in COS cells which is an established system. EBV vectors might be used, however, if a human cell line is necessary to provide the expression. Usually they are used to establish cell lines that contain multiple episomal (extrachromosomal) copies of the introduced gene (Sambrook *et al.*, 1989). These vectors are maintained at a lower copy number than SV40 vectors so the health of transfected cells is not affected in the long term (Sugden *et al.*, 1989). EBV vectors are also used in the study of DNA replication. Often, EBV vectors are used in cell lines which express the EBNA-1 protein constitutively from the host genome such as Raji cells because the vectors replicate efficiently in this background. Some research has shown that vectors which contain both the *oriP* and *EBNA-1* sequences can replicate extrachromosomally in a human cell line which does not already express the viral protein (Krysan *et al.*, 1989). EBV vectors are usually used in human cell lines but they have been reported to replicate in monkey and dog cells as well (although not rodent cells) (Yates *et al.*, 1985).

Vectors carrying the *oriP* and *EBNA-1* sequences should be retained within the nucleus. Origins of replication from yeast and from prokaryotes were isolated by inserting genomic sequences into a bacterial-type plasmid along with a selectable marker (Yasuda and Hirota, 1977; Stinchcomb *et al.*, 1979). This strategy originally failed in the isolation of mammalian ARSs perhaps because yeast undergo closed mitosis while mammalian cells undergo open mitosis (Biamonti *et al.*, 1985). (In yeast, the nuclear envelope does not break down during mitosis, so all

DNA molecules are retained within the nucleus whether or not they interact with the mitotic spindles. In mammalian cells, the nuclear envelope does break down, therefore any DNA not linked to a centromere can be lost to the cytoplasm). The EBNA-1 protein may interact with the proteins of the nuclear envelope or matrix. One hypothesis is that when the nucleus is reconstituted at the end of mitosis, the protein is "dragged back" bringing the vector DNA to which it is bound with it. Krysan *et al.* (1989) were able to use the essential replication sequences of EBV to isolate a putative human ARS element on a replication-defective circular mammalian shuttle vector. By deleting the dyad region and leaving the *EBNA-1* gene intact, a human genomic fragment was isolated which apparently fulfilled the replication function. This approach could feasibly be extended to the isolation of a human centromere. However, the deletion of an EBV sequence may not be necessary. The *EBNA-1* gene can maintain a vector in the mammalian cell, but it does not act like a centromere. EBV vectors are maintained at about ten copies per cell while a centromere maintains a single copy (Sugden *et al.*, 1985). Presumably, if an EBV vector contained a functional centromeric region, it would decrease in copy number. A linear EBV vector might also increase in stability if it contained a centromere.

1.1.3 Yeast artificial chromosomes (YACs)

YACs are eukaryotic vectors which have revolutionized the cloning of DNA from the genomes of mammals and higher plants. Since their development by David Burke *et al.* (1987), YACs have provided a link between the genetic linkage maps and the physical maps of mammalian chromosomes (reviewed in Schlessinger, 1990).

In general, fragments of genomic DNA of hundreds of kilobases which have been generated by cleavage are sealed between two vector arms. One of the vector arms contains a *Saccharomyces cerevisiae* ARS and a centromere, but both contain a selectable marker and end in a telomeric sequence. A bacterial origin of replication and selectable marker are included in the vector sequences. This allows the vector arms to replicate in bacteria as a plasmid (the telomeres are separated by a stuffer fragment before the genomic DNA is cloned into it).

In yeast, the YAC is a relatively stable linear molecule. The vector sequences total about 10kb, but the genomic DNA insert can approach 1Mb in size. In practice, the stability of YACs increases with the size of the DNA insert being carried (Murray and Szostak, 1983).

1.1.4 Requirements for a mammalian artificial chromosome or linear vector

The essential elements of a YAC are the telomeres, a centromere, selectable markers and an autonomously replicating sequence (ARS). Although *S.cerevisiae* is far removed in evolution from the mammalian cell, the YAC still provides us with a framework around which to construct a mammalian artificial chromosome. YACs were, in fact, used to isolate telomeres from human cells by a "half-YAC" cloning procedure. One vector arm of an existing YAC vector was directly ligated to fragments of mammalian DNA without a second arm being present. Yeast cells transformed with the ligated DNA were found to contain some YACs where a human (TTAGGG)n sequence had been the substrate for the addition of yeast TG_{1-3} repeats. By this criterion, the human repeats were considered to have provided telomeric function in yeast (Brown, 1989; Cross *et al.*, 1989; Reithman *et al.*, 1989).

S.cerevisiae-based YACs do not replicate extrachromosomally in mammalian cells. Usually, they integrate into the genome and often rearrange (Huxley and Gnirke, 1991). The reasons for the inability of YACs to survive extrachromosomally in mammalian cells are open to conjecture. For instance, the telomeres of the YAC may not be recognized by cellular proteins which might specifically interact with the telomeres in order to prevent integration into the host genome. It is possible that the mammalian equivalents of the yeast ARS, centromere and telomere elements are required for extrachromosomal replication and survival of such artificial chromosomes.

1.2 AUTONOMOUSLY REPLICATING SEQUENCE (ARS) ELEMENTS

An ARS may be defined as a DNA element which will promote extrachromosomal replication of a plasmid or linear vector. A proposed mammalian linear vector must contain a DNA sequence with this function, or it will eventually be diluted out of a growing population of cells.

Among eukaryotes, the best described ARS elements are those of *S.cerevisiae* and of the viruses, notably SV40. In these cases, replication is initiated at specific origin sequences which are recognized by initiator proteins (either cellular, or in the case of SV40 T-antigen, encoded by the virus). The ARS must provide a region where the DNA duplex can be unwound so that the elongation complex can form. The elongation complex of the cell consists of DNA helicases (which unwind the DNA), DNA primase, DNA polymerases and accessory proteins (Diffley and Stillman, 1990).

The yeast ARS elements were isolated as sequences which promoted extrachromosomal replication of plasmids carrying suitable selectable markers in *S.cerevisiae*. These ARS elements were demonstrated to act as origins of replication on plasmids by Brewer and Fangman (1987) using two-dimensional electrophoresis. The yeast chromosomal origins of replication have been subsequently analysed and have also been found to act as ARS elements on plasmids. However, all yeast origins of replication are ARSs but not all ARSs initiate replication on the chromosome (Fangman and Brewer, 1991). Two generalizations can be made about all *S.cerevisiae* ARS elements. First, they all contain an AT-rich 11bp consensus sequence [A/T]TTTATRTTT[A/T] (where R is a purine) which is essential for ARS function. Secondly, a flanking sequence, positioned 3' to the T-rich strand and ranging from 16bp to over 100bp, is also required for full ARS activity (Newlon, 1988; Umek *et al.*, 1989). Several groups of researchers have now isolated proteins which bind to the ARS consensus sequence (ACS). Firstly, three groups have described a 67kDa single-stranded binding protein which binds to the T-rich strand of the ACS (Hofmann and Gasser, 1991; Schmidt *et al.*, 1991; Kuno *et al.*, 1990). Secondly, Bell and Stillman (1992) have isolated a complex of six polypeptides that bind specifically to the ACS in its double-stranded form. The binding of this complex, which has been termed the ORC (origin recognition complex), is ATP-dependent, although as yet, DNA helicase, protein kinase, or ATPase activities have not been detected which might explain this requirement. The affinity of binding *in vitro* of both the single-stranded binding protein and the ORC is diminished by point mutations in the ACS which also eliminate or reduce ARS function *in vivo*. Diffley and Cocker (1992) have found supporting evidence that the ORC binds origins *in vivo*. This evidence has caused speculation that these proteins are involved in the initiation of DNA replication and that the ORC represents the initiator protein complex of yeast. The 67kDa single-stranded binding protein may be a secondary factor which binds the DNA after it has been unwound by the ORC.

Viral origins of replication have been briefly described above. Replication of SV40 has been studied in depth as a model for the initiation of replication in mammalian cells since the host elongation complex is utilized, but in an extrachromosomal context (Diffley and Stillman, 1990). SV40 and EBV both begin the replication of their genomes by the binding of virally-encoded initiator proteins (the large T antigen and EBNA-1 respectively) to specific sites. The T antigen binds to a set of pentanucleotides in the SV40 origin and unwinds the region called the early palindrome. Additionally, the T antigen also induces an ATP- and temperature-dependant structural deformation in an AT-rich part of the origin (Borowiec *et al.*, 1990). Initiation of replication in EBV occurs in an analogous fashion. The EBNA-1 protein binds to the direct repeats of the *oriP* region (illustrated in figure 1.1) region and initiates replication in the dyad element, presumably by causing a localized unwinding (Gahn and Schildkraut, 1989). Once regions of DNA have been unwound, host elongation complexes can form at the sites and proceed with replication.

The evidence gathered to date on mammalian chromosomal origins of DNA replication has shown significant differences from those of yeast and SV40. The mammalian genome ($\sim 6 \times 10^9$bp) is much larger than those of prokaryotes ($\sim 4 \times 10^6$bp for the *E.coli* chromosome) and viruses ($\sim 1.7 \times 10^5$bp for the EBV genome). If DNA replication was initiated from one origin site, the cell cycle would be unfeasibly long. Calculations made on the basis of the size of the genome, the length of the DNA synthesis or S phase of the cell cycle and the activity of DNA polymerase predict that origins of replication occur once every 100kb in the human genome. This frequency was, indeed, visualized by Huberman and Riggs (1968) who studied stretched eukaryotic chromosomes which had been labelled with (H^3) thymidine. The DNA fibres were too small to be seen directly, but exposure to photographic emulsion produced silver grain tracks which could be seen by light microscopy. From studying the images of the replication forks, they drew several conclusions: (1) each chromosome fibre contains multiple growing points along its length; (2) most replication forks arise from bidirectional origins of replication; (3) initiation sites are spaced, on average, about

100kb apart; (4) replication forks move at 3kb/min (slower than in bacteria); (5) clusters of 5-10 adjacent origins initiate simultaneously. Although these observations were informative, they did not reveal whether replication is initiated at specific sequences analogous to those of bacteria and viruses. A full discussion of the initiation of chromosomal DNA replication is outside the scope of this work, but a few examples are briefly described below.

The best characterized chromosomal origin of replication in mammalian cells is a 28kb region downstream of the Chinese hamster dihydrofolate reductase (DHFR) gene. One study has shown that replication initiates at many sites throughout this 28kb region (Vaughn *et al.*, 1990), while another demonstrates the formation of a single bidirectional replication bubble within a specific sequence of 450bp (Burhans *et al.*, 1990). One model reconciles these two contrasting observations by proposing that many bidirectional replication origins form in the region and meet and fuse to create a single bubble covering many kilobases. Replication then proceeds from this region in two directions (Linskens and Huberman, 1990).

A putative origin which has been studied is the locus of the c-*myc* proto-oncogene. Vassilev and Johnson (1990) mapped an origin to a 2kb region centred 1.5kb upstream of the promoter of the c-*myc* gene. However, the c-*myc* locus also acts as an origin on extrachromosomally replicating plasmids. This activity is apparently dependent on the binding of the c-*myc* protein itself to the c-*myc* locus. The c-*myc* protein also promotes the initiation of DNA replication at other sites where it binds specifically (Iguchi-Ariga *et al.*, 1987 and 1988; Ariga et al., 1989). Most of the research done on c-*myc* has concerned its role as a transcription factor, but its activation of origins of replication is similar to other proteins. The SV40 T-antigen, for instance, also acts as a transcriptional enhancer of the viral genome (DePamphilis, 1988). It is tempting to hypothesize that the c-*myc* protein has a similar biological function to SV40 T-antigen in transcriptional activation and the initiation of DNA replication.

The laboratory of Michele Calos has isolated fragments of human genomic DNA which appear to act as ARS elements in replication-deficient EBV plasmids. Studies using two-dimensional electrophoresis have demonstrated the formation of replication forks and bubbles in these fragments (Krysan and Calos, 1991). Replication initiates at many sites within a given fragment, but also at sites within the vector sequences. However, the initiation of replication does not occur in the absence of linked human DNA. The sizes of the fragments which have been tested for ARS function range from 1kb to 21kb, but the DNA sequences have not yet been published. However, a positive correlation has been found between the size of the DNA fragment and the efficiency of autonomous replication (Heinzel *et al.*, 1991). There is no evidence that these genomic fragments function as origins of replication in the chromosomes.

A generalization that may be made about mammalian origins of replication is that replication does not initiate at a specific origin sequence. However, specific DNA sequences may be required in order to bind initiator proteins. The initiator protein(s) may then direct the assembly of the initiation/elongation complex at any site in the surrounding DNA (within a certain range).

Although the isolation of putative ARSs is an important step in the development of linear vectors, a necessary prerequisite is a linear vector bounded by telomeres which will escape integration into the genome.

1.3 TELOMERES

The main functions of telomeres involve complete replication of the DNA, maintaining stability of chromosomes in the nucleus and, possibly, promoting the pairing of homologous chromosomes during meiosis (Zakian, 1989). Telomeres may also provide binding sites for the chromosomes in the nuclear matrix (De Lange, 1992).

Much of the current knowledge of telomeric structure and replication has been obtained through the study of ciliated protozoa such as *Tetrahymena*, and also the study of yeast chromosomes. It is now known that the general structure of telomeres is conserved throughout the ciliates and in yeast (Zakian, 1989). Further work has shown that the conservation of the telomeric consensus sequence extends into humans (Moyzis *et al.*, 1988) and all vertebrates (Meyne *et al.*, 1989).

Since this vector will require working telomeres, a description is required of their structure, replication and of the fate of cloned telomeres which have been reintroduced into eukaryotic cells on linear vectors.

1.3.1 Function of telomeres

The necessity of telomeres on chromosomes was illustrated by the pioneering work of Muller and McClintock who demonstrated the consequences of the deletion of telomeres. Muller analyzed the chromosomal rearrangements which appeared after X-irradiation of *Drosophila melanogaster* and found that deletions and inversions at the terminus were recovered very rarely (Muller, 1938; Muller and Herskowitz, 1954). Muller argued that a specialized terminal structure which he named a telomere is required for chromosomes to be maintained stably in the nucleus. McClintock investigated, in *Zea mays*, the fate of chromosome ends created by breakage. She studied dicentric chromosomes which can be pulled in opposite directions during meiosis and broken. The broken end was often reactive and fused with other broken ends to create new dicentric chromosomes and continue a "breakage-fusion-bridge" cycle. However, the natural chromosomes which were capped by telomeres were stable and did not fuse with either broken ends or other telomeres (McClintock, 1941 and 1942). Linear plasmids without telomeric sequences which are introduced into yeast are destroyed by a mixture of recombination and exonuclease action in the cell. These early experiments showed that the telomere acts as a protective cap for the chromosome (see Zakian, 1989 for review).

However, some of Muller's conclusions have been contradicted by recent work in *Drosophila* where deletions of the chromosomes at the terminus were recovered. A strain of *Drosophila, mu-2*, has a high mutation rate in its genome. If this strain is X-irradiated, chromosomes which lack telomeres can be found to exist without fusing with other non-telomeric ends. The chromosomes did reduce in size over successive mitotic cycles owing to incomplete end replication. Later work has shown that the ends of the terminally deleted chromosomes have been healed by the addition of a transposable element, HeT-A. No telomeric sequences have been identified for *Drosophila* as yet and the HeT-A repeated element may fulfil this role (Levis *et al.*, 1989; Biessman *et al.*, 1990 and 1992).

As well as possibly preventing fusion and exonucleolytic degradation of the chromosome end, the telomere may also prevent the loss of genetic information during DNA replication (incomplete end replication). Olovnikov first described this problem using the term, marginotomy (1973). DNA replication is carried out by template-dependent DNA polymerase enzymes which only synthesize DNA in the 5' to 3' direction and usually require a RNA primer. The RNA primer is removed by the RNaseH activity of DNA polymerase. The consequence is a gap left at the 5' end of a strand after the excision of the RNA primer. Over successive generations, the gradual whittling away of the chromosome would lead to the eventual loss of genetic information (Watson, 1972; Olovnikov, 1973). The specialized structure of a telomere apparently circumvents this problem. When the work involving the *Drosophila mu*-2 strain, described above, is considered, the prevention of incomplete end replication may be a more realistic function of telomeres than the prevention of fusion into other chromosomes.

The mechanism by which telomeres ensure the full replication of the chromosome is not yet definitely known. There is a great deal of evidence for a process which involves a ribonucleoprotein, telomere terminal transferase, or telomerase, which has been isolated and characterized by Greider and Blackburn (1985, 1987, 1989) and by Morin (1989). Telomerase synthesizes telomeric repeats which are added onto the 3' end of the G-rich strand (see figure 1.3). However, some evidence is available for a mechanism which depends upon recombination between telomeres (Morin and Cech, 1988; Pluta and Zakian, 1988; Wang and Zakian, 1990). These mechanisms will be described in greater detail in section 1.3.4, Replication of telomeres.

Telomere-telomere associations may occur during interphase and mitotic prophase when, in *Drosophila* cells, telomeres are positioned close to the nuclear envelope and 180^{o} away from their respective centromeres (Agard and Sedat, 1983). Other observations have shown that the telomeres are tightly associated with the nuclear matrix (De Lange, 1992). It is not yet clear if the observations made so far are widespread among species.

1.3.2 Structure of telomeres

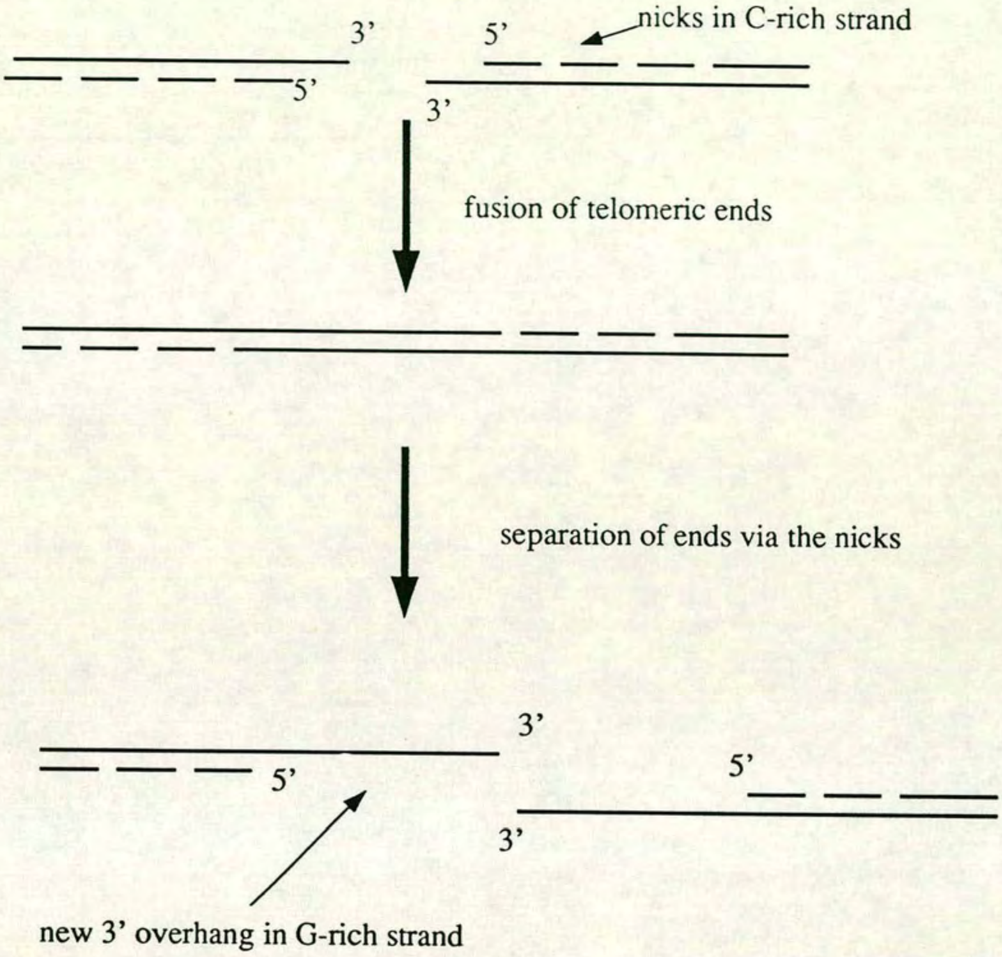
Most eukaryotic telomeres cannot be sequenced directly because of the relative low abundance of telomeric DNA in the nucleus. However, much of the early work on telomeres was carried out in the ciliated protozoa such as *Tetrahymena*. The ciliated protozoa contain two types of nuclei, the micronucleus and the macronucleus. The micronucleus is transcriptionally inactive, but is a structurally conventional diploid nucleus. The macronucleus is formed at the end of a complex pathway which involves chromosome breakage, DNA deletion and rejoining, DNA replication and, most importantly from our point of view, telomere formation (Yao, 1981). These organisms contain linear DNA molecules of a small size (~21kb) at a high copy number in the macronucleus. The biological reason for the formation of the macronucleus may lie in the physical size of ciliate protozoan cells which can approach 3mm. The protozoan cell may require many transcriptionally active copies of certain genes. The protozoan uses the production of high copy number "micro-chromosomes" which only carry the genes which it requires in high doses such as the ribosomal genes. A substantial amount of unique germ-line-specific DNA sequence which is present in the micronucleus is absent from the macronucleus. The main task of the macronuclear genome may be to provide the high rate of transcription that is required from genes involved in the vegetative growth of protozoans of this size. The creation of these small DNA elements is probably more efficient than maintaining polyploidy for its entire genome since DNA not required for vegetative growth is not maintained at a high copy number. The advantage of studying telomeres in the macronuclei of ciliates is the high rate of telomere formation and healing which occurs as well as their high relative abundance as a proportion of the genome (Blackburn and Karrer, 1986).

One of the small DNA molecules present in the *Tetrahymena* macronucleus encodes the ribosomal DNA. These telomeres were the first to be sequenced (Blackburn and Gall, 1978). The studies showed that the rDNA termini carried a tandem repeat of CCCCAA (or TTGGGG on the complementary strand), the number of which varied between individual termini, ranging from 20 to 70 copies. The CCCCAA sequence runs from 5' to 3' towards the centre of the molecule and single base gaps were found at the C-rich strand in the terminal 100 bp.

The nicks in the C-rich strand which are a conserved feature in many species may be part of a mechanism to resolve fused telomeres. If the telomeres of two chromosomes fuse, the opposing C-rich strands will both contain the breaks. The chromosomes should then be able to separate at these breaks during the next mitosis producing overhangs on the 3' end of the G-rich strand in the process (see figure 1.2).

Figure 1.2

Possible mechanism for resolving fused telomeres



Similar tandem repeat sequences were found in the termini of macronuclear molecules of the ciliate *Oxytrichia* (Klobutcher *et al.*, 1981; Pluta *et al.*, 1982). Direct sequencing of the termini showed that virtually all had the sequence:

$$\text{........}G_4T_4G_4T_4G_4T_4G_4T_4G_{4\ OH}\ \ 3'$$
$$\text{........}C_4A_4C_4A_4C_4 \qquad\qquad\qquad 5'$$

The 3' overhang of the G-rich strand was first found in hypotrichous ciliates such as *Oxytrichia*, but it has been speculated that this structure may be a conserved feature of telomeres (Henderson *et al.*, 1989).

The presence of small linear extrachromosomal DNA molecules in the ciliates was convenient, but the study of telomeres of other eukaryotes is more difficult. The ligation of telomeres onto linear yeast plasmids and cloning in yeast proved to be an effective means of analysis. This strategy was first used to study a telomere cloned from the yeast *Saccharomyces cerevisiae* (Blackburn and Szostak, 1984; Shampay *et al.*, 1984). The genomic DNA was digested with a restriction enzyme and the fragments were ligated to a linearized yeast plasmid vector carrying a selectable marker. DNA molecules carrying the selectable marker but which were not bounded by ligated telomeric sequences were degraded or integrated into the genome in the yeast cell. After transformation into yeast, some of the surviving transformants contained linear vectors bounded by telomeric sequences. As part of a yeast plasmid, telomeric DNA could be easily manipulated. The telomere was subcloned into a bacterial plasmid to obtain sufficient quantities of DNA for sequencing. The telomeric sequence of the G-rich strand in *S.cerevisiae* was found to be $TG_{1\text{-}3}$. The C-rich strand also contained single strand breaks in the most distal part of the terminus like those found in *Tetrahymena* rDNA (Shampay *et al.*, 1984). The linear vector which carried the yeast telomere also has a *Tetrahymena* rDNA terminus at the other end (containing the TTGGGG repeat). After being subcloned and sequenced, this telomere was found to have been modified in yeast. Yeast $TG_{1\text{-}3}$ repeats had been added to the *Tetrahymena* repeats showing that the latter could also support telomere function in *S.cerevisiae*.

After cloning a putative telomeric fragment, its terminal location on authentic chromosomes may be tested by sensitivity to the exonuclease *Bal*31. The enzyme *Bal*31 degrades the 5' and 3' ends of double-stranded DNA in a progressive manner. Total genomic DNA is digested with *Bal*31 over increasing time periods, then digested with a restriction enzyme and analyzed by gel electrophoresis and Southern blotting. The genomic DNA is then probed with the region of the putative terminal fragment lacking terminal repeats. The sizes of terminal genomic DNA fragments decrease owing to *Bal*31 digestion while internal DNA fragments are not affected (Cooke *et al.*, 1985).

By the use of linear plasmids and artificial chromosomes (Murray and Szostak, 1983; Cross *et al.*, 1989; Brown, 1989; Riethman *et al.*, 1989), the telomeric sequences from a number of species have been cloned and sequenced. The consensus sequence of most telomeres studied to date is 5' - T_{0-4} A_{0-1} G_{1-8} - 3'. In most organisms, the telomeric repeat is precise while in some, such as *Saccharomyces*, the repeat is heterogeneous, while another exception is *Paramecium* where the telomeric sequences consist of a mixture of two types of repeat. These exceptions will be discussed in section 1.2.4 Mechanism of telomerase. It is worth noting that one repeat sequence, TTAGGG, is present in organisms as evolutionarily wide ranging as the flagellated protozoan, *Trypanosoma*, and all vertebrates tested so far (Meyne *et al.*, 1989; Moyzis *et al.*, 1988). The orientation for all the sequences is the same in that the G-rich strand runs 5' to 3' towards the terminus of the chromosomes.

Single base gaps are also present in the C-rich strand in many of the organisms. The presence of the gaps can be demonstrated directly by sequencing (as in *Tetrahymena*, Blackburn and Gall, 1978) or inferred from the ability of DNA polymerase to nick translate sequences *in vitro* (Zakian, 1989).

The Southern blotting data has revealed a wide range in the actual amount of repeated DNA at individual telomeres as well as in different organisms. The average telomere size in different organisms ranges from 20bp for the linear molecules in the macronuclei of *Oxytrichia* to 2-20kb in humans (Cross *et al.*, 1989) and over 100kb in mice (Kipling and Cooke, 1990; Starling *et al.*, 1990). In addition, the amount of telomeric DNA varies between individual telomeres, illustrated by gel electrophoresis of telomeric restriction fragments. A smeared band is produced on probing with a telomeric sequence (Allshire *et al.*, 1988).

It is likely that telomere length is regulated by the cell. A simple balance between a constant rate of telomere elongation and loss (e.g. due to incomplete replication) is unstable over a long period. Any alteration to either rate would result in steady growth or shrinkage of the chromosome until it became either too large to be stable or it became small enough for genetic information to be lost from previously sub-telomeric regions. (See section 1.3.6).

1.3.3 Telomere-associated sequences

The chromosomes usually contain middle-repetitive elements called telomere-associated sequences in addition to the simple repeats found in the most distal regions. Although these sequences have been detected in several species (reviewed by Zakian, 1989), only those of *S.cerevisiae* which have been studied in the greatest detail and humans which are the most relevant to this work will be discussed here.

In yeast, there are two classes of telomere-associated sequences, Y' and X elements (Chan and Tye, 1983a and b), which are more complex than the simple telomeric TG_{1-3} repeated sequence. The Y' element is a highly conserved 6.7kb long sequence which is situated immediately internal to the TG_{1-3} sequence. It is present in one to four copies at the ends of many, but not all, yeast chromosomes. The X elements are less highly conserved and range from 0.3-3.7kb in length. While the Y' elements are tandemly repeated, the X elements are present as a single copy just internal to the Y' element(s) (Chan and Tye, 1983a). Unique DNA then occurs proximal to the X element (Chan and Tye, 1983a). Tracts of TG_{1-3} DNA of 50-130bp are also found embedded in the telomere-associated sequences, both at Y'-Y' and at Y'-X junctions (Walmsley *et al.*, 1984).

It is difficult to assign any necessary function to the telomere-associated sequences of yeast. Both the X and Y' elements were originally isolated as sequences containing ARSs (Chan and Tye, 1980), however they may not be essential in chromosome replication. Artificial chromosomes which lack Y' ARSs replicate and segregate through meiosis as well as artificial chromosomes which do possess these elements (Wellinger and Zakian, 1989). Additionally, the X and Y' elements are not required to form new telomeres in *S.cerevisiae* (Murray *et al.*, 1988; Pluta and Zakian, 1989; Wellinger and Zakian, 1989).

The major feature of the X and Y' elements is the high rate of recombination found at this region (Dunn *et al.*, 1984). Recombination in the telomere-associated sequences is probably the cause of the variability in the organization of these regions (Dunn *et al.*, 1984; Zakian and Blanton, 1988). The Y' elements not only vary between telomeres in the number of copies, but they also vary in their presence. Pulsed field gel analysis showed that, in *S.cerevisiae*, some chromosomes lack Y' elements, or, in the case of chromosome I, lack both X and Y' elements (Zakian and Blanton, 1988).

Although telomere-associated sequences may not have an essential role in yeast chromosomes, they may be involved in telomere acquisition by recombination and in chromosome pairing.

Analysis of telomere-associated sequences in human chromosomes has revealed a similar situation to that in yeast. A subterminal sequence isolated by Cross *et al.* (1990) was located to human telomeres by *in situ* hybridization. The subterminal sequence itself consists of a repeated 28bp monomer unit. The sequence was identified in over 50% of human telomeres, but at varying signal strengths between individual telomeres. Earlier work by Cooke and Smith (1986) showed that the terminal sequences of human chromosomes are highly variable.

No function can, as yet, be assigned to the telomere-associated sequences of human chromosomes and they may be unnecessary for the formation of telomeres. Recent work which studied the healing of a chromosome 16 breakpoint has shown that such subterminal sequences are not required for telomere formation *in vivo* (Wilkie *et al.*, 1990). Similar conclusions may be drawn from the work of Farr *et al.* (1991) who successfully created a new telomere which lacks telomere-associated sequences on the X chromosome by a telomere-mediated breakage. Both of the above cases are discussed in more detail in sections 1.3.5 and 1.3.8.

1.3.4 Replication of telomeres

The ends of linear chromosomes cannot be fully replicated by conventional DNA polymerases for reasons described earlier (Watson, 1972; Olovnikov, 1973). Telomeres, however, enable the complete replication of eukaryotic chromosomes. The most widely held theory which has been suggested to explain the replication of telomeres involves the enzyme, telomerase (see Greider, 1990 for review).

The activity of the telomere-specific replication enzyme, telomere terminal transferase or telomerase, has been identified in ciliates (e.g. *Tetrahymena*, Greider and Blackburn, 1985) and in human immortal cell lines e.g. HeLa cells (Morin, 1989) and 293 cells (Counter *et al.*, 1992). Telomerase has been shown *in vitro* to processively elongate the G-rich strand of a telomere by adding one base at a time. The enzyme of *Tetrahymena* adds TTGGGG repeats to the 3' end while the HeLa cell enzyme adds TTAGGG repeats. The C-rich strand can be completed by RNA primers and DNA polymerase, maintaining a gap at the 5' end. Known telomerases are ribonucleoproteins with an essential RNA component which contains a sequence complementary to the G-rich strand of the telomere. For instance, the *Tetrahymena* telomerase RNA contains the sequence 5'-CAACCCCAA-3' (Greider and Blackburn, 1987 and 1989). This activity is inhibited *in vitro* by the addition of DNA oligonucleotides complementary to this sequence and RNase H which cleaves only RNA-DNA duplexes suggesting that this sequence provides a template for the 5' to 3' synthesis of the G-rich strand. Research has shown that *in vitro* the *Tetrahymena* telomerase can use *Oxytrichia* (TTTTGGGG), and *Saccharomyces* (TG_{1-3}) single-stranded repeat sequences as substrates for elongation (Greider and Blackburn, 1985 and 1987). However, whatever the substrate telomeric sequence, the *Tetrahymena* telomerase always adds the *Tetrahymena*-specific repeat TTGGGG/CCCCAA.

Mutations have been made in the cloned gene for the *Tetrahymena* telomerase RNA in this sequence by Yu *et al.*, (1990). After transformation into *Tetrahymena* on a high copy number plasmid, the telomeres were found to have an altered repeat sequence corresponding to the mutations. (The endogenous telomerase was probably still active but the mutant enzyme was present in an excess amount.) This work demonstrates that the RNA component supplies the sequence information for d(TTGGGG)n synthesis and that telomerase is directly involved in telomere synthesis *in vivo*.

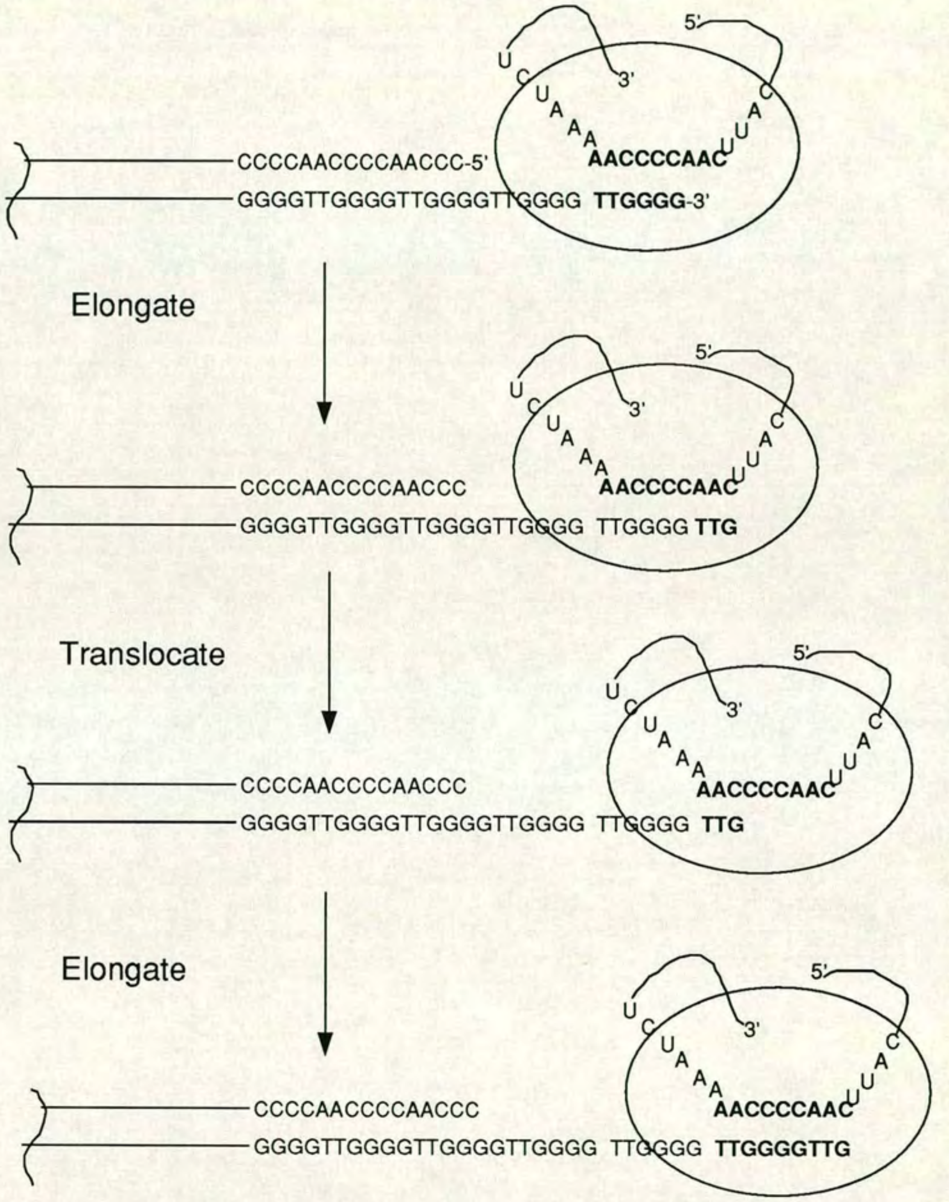
The formation of new telomeres has also been studied in yeast. Although the *in vitro* protein activity of telomerase has not yet been detected in *Saccharomyces* as it has in ciliates, there is molecular genetic evidence for a yeast telomerase. The mutation *est1* (*e*ver *s*horter *t*elomeres) produces a phenotype where the telomeres get progressively shorter by incomplete end replication and the cells enter senescence, eventually becoming inviable (Lundblad and Szostak, 1989). Telomerase activity is highly unstable and the proteins have not been successfully purified from ciliates (Greider and Blackburn, 1987). The amino acid sequence of the *EST1* gene has, however, revealed similarities to RNA dependant polymerases including the reverse transcriptases (Lundblad and Blackburn, 1990). Telomerase is formally a reverse transcriptase since an internal RNA template is copied into DNA. This sequence similarity has led to speculation that the *EST1* gene encodes a protein component of the yeast telomerase.

1.3.5 Mechanism of telomerase

The mechanism of telomere elongation by telomerase is now believed to be processive rather than distributive. Telomerase recognizes the telomeric G-rich strand, then the 3'-most nucleotides are hybridized to the internal RNA template sequence in the enzyme. After, the G-rich strand is elongated to the end of the RNA template, a translocation of telomerase repositions the enzyme with respect to the telomeric DNA repeat sequence allowing a further round of elongation (see figure 1.3). A distributive mechanism of elongation would involve dissociation of the enzyme from the primer G-rich strand after a single repeat was added followed by rebinding for a further round of elongation. Greider (1991) showed that the *Tetrahymena* telomerase enzyme synthesizes approximately 500 nucleotides before dissociating from the primer strand which indicated that the elongation was processive.

Initially, work carried out *in vitro* suggested that only molecules terminating in at least eight bases of single-stranded telomeric DNA could be elongated by telomerase (Greider and Blackburn, 1987). However, more recent work suggests that internal TG-rich, i.e. telomere-like,

Figure 1.3 <u>Processive elongation of telomeres by telomerase</u>



The 3'-most nucleotides of the TTGGGG strand are hybridized to the internal RNA template. The sequence TTG is added one base at a time. Translocation repositions the telomerase with respect to the TTGGGGTTG sequence. A further round of elongation can now occur. (From Greider and Blackburn, 1989).

sequences may be recognized by telomerase prior to adding telomeric repeats onto a completely non-TTAGGG- or TTGGGG-like sequence.

Wilkie *et al.* (1990) described an α-thalassaemia mutation associated with a truncation of the short arm of chromosome 16 to a site 50kb distal to the α globin genes. Telomeric repeats of (TTAGGG)n had been added directly to the breakpoint. The mutation is stably inherited demonstrating that TTAGGG repeated sequence can adequately heal a broken chromosome end. Existing models for the mechanism of telomerase action could not explain the apparent *de novo* telomere formation at a non-telomeric end. A TTAGGG-like sequence is, however, present 25bp proximal to the breakpoint in the truncated chromosome 16. The terminal breakpoint DNA including the TTAGGG-like sequence of the truncated chromosome were used as a test substrate for telomerase from *Tetrahymena* (Harrington and Greider, 1991) and from HeLa cells (Morin, 1991) in later experiments. Telomerase repeats were successfully added in both cases.

The models for the mechanism of telomerase were altered to accommodate these results. Telomerase may bind to the internal telomeric-like sequence and then slide to the end of the DNA molecule where telomeric repeats are added. Alternatively, telomerase may remain bound to the recognized sequence and then bind the end of the DNA "looping out" the intervening sequences. These models are illustrated in figure 1.4.

There are some aspects of telomere formation in yeast which are not easily explained by activity of telomerase. The *in vitro* properties of telomerase are ideal to explain the replication and extension of telomeres in organisms where the telomere repeat is precise such as ciliate protozoa and vertebrates. However, the telomerase models must be modified to explain the replication of telomeres in organisms where the repeat is complex such as *Saccharomyces* (TG_{1-3}), *Schizosaccharomyces* ($T_{1-2}ACC_{0-1}G_{1-6}$) and the slime mould *Dictyostelium* (TG_{1-8}). Yu and Blackburn proposed an explanation based on their work with mutant telomerase enzymes (1991). One of the mutant enzymes synthesized repeats which contained a variable number of G residues (while the other bases remained constant) like the repeats of *Saccharomyces* and *Dictyostelium.* On the basis of this finding, they proposed that the template of the enzyme slips or stutters at this position of the DNA repeat, the result being a repetition of the G residue. It has been proposed on

Figure 1.4 Three models for telomerase elongation of non-telomeric termini

(From Harrington and Greider, 1991).

The telomerase RNA template region and complementary telomeric sequence are shown in bold.

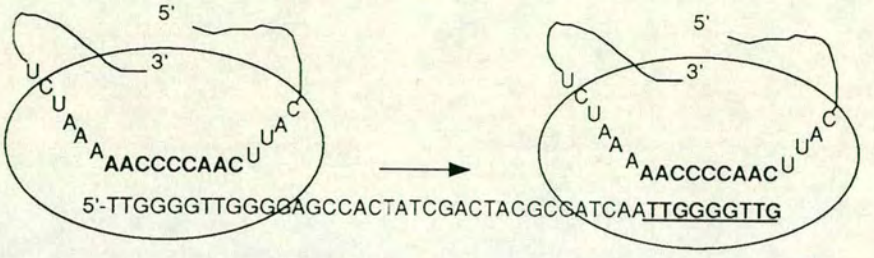
(a) Telomerase binds to the 5' telomeric or G-rich sequence through hybridization of the internal RNA template. After recognition, the complex would track to the 3' end of the oligonucleotide an dinitiate d(TTGGGG) synthesis.

(b) Telomerase RNA recognizes 5' G-rich sequences as in (a). Non-telomeric sequences are looped out, retaining hybridization of the 5' end of the oligonucleotide, while bringing the 3' end close to the RNA template to allow d(TTGGGG) synthesis.

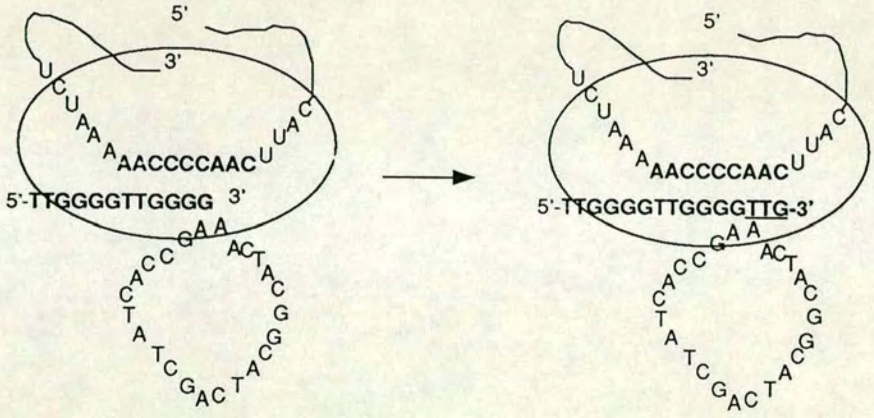
(c) Telomerase has an RNA-independent ability to recognize and bind G-rich structures which positions telomerase RNA at the 3' terminus of the oligonucleotide to initiate telomere addition. This model might also require tracking of telomerase to the 3' end of the oligonucleotide after binding.

Figure 1.4 Telomerase elongation of non-telomeric DNA

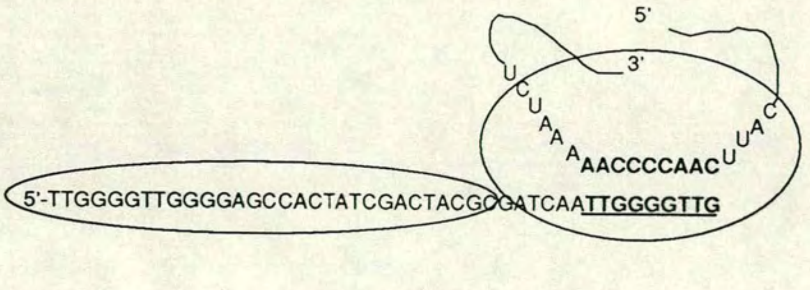
(a) Bind and slide



(b) Loop-out



(c) RNA -independent recognition



the basis of work carried out on mutant telomerase enzymes that in these organisms, slippage of the template occurs during synthesis of the repeats. In the case of *Paramecium* and *Plasmodium*, the telomeres consist of a mixture of two repeat sequences of the same length but which differ by one nucleotide. From their work on *Tetrahymena* which contained active mutant and wild type telomerases, Yu and Blackburn proposed that these two organisms both contain two telomerase enzymes which have correspondingly different RNA templates (1991).

The work of Zakian's group (reviewed in Zakian, 1989) cast some doubt on the activity of telomerase in yeast, instead suggesting that most new telomere formation was the result of non-reciprocal recombination events. They proposed models where one end, 3' or 5', invades an internal site in another telomere where it could be elongated on the existing template provided by the complementary strand (Wang and Zakian, 1990). However, although recombination probably can take place between telomeres, it has been found to be unable to rescue the *est1* senescence mutation which involves the gradual and inexorable loss of telomeres (Greider, 1990).

Doubt was also cast on the ability of telomerase to form telomeres *de novo* in the development of the ciliate macronucleus where telomeres are formed at sites which lack any telomeric repeats. However, recent work by Yu and Blackburn (1991) showed that telomerase is involved. A mutant telomerase template RNA was used to monitor the addition *in vivo* of telomeric repeats *de novo* onto the non-telomeric ends generated in the development of the *Tetrahymena* macronucleus. The new telomeres contained mutated repeats demonstrating that the mutant telomerase has synthesized them.

1.3.6 The physiology of telomeres

The telomeres of unicellular organisms such as protozoa and yeast have been hypothesized to be dynamic structures. Processes which extend or shrink the telomeres according to the physiological state of the cell would help to explain the range of telomere sizes seen on electrophoresis.

Some ciliated protozoa display a gradual increase in size in the terminal restriction fragments of their chromosomes when the cells are grown in continuous log-phase. In *Trypanosoma*

(Bernards *et al.*, 1983) and *Tetrahymena* (Larsen *et al.*, 1987), terminal restriction fragments, i.e. the telomeres, show a 3-10bp increase per generation which may be due to the action of telomerase. The telomeres also undergo very steep decreases in size during growth.

The telomeres of *S.cerevisiae* do not show a similar increase in size during logarithmic growth, but can change in response to the introduction of excess telomeric sequences on linear and circular plasmids. Genetic mutations which also cause changes in telomere length may be connected with this phenomenon. Runge and Zakian (1989) found that the introduction of excess telomere sequences on high copy number plasmids caused an increase in the overall length of tracts of TG_{1-3} DNA which were situated at the telomeres. This increase in length was proportional to the number of copies of the plasmids in the cell. They hypothesized that telomere growth was normally constrained by a limiting factor(s) in the cell. The excessive amounts of telomeric DNA competed for this factor and exposed the telomeres to elongation processes without any negative regulation.

Phenotypes have also been described which relate directly to altered telomere length in *S.cerevisiae*. The product of the *RAP1* gene is a telomere binding protein (Conrad *et al.*, 1990) which is believed to be important in maintaining telomere length. The telomeric binding sites of RAP1, GGTGTGTGGGTGT, are found at an average density of one site per 40bp of telomeric poly TG_{1-3} DNA (Shampay *et al.*, 1984). Mutations which reduce RAP1 function cause a shortening of the telomeres *in vivo* (Conrad *et al.*, 1990; Lustig *et al.*, 1990) as do mutations in RAP1 binding sites (Lustig *et al.*, 1990). Conrad *et al.* (1990) also found that overproduction of RAP1 causes chromosome loss and changes in telomere structure which suggests that the interaction of RAP1 with telomeres is complex.

The *cdc17* mutation has a lethal stage-specific effect in the cell cycle (Hartwell, 1973). However, strains which possess temperature sensitive *cdc17* mutations show telomere-related phenotypes when grown at permissive temperatures. During growth, the telomeres are longer than in *CDC17* strains owing to the addition of sequence at the telomeric TG_{1-3} DNA tracts (Carson and Hartwell, 1985). The CDC17 protein was later found to be DNA polymerase 1 of *S.cerevisiae* (Carson, 1987) which could explain the effects of a mutation in this gene on telomere metabolism. (Although the enzyme telomerase is thought to elongate the extreme termini of the G-rich strand of the telomeres, DNA polymerase is still required to complete the complementary strand.) However, it is difficult to explain how a loss-of-function mutation which is indicated by the temperature-sensitive lethality could produce a gain-of-function phenotype at a permissive temperature

Two *S.cerevisiae* mutations, *tel1* and *tel2*, have been described which cause telomeres to be reduced to a shorter length (Lustig and Petes, 1986). Little has been hypothesized about these two mutations, but the proteins which are responsible may regulate telomere length by binding to the termini and displacing telomerase.

The *est1* mutation, described earlier, produces a phenotype in which the telomeres shrink on each round of replication in a fashion that is consistent with incomplete end replication (Lundblad and Szostak, 1989).

The existing evidence leads us to think that telomere length, in *S.cerevisiae* at least, is regulated by processes which elongate and shorten telomeres. Regulation of telomere length must, therefore, require some form of feedback where the cell's physiological controls can recognise the existing telomere size.

1.3.7 Telomere-binding proteins

Telomerase and the yeast protein, RAP1, are probably the most extensively studied proteins which interact with or bind the telomeres (see sections 1.3.4, 1.3.5 and 1.3.6). The DNA of the telomeres is also likely to be associated with specialized structural proteins. A major feature of telomeres is their resistance to fusion and exonucleolytic degradation (Zakian, 1989). Proteins which are bound to the telomeric DNA may be responsible for the stable nature of the telomere. There are several lines of evidence mainly from work done in ciliates and yeast which suggest that telomeric DNA is bound to non-histone proteins.

Digestion of nuclei from a number of organisms with micrococcal nuclease has shown that the telomeric DNA is packaged into non-nucleosomal complexes. Micrococcal nuclease cuts chromatin between packaged nucleosomes which releases the DNA as fragments of 200bp which can be visualized by electrophoresis as an ordered array of fragments which increase by 200bp intervals. However, hybridization with telomeric repeat probes reveal the telomeric DNA to be packaged into complexes which are a different size from the conventional nucleosome. For instance, the complex in *Tetrahymena* and in *Saccharomyces* is larger and more variable at 400-800bp (Blackburn and Chiou, 1981; Wright *et al.*, 1992) while in Oxytrichia the complex is 100-150bp (Gottschling and Cech, 1984). The telomeric DNA within these non-nucleosomal complexes is protected from the action of BAL31 exonuclease and methylation agents (Gottschling and Zakian, 1986; Price, 1990). A common feature of some telomere-binding proteins which have been isolated is very tenacious, although non-covalent, binding of the telomeric DNA. The proteins bind telomeric DNA in conditions of high salt under which the histones elute, a property which has been useful in their purification (Gottschling and Zakian, 1986; Price, 1990).

Proteins that co-purify with chromosome ends have been isolated from a number of species. RAP1, and the telomere-binding factors TBFα and TBFβ of *Saccharomyces* all bind to the telomere sequences (Conrad *et al.*, 1990; Liu and Tye, 1991). TBFβ only binds to the TG_{1-3} repeated sequences, while TBFα binds to the junction between the subtelomeric X sequence and the TG_{1-3} repeats. TBFα also binds to the TTAGGG repeat sequence which is found within the X element and in vertebrate telomeres (Liu and Tye, 1991). A 98kDa heterodimeric protein has been isolated which binds to the telomeric T_4G_4 sequences of *Oxytrichia* (Gottschling and Zakian, 1986; Price and Cech, 1989). The efficient binding of this protein requires not only the repeated sequence itself, but both the 3' single-stranded overhang of $(G_4T_4)_2$ which is a feature of *Oxytrichia* telomeres and the adjacent duplex region (Gottschling and Zakian, 1986). A 50kDa protein was localized to the telomeres of the related ciliate, *Euplotes,* which has the same telomeric repeat as *Oxytrichia* (Price, 1990). A 10kDa protein which binds specifically to TTAGGG repeated sequence has been isolated in the slime mould *Physarum* which has the same telomeric repeat as vertebrates (Coren *et al.*, 1991). Finally and, most recently, a 50kDa protein, TRF (TTAGGG repeat factor), has been detected in nuclear extracts from human, mouse and monkey cells. TRF specifically binds double-stranded TTAGGG repeats in linear and circular DNAs. However, TRF does not recognize single-stranded TTAGGG repeats or related repeated sequences such as TAGGG, TTTAGGG, TTGGGG, TTTTAGGG and TTAGGC (Zhong *et al.*, 1992).

In most of the organisms studied so far, little is known about the function of the telomere-binding proteins. The only binding protein for which there is any evidence for a possible function is RAP1 of *Saccharomyces* (described in section 1.3.6). RAP1 has been shown to be essential for growth (Shore and Nasmyth, 1987), but both over- (Conrad *et al.*, 1990) and under-expression (Conrad *et al.*, 1990; Lustig *et al.*, 1990) of RAP1 result in altered telomere length suggesting that the protein is involved in the regulation of telomere size. Liu and Tye (1991) speculated that TBFα may act as an anchoring protein for telomerase in yeast, since its binding site on the chromosome is the most distal point before the telomeric TG_{1-3} repeats begin.

The products of the genes responsible for the *tel1* and *tel2* phenotypes (see previous section) may also bind to the telomeres (Lustig and Petes, 1986). However, the proteins encoded by these genes have not yet been isolated.

The simplest hypothesis to explain the function of telomere-binding proteins is that they specifically bind the DNA into a non-nucleosomal complex which acts as a physical "cap" for the end of the chromosome. Apparently, the proteins are specific to terminal repeats and have not been localized to the sub-terminal regions. It is, therefore, possible that cloned telomeric repeats may be adequate to bind telomere-specific proteins *in vivo* and, thereby, maintain a stable DNA terminus.

1.3.8 Telomeres in somatic, germline and immortal cells

Organisms which replicate by binary fission such as yeast and ciliate protozoans are effectively immortal and may require constant expression of telomerase and regulation of telomere length. Comparisons between the telomeres in somatic, immortalized and germline cells in humans have led to speculation that telomerase is not present in somatic cells in higher organisms.

The first hints that this might be the case were obtained from McClintock's work in maize. While studying the fate of broken chromosome ends, she noted that the broken ends in some tissues were stabilized or "healed". These ends had presumably obtained new telomeres. However, these healing events only took place in cells from germline tissues of the maize plant and not from the somatic tissues (1941 and 1942).

A study on the ageing of primary human fibroblast cells was carried out by Harley *et al.* (1990). Telomere length in fibroblasts which were passaged *in vitro* was found to decrease by an average of 50bp/generation. The mean telomere length decreased by 2kb before the cultures senesced and ceased division. The researchers proposed that telomere shortening may also occur *in vivo*. Lindsey *et al.* (1991) found that the telomeres of primary epithelial cells showed a significant decrease with increasing age of individuals. In contrast, transformed cells do not suffer from telomere shortening during passage *in vitro* (De Lange *et al.*, 1990), although the telomeres are shorter than in somatic cells from the outset (De Lange *et al.*, 1990; Hastie *et al.*, 1990). Additionally, the telomeres in human sperm cells are longer than those of blood cells which could mean that the former cells contain active telomerase (Cooke and Smith, 1986; Allshire *et al.*, 1989; Cross *et al.*, 1989).

These observations have led to the theory that the expression of telomerase is in some way developmentally regulated in higher organisms. According to this hypothesis, in germline cells, telomerase is expressed and maintains telomeres at a constant size. In somatic cells, however, telomerase is no longer present. Although the telomeres gradually decrease over the lifetime of somatic cells owing to incomplete replication at the chromosome ends, the telomeres are sufficiently long to forestall any damage to the internal coding sequences. It has since been speculated that telomere loss represses (by a position effect) a gene which regulates a senescence program (Greider, 1990; Wright and Shay, 1992).

Since immortalized cell-lines do not suffer telomere loss, presumably the process of immortalization reactivates telomerase expression. The mechanism for maintaining telomere length equilibrium is regained but reset to a shorter size than that seen *in vivo* for the same cell type (Counter *et al.*, 1992). Theoretically, telomerase in such cell types should be capable of maintaining reintroduced telomeres or telomeric fragments at a stable length by the addition of TTAGGG repeats until the telomere length is the same as that of telomeres on the endogenous chromosomes in the cell.

1.3.9 Reintroduction of cloned telomeres into immortalized cells

The current knowledge which has been gained concerning the replication and elongation of telomeres in mammalian cells allows the design of experiments where telomeric fragments may be reintroduced. Such experiments were carried out in lower eukaryotes at a relatively early stage in research on telomeres, often using linearized yeast plasmids. The methods involved ligating restriction fragments from an organism's nuclear DNA to a linearized plasmid. The entire ligation mix was then used to transform yeast. A mixture of DNA molecules was obtained which included linear plasmids capped by telomeres as well as recircularized and integrated plasmids. As described previously, telomeres from the ciliate protozoa and from yeast itself were isolated and cloned in this manner. The actual termini of all telomeric sequences are modified in yeast by the addition of yeast-specific repeats (Szostak and Blackburn, 1982; Brown, 1989; Cross *et al.*, 1989; Riethman *et al.*, 1989).

Although reintroduced telomeres from a variety of organisms have functioned as part of linear vectors in yeast, similar experiments have not yet been successful in mammalian cells. However, the work done on the α-thalassaemia breakpoint demonstrates that telomerase in immortal human cells is capable of synthesizing functional telomeres from a minimum of telomeric sequence (Wilkie *et al.*, 1990; Harrington and Greider, 1991; Morin, 1991).

Human telomeric sequences were cloned by functional complementation using YACs which were equipped with only one telomere (Brown, 1989; Cross *et al.*, 1989; Riethman *et al.*, 1989). These sequences fulfil the known criteria for being telomeric in the genome. They show telomeric function in yeast, sensitivity to digestion with *Bal*31 and hybridization to *Bal*31-sensitive restriction fragments in a Southern blot of human genomic DNA. The sequence consists of the TTAGGG repeat which is the sequence added by the HeLa cell telomerase. Additionally, in a Southern blot of genomic DNA the telomeric fragment also detects restriction fragments of about 10kb for blood and of about 15kb for sperm (Cross *et al.*, 1989).

Recently, a 500bp tract of (TTAGGG)n was tested for its ability to form a functional telomere after reintroduction into mammalian cells. (Larger telomeric fragments were not available because the "half-YAC" cloning procedure severely shortens the telomere from an average of 10kb to a few kilobases.) The repeats were used as a single artificial telomere in an integration vector which contained the HisD selectable marker. After being cut with a restriction enzyme to expose the telomere, the vector was transfected into a hybrid cell line containing a human X-chromosome on a hamster background. After transfection, a breakage event in the human X-chromosome was detected. The vector had integrated into an internal site of the long arm, breaking off the distal part. The new terminus of vector DNA was stable and had been elongated, presumably by telomerase (Farr *et al.*, 1991).

It may be concluded that a reintroduced telomere may survive in a mammalian cell providing it is present in a "stable" environment (i.e. attached to a centromere and a selectable marker). The telomere-mediated breakage described above is one way of studying the fate of reintroduced cloned telomeres. A further step is the introduction of telomeres on a linear vector which is designed to replicate episomally (without integrating into the host chromosomes). Specifically, the next step is the construction of a telomere-bounded artificial chromosome for mammalian cells. The construction of artificial chromosomes, or any vectors which replicate independently of the mammalian host chromosomes is intertwined with the isolation of centromeres or functional analogues. The centromere of the eukaryotic chromosome prevents loss of the DNA molecule during successive divisions of the cell. Plasmids or linear vectors which are introduced into mammalian cells will usually be integrated into the host chromosomes by the recombination enzymes of the cell (or lost entirely from the nucleus) unless a centromeric sequence is included.

1.4 CENTROMERES

The centromere of a eukaryotic chromosome binds the kinetochore which in turn binds spindle microtubules during mitosis. The centromere is also the last point of sister chromatid attachment before anaphase and ensures their proper assortment during mitosis and meiosis.

1.4.1 Centromeric sequences in yeast

The centromeres of the budding yeast, *Saccharomyces cerevisiae*, are the smallest among eukaryotes and are the best understood. The centromeric elements of *S.cerevisiae* were first cloned and characterized by Clarke and Carbon (1980). The centromeres of yeast consist of three regions, designated CDE (centromere DNA element) I, II and III, all of which are contained within a DNA segment of about 125bp (the centromeres vary between chromosomes, see figure 1.5). CDE II contains ~88bp of 93% AT-rich DNA. CDE III is a very tightly conserved AT-rich sequence (TGTTT [T/A] TGNTTTCCGAAANNNAAAAA). This sequence is a protein binding site and essential for centromere function. Recently, a protein complex, CBF3, has been identified which binds specifically to the CDEIII sequence (Lechner and Carbon, 1991). Further *in vitro* studies have demonstrated that CBF3 is involved in kinetochore function in *S.cerevisiae*. The CBF3 proteins can link CEN DNA to microtubules and also contain a microtubule-based motor (Hyman *et al.*, 1992). CDE I is a less tightly conserved AT-rich sequence (PuTCACPuTG) and believed to be the least essential of the three regions for centromere function (Clarke, 1990). During the characterization of the CEN sequences, certain deletions proved to cause meiotic-specific defects in segregation while mitotic segregation was unaffected (Clarke and Carbon, 1983). This led to speculation that different portions of the sequence are involved in segregation at different stages in the life cycle.

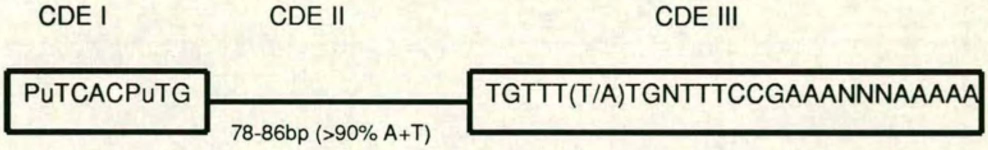
The isolation of telomeres and functional centromeric elements (CEN) enabled the construction of yeast artificial chromosomes (YACs) in *S.cerevisiae* (Murray and Szostak, 1983; Burke *et al.*, 1987). Sequences which enabled plasmids to replicate autonomously in yeast (ARS elements) had been isolated previously (Chan and Tye, 1980, 1983a and 1983b). When Murray and Szostak (1983) constructed mitotically stable long (55kb) linear plasmids using ARS elements, CEN sequences, selectable markers and telomeres, there appeared to be a direct correlation between the length of the linear plasmid and its mitotic stability in the yeast cell. Although functional linear plasmids existed which did not possess CEN or yeast-specific ARS sequences such as those used in the studies of ciliate telomeric sequences, these DNA molecules were not as stable as the YACs which contained functional centromeres.

The centromeres of the fission yeast, *Schizosaccharomyces pombe*, are now the subject of a great deal of study. Unlike those of *Saccharomyces*, the centromeres of *S.pombe* bind multiple microtubules. The *S.pombe* centromere also captures free microtubule ends at the beginning of

Figure 1.5

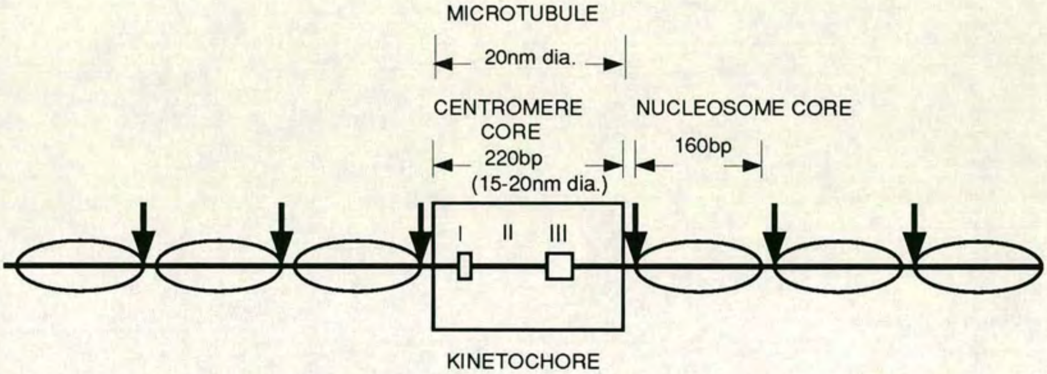
Centromeres of *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*

Consensus sequence of functional centromeric DNA in *S.cerevisiae*



(Clarke, 1990)

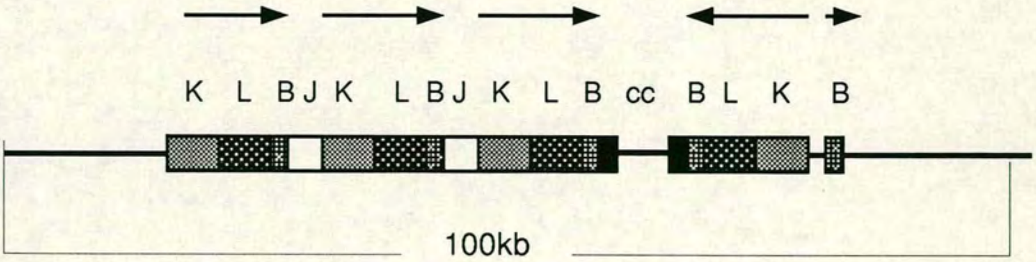
Schematic representation of an *S.cerevisiae* kinetochore (*CEN11*)



(Adapted from Bloom *et al.*, 1983)

Structure of the *S.pombe cen2* region (adapted from Clarke and Baum, 1990)

arrows indicate the orientation of the repeats



mitosis and releases them at the end (Ding *et al.*, 1993). These two features of *S.pombe* are shared with the centromeres of more complex eukaryotic cells, so this organism may provide a better model than *S.cerevisiae*.

S.pombe centromeres contain several classes of untranscribed repeated DNA sequences and span many kilobases (40kb for chromosome I and 80kb for chromosome II). The first *S.pombe* centromere to be described was designated *cen2* from chromosome II (Clarke and Baum, 1990). The *cen2* region is divided into four 14kb tandem repeated units (see figure 1.5). Each 14kb unit is, in turn, composed of repeated elements K (6.4kb), L (4.5kb) and B (<1kb) (Fishel *et al.*, 1988). Another repeat, J, separates the three leftmost 14kb "K-L-B" units from each other. Portions of K, L and B are found on all three *S.pombe* chromosomes but J is found only in the *cen2* and *cen3* regions. Two of the K-L-B units in the *cen2* region are organized into part of a 31kb inverted repeat that flanks a 7kb central core which is unique to centromere 2 (Fishel *et al.*, 1988; Chikashige *et al.*, 1989). The DNA sequences of K and L have not revealed any distinguishing features. The sequence of the central core has also not revealed many clues as to its role in the centromere. The central core does not cross-hybridize to *cen1* or *cen3* DNA, it is specific to the *S.pombe* genome and, like the K-L-B repeats, is not transcribed (Clarke and Baum, 1990). The *cen2* core sequence is AT-rich and contains tRNA genes (Takahashi *et al.*, 1991) but is otherwise undistinguished. Although the organization of the repeated elements differs between centromeres, the inverted repeat structure is common to *cen1* and *cen3* as well (Clarke and Baum, 1990).

The functional significance of the elements of *cen2* was studied by reintroduction of structurally defined portions back into *S.pombe* on plasmid vectors. The transformants were then assayed for proper mitotic and meiotic segregation of the minichromosomes. The main criteria for centromeric function are stability of the vector through mitosis, a single copy per cell and "2+:2-" segregation* through meiosis. The smallest fragment of the *cen2* region that can provide anything approaching centromeric function contains the 7kb central core, 3.5kb of inverted repeats to the left

---

* During meiosis, the homologous chromosomes are separated at meiotic division I, so that two out of the four haploid daughter cells receive a chromatid from homologue A while the other two cells receive a chromatid from homologue B. A vector with full centromeric capability will segregate like an aneuploid chromosome through meiosis and will only be received by two of the cells, hence 2+:2- segregation.

and right of the core, and an entire B-K-L unit plus a second B repeat to the right of the core. Minichromosomes containing this fragment of *cen2* are mitotically stable in *S.pombe* and segregate predominantly 2+:2- through meiosis. However, a high degree of premature sister chromatid separation is found in the first meiotic division. Apparently, full centromeric function is obtained only with the entire *cen2* region (Clarke and Baum, 1990). Unfortunately, work on the *cen2* region DNA is hampered by the instability of the complex inverted repeat arrangement in bacterial plasmids. The elements of the *S.pombe* centromere can be manipulated in *S.cerevisiae* as a whole in a YAC, or separately from each other in bacterial plasmids (Hahnenberger *et al.*, 1989; Clarke and Baum, 1990).

The complexity of centromeres is illustrated by the lack of cross-species function e.g. *S.pombe* centromeres do not work in *S.cerevisiae* cells and *vice versa.* When dicentric chromosomes are formed in eukaryotic cells (by X-irradiation for instance), they are broken into two during mitosis by the centromeres being pulled to opposite poles. However, when *S.pombe* centromeres are introduced into *S.cerevisiae*, YACs carrying *S.pombe* centromeres do not delete, rearrange or act like dicentrics. On the other hand, plasmids carrying *S.cerevisiae* CEN sequences are mitotically unstable in *S.pombe* and do not show any centromeric activity (Hahnenberger *et al.*, 1989; Mann and Davis, 1983). These observations stand in contrast to the situation in telomeres. Telomeric structure is similar between species and this is reflected in the functional replacement of telomeres from one species with those of another (in yeast at least) (Zakian, 1989). It may be concluded that centromere function is of greater complexity than telomere function. However, although centromeric sequences differ between the species studied to date, the involvement of proteins which are conserved through evolution cannot yet be ruled out.

1.4.2 Centromere proteins in mammalian chromosomes

The essential DNA sequences of yeast centromeres are known in detail, while the study of the proteins which are involved is in its early stages. The essential sequences of the centromeres of, for instance, mammals are still the subject of speculation (reviewed in Willard, 1990). Many centromere-associated proteins have, however, been identified.

These centromere proteins (CENPs) were identified by autoimmune sera from patients with the CREST syndrome of progressive systemic sclerosis (Moroi *et al.*, 1980). CENP-A (17-19kDa) is a centromere-specific histone (Palmer *et al.*, 1990). CENP-B (80kDa) is confined to the central chromatin domain and is a satellite DNA binding protein (Earnshaw *et al.*, 1987). CENP-C (140kDa) and CENP-D (50kDa) are both located in the kinetochore but CENP-C, which is a component of the inner kinetochore plate (Saitoh *et al.*, 1992) is the only protein which is specific to active centromeres (Earnshaw *et al.*, 1989). CENP-E (312kDa) was first localized to the centromere regions of chromosomes in prometaphase (Yen *et al.*, 1991). It has since been identified as a kinesin-like motor protein (Yen *et al.*, 1992). Other proteins have been localized to the pairing domain and link the sister chromatids at the primary constriction. Immunochemistry has shown that the centromere-specific proteins are functionally significant. The injection of anti-centromere antibodies into mammalian tissue culture cells causes the disruption of chromosome movements during mitosis although the specific CENP antigens which are responsible have not yet been identified (Bernat *et al.*, 1990).

The INCENPs (INner CENtromere Protein) are antigens which are bound between sister chromatids at the centromere as well as other points where they make close contact. These proteins are unusual in that at the onset of anaphase, they leave the chromosomes. One component co-localizes with the region of overlap of the microtubules in the central spindle and as the overlap decreases through anaphase, the INCENPs become more concentrated in this region. Another INCENP protein localizes to the cell cortex in the region of the forming contractile ring. These observations have suggested that the INCENPs are involved in three mitotic processes. Since they are released from the chromosomes at the start of anaphase, they may have a role in the regulation of sister chromatid pairing. The localization of the INCENPs to the overlapping ends of the polar microtubules suggests a possible role in spindle microtubule movement. Thirdly, the INCENPs might be involved in the formation of the contractile ring during cytokinesis (Cooke *et al.*, 1987).

The centromeric proteins referred to as CLiPs (Chromatid Linking Proteins) may be involved in sister chromatid pairing. The CLiPs are identified by autoimmune sera as a filamentous structure that appears to traverse the two chromatids perpendicular to the long axis of the chromosome (Rattner *et al.*, 1988).

The kinetochore, where CENPs C and D are situated, is the proteinaceous portion of the centromere and is almost certainly responsible for mediating chromosome movement. The kinetochore's appearance has been described from cytological studies although little is known about the fine molecular structure. Under the electron microscope, the kinetochore is trilaminar and aligned as two parallel plates on either side of the primary constriction (Comings and Okada, 1971). The kinetochore has been identified visually as the binding site of the spindle microtubules. Unlike the *S.cerevisiae* or *S.pombe* centromeres which can bind only one or very few microtubules, the mammalian kinetochore can bind around a dozen (Mitchison and Kirschner, 1985; Hayden *et al.*, 1990).

Originally, the capture and binding of the ends of the spindle microtubules was thought to be the only active role of the centromere in mitosis. However, there is a great deal of evidence that force-producing mechano-chemical motors are present in the kinetochore. At prometaphase, the chromosomes move back and forth between the two spindle poles until they are centred on the metaphase plate. At anaphase, the sister chromatids split and then the chromosomes progress towards the poles. During chromosome movements, the microtubules change length by adding or losing sub-units at the ends attached to the kinetochores. Microtubule-based motors are also situated in the kinetochores which enable the centromeres to slide over the microtubules. Dynein and kinesin have both been localized to the kinetochore. Dynein is a "minus-end-directed" motor - effectively, it moves along the microtubule towards the spindle pole end. Kinesin, on the other hand, is "plus-end-directed" - it moves along the microtubule towards the growing end (Hyman and Mitchison, 1991).

1.4.3 DNA sequences at the human centromere

The centromeric regions of all mammals consist of satellite DNA sequences with very few common motifs. The "primary constriction", the cytological location of the centromere in a metaphase chromosome, is mostly α-satellite DNA in humans as well as the other primates. The α-

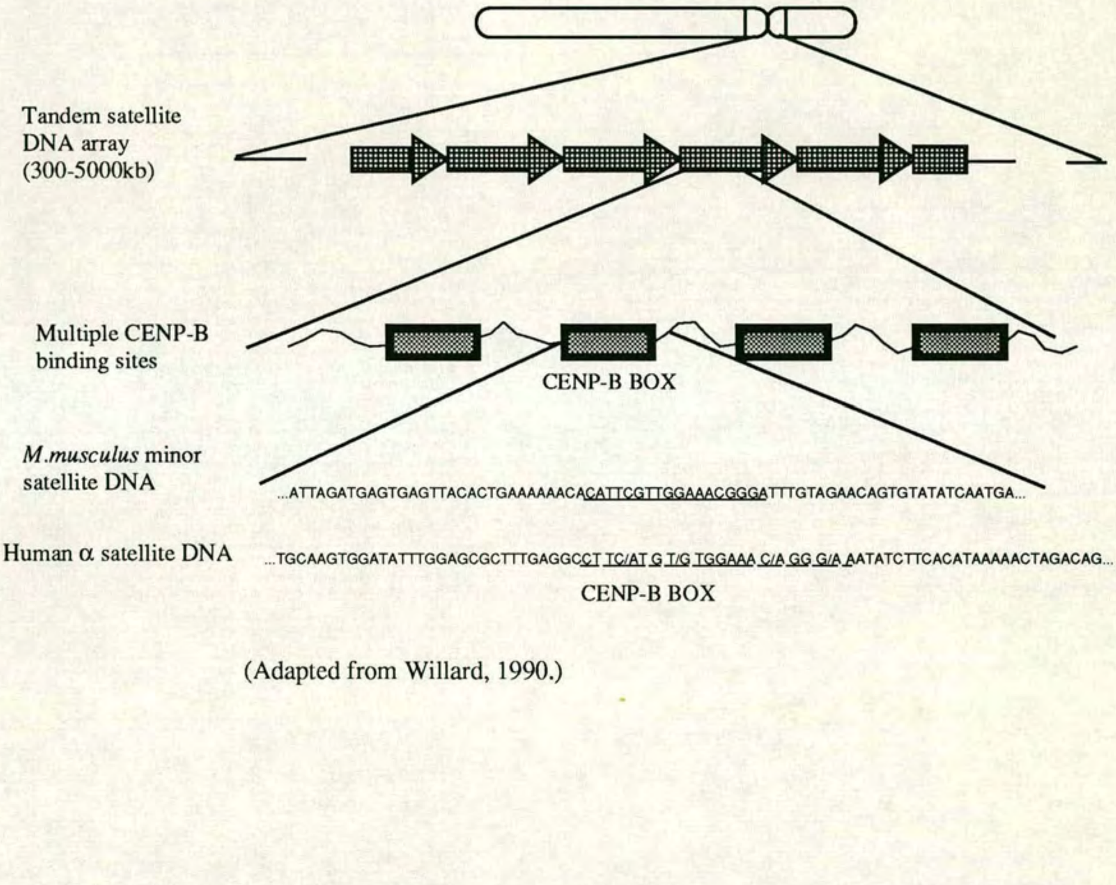
satellite sequence is a 171bp AT-rich monomer (see figure 1.6) which is repeated tandemly. The length of α-satellite tracts varies between chromosomes and ranges from 300-5000kb. In contrast to telomeric repeats, α-satellite DNA has a complex organization which is illustrated in figure 1.6. The 171bp monomer shows a great deal of divergence within a given chromosome as well as between chromosomes. The organization of the repeat arrays in each chromosome is directly related to the divergence between different monomers. At the level of the DNA sequence, the 171bp monomer is repeated n number of times. However, the sequence divergence between consecutive monomers ranges from 20-40%. The n+1 monomer has virtually the same sequence (<2% divergence) as the first of the previous n monomers. The result is a higher-order repeat unit (repeated x number of times per chromosome) which consists of n monomers with inter-monomer divergence. The centromere is made up of hundreds of kilobases of the higher-order repeat unit. Both the number of 171bp monomers (n) and the number of higher-order repeat units (x) can be chromosome-specific (Willard and Waye, 1987).

This pattern of chromosomal organization is common throughout the mammals although the actual satellite sequences found in different orders, e.g. rodents and primates, are mainly unrelated. Presumably, this pattern arose by repeated duplications and divergence of the original monomer. One portion of the α-satellite monomer is common to the repeat found at the centromeres of other mammalian orders. A sub-set of the α-satellite monomers and the minor satellite (the repeat found at the centromeres of *Mus musculus*) contain the CENP-B box, the binding site for a centromere-specific protein (Masumoto *et al.*, 1989) The importance of the CENP-B box has been questioned since it has not yet been detected in the α-satellite DNA of the human Y chromosome (Tyler-Smith and Brown, 1987). Correspondingly, the CENP-B protein has not yet been detected in the centromere of the Y-chromosome (Pluta *et al.*, 1992)).

Although a great deal of effort has gone into the characterization of the α-satellite arrays at centromeres and the interaction with CENP-B, it is still unclear what DNA sequences are essential for a functional human centromere. The apparent lack of the CENP-B protein at the Y-chromosome (Pluta *et al.*, 1992) and the presence of CENP-B at inactive centromeres (Earnshaw *et al.*, 1989) points towards an indirect role for α-satellite DNA and CENP-B in the formation of centromeres.

Figure 1.6

The organization of satellite DNA at the primary constriction of mammalian chromosomes



(Adapted from Willard, 1990.)

Recent work by Masumoto's group (unpublished results) provides evidence for the involvement of CENP-B in folding the DNA of the primary constriction into a form of heterochromatin which may be specific to centromeres. A possible hypothesis is that the CENP-B: α-satellite DNA interaction creates a structure of regularly spaced "folds" or domains upon which the kinetochore forms. However, this model does not account for the presence of the kinetochore on only a limited portion of the large α-satellite array which is present at the centromere.

The organization of the DNA at the primary constriction has been elucidated from cloned fragments since the amount of α-satellite DNA present is so large. One model of centromere organization involves other satellite sequences which are present at the centromeres although not in as great an abundance as α-satellite. Arrays of β-satellite (Waye and Willard, 1989) and satellites I, II and III (Prosser *et al.*, 1986) are also found at the centromeres of human chromosomes. Grady *et al.* (1992) studied the conservation of a sequence (GGAAT)n which forms the core of satellites II and III. They isolated the (GGAAT)n sequence on the basis of its evolutionary conservation. This repetitive sequence hybridized to chicken, maize, *Drosophila* and sea urchin DNAs as well as to orang-utan and human DNA. The α-satellite sequence is not conserved to this extent (Willard, 1990). Major clusters of this sequence were located at the centromeres of both human and *Drosophila* chromosomes. Evidence of G:A pairing was also found which could create unusual 3-dimensional structures from this repeat sequence. Plasmids carrying the repeat arrays bound proteins from nuclear extracts in DNA-mobility-shift assays. It has been hypothesized that simple sequence repeats such as those described by Grady *et al.* are essential for the formation of the kinetochore. The chromatin folding in the centromeric region could position these simple sequences on the outer surface of the chromosome where they may interact with kinetochore proteins.

1.4.4 Reintroduction of centromeric DNA

Recently, experiments have been carried out on the transfection of α-satellite DNA into mammalian cells to investigate centromere formation (Haaf *et al.*, 1992). African Green monkey (AGM) cells were transfected with a 15.9kb fragment of human α-satellite DNA from chromosome 17 cloned into a cosmid. Analysis of one of the transfected cell lines revealed that the α-satellite fragment had integrated into some of the chromosomes. At multiple integration points, the satellite DNA had been amplified into large blocks of 20-200kb in length. The total amount of introduced α-satellite DNA was estimated to exceed several thousand kilobases. Large arrays of human α-satellite DNA had formed in chromosomes which already contained normal AGM centromeres and in microchromosomes present in the cells which were acentric. The sites of integration could be identified by *in situ* binding of CREST sera and hybridization to DNA probes of the α-satellite repeat.

The chromosomes which contained both normal AGM centromeres and integrated α-satellite DNA showed bridging separating sets of chromatids during anaphase. (When chromatids are attached simultaneously to opposite spindle poles, they appear as anaphase bridges under cytological examination.) Both phenomena are symptoms of malsegregating chromosomes. It could be argued that the mitotic instability of chromosomes containing transfected α-satellite DNA may be due to inherent instability caused by large amounts of interstitial α-satellite DNA rather than any centromeric function of the DNA. The microchromosomes with integrated α-satellite DNA and no AGM centromere segregated faithfully through mitosis showing that there is no inherent instability of the integrated satellite DNA.

This work suggests that α-satellite DNA provides the primary sequence information for centromere protein binding and could also provide some of the functions of the centromere, notably the binding of spindle microtubules during mitosis.

1.4.5 A model for the mammalian centromere

ATP requiring activities of the centromere such as microtubule movement are probably intrinsic to the kinetochore proteins (Hyman and Mitchison, 1991). However, the centromeric DNA sequence is important otherwise the kinetochore could form at any site along the chromosome length instead of specifically at satellite DNA sequences. The only obvious common factor between the centromeres of higher eukaryotes may be the presence of repeating sub-units of repetitive sequences. An interesting model has been proposed by Zinkowski, Meyne and Brinkley (1991) which is consistent with much of what is known about centromeres and kinetochores. From their data, the kinetochore is described as being intimately connected with the chromatin fibre along the length of the satellite sequences as opposed to just sitting discretely on top of the primary constriction. Kinetochore sub-units, each of which can bind a spindle microtubule, are formed in conjunction with blocks of satellite DNA. When the chromosome fully condenses at metaphase, the sub-units also condense into the compact trilaminar kinetochore plate seen down the electron microscope. However, one discrepancy which has yet to be fully explained is that the number of microtubules bound by the mammalian kinetochore is not much more than the number bound by the *S.pombe* centromere although the apparent amount of centromeric DNA required is many times greater.

The possibility that the centromere consists of a concentration of sub-units makes feasible the isolation of a functional centromeric unit on a linear vector or artificial chromosome. The size of the mammalian centromere ranges from 300 to 5000kb as mentioned before. It would be impractical to try and create artificial chromosomes by attempting to manipulate DNA fragments of that size range. However, maybe only a few sub-units of the "full" centromere are necessary to function in a DNA vector of a limited size. Realistically, only the DNA component of the centromere could be isolated for the purposes of vector construction. The assumption would have to be made that once separated from the rest of the chromosome, a limited amount of centromeric, i.e. satellite, DNA would be capable of reconstituting a kinetochore sub-unit by binding to proteins such as CENP-B.

The bearing that the results of Haaf *et al.* (1992) have on the construction of artificial chromosomes are limited. The original transfected α-satellite 15.9kb clone was amplified into megabase-sized arrays before it showed any centromeric function. An artificial chromosome requires a working centromeric element which is ready to function as soon as it is transfected into the cell. Unanswered questions still remain concerning DNA which is essential for centromere function. The results of Warburton and Willard (1990) and Haaf *et al.* (1992) apparently demonstrate that only α-satellite DNA is required for centromere formation. However, are simple sequences which are interspersed within the α-satellite arrays also important as suggested by Grady *et al.* (1992)? Why is the CENP-B protein found at both active and inactive centromeres, but

CENP-C only at active centromeres? What is the minimum amount of DNA from the primary constriction (α-satellite and/or other sequences) that will support kinetochore formation and function? Can a putative centromeric sequence be introduced into a mammalian cell and exhibit centromeric properties, i.e. binding of spindle microtubules, without being amplified and rearranged?

1.5 CONSTRUCTION OF A LINEAR VECTOR OR ARTIFICIAL CHROMOSOME

1.5.1 Cloning yeast centromeres

The question arises whether centromeric elements from mammalian chromosomes can be isolated in the same way as they were from budding and fission yeast. Unfortunately, it is not straightforward to apply the same methodology to the more complex higher eukaryotes. Relatively simple plasmids that could replicate in *S.cerevisiae* were in use even before YACs were developed, which helped the isolation of telomeres and centromeres (Murray and Szostak, 1983). The first yeast centromere was, in fact, cloned on a circular plasmid by chromosome walking (Clarke and Carbon, 1980). The LEU2 and CDC10 genes are located on either side of the centromere of chromosome III. The intervening DNA region of 25kb was cloned by overlap hybridization into plasmids which were tested for mitotic stability. This procedure cannot be extended to mammalian chromosomes because hundreds to thousands of kilobases of repetitive DNA separate the nearest genetic markers that bound the centromeres.

Hahnenberger *et al.* (1989) designed a general scheme for the testing of large putative centromeric DNA fragments when they constructed mini-chromosomes based on YAC vectors which contained *S.pombe* centromeres. They adapted a pre-existing YAC vector, pYAC4 (Burke *et al.*, 1987), by inserting a *S.pombe* ARS element and a *S.pombe* selectable marker to create an *E.coli*: *S.cerevisiae*: *S.pombe* circular shuttle vector, pMB-1. They subsequently constructed a linear YAC library of the *S.pombe* genome using pMB-1. The studies of Fishel *et al.* (1988) had already shown that the centromere regions of *S.pombe* reside on three large *Sal*I restriction fragments. On this basis, size-fractionated *Sal*I-genomic restriction fragments were used as the source of the DNA for the library. After transformation into *S.cerevisiae*, the resulting mini-chromosomes were introduced into *S.pombe* to assay them for centromere function. The assay for mitotic stability involved negative selection of the selectable marker to look for minichromosome loss. Meiotic stability of the mini-chromosomes was also assayed by looking for 2+:2- segregation of the selectable marker in the formation of tetrad spores.

After transformation into *S.pombe*, intact minichromosomes were recovered which contained unrearranged *cen1* and *cen3* DNA fragments. Although the *cen1*-containing minichromosome was linear, only a circular *cen3*-containing minichromosome was found. Both minichromosomes, however, showed significantly lower rates of loss from the transformed cells than linearized pMB-1 which did not contain *S.pombe* DNA. (pMB-1 was only recovered in recircularized forms from *S.pombe*.)

1.5.2 Isolation of putative centromeric DNA fragments

Hahnenberger *et al.* (1989) obtained isolated centromeres from *S.pombe* by constructing a genomic library in *S.cerevisiae* YAC vectors. This procedure may not be applicable to mammalian centromeric DNA. The large arrays of repeated DNA which is found at human centromeres, such as α-satellite, show structural instability when cloned in YAC vectors and are subject to deletions (Neil *et al.*, 1990). It is possible that large arrays of repeated DNA may be more stable in the recombination-deficient strain of *S.cerevisiae*, *rad52* (Klein, 1988).

Alternatively, initial experiments may be attempted using the clones of limited size which are available in bacterial vectors such as the 15.9kb fragment of α–satellite DNA cloned in a cosmid used by Haaf *et al.* (1992) in their work. The use of smaller available cloned centromeric DNA fragments would, initially at least, bypass manipulation of YAC vectors.

1.5.3 Mammalian linear vectors

The approach taken by Hahnenberger *et al.* cannot be easily applied to the study of putative centromeric DNA fragments from human chromosomes. They constructed their own *S.pombe* genomic library in an adapted YAC vector, pMB-1, which could shuttle from *E.coli* to *S.cerevisiae* to *S.pombe*. In order to apply a similar scheme to the analysis of a mammalian centromere, an *E.coli*: yeast: mammalian shuttle vector would be required. This would entail the insertion of a selectable marker for mammalian cells (such as neomycin resistance) and an ARS element.

One major pitfall in this strategy is the lack of defined mammalian ARS elements. Although mammalian origins of replication have been described, such as the downstream region of the DHFR locus (Vaughn *et al.*, 1990; Burhans *et al.*, 1990) and the c-*myc* locus (Vassilev and Johnson, 1990), this area is still controversial. Recently, Caddle and Calos (1992) were unable to find any difference in the efficiency of replication initiation in the context of a plasmid between the DHFR region and randomly isolated fragments of human genomic DNA which had been described previously (Krysan *et al.*, 1989; Heinzel *et al.*, 1991). Caddle and Calos were also unable to detect initiation from the c-*myc* locus. Owing to the uncertainty and lack of reproducible results between different laboratories, the use of replication sequences from eukaryotic viruses may be the most reliable option in the short term.

Another problem which will have to be surmounted may be telomeres and their replication. Hahnenberger *et al.* (1989) transformed *S.pombe* with total genomic DNA from the *S.cerevisiae* transformants. They found that although the adapted YAC vectors were replicated in *S.pombe*, the frequency of recircularization was high. The presumed reason was that the *S.cerevisiae* telomeres, with which the YACs were equipped, did not efficiently act as substrates for the *S.pombe* telomerase enzyme (Sugawara and Szostak, 1986). It is highly probable that this problem also applies to the introduction of YAC vectors into mammalian cells, although integration occurs rather than recircularization (Huxley and Gnirke, 1991). The first step, therefore, is to construct a linear vector with telomeric sequences, (TTAGGG)n, which are efficient substrates for the human telomerase enzyme (Morin, 1989).

The DNA of a recombinant YAC is usually rescued from yeast as part of a total genomic DNA extract. This complication makes further manipulation of the YAC difficult. If the YAC vector approach is taken, it may be necessary to include internal human telomeres. The human telomeres could be exposed by restriction digestion of the genomic DNA extract from yeast prior to transfection into mammalian cells.

1.5.4 Circular vectors for assaying putative mammalian centromeres

Hahnenberger *et al.* were primarily looking for linear minichromosomes, but they were able to assay the circular *cen3*-containing minichromosome for centromere function as well. The other recircularized DNA molecules that were found in the *S.pombe* transformants had deleted the *cen* sequences, so these were lost from the cells at a much higher rate than the *cen*-containing minichromosomes. These recircularized and deleted forms made up most of the background of the transformants. Although these were found to be unstable, initially the rearranged forms could only be identified by pulsed field gel analysis of the extracted total genomic DNAs of the transformants. Predictions concerning a similar experiment in mammalian cells are premature, but analysis of the DNA of transformed colonies by Southern hybridization may be the only definitive way of distinguishing between linear and circular minichromosomes, and circular molecules which contain deletions. Theoretically, there may be no reason why a mammalian centromeric element be assayed for function as part of a circular rather than a linear vector.

Circular as opposed to linear vectors could be feasibly used to assay centromere function in mammalian cells which would make telomeres unnecessary. However, the circular plasmids which are used in mammalian cells such as EBV vectors tend to show stable maintenance over many generations without requiring centromeric sequences (Sugden *et al.*, 1985; DuBridge *et al.*, 1987; Krysan *et al.*, 1989), so screening for circular molecules which contain centromeres may be difficult if the centromeric DNA is deleted at a high frequency. A simplistic solution to this dilemma might be the use of a circular vector which can replicate in mammalian cells, but which lacks nuclear retention sequences such as those of EBV (Krysan *et al.*, 1989). Therefore, only circles which contain DNA fragments which functioned as centromeres would, theoretically, be retained. The drawback with this idea is the lack of defined mammalian origins of DNA replication and ARS elements (Caddle and Calos, 1992).

1.5.5 Telomere-capped linearized *E.coli*: mammalian cell shuttle vectors

The main source of large cloned human DNA fragments for analysis in linear vectors is likely to be a YAC vector genomic library. It may seem sensible, therefore, if the actual YAC vector has the potential to be converted into a mammalian artificial chromosome, for instance, by the exposure of internal human telomeres by restriction digestion.

Recently, a new system for the cloning of large (300kb) fragments of human DNA in *E.coli* was reported (Shizuya *et al.*, 1992) which involves bacterial artificial chromosomes based upon the single copy F factor plasmid. If large fragments of centromeric DNA are stable in "BAC" vectors, these may also be suitable for modification into mammalian linear vectors.

The purpose of this project is to test whether reintroduced human telomeric sequences will function, i.e. act as substrates for human telomerase, when situated at the termini of a linearized vector. The vector does not have to be YAC-based since large fragments of genomic DNA may not necessarily be cloned into it. At this time, the most suitable vector to use in this work is a circular *E.coli*: mammalian cell shuttle vector which uses EBV replication sequences.

CHAPTER 2 MATERIALS AND METHODS

MATERIALS AND METHODS

2.1 MAMMALIAN CELL CULTURE

2.1.1 Cell Line

Raji A human B-lymphocyte cell line transformed with Epstein-Barr virus (Pulvertaft, 1965) which grows in suspension. Raji cells were cultured in RPMI 1640 (+ 10% FCS) medium.

2.1.2 Maintenance of cells and cell lines in culture

Cells were grown in RPMI 1640 media (from Flow Laboratories) supplemented with 10% foetal calf serum (FCS) which had been inactivated for 30 min. Penicillin and streptomycin (Gibco Bio-cult) were added to media at 100 units/ml and 100μg/ml respectively. All cultures were grown at 37^{o}C in 25,or 80cm^2 plastic flasks (Nunc) in a 5% CO_2 atmosphere.

Cells were routinely fed or sub-cultured twice weekly. Raji cells were sub-cultured by 1:5 dilutions into fresh media. Cells were pelleted during harvesting at 800rpm in a Sorvall RT6000B centrifuge. When hygromycin was applied to cells, the selection was maintained at 200μg/ml. Fresh antibiotic was added whenever the cells were fed with fresh media. After transfected cells were plated out, they were subsequently sub-cultured into fresh flasks. This was not done, however, in the experiment which resulted in transfected cell line Raji-p220.2(19.6).

Frozen stocks were made by resuspending harvested cells in 90% FCS/10% DMSO (dimethylsulphoxide) at not less than 10^6 cells/ml. The cells were transferred to polypropylene vials in aliquots of 0.5ml and frozen overnight at -70^{o}C in a polystyrene box. The vials were then stored in liquid nitrogen the following day. Cell lines were recovered from frozen by resuspending rapidly thawed cells in 10ml of medium and seeding in a fresh tissue culture flask.

2.1.3 Synchronization of Raji cells in G2/M phase

Electroporation of cells which were synchronized at the transition from G2 to mitosis in the cell cycle was tried in order to improve the efficiency of transfection. Recent work by Teshigarawa and Katsura (1992) stated that Raji cells were transfected more efficiently if they were electroporated at this stage.

Firstly, dividing cells were synchronized by addition of thymidine to the medium to a concentration of 3.3mM. Thymidine blocks the cell cycle at the boundary between the G1 and synthesis phases (Goldstein *et al.*, 1989). The cells were exposed to thymidine for 16-18 hours to allow the block to affect most of the cells. The cells were counted at intervals over the period surrounding the cell cycle blockade and subsequent release. They were released from the blockade by being washed in PBS and then transferred to fresh medium. The cells were electroporated 6-8 hours later.

The thymidine was found to be effective by counting cells which were blocked and then released over a 5 day period. The increase in cell numbers was compared with the growth of an unblocked population of cells over the same period. This data is described in section 3.2.4.

2.1.4 Electroporation of DNA into mammalian cells

DNA vectors were introduced into cells by the procedure termed electroporation (Potter *et al.*, 1984). Cells are permeabilized by a high voltage pulse which allows the DNA to enter the cytoplasm. Attempts were made to transfect a different cell line, human embryonic kidney cells, by calcium phosphate co-precipitation early on in the project. This method was found to be deleterious to the cells, however, and was discontinued.

Cells which were doubling their numbers at a constant rate, were harvested for electroporation. After counting using a Coulter Counter, the cells were resuspended in cold PBS at about 10^6-10^7/0.5ml. 0.4ml of cells were transferred to a cold Bio-Rad "Gene Pulser" cuvette (0.4cm gap) and mixed with 10µg of the DNA to be transfected. The Raji cells were pulsed by 0.3kV with a capacitance of 500µF. The conditions for Raji cells were based upon those described by Teshigarawa and Katsura (1992) The cells were then left on ice for 10 min before being plated out in fresh prewarmed RPMI/10% FCS medium.

2.1.5 Extraction of Total Genomic DNA From Cells Grown in Culture

10^6-10^7 cells were harvested for a genomic DNA extraction. The cells were washed in PBS and resuspended in TEN (10mM Tris.HCl pH 8.0, 10mM EDTA, 10mM NaCl). SDS was added to 1% and the suspension was left to stand for 5-12 hrs. After precipitation with 2 volumes of 100% ethanol, the DNA was spooled out using a sterile loop. The DNA was partially dried under vacuum and redissolved in 1ml of TEN. Proteinase K was added to 20µg/ml and the DNA was incubated at 37°C overnight. The DNA was reprecipitated with 2 volumes of ethanol, then spooled out with a sterile inoculation loop into a microfuge tube containing 1ml of 70% ethanol. The DNA was pelleted in a microfuge and the supernatant was removed. The pellet was dried under vacuum and redissolved in TE.

2.1.6 Extraction of Low Molecular Weight DNA From Mammalian Cells

A modification of the Hirt cell lysis procedure (which was originally designed to extract extrachromosomal viral DNA) was used to extract plasmid and vector DNA from mammalian cells (Hirt, 1967). The procedure involves the preferential precipitation of high molecular weight cellular DNA along with cellular debris in the presence of SDS and potassium. The supernatant which remains contains low molecular weight DNA which was not precipitated which includes sheared genomic chromosomal DNA and transfected DNA molecules.

About 10^6 cells were harvested and resuspended in 0.8ml of Hirt solution (0.6% SDS, 10mM Tris.HCl pH 7.5, 10mM EDTA). The cells were left to stand for 15 min at room temperature. 200µl of 5M NaCl were added and the suspension was mixed well. The mixture was then tipped into a 1.5ml microfuge tube which was incubated at 70°C for 15 min. 200µl of 5M potassium acetate (pH 9.5) were added and vigorously mixed with the cells. The tube was then placed on ice for 1 hour before being centrifuged at 4°C for 20 min or until a tight pellet formed. A 100µl aliquot of the supernatant was transferred to a fresh tube to which 300µl of TE and 1ml of ethanol was added. The DNA was precipitated at -20°C for a minimum of 1 hour. After precipitation, the DNA was washed in 70% ethanol and resuspended in 60µl of TE.

2.2 BACTERIAL CELL CULTURE

2.2.1 Media and Additives

All media was sterilized by autoclaving.

L-Broth and Agar

Per litre: 10g tryptone (Difco), 5g yeast extract (Difco), 10g NaCl, 2.46g $MgSO_4$ pH 7.2. L agar contains in addition 15g agar/litre.

SOC Broth

Basic media per litre: 20g tryptone, 5g yeast extract, 0.59g NaCl, 1.8g KCl.

Before use, filter sterilized $MgCl_2$ and $MgSO_4$ were added to a final concentration of 10mM, and glucose was added to 20mM from a filter sterilized 2M stock.

Media Additives

When appropriate, antibiotics were added to media and agar. Ampicillin was used at a concentration of 50μg/ml. Kanamycin was used at a concentration of 25μg/ml.

Bacteria were grown at $37^{o}C$, with good aeration for liquid cultures. Bacterial stocks were stored frozen at $-70^{o}C$ in 15% glycerol.

2.2.2 Bacterial Strains Used

All bacterial strains used were derivatives of *E.coli* K-12.

DH5α: DH5α, a derivative of DH1 (Hanahan, 1983), is a high efficiency transformation strain for pUC-based plasmids. Its genotype is F^-, recA1, endA1, gyrA96, thi-1, hsdR17, (rk^-, mk^+), supE44, relA1, λ-, mcrB(-), Δ(argF-laczya)U169, φ80dlacZΔM15.

XL1-Blue: This strain is a host for pUC-based plasmids. Its genotype is recA1, endA1, gyrA96, thi, hsdR17, (rk^-, mk^+), supE44, relA1, λ-, (lac), {F', proAB, $lacI^qZ\Delta M15$, Tn10(tet^R)} (Bullock *et al.*, 1987).

Both DH5α and XL1-Blue are deficient in the recA recombinase enzyme which makes them more tolerant of repeated sequences cloned into plasmid vectors.

2.3 MANIPULATION OF PLASMID DNA

2.3.1 Preparation of Frozen Cells for Electroporation

1 litre of L broth was inoculated with 0.1ml fresh overnight XL1-Blue or DH5α culture and the cells were grown at 37^{o}C with vigorous shaking to an abs$_{600nm}$ of 0.5-1. The cells were harvested by being cooled for 15-30 min on ice and followed by centrifugation at 4000x g_{max} for 15 min at 4^{o}C. All subsequent steps were performed at 4^{o}C. The cells were then resuspended in 1 litre H_2O and repelleted as before. The cells were then resuspended in 500ml H_2O and repelleted as before. The cells were resuspended in 100ml H_2O and repelleted as before. Then the cells were resuspended in 20ml 10% glycerol, repelleted as before and the cells resuspended to a final volume of 2-3 ml in 10% glycerol. The cell concentration should be approximately 3x10^{10} cells/ml. 40μl aliquots were quickly frozen in dry ice and stored at -70^{o}C.

2.3.2 Electroporation of XL1-Blue and DH5α Cells

An aliquot of the cells prepared for electroporation was thawed at room temperature and then placed on ice. Plasmid DNA (<1μg in 1-2μl TE or H_2O) was added to the cells, mixed well and the mixture left on ice for 1 min. The Gene Pulser apparatus was set to 25μF and 2.5kV and the Pulse controller to 200Ω. The mixture of cells and DNA was transferred to a cold, 0.2cm gap electroporation cuvette, the cuvette placed in a cold safety chamber slide and the slide pushed into the chamber until the cuvette was positioned between the contacts at the base of the chamber. the cells were pulsed once which should produce a pulse with a time constant of 4.5 to 5 msec (the field strength was 12.5kV/cm). After removal of the cuvette from the chamber, 1ml of SOC was immediately added and the cells quickly resuspended with a pastette. The cell suspension was then transferred to a 17x100mm polypropylene tube and incubated for 60 min at 37°C with shaking at 225 rpm. The cells were plated out onto L agar containing the necessary selection.

2.3.3 Large-scale Plasmid Preparation

Plasmid DNA was isolated by a modified version of the alkaline-lysis method of Birnboim and Doly (1979).

500ml broth containing the necessary antibiotic selection was inoculated with a single colony and the culture grown overnight. The cells were pelleted by centrifugation at 5000rpm for 10 min at 4°C and resuspended in 20ml solution I (25% glucose, 50mM Tris-HCl pH 8.0, 10mM EDTA pH 8.0) and lysozyme was added to 0.5mg/ml. After 10 min at room temperature, 40ml fresh solution II (0.2M NaOH, 1% SDS) was added and mixed gently. After 5 min on ice, 20ml of cold solution III (3M potassium acetate pH 4.8 with acetic acid) was added and the lysate left on ice for a minimum of 15 min. Cell debris was pelleted at 5000rpm for 5 min and the supernatant strained through Whatman (113^{V}) folded filter paper. DNA was precipitated by the addition of 0.6 vols. of propan-2-ol and collected by pelleting at 4000rpm for 10 min at 4°C. The pellet was washed with 70% ethanol and resuspended to 13ml with TE. 14.58g caesium chloride and 1.5ml of 10mg/ml ethidium bromide (EtBr) were added. The plasmid DNA was banded by centrifugation in an 18ml polyallomer tube at 40000rpm, 20°C overnight in a vertical rotor (which 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 EtBr which can be intercalated into the molecule. The plasmid band was collected under UV light (300nm) with a syringe and the EtBr removed by a series of extractions with CsCl-saturated-propan-2-ol until no colour remained in the aqueous phase. The DNA was then precipitated with 3 vols. of 70% ethanol.

2.3.4 Small-scale Plasmid Preparations

4ml of a fresh overnight culture was pelleted by centrifugation at 3000rpm for 5 min and resuspended in 0.6ml of STET (8% sucrose, 5% Triton X-100, 50mM EDTA pH 8.0, 50mM Tris-HCl pH 8.0). After being transferred to a microfuge tube, a few flakes of lysozyme were added and the cells left on ice for 3 min. The tube was then placed open-capped in a 95°C waterbath for 3 min. The lysate was placed on ice for 5 min before pelleting the cell debris by centrifugation for 15 min at 4°C. The pellet was removed using a toothpick. An equal volume of 5M lithium chloride was added and the tube was incubated on ice for 5-10 min. Precipitated debris was pelleted by centrifugation for 10 min after which the supernatant was transferred to a fresh tube. The DNA was precipitated by the addition of a 0.6 vol. of cold propan-2-ol followed by centrifugation for 10 min. The DNA was washed in 70% ethanol, dried under vacuum and resuspended in 100μl of TE.

2.4 CALCULATION OF DNA CONCENTRATIONS

Plasmid DNA concentrations were calculated from the absorbance reading at 260nm on a spectrophotometer.

10μl of the DNA sample were diluted 1/50 in 500μl of dH_2O in order to take absorbance readings. The reading was multiplied by a factor of 2.5 to obtain the concentration in μg/μl.

1 abs_{260} unit = 50μg DNA/ml

abs. reading = conc. x dilution factor / 50μg DNA/ml

=>abs.reading x 1/dilution factor x 50μg DNA/ml = conc.

=>abs. reading x 50 x 50 = conc. μg/ml => **abs. reading x 2.5 = conc. μg/μl**

The absorbance readings were not accurate enough for estimating the small amounts of DNA required for random labelling reactions. However, λ control DNA was supplied in the Boehringer Mannheim random prime labelling kits at 12.5ng/μl. In order to estimate the 25ng of DNA required for a labelling reaction, probe DNA was compared to 2μl of the supplied λ DNA by agarose gel electrophoresis.

2.5 MANIPULATION OF DNA BY ENZYMES

2.5.1 Restriction endonuclease digestion

Digests were carried out according to the manufacturers' specifications using the A, B, H, L and M buffers provided by Boehringer Mannheim, except when a special buffer was required.

Genomic digests were carried out overnight with 2-4 units of enzyme/μg DNA. Reactions were stopped by addition of stop buffer.

2.5.2 Dephosphorylation of DNA

To prevent unwanted ligation of certain DNA fragments, 5'- phosphate groups were removed by bacterial alkaline phosphatase (BAP). 1 unit of BAP was incubated with digested DNA at 65°C for 30 min. To remove the BAP, proteinase K was added to 100μg/ml and the incubation continued for 30 min at 50°C. The DNA was extracted with phenol and then chloroform. The DNA was precipitated and centrifuged, the pellet washed in 70% ethanol and the DNA resuspended in TE.

2.5.3 Ligation of DNA molecules

T4 DNA ligase, isolated from *E.coli* infected with bacteriophage T4, catalyses the formation of a phosphodiester bond between 3'-OH and 5' phosphate groups of DNA (Weiss *et al.*, 1968). This enzyme can, therefore, be used to join complementary cohesive termini of dsDNA molecules. Ligations were carried out in 50mM Tris-HCl pH 7.5, 10mM $MgCl_2$, 1mM spermidine, 100μg/ml BSA, 1mM ATP, 10mM DTT. Reactions were carried out using 1 unit of enzyme at 14°C.

2.5.4 Ligation of Linkers

Oligonucleotides were supplied after synthesis in an ammonium hydroxide solution. To precipitate the DNA, 35μl of 3M NaOAc and 700μl of cold ethanol was added to 350μl and the mixture left for 30 min at -20°C. The DNA was then pelleted by centrifugation for 10 min and the pellet washed twice with 80% ethanol before resuspension of the DNA in 200μl of TE. The DNA concentration was estimated by measuring the absorbance at 260nm.

If necessary, the oligonucleotide linkers could be phosphorylated using the enzyme T4 polynucleotide kinase (PNK). PNK can catalyse the transfer of the γ-phosphate of ATP to a 5'OH terminus of DNA or RNA (Richardson, 1971). 1μg of single-stranded linkers were incubated in 1x

PNK buffer (50 mM Tris.HCl pH 7.6, 10mM $MgCl_2$, 5mM DTT, 0.1mM spermidine, 0.1mM EDTA pH 8.0), 10mM ATP and 10 units of PNK at 37°C for 30 min.

Oligonucleotides were annealed by first heating to 80°C for 10 min, then incubated at room temperature for 60 min. About 1µg of annealed linkers were added to the ligation reaction in order to provide a large excess of linker ends. Ligations were carried out as described above.

2.5.5 Filling of recessed termini to create blunt-ended molecules

In order to ligate DNA fragments with incompatible cohesive ends, the termini were converted to blunt ends by filling the recessed ends. Usually, the reaction was carried out in the same reaction mix as the restriction digestion. 4µl of each dNTP (from a 0.5mM stock) was added to the reaction (the amount of each dNTP in the reaction was 2pmol). 1 unit of Klenow enzyme was added and the reaction was incubated at room temperature for 30 min. The reaction was stopped by incubation at 65°C for 5 min.

2.6 SEPARATION OF DNA BY ELECTROPHORESIS

DNA fragments were separated by size by electrophoresis through agarose gels. Usually stop mix was added to DNA samples prior to loading onto the gel. This stop mix acts as a visible marker and migrates in front of the DNA. After electrophoresis, gels were placed on a UV transilluminator and photographed using a Polaroid MP4 Land camera fitted with a red filter using Kodak TMAX Professional film 4052 with an exposure time of 15 sec. The film was developed using a X1 X-OGRAPH automatic film processor.

2.6.1 Solutions and Buffers

50x TAE

Per litre: 242g Tris base, 57.1ml glacial acetic acid, 100ml 0.5M EDTA (pH 8.0)

5x TBE

Per litre: 54g Tris base, 275g boric acid, 20ml 0.5M EDTA (pH 8.0)

10x stop mix: 15% Ficoll, 0.5% bromophenol blue dye, 200mM EDTA (pH 8.0).

2.6.2 Conventional Agarose Gel Electrophoresis

Conventional horizontal agarose gel electrophoresis was used to separate DNA fragments of up to 50kb. Agarose concentrations of between 0.7% and 2% were used depending upon the size range of fragments to be resolved. Gels were made and run in 0.5x TAE buffer, and stained with ethidium bromide before photography. The gels were run at various voltages and for various times as required.

2.6.3 Polyacrylamide Gel Electrophoresis

Sequencing reactions were run on 5% polyacrylamide gels. The gels were cast and run in a Bio-Rad sequencing apparatus which was assembled, and the gel poured, following the manufacturer's recommendations. The glass plates were cleaned thoroughly before use and the bottom plate was siliconized. The glass plates used were 21 x 53cm and were separated by a wedge spacer which was 0.25mm thick at the top and 0.75mm at the bottom. Pre-weighed and pre-mixed 19:1 acrylamide/Bis was used to make a 30% stock. For a 5% gel, a 100ml volume was mixed which also contained urea to 7M. 130µl of 10% (w/v) ammonium persulphate and 50µl of TEMED (NNN'N'-tetramethylethylethlenediamine) were added to a 10ml aliquot which was used to set the base of the gel following the manufacturer's instructions. When the base of the gel had set, 560µl of ammonium persulphate and 28µl of TEMED were added to the remaining 90ml of acrylamide which was then poured immediately. After the gel had set (about 1 hour), it was pre-run in 1x TBE for about 30 min until the gel temperature had stabilized at 50^{o}C. The gels were run at 50W for the required time. After running, the gels were fixed in 10% acetic acid, 10% methanol for 20 min to remove the urea before drying down under vacuum, and autoradiography.

2.6.4 Preparative Agarose Gel Electrophoresis

To isolate specific fragments of DNA for cloning or for radiolabelling, DNA was separated on agarose gels as described in 2.6.2 except that low melting agarose was used. The gel slice containing the required DNA fragment was cut out under long-wave UV light and the DNA purified using the method of Heery *et al.* (1990). The gel slice was placed in a punctured 0.5ml microfuge tube that had been plugged with 4mm of glass wool. This was placed inside a 1.5ml microfuge tube and centrifuged for 10 min at 6000rpm (or the low speed setting) in a microfuge. The eluate which contained the DNA was collected in the bottom tube. The DNA was precipitated with a 0.6 vol. of propan-2-ol, washed with 70% ethanol and resuspended in TE.

2.7 SOUTHERN TRANSFER OF DNA

DNA was transferred to nylon membranes (Hybond-N) using the method of Southern (1975). Agarose gels were subjected to 60 mJ of UV light in order to introduce thymidine dimers into the DNA which increases the efficiency of transfer of large DNA molecules. After photography, gels were denatured by shaking in denaturing solution (1.5M NaCl, 0.5M NaOH) for 60 min and then neutralized by shaking in neutralizing solution (3M NaCl, 0.5M Tris.HCl pH 5.0) for a further 60 min. The DNA was capillary blotted onto nylon membrane with 20x SSC (0.3M Tri-sodium citrate, 3M NaCl pH 7.0). The gel was placed onto Whatman 3MM paper supported on a persphex sheet with the paper trailing into a reservoir of 20x SSC below. A piece of nylon membrane, pre-wet with 2x SSC and cut to the exact size of the gel, was placed on top, care being taken to ensure that no air bubbles were trapped between the gel and the membrane. Two pieces of 3MM pre-wet with 2x SSC were placed on top followed by a stack of dry paper towels and finally a persphex plate which was weighted down. After transfer overnight, the nylon membrane was washed in 2x SSC and allowed to air dry. The DNA was covalently bound to the membrane by baking for two hours at $80^{o}C$ under vacuum.

Probe was removed from nylon membranes by pouring a boiling solution of 0.1mM EDTA, 0.1% SDS over the membrane and allowing to cool to room temperature.

2.8 RADIOLABELLING OF DNA

2.8.1 Random Priming

DNA probes were labelled by random priming using the method of Feinberg and Vogelstein (1983 and 1984). Random hexanucleotides were annealed to denatured probe DNA. The hexanucleotides can then act as primers for the synthesis of the complementary strand by Klenow enzyme from the 3' OH termini. The reaction is carried out in the presence of ^{32}P-labelled dCTP and unlabelled dATP, dGTP and dTTP. Subsequent heat denaturation produces single-stranded radiolabelled DNA ready for hybridization.

DNA for random priming was dissolved in TE. A commercially available random priming kit from Boehringer Mannheim was used. 25ng DNA, denatured by boiling for 10 min, was added to a reaction mixture containing dATP, dGTP and dTTP (all at 25μM), buffer and random hexanucleotides. 5U of Klenow enzyme were added, followed by 30μCi α-^{32}P dCTP (3000Ci/mmol, 10mCi/ml). The reaction volume was 20μl. After incubation at 37°C for 2-5 hrs, the percentage incorporation of radioactive nucleotide into the DNA was estimated from the proportion of counts precipitated onto a Whatman GF/A filter by 10% TCA which quantitatively precipitates oligonucleotides of >20 bases.

Proteins and unincorporated nucleotides were removed from radiolabelled probes by passing the probe down a commercial Pharmacia Sephadex G-50 column (following the manufacturer's instructions), DNA being excluded from the gel matrix (Sambrook *et al.*, 1989).

2.8.2 End-labelling

Oligonucleotides could be end-labelled using the same reaction as that used to phosphorylate linkers.

The reaction mix consisted of 5μg of oligonucleotide, 1x PNK buffer (50mM Tris.HCl pH 7.6, 10mM $MgCl_2$, 5mM DTT, 0.1mM spermidine, 0.1mM EDTA pH 8.0), 20μCi γ-^{32}P ATP and 10 units of PNK in a total volume of 20μl. Incubation was at 37°C for 30 min.

2.9 NUCLEIC ACID HYBRIDIZATION

2.9.1 Hybridization solutions

Solution A

0.5M Na_2HPO_4 pH 7.2 with orthophosphoric acid, 7% SDS, 0.5% dried milk powder, 10mM EDTA (amended from the hybridization solution described in Church and Gilbert, 1984). This solution was used for hybridizations to Hybond N using random primed DNA probes. 10µg/ml denatured salmon sperm DNA were added to the solution as a competitor when Hybond N+ was used.

"Quick Hyb"

0.1% SDS, 0.1% sodium pyrophosphate, 0.05% BSA, 0.05% PVP, 0.05% Ficoll, 5x SSC. This solution was used for hybridizations using radiolabelled DNA oligonucleotides.

2.9.2 Hybridization Protocols

Membranes were hybridized in glass Hybaid cylinders or sealed plastic bags overnight in the appropriate solution containing radiolabelled probe. Membranes were prehybridized for a minimum of 4 hrs before addition of probe. Prehybridization and hybridization were carried out at 65-68^{o}C for random primed DNA probes and at 50^{o}C for oligonucleotide probes.

After hybridization, membranes were washed to remove non-specifically bound probe. When using random primed probes, membranes were washed once at room temperature in 2x SSC, 0.1% SDS for 5 min, once in 2x SSC, 0.1% SDS at 65^{o}C for 30 min and two times in 0.1x SSC, 0.1% SDS for 30 min at 65^{o}C. When using end-labelled oligonucleotide probes, membranes were washed once at room temperature for 30 min in 4x SSC followed by a stringent wash for 30 min at 65^{o}C in 4x SSC, 0.1% SDS. After washing, membranes were placed in sealed plastic bags ready for autoradiography.

2.10 AUTORADIOGRAPHY

Autoradiography was carried out using X-ray film (Kodak X-OMAT) in cassettes which contained intensifying screens except when exposing sequencing gels. Generally, membranes were exposed overnight at -70^{o}C in the first instance and further exposures done for different lengths of time as required. Sequencing gels were exposed overnight at room temperature. The X-ray films were developed in the automatic X-ray film processor.

2.11 SEQUENCING

DNA was sequenced using the dideoxy chain termination reaction (Sanger *et al.*, 1977) with the commercially available T7 Sequencing kit (Pharmacia). Double-stranded sequencing was carried out using the following protocol.

2μg of plasmid DNA were suspended in 8μl of TE to which 2μl of 2M NaOH were added and mixed. After being incubated for 10 min, the DNA was precipitated by the addition of 3μl of 3M sodium acetate, 7μl of dH_2O and 60μl of 100μl 100% ethanol. The DNA was pelleted by spinning in a microfuge for 5 min and then washed with 70% ethanol. The pellet was dried under vacuum and dissolved in 7μl of dH_2O. 1μl of primer DNA and 2μl of reaction buffer (from the kit) were then added to the 7μl of denatured DNA. The mix was incubated for 5 min at 65^{o}C and allowed to cool to room temperature slowly.

Reaction termination tubes were set up by placing 2.5μl of each dideoxy nucleotide from the kit into separate tubes (each labelled accordingly) which were left at 37^{o}C to prewarm.

To the 10μl of annealed DNA, the following were added: 1μl of 0.1M DTT; 2μl of labelling mix (stock solution in kit diluted 1/5); 12μCi of ^{35}S-dATP; and 2μl of Sequenase enzyme (diluted 1/8 using the enzyme dilution buffer from the kit). The mixture was incubated at room temperature for 2-5 min.

3.5μl of the labelling reaction were added to each termination tube, mixed and incubated for 5 min at 37^{o}C. The reactions were stopped by the addition of 4μl of stop solution from the kit. The terminated reactions were heated to 75^{o}C before loading onto a polyacrylamide gel to denature the DNA.

CHAPTER 3 RESULTS

3.1 VECTOR CONSTRUCTION

3.1.1 Introduction

An EBV-based plasmid, p220.2 (figure 3.1), was chosen as the basis for a linear vector for mammalian cells. p220.2 (obtained from Liz Thompson, Experimental Studies Section, MRC Human Genetics Unit) and has been described in published literature (DuBridge *et al.*, 1987; Krysan *et al.*, 1989).

p220.2 is a 9kb plasmid which is a bacterial-mammalian cell shuttle vector. It contains the *EBNA-1* gene and the *oriP* element from the EBV genome as well as a bacterial replication origin and the ampicillin resistance selectable marker (Ap^R). A selectable marker for mammalian cells is also present in the form of the hygromycin resistance gene (*hph* - hygromycin phosphatase) using the herpes simplex virus *thymidine kinase* (*tk*) promoter and termination sequences. The restriction site polylinker from pUC12 is present in the *tk* termination sequences. p220.2 is reportedly competent to replicate in immortalized human cells (Haase and Calos, 1991; Krysan and Calos, 1991).

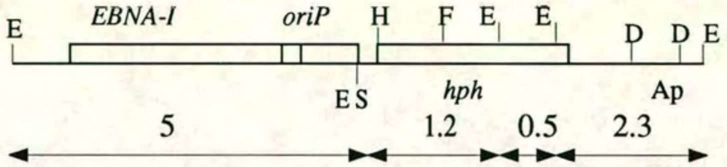
It was proposed to modify p220.2 into a precursor of a linear vector by the introduction of human telomeric sequences such that an inverted repeat was formed by the two opposing tracts (illustrated in the structure of p220.2TC in figure 3.1). The initial source of the telomeric repeats was the telomeric DNA subclone, pHutel-2-end (Cross *et al.*, 1990).

The two telomeric sequences when adjacent to each other in opposing orientations make up an inverted repeat which normally is unstable in bacteria which have functional recombination enzymes. The most probable reason for this instability may be the ability of inverted repeats, or perfect palindromes, to form cruciform structures. It is thought that cruciforms are recognized by the bacterial recombinases which then excise such structures from DNA molecules (Leach and Stahl, 1983).

In order to select for retention of the telomeric DNA, a selectable marker (kanamycin resistance - Kan^R) was used as the stuffer fragment. The stuffer fragment served two purposes: (1)

Figure 3.1 Restriction sites of p220.2, p220.2TC, pET and pIT

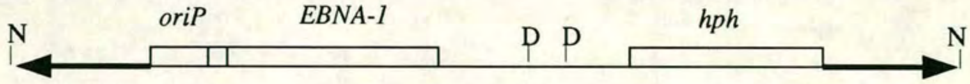
p220.2 9kb (linear restriction map)



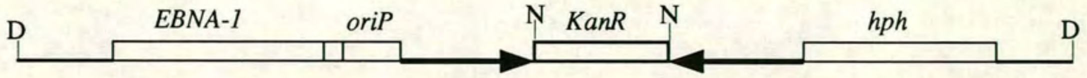
the sizes of the *Eco*RI restriction fragments of p220.2 are given in kb

p220.2TC ~11kb (linear restriction map)

pET ~10kb



pIT ~10.5kb



B - *Bam*HI C - *Cla*I D - *Dra*I E - *Eco*RI F-*Sfu*I
H - *Hin*dIII N- *Not*I S- *Sal*I X - *Xho*I

Bold arrows indicate telomeric sequences.

the fragment physically separated the two tracts of telomeric DNA, thereby reducing the possibility of a cruciform structure and subsequent deletion, and (2) if deletions still occurred, the presence of the kanamycin resistance marker should select plasmids in the population which retained the cassette.

A major reason for using a bacterial mammalian shuttle vector is that, despite the difficulties of constructing a stable molecule in *E.coli*, large amounts of highly purified plasmid vector can be obtained by relatively simple DNA extraction methods. The efficiency of transfection of mammalian cells often depends upon the purity and concentration of the DNA.

The plasmid can be linearized by an appropriate restriction digestion to expose the telomeric sequences before transfection into mammalian cells. Other vector molecules would also be transfected into the cells as controls. The controls would include the original unmodified circular plasmid, the circular plasmid with internal telomeric sequences and a linear plasmid which possessed telomeric sequences at internal sites.

Attempts were also made to convert another EBV-based plasmid, pLIB16, into a linear vector. pLIB16 was a gift from Michele Calos (Krysan *et al.*, 1989). pLIB16 (23kb in size) differs from p220.2 in that the *oriP* element contains a deletion which destroys the ability of *oriP* to initiate replication. pLIB16 also contains a 14kb human genomic DNA fragment which has been cloned into the *Hin*dIII site of the polylinker. The human DNA fragment has been reported to act as an ARS element in immortalized human cells (Krysan *et al.*, 1989). Therefore, a linear vector based upon pLIB16 might demonstrate replication from a human ARS element along with the possession of telomeric sequences. It was not known if the EBV origin was able to initiate replication in a linear DNA molecule. The work of Krysan and Calos, (1991) found evidence that putative ARS elements such as that in pLIB16 are able to initiate replication bubbles in the surrounding vector sequences as well as within its own DNA. This property might be able to bypass any possible problems with using the *oriP* in a linear vector.

3.1.2 Construction of pEND2

The plasmid, pEND2, was the source of telomeric DNA fragments in the construction of linear vectors. pEND2, which was constructed by Dr Cooke, is a derivative of pHutel-2-end which contains a ~2.0kb telomeric DNA fragment (figure 3.2 a and b, lane 1; Cross *et al.*, 1990). This fragment contains over 1kb of subtelomeric DNA which were removed to facilitate the manipulation of the telomere repeats. The subtelomeric DNA consists of a complex repeat unit which contains a *Hae*III site (Cross *et al.*, 1990). Restriction site mapping and DNA sequencing had shown that a fragment of ~0.5kb was present which consisted solely of TTAGGG repeats. The 0.5kb telomeric fragment was isolated by cleaving pHutel-2-end with *Eco*RI and *Hae*III and identified by comparison with pGEM7 which had been similarly digested (figure 3.2a, lanes 2 and 3). The *Eco*RI-*Hae*III restriction pattern of pHutel-2-end differs from that of pGEM7 by the presence of additional bands. By comparison with an appropriate DNA size marker, the 0.5kb fragment of TTAGGG repeats could be identified. Southern blot analysis of the gel to a (TTAGGG)n DNA probe showed that the correct fragment had been sub-cloned. None of the other *Eco*RI-*Hae*III fragments hybridized to the oligonucleotide probe. After excision from an agarose gel, the fragment was ligated to pGEM7 (which had been cleaved with *Eco*RI and *Sma*I) to create pEND2 (figure 3.2a, lane 4). The telomeric DNA fragments of pHutel-2-end and pEND2 are compared in figure 3.2.a, lanes 1, 3 and 4, and 3.1b, lanes 1, 2 and 3.

3.1.3 Sequence of the (TTAGGG)n fragment (figure 3.3)

The telomeric DNA fragment of pEND2 was sequenced in order to verify the restriction sites present at the junctions with the polylinker on either side. In addition, the number of TTAGGG repeats had yet to be found. The fragment had been inserted into the polylinker sequence of the plasmid vector pGEM7 which is bounded by binding sites for the "forward" and "reverse" M13 sequencing primers which enabled sequencing of the subcloned DNA fragment from both directions. The telomeric fragment was inserted such that the 3' end of the G-rich strand was next to the *Eco*RI site of the polylinker. The other end of the fragment had been blunt-end-ligated into the *Sma*I site.

Figure 3.2 **The construction of pEND2 (carried out by H.Cooke)**

The ~2.0kb *Xho*I-*Bam*HI DNA fragment in pHutel-2-end, shown in lane 1, contained subtelomeric sequences as well as TTAGGG-like repeats.

(a) Restriction enzyme digestion patterns of plasmids used in the construction of the telomeric DNA clone, pEND2, are shown in an ethidium bromide-stained agarose gel. The 0.5kb *Hae*III-*Eco*RI (TTAGGG)n fragment in Hutel-2-end (lane 3) was identified by its absence from the pattern of a similar digestion of pGEM7 (lane 2). Details of the identification of the fragment are described in the text. Digestion of pEND2 shows that the 0.5kb fragment was successfully subcloned into pGEM7 without subtelomeric sequences (lane 4). (After excision of telomeric DNA sequences, the pGEM7 vector DNA is seen as a 3kb band in lanes 1 and 4.)

(b) The autoradiograph shows the same gel* blotted onto Hybond membrane and hybridized to a $(TTAGGG)_4$ oligonucleotide probe labelled with γ-^{32}P-ATP. The DNA fragments which contain the TTAGGG repeat on the gel (a) are indicated by arrowheads. The exposure time was 1 hour.
The fragment sizes of the λHHR DNA size marker are indicated.

Legend

(a)

1. pHutel-2-end digested with *Xho*I and *Bam*HI
2. pGEM7 digested with *Eco*RI and *Hae*III
3. pHutel-2-end digested as above
4. pEND2 digested with *Xho*I and *Hin*dIII
5. λHHR size marker (λ DNA digested with *Hin*dIII plus λ digested with *Eco*RI and *Hin*dIII)

(b)

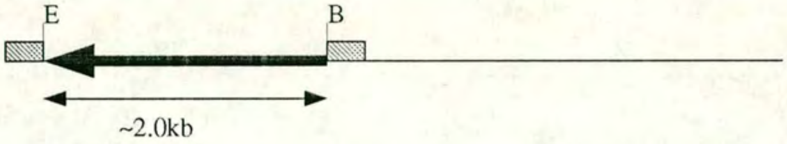
1. pHutel-2-end digested with *Xho*I and *Bam*HI
2. pHutel-2-end digested with *Eco*RI and *Hae*III
3. pEND2 digested with *Xho*I and *Hin*dIII

*Figure (b) is reproduced at a different magnification to figure (a).

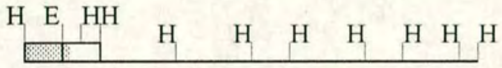
Figure 3.2c Linear restriction maps of pHutel-2-end, pGEM7 and pEND2

The shaded box indicates the polylinker region. The T7 promoter of the pGEM7 sequence is at the leftmost position of the maps.

1. pHutel-2-end *Eco*RI (E) and *Bam*HI (B) sites

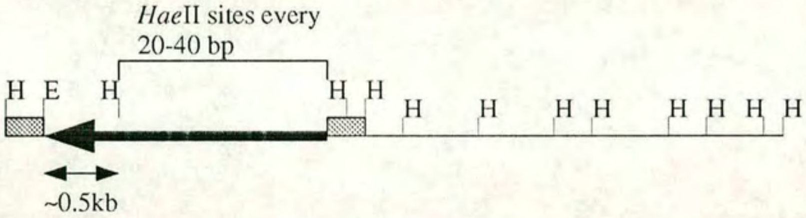


2. pGEM7 *Eco*RI (E) and *Hae*III (H) restriction sites



Only the major HaeIII fragments have been shown. The fragments which are visible in figure 3.2 have the calculated sizes: 657, 458, 434, 328, 289, 267, 174, 142 and 102bp. (Sizes were obtained from the Promega catalogue restriction map of pGEM7.)

3. pHutel-2-end *Eco*RI (E) and *Hae*III (H) restriction sites



4. pEND2 *Xho*I (X) and *Hin*dIII restriction sites

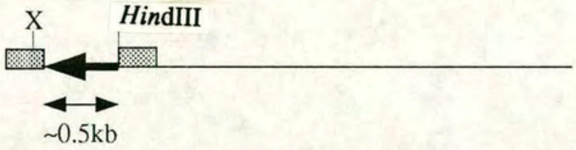


Figure 3.3 Sequence of the 0.5kb (TTAGGG)n fragment from pEND2

*Xba*I *Xho*I *Eco*RI Yeast repeats TTAGGG-like repeat
5' **TCTAGA CTCGAG GAATTC** GCCACA CCCCAC ACCTAA 3'

(a) The sequence of the polylinker-(TTAGGG)n fragment junction obtained using the "forward" sequencing primer. This direction leads into the CA-rich strand of the telomeric sequence.

*Bam*HI *Hin*dIII *Cla*I *Sfu*I Sub-terminal sequence

5' **GGATCC AAGCTT ATCGAT TTCGAA** CCCCC GCGCCG CCTTTG CGAGGG

TGGAGT TGCC

TTAGGG TTAGGG TTAGGG TTAGGG TTAGGG TTAGGG TTAGGG TTAGGG

TTAGGG TTAGGG TTAGGG TTTAGGG TTAGGG TTAGGG TTAGGG TTAGGG

TTAGGG TTAGGG TCAGGG TCAGGG GTAGGG TCAGGG GTAGGG TCAGGG

GTAGGG GTAGGG GTAGGG TCAGGG TTAGGG TTAGGG TTAGGG TTAGGG

TTAGGG TTAGGG TTAGGG TCAGGG TTAGGG TTAGGG TTAGGG NNAGGG

GTAGGG GTAGGG GTAGGG TTAGGG TTAGGG TTAGGG TTAGGG TTAGGG

TTAGGG TTAGGG TTAGGG GTAGGG GTAGGG GTAGGG TTAGGG TTAGGG

TTTAGGG TTAGGG TTAGGG TTAGGG TTAGGG TTAGGGG TTAGGG TTAGGG

TTAGGG TTAGGG TTAGGG TTAGGG TTAGGG TTAAGGG TTAAGGG

Yeast repeats *Eco*RI
TTAGGT GTGGGG TGTGGC **GAATTC** 3'

(b) The sequence of the pEND2 telomeric DNA insert obtained using the reverse sequencing primer. This direction provides the G-rich TTAGGG sequence but the actual data was obtained from overlapping sequences from both forward and reverse directions. TTAGGG repeats are double-underlined, remaining subterminal telomeric sequences are underlined and restriction sites are indicated by bold type. (EMBL accession number: Z19866.)

It was important to know the orientation of the cloned telomeric sequence so that during construction of a linear vector the telomeric fragments would retain the same orientation as chromosomal telomeres, i.e. the 3' end of the G-rich strand is the most distal part of the chromosome.

The telomeric DNA fragment is 524bp long and consists mainly of TTAGGG repeats (figure 3.3). However, point mutations which were probably present in the DNA prior to the initial cloning from human cells have created a large number of variants such as TCAGGG and GTAGGG. The G-rich strand runs 5' to 3' towards the *Eco*RI site of the polylinker, therefore the "*Eco*RI end" of the fragment should be exposed in a proposed linear vector (figure 3.1 220.2TC). The 12bp of yeast DNA are presumed to be a remnant of the yeast telomeric repeats added during the original isolation of the human telomeric fragment (Cross *et al.*, 1989).

3.1.4 Construction of p220.2TC

The plasmid, p220.2TC was constructed by inserting a "telomere cassette" into the existing vector, p220.2 (DuBridge *et al.*, 1987; Krysan *et al.*, 1989). The telomere cassette is an arrangement of two telomeric sequences in opposing orientations which border a removable stuffer fragment (see figure 3.1 p220.2TC).

The source of the Kan^R gene for the telomere cassette was the plasmid pUC4KN which is a derivative of pUC4K (from Pharmacia). Both plasmids contain the gene within a symmetrical polylinker. To create pUC4KN, the *Pst*I sites in pUC4K were replaced by *Not*I sites via the insertion of linker oligonucleotides.

The sequence of the oligonucleotide was 5' AGCGGCCGCTTGCA 3'. The oligonucleotide was able to self-anneal to produce a linker which contained a *Not*I site bounded by cohesive ends which were complementary to those produced by *Pst*I digestion. The annealed linker is shown below (the *Not*I site is in bold face):

```
5'         AGCGGCCGCTTGCA 3'
3' ACGTTCGCCGGCGA         5'
```

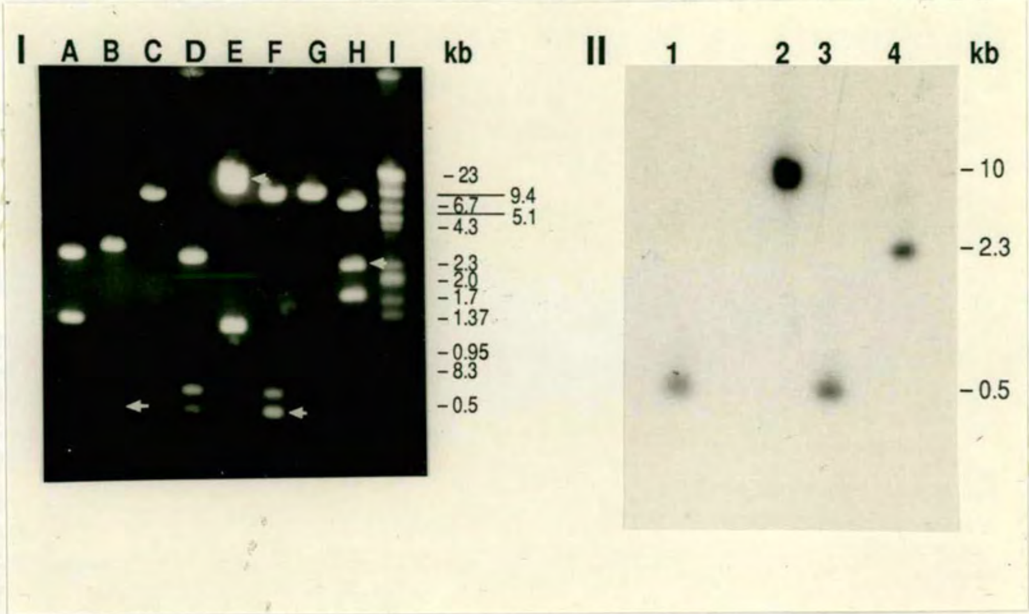
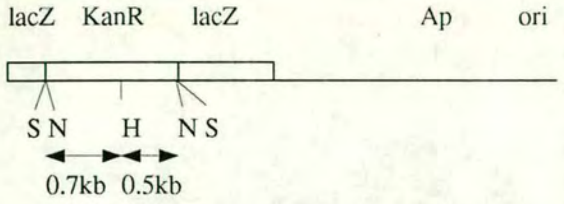
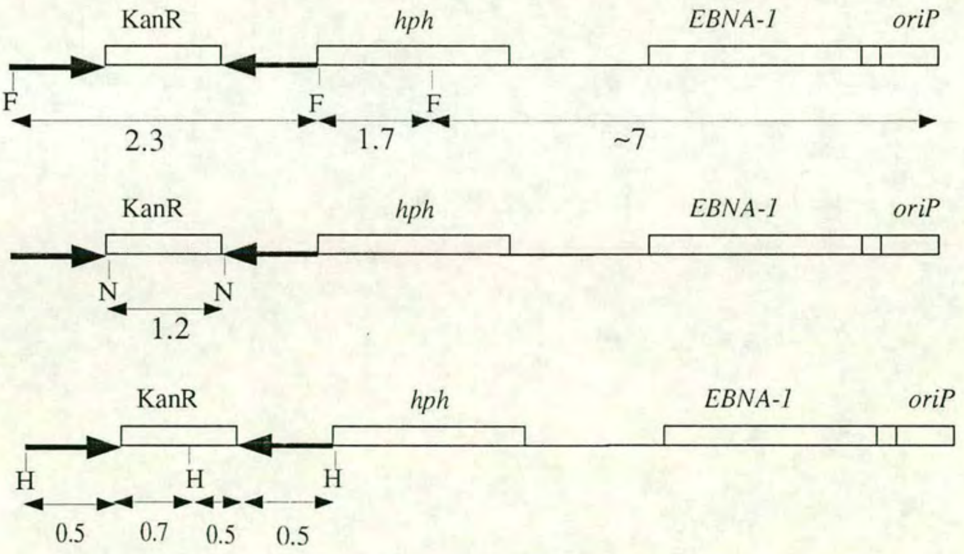

I
A
B
C
D
E
F
G
H
I
kb
- 23
- 6.7
9.4
5.1
- 4.3
- 2.3
- 2.0
- 1.7
- 1.37
- 0.95
- 8.3
- 0.5
II
1
2
3
4
kb
- 10
- 2.3
- 0.5

Figure 3.4 III Linear restriction sites of pUC4KN and p220.2TC

pUC4KN 3.9kb: restriction sites are shown for *Sal*I (S), *Not*I (N) and *Hin*dIII (H).



p220.2TC ~11kb (linear restriction maps): restriction sites are shown for *Sfu*I (F), *Not*I (N) and *Hin*dIII (H). Fragment sizes are shown in base pairs (not to scale).



B - BamHI C - ClaI D - DraI E - EcoRI F-SfuI
H - HindIII N- NotI S- SalI X - XhoI

Bold arrows indicate telomeric sequences.

The presence of the rare-cutter *Not*I sites on either side of the stuffer fragment enable its excision from p220.2TC, exposing the telomeric sequences without cleaving anywhere else in the vector.

The telomeric fragment was released from pEND2 by cleavage with *Xho*I and *Hin*dIII (figure 3.4 I, lane B and 3.4 II, lane 1). The Kan^R stuffer fragment was excised as a 1.2kb from pUC4KN by cutting at the *Sal*I sites in the symmetrical polylinker (which is shown by the equally sized *Not*I cleavage products in figure 3.4 I, lane A). The telomeric fragments were ligated into the *Hin*dIII site of p220.2 (figure 3.4 I, lane C) and the stuffer fragment was ligated to the *Xho*I-cohesive ends of the telomeric DNA in a one-step reaction. The reaction took place in the presence of *Xho*I to select against self-ligation of *Xho*I-cohesive ends (*Xho*I- and *Sal*I-cohesive termini are compatible, but the site which is formed at the ligation is not cleaved by *Xho*I).

The structure of p220.2TC was checked by restriction digests which released the telomeric and stuffer fragments from the rest of the vector. Lane E shows that the 1.2kb Kan^R stuffer fragment is released by digestion with *Not*I while the 10kb fragment retains the telomeric sequences (3.4 II, lane 2). Lane D shows that excision of the Kan^R fragment with *Not*I followed by digestion with *Hin*dIII cleaves the 1.2 kb fragment into 0.5 and 0.7 kb fragments. This pattern may be compared with lane F where cleavage of the telomeric sequences from p220.2TC with *Hin*dIII after excision of the Kan^R fragment produces a similar pattern. The 0.5kb (TTAGGG)n fragments occupy the same position as the 0.5kb *Not*I-*Hin*dIII fragment of Kan^R (figure 3.4 II, lane 3). The entire telomere cassette could be excized from p220.2TC as a single 2.3kb *Sfu*I fragment (figure 3.4 I, lane H and 3.4 II, lane 4). A comparison with the digestion of p220.2 with *Sfu*I (figure 3.4 I. lane G) localized another *Sfu*I site to the region of the hygromycin resistance gene.

3.1.5 Modifications of p220.2TC

Prior to transfection into human cells, p220.2TC was modified in two ways (figure 3.1). The plasmid was converted into a linear molecule with exposed telomeric DNA sequences by cleaving with *Not*I to excise the Kan^R stuffer fragment. This linear plasmid was termed p220.2ET (for External Telomeres) and possessed terminal telomeric DNA sequences in the correct orientation, i.e. the G-rich strand running 5' to 3' to the terminus. p220.2ET (hereafter referred to as pET) was the linear vector being tested in this project. The alternative was cleavage of p220.2TC with *Dra*I which cleaves in the ampicillin resistance marker and the bacterial origin on the other side of the circular molecule. This linear molecule, termed p220.2IT (for Internal Telomeres), did not have exposed telomeric sequences and was used as a negative control for telomere function. Since the EBV replication sequences were intact, p220.2IT (hereafter referred to as pIT) should be as competent as pET to be replicated and maintained in the cell by the EBNA-1 protein. p220.2 was used as a positive control for replication and maintenance. p220.2TC was used as a control for the effect of internal telomeric sequences in a circular vector upon replication.

3.1.6 Subcloning of the telomere cassette from p220.2TC into pLIB16

I attempted to convert pLIB16, a derivative of p220.2, into a vector which could be used in a linear form by inserting the telomeric sequence cassette from p220.2TC.

Most of the restriction sites in the polylinker are present elsewhere in pLIB16 which made further manipulations of the plasmid difficult. The *Sal*I site in the polylinker and a *Cla*I site between the ampicillin resistance marker and the *EBNA-1* sequence were the only useful unique sites in pLIB16.

The telomere cassette could be excised from p220.2TC as a single 2.3kb fragment by using *Sfu*I (see figure 3.2 and 3.4). The telomere cassette was inserted into the *Cla*I site of pLIB16. The termini produced by cleavage with restriction enzymes *Cla*I and *Sfu*I are compatible for the ligation of cohesive ends. Cells transformed with the ligation mixture were plated onto selective media containing ampicillin and kanamycin. Recombinant colonies were obtained at a low frequency.

Analysis using restriction digestion showed that the recombinant plasmids contained substantial deletions. The total size of the recombinant plasmids was 4.5kb, while the predicted product of the ligation would be 25.3kb. Since the colonies were kanamycin resistant, presumably the telomere cassette, which contained the kanamycin resistance marker, had ligated to pLIB16.

Figure 3.5 **The insertion of the telomeric DNA cassette into pLIB16**

Agarose gel showing the restriction patterns of the products of the ligation of the telomere cassette (lane B) into pLIB16 (lane A). None of the recombinant plasmid products (lanes C-F) are larger than pLIB16. If the 2.3kb telomeric DNA cassette had stably ligated to pLIB16, a *Hin*dIII restriction digestion would produce fragments of 5, 6 and 14kb.

The fragment sizes of the λHHR DNA size marker are indicated.

Legend

A. pLIB16 cut with *Cla*I

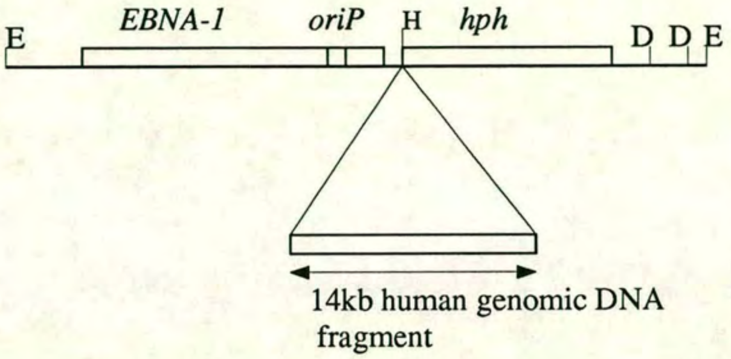
B. 2.3kb *Sfu*I DNA fragment containing the telomeric DNA cassette

C. - F. recombinant plasmids cut with *Hin*dIII

G. uncut recombinant plasmid

H. λHHR size marker

Figure 3.5b Linear restriction map of pLIB16



pLIB16 is a derivative of p220.2. The dyad repeat region has been deleted from the *oriP* element rendering it inactive. A 14kb human genomic DNA fragment which has been reported to be a putative autonomously replicating sequence has been inserted into the *Hin*dIII site (H). The *Dra*I sites (D) and two of the *Eco*RI sites (E) of the original vector have been shown (compare with figure 3.1).

However, the presence of the 14kb human DNA fragment in pLIB16 together with the repetitive DNA of the telomeric sequences may have created an unstable molecule leading to an intramolecular recombination and deletion event. The *Hin*dIII restriction patterns of the recombinant plasmids are shown in figure 3.5.

These transformations were carried out in both DH5α and XL1-Blue *E.coli* strains which are deficient in the recA recombinase enzyme. Often this deficiency in a strain can prevent deletion of repeated sequences from bacteria. However, in this case, deficiency in one recombinase was apparently not sufficient to prevent deletion of sequences. An *E.coli* strain which is deficient in both recB and recC recombinase activities may have been able to propagate plasmids containing large genomic DNA fragments and telomeric repeat tracts.

3.1.7 Discussion of vector construction

The introduction of the two telomeric DNA fragments into p220.2 proved to be very difficult. The construction of the telomeric DNA cassette of p220.2TC by a simultaneous ligation with a Kan^R stuffer fragment was the only successful strategy out of several which were tried.

These strategies and their outcomes are summarized below. It was originally planned to adapt the plasmids pDY- and pLIB16 into linear vectors (both plasmids were made by Krysan *et al.*, 1989 and obtained by Howard Cooke). pDY- is a derivative of p220.2 from which the dyad repeat region had been deleted. This plasmid was not able to replicate in mammalian cells owing to this deletion, although it could still be retained within dividing cells for up to 2 weeks. pLIB16 (described earlier in section 3.1.1) contained a 14kb genomic DNA fragment which reportedly acted as a functional replication origin (Krysan *et al.*, 1989). The aim of using these two vectors was the planned testing of a replicating linear vector based upon pLIB16 against a non-replicating linear vector based upon pDY-. The 1.7kb telomeric DNA fragment from pHutel-2-end was chosen as the initial source of telomeric DNA and was also present in another clone, pEND2R, in the reverse orientation. The initial aim was to construct linear vectors capped with telomeric sequences in different orientations (to provide negative controls for telomere function). One version of the vectors would contain telomeric DNA sequences with the G-rich strand running 5' to 3' towards the terminus (as in human chromosomes), while the other version would contain telomeric DNA where the C-rich strand ran 5' to 3' towards the terminus.

(i) Dephosphorylated telomeric sequences were ligated to pDY- and pLIB16. A fragment of lambda DNA was prepared as a removeable stuffer by ligating oligonucleotide linkers containing "rare-cutter" sites onto the termini. The prepared stuffer fragment was then ligated to the adapted EBV vectors. The products of this ligation were transformed into bacteria. The recombinant plasmids obtained from the transformations were examined by restriction digestion. One class of plasmids retrieved from the transformation were the same size as pDY- (9kb) and did not contain telomeric sequences, or (in the case of the pLIB16 ligations) human genomic DNA. The other class of plasmids appeared to be either pHutel-2-end or pEND2R.

(ii) Oligonucleotides consisting of TTAGGG and CCCTAA repeats had been designed which be annealed and polymerized by ligation to form short telomeric structures complete with 3' overhangs. The aim of the concept was to construct linear vectors *in vitro* and not via bacteria. This would avoid the problems of recombination associated with cloning in bacteria. (However, the purification of the ligation products away from agarose to a degree high enough for transfection might have been a problem.) The oligonucleotides would be annealed, phosphorylated and ligated together followed by isolation of the higher molecular weight products by agarose gel electrophoresis. These would be purified and ligated onto the termini of linearized pDY- or pLIB16. These ligation products would be transfected directly into mammalian cells. In practice, however, the polymerization of the oligonucleotides was not efficient.

(iii) An attempt was made to construct derivatives of pDY-and pLIB16 containing one telomeric tract only. These constructs might have been stable in bacteria. Prior to transfection into mammalian cell, the construct would be linearized and a second telomeric DNA fragment would be ligated onto the uncapped terminus. However, the ligation of a single telomere into the plasmids failed to produce anything other than deleted recombinant plasmids after transformation.

(iv) After the construction of pEND2, the building of a telomeric cassette in the vector pGEM7 was attempted. Once assembled, the cassette could be excised and moved into the vector of choice as a complete unit. It was hypothesized that a physical separation between the two telomeric DNA fragments might improve their stability as an inverted repeat (see section 3.1.3). However, the human cDNA fragment failed to prevent deletions from recombunant plasmids. In the event, a second telomeric tract could not be inserted such that it opposed another on the other side of a 1.5kb cDNA stuffer fragment. The plasmids retrieved from the transformations contained either only one telomeric DNA tract or none at all.

When initial attempts to transfer the 1.7kb telomeric DNA fragment of pHutel-2-end into the EBV vectors were unsuccessful, the sub-telomeric sequences were hypothesized to have been a possible factor in the deletion events. These were removed in the construction of pEND2.

The telomere cassette which is present in p220.2TC has proven very stable in *E.coli* while selection of the Kan^R fragment is maintained. It was conjectured that selection for kanamycin resistance might simply result in a population of plasmids where the telomeric sequences had been deleted from the cassette leaving the selectable marker behind. However, this was not the case and plasmids were selected which contained the entire cassette without rearrangements. It is not certain whether or not the physical separation of the telomeric sequence tracts played any role in improving the stability of the construct.

Originally, it was planned to sub-clone the telomere cassette from one type of vector into another. The excision of the entire cassette is possible by cleaving p220.2TC with the enzyme, *Sfu*I. This has, in fact, been done during the construction of a linear vector for a mycelial fungus, *Podospora anserina* which contains the same telomere repeat sequence as vertebrates (Javerzat *et al.*, 1993). However, attempts to move the cassette into a vector, pLIB16, which already contained a large fragment of human genomic DNA were not successful. Substantial deletions in the recombinant plasmids may have been caused by recombination between the human DNA sequences, or by the overall size of the construct. It is unlikely that the sequences were inherently detrimental to the cells, thus favouring their deletion since the human sequences, including the 14kb genomic fragment are stable when separately cloned in plasmids. This likelihood of recombination may hamper the use of the telomere cassette in a vector which contains a cloned putative centromeric sequence, or, as in the case of pLIB16, a putative human ARS element.

3.2 CHARACTERIZATION OF CELL LINES

3.2.1 Introduction

A human cell line was required to act as the host for the modified EBV vectors. Raji cells were selected as a host cell line because they were used in the early research on EBV essential replication sequences (Sugden *et al.*, 1985).

A great deal of effort was invested in transfections of a human embryonic kidneys cell line which was initially chosen as the host system for the EBV-based vectors. They were selected on the basis of the work of Krysan *et al.* (1989) who used a suspension-adapted derivative of this cell line for their work. (They used calcium phosphate co-precipitation in their transfections.) Although both different conditions of electroporation and different methods of transfection were tried, cells which were resistant to hygromycin at 200µg/ml did not arise from the transfected populations. Eventually, further work on this cell line was discontinued and Raji cells were used as the host system.

Raji is a human cell line which grows in suspension and is derived from the transformed cells of Burkitt's lymphoma (Pulvertaft, 1967). These cells are lymphoblasts which have been transformed during an Epstein-Barr virus infection. Raji cells contain at least one chromosomally integrated copy of the viral genome as well as multiple episomally replicating copies (Jankelevich *et al.*, 1992).

Raji cells constitutively express the EBNA-1 protein which is believed to be an initiator of DNA replication. It seemed likely that since the Raji cell contains a constant level of the EBNA-1 protein, transfected plasmids which contained the *oriP* sequence would be replicated efficiently. The same plasmid might not replicate in an EBNA-1-negative cell line if it is dependant upon the expression of the protein from sequences *in cis*. An adequate level of the protein might not be present in the cell after transfection in order to initiate replication from *oriP*.

3.2.2 Electroporation of Raji cells

Electroporation conditions which apply to Raji cells have been published (Teshigarawa and Katsura, 1992). An electroporation viability-curve was obtained by varying the voltage at which the cells were pulsed while keeping the capacitance constant at the suggested value of 500µF (see table 3.1).

Table 3.1

voltage (kV)	0.0	0.05	0.1	0.2	0.3	0.4
viable cells ($x10^5$)	23	18	22	17	13	6

A certain amount of variation occurred between different experiments, but, on average, the conditions of 0.3kV pulses at 500µF killed 50% of the cells (see graph in figure 3.6).

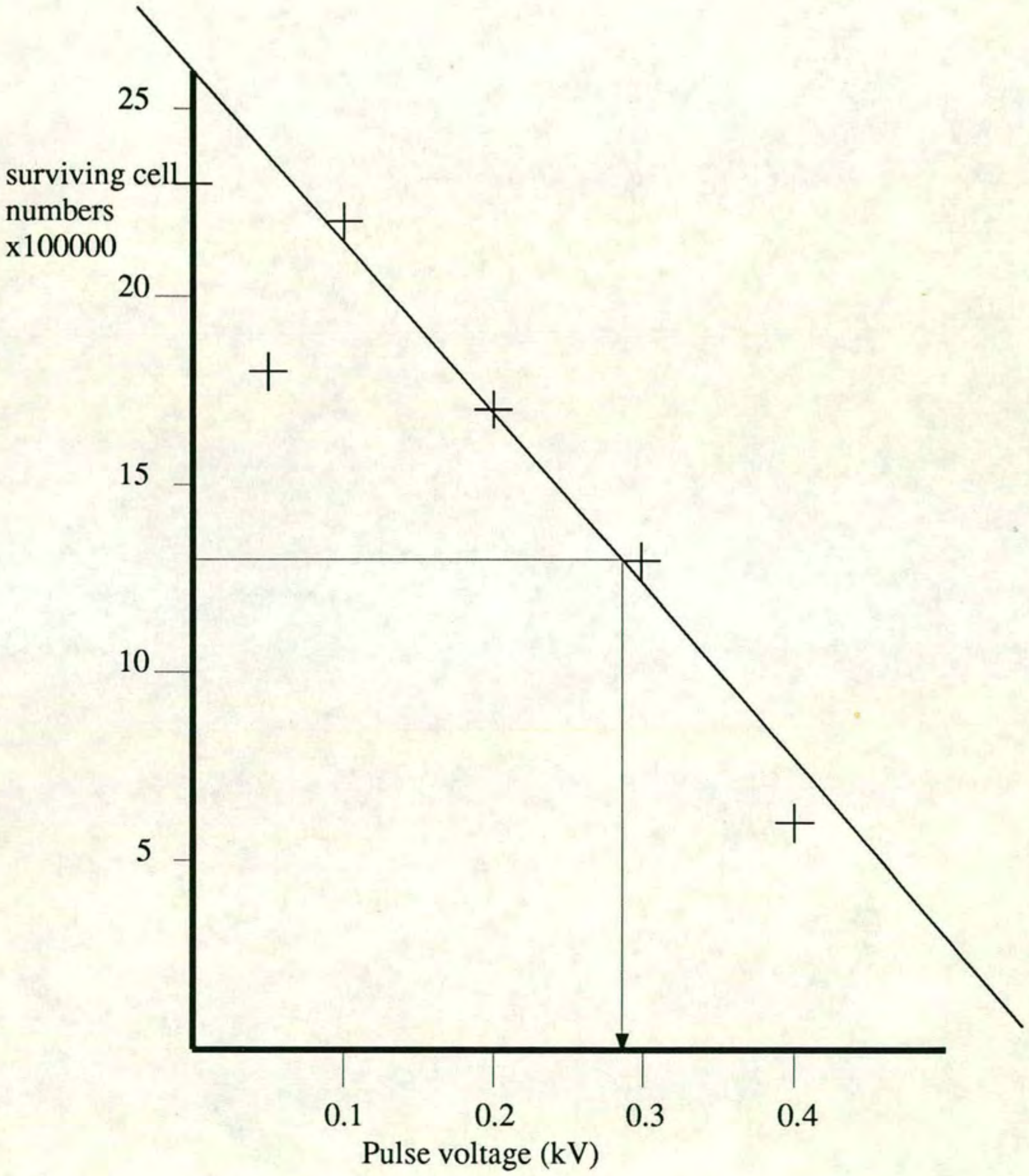
3.2.3 Hygromycin sensitivity

Raji cells were plated in different concentrations of hygromycin B (HmB) in order to test their sensitivity to the antibiotic. Cells still proliferated at 50µg/ml HmB while 100µg/ml was 100% lethal. On this basis, a hygromycin concentration of 200µg/ml was chosen as a suitable level of selection when required. This level of selection is consistent with the hygromycin concentrations used in published work (Blochlinger and Diggelman, 1984).

Depending upon the duration of the transfection assay which was being executed, antibiotic selection was not always required. Krysan *et al.* (1989) had shown in their work that in a short term

Figure 3.6 Electroporation viability of Raji cells

Raji cells were electroporated at a range of voltages with a constant capacitance of 500μF. The voltage which resulted in the survival of 50% of the cells was estimated from the graph to be approximately 0.3kV.



assay, e.g. 20-30 days, antibiotic selection did not need to be applied in order to maintain the transfected EBV plasmids in the cell population. Presumably, this is an effect of the EBNA-1 protein which is a putative "nuclear retention" factor as well as a replication enhancer and prevents the loss of DNA which is linked to *oriP* from the nucleus during successive cell divisions (see introduction section 1.1.2). In the longer term, however, sub-culturing may reduce the proportion of transfected cells compared to untransfected cells in the population and antibiotic selection is required. A technical advantage of not applying selection in a short term assay is the ability to sample the cells at regular intervals without the possibility of a lag phase in population growth which occurs when an antibiotic kills a substantial number of the cells.

Since Raji cells grow in suspension, clonal lines cannot be obtained by isolating colonies. Although not done in this work, clonal lines may be established by serial dilutions of the transfected cells.

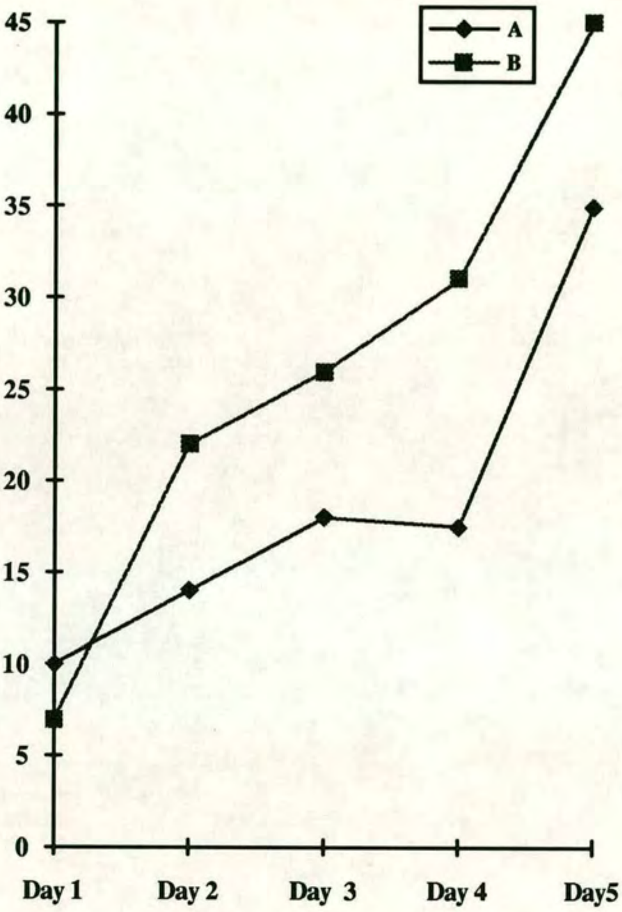
3.2.4 Blockade of cell cycle by thymidine

The establishment of cell synchrony prior to electroporation was achieved by the addition of 3.3mM thymidine to the medium (see methods and materials section 2.1.3). The ability of thymidine to block the cell cycle was checked by comparing the growth of two populations of Raji cells, one of which was exposed to thymidine for about 36 hours. The thymidine blockade was applied for twice the normal amount of time in order to detect an appreciable difference in proliferation. The cell numbers are listed in table 3.2 and a graph of the data is shown in figure 3.7.

Table 3.2 Blockade of cell cycle using thymidine

	A (thymidine present)	**B (no thymidine)**
Day 1 (thymidine added)	1.0 x10^6 cells	0.7 x10^6 cells
Day 2	1.4 x10^6 cells	2.2 x10^6 cells
Day 3 (thymidine removed)	1.8 x10^6 cells	2.6 x10^6 cells
Day 4	1.75 x10^6 cells	3.1 x10^6 cells
Day 5	3.5 x10^6 cells	4.5 x10^6 cells

Figure 3.7 Thymidine-induced blockade of cell cycle



Y axis - cell numbers x10^5

A - cells treated with 3.3mM thymidine

B - cells not treated with thymidine

Thymidine was present in the medium of the "A" population from day 1 to day 3.

Graph showing the inhibition of cell proliferation by the addition of 3.3mM thymidine to the culture medium. Thymidine blocks the cell cycle at the G1/S phase transition. After release from the blockade, the the cells complete S and M phases, and cell division continues.

The cells which were not exposed to thymidine displayed steady growth and doubled about once every one and a half days. The cells which were exposed to thymidine proliferated at a slower rate, followed by a slight decrease. Thymidine blocks the cell cycle at G1/S phase rather than preventing mitosis, so cells which had already commenced G2 phase completed mitosis as normal before being blocked. This "lag" in the action of the thymidine block results in a slight increase in the cell numbers before the entire population of cells is blocked. The decrease in the numbers of thymidine-treated cells was probably due to the relatively long period of exposure to the blockade which may have resulted in decreased viability.

Electroporation of synchronized cells was only carried out a few times since it did not appear to improve cell viability after transfection.

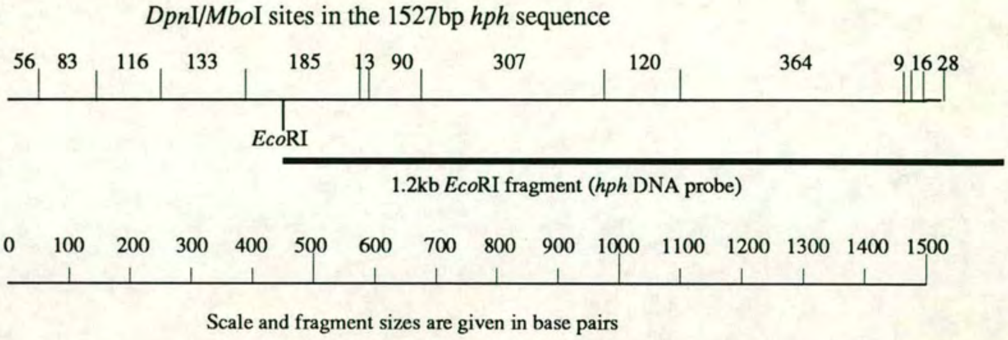
3.3 RESULTS OF TRANSFECTIONS

3.3.1 Introduction

DNA vectors were electroporated into dividing Raji cells (see Materials and Methods section 2.1.4). The cells were simultaneously harvested and sub-cultured at intervals of 3-5 days after transfection by removal of 4/5 of the total medium. (Since Raji cells grow in suspension, spent medium and excess cells can be removed and fresh medium supplied with minimal disturbance to the cells.) Low molecular weight DNA was obtained from the harvested cells by the Hirt method.

The Hirt-extracts were used to provide a time course via a Southern blot showing the fate of the introduced DNA vectors. Total genomic DNA extracts were also obtained in certain cases at the later time points. A 1.2kb *Eco*RI fragment of the hygromycin resistance gene (see figure 3.8) was used as a DNA hybridization probe. The radiolabelled DNA probe also included 3ng of labelled λ DNA which was included to visualise the λ size marker in the autoradiograph. The DNA was examined by digestion with restriction enzymes including *Mbo*I. *Mbo*I recognises the site GATC, but cannot cleave it unless it is unmethylated. DNA which has been replicated in mammalian cells is not methylated at this site while adenosine residues in GATC sites which have been replicated

Figure 3.8 The *Dpn*I/*Mbo*I restriction map of the *hph* sequence



The fragments of the *hph* sequence which are detected by the 1.2kb *Eco*RI probe have the following sizes in bp: 9, 13, 16, 28, 90, 120, 185, 307 and 364. Only the two largest fragments are visible in the autoradiographs. The information was obtained from a linear mapplot of the hygromycin phosphatase sequence in the EMBL databank (accession number K01193) using the University of Wisconsin Genetics Computer Group program.

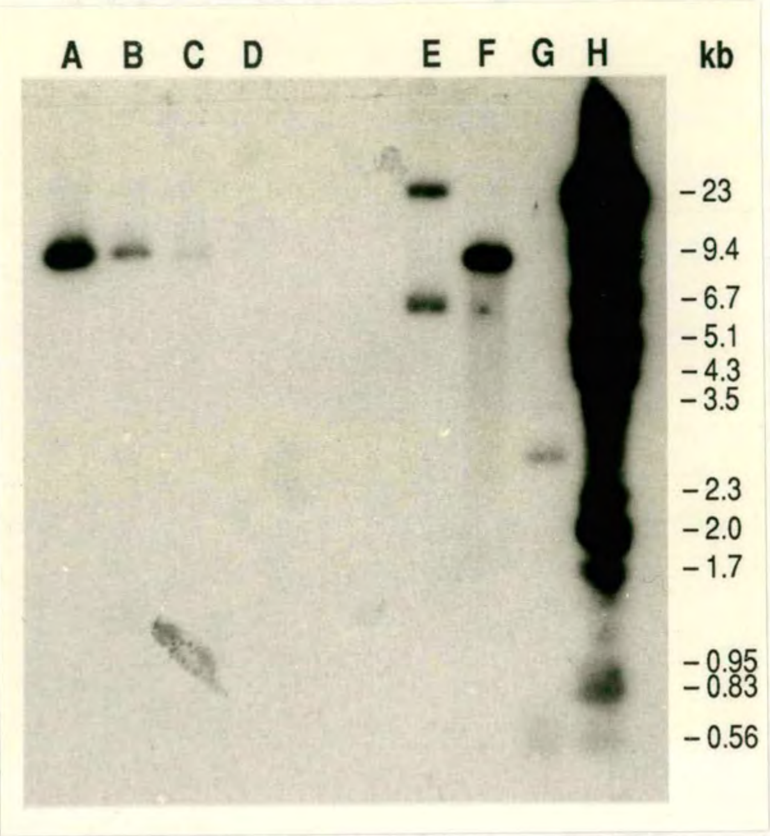
in *dam*$^+$ bacteria are methylated. Therefore, a relatively simple test for the replication of transfected DNA in mammalian cells is sensitivity to the restriction enzyme, *Mbo*I. (*Mbo*I has two isoschizomers, *Dpn*I and *Sau*3A1: *Dpn*I only cleaves DNA which is methylated at the site, while cleavage by *Sau*3A1 is unaffected by adenosine methylation.) The restriction patterns were checked against 25pg of the appropriately digested or undigested plasmid which was loaded onto the gel with 10μg of *Hin*dIII-digested human genomic DNA.

The Hirt-extracted DNA consists mainly of low-molecular-weight DNA (owing to the removal of the cellular DNA by preferential precipitation using SDS and NaCl see Materials and Methods section 2.1.6 and Hirt, 1967): episomal transfected vectors, sheared chromosomal DNA, RNA and, in the case of Raji cells, EBV genomic DNA. If the vectors have integrated into the host genome and rearranged, studies of Hirt DNA extracts might not show this. It is also possible that plasmid DNA actually replicated after integration and restriction enzyme digestion might create the impression of extrachromosomal replication. Southern blot hybridization to a total genomic DNA extract, however, should reveal integration of the vector sequences.

3.3.2 Test for contamination of Hirt DNA samples with exogenous plasmid

A control blot was carried out in order to find the extent to which DNA that did not enter cells during an electroporation contaminated subsequent Hirt extractions. An attempt was made to simulate four successive harvests and Hirt extractions from a flask of cells where contaminating extracellular plasmid DNA was present. 10μg of p220.2 plasmid DNA was mixed with 4×10^6 cells in 10ml of media in a tissue culture flask. The cells were centrifuged and washed with PBS. A Hirt extraction was then performed on 10^6 cells. The remainder of the cells were resuspended in fresh media in the same flask. The majority of the p220.2 DNA was removed from the environment of the cells by the two centrifugation steps. The washing and Hirt extraction steps were then repeated on the cells in the flask such that four "mock" Hirt DNA samples had been obtained. These DNA samples contained the same amount of low molecular weight human cellular DNA. It was likely that p220.2 plasmid DNA which had persisted through the washes was also present. The DNA samples

Figure 3.9 **Test for contamination of Hirt DNA samples by untransfected plasmid DNA**



The figure shows the autoradiograph resulting from the hybridization of the *hph* probe to Hirt DNA samples obtained from cells which were mixed, but not electroporated, with p220.2 plasmid DNA. The plasmid DNA is visible in lanes A-D as a 9kb band. The exposure time was overnight.

Legend

A. 1st Hirt DNA sample digested with *Hin*dIII
B. 2nd " " "
C. 3rd " " "
D. 4th " " "
E. uncut p220.2 DNA with 10μg *Hin*dIII-digested total genomic human DNA
F. *Hin*dIII-digested p220.2 " " " "
G. *Dpn*I-digested p220.2 " " " "
H. λHHR size marker

were digested with *Hin*dIII and subjected to analysis by Southern blotting using *hph* as a DNA probe. The resulting autoradiograph is illustrated in figure 3.9.

The blot shows that exogenous plasmid, visible as a 9kb band, disappeared from the Hirt DNA by the fourth extraction. This result can be applied, to a large extent, to the situation where a transfected population of cells was being subcultured and harvested. If this result is used as a guide, it is unlikely that exogenous plasmid DNA would contaminate Hirt DNA samples extracted towards the end of a long term assay. It may be argued that in cases where the electroporated vector failed to replicate, the DNA had not entered the cells. However, a comparison of figure 3.9 with, for instance, figure 3.18 (a transfection with p220.2) shows a much greater yield of plasmid DNA after electroporation, although the sequences are eventually lost. If the electroporation had not allowed the DNA to enter the cells, presumably the hybridization signals due to contaminating exogenous DNA would be similar to those seen in figure 3.9.

3.3.3 Transfection of Raji cells with p220.2

The circular positive control, p220.2, was transfected into Raji cells as a preliminary test of the cell line's ability to support replication of the control vector. The transfected cell line was designated Raji-p220.2 (19.6).

The transfected cells were initially grown without selection. Cells were harvested from the population at 3, 5, 10, 15, 20, 25 and 30 days after electroporation. Low-molecular-weight DNA was obtained from these harvested cells by Hirt extraction. After 30 days, selection was applied to the cells at a concentration of 200μg/ml HmB. Two further harvests were taken at 48 and 53 days. Both Hirt-extracted DNA and total genomic DNA was obtained from the cells harvested at these two time points.

Hirt-extracted DNA was digested with restriction enzymes and analysed by Southern blots (figures 3.10 and 3.11). *Mbo*I was used for the reasons described above. *Hin*dIII was used to identify any possible deletions from the 9kb linear fragment of p220.2 (*Hin*dIII cleaves the plasmid once, figure 3.4 I, lane C). *Dra*I and *Hin*dIII were used in a double digest because two distinct fragments are produced which can be detected by the *hph* probe: a 3kb band contains most of the *hph* sequence; a 5.5kb fragment which contains the EBV sequences was detected less strongly by the same probe possibly owing to an overlap with the probe DNA fragment.

(i) *MboI* digestion of Hirt DNA (figure 3.10)

A control was included which contained p220.2 cleaved to produce the *MboI/DpnI* restriction pattern (fig. 3.10 lane L). There are no bands which hybridize strongly to the *hph* probe in lane A which contains Hirt DNA from untransfected cells. A strong signal in the 3 day time point (fig. 3.10 lane B) is present near the top of the track which probably corresponds to uncut bacterially replicated p220.2. A strong signal from a smaller band is also present which may indicate that some of the vector sequences had become sensitive to *MboI* at this time. The strongest hybridizing band of the control at <0.5kb (the *DpnI/MboI* restriction map of the *hph* sequence shown in figure 3.8 contains a doublet of 307 and 364bp) in lane L is present in the DNA from all the time points from 10 days onwards (fig. 3.10 lanes C, D, E, F, G and H). The largest *DpnI/MboI* fragment of the *hph* sequence is 364bp. A less strongly hybridizing control band of 2.8kb in lane L is visible at the the time points from 25 days onwards (fig. 3.10 lanes E, F, G and H). This band is not part of *hph* but probably another adjacent fragment of p220.2 which overlaps the probe fragment. (This same band may be present in the 15 day time point in fig. 3.10 lane D, but it may have run at an apparently higher molecular weight owing to aberrant running of the DNA.) An anomaly in the time points from 30 days onwards is the presence of hybridizing bands which correspond to uncut plasmid (seen in the control lane J) indicating the presence of *MboI*-resistant bacterially replicated DNA. The presence of these bands is discussed at the end of this section.

(ii) *DraI+HindIII* digestion of Hirt DNA (figure 3.11)

Control lanes containing uncut p220.2 (fig. 3.11 lane I) and p220.2 cleaved with *DraI+HindIII* (fig. 3.11 lane J) were included. The two bands present in the control lane are about 3 and 5.5kb in size. There are no visible hybridizing bands present in lane A which contains Hirt

Figure 3.10 Southern blot analysis of Hirt-extracted DNA from Raji-p220.2(19.6)

(a) Ethidium bromide-stained gel of *Mbo*I-digested Hirt DNA samples extracted over a time course after transfection of Raji cells with the plasmid vector, p220.2 (see text in section 3.3.1).

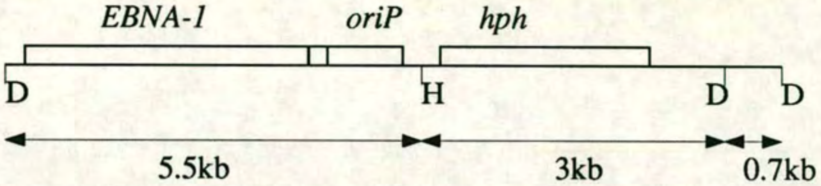
(b) The gel was blotted onto Hybond membrane and hybridized to the *hph* DNA probe labelled with ^{32}PdCTP. The autoradiograph shows sequences that hybridize to *hph* becoming *Mbo*I-sensitive with time (see text in section 3.3.1 and 3.3.3). The autoradiograph shown is a composite of films which were exposed to the membrane for different periods of time:
lanes A and C-H were exposed for 3 days; lane B was exposed overnight; lanes I-M were exposed for 3 hours.
The fragment sizes of the λHHR DNA size marker are indicated.

Legend
A. DNA from untransfected cells digested with *Mbo*I restriction enzyme
B. DNA extracted at 3 days post-electroporation "
C. " " 10 " " " "
D. " " 15 " " " "
E. " " 25 " " " "
F. " " 30 " " " "
G. " " 48 " " " "
H. " " 52 " " " "

I. " " 3 " " " uncut

J. p220.2 uncut plasmid loaded with 10μg *Hin*dIII-digested total genomic human DNA
K. p220.2 plasmid digested with *Dra*I + *Hin*dIII " " " "
L. p220.2 plasmid digested with *Dpn*I " " " " "
M. λHHR size marker

Figure 3.11c The linear *Dra*I + *Hin*dIII restriction map of p220.2



The 1.2kb *Eco*RI fragment of the *hph* sequence was able to detect the 3kb and the 5.5kb *Dra*I (D) and *Hin*dIII (H) fragments. The latter produced a weaker signal as the hybridization was probably the result of a slight overlap around the *Hin*dIII site. The 0.7kb fragment was not detected by the probe.

DNA from untransfected cells. Only in the 30 day time point (fig. 3.11 lane F) do the two bands which are present correspond closely in size to the control bands. This aberration in the *Dra*I+*Hin*dIII restriction patterns of the transfected p220.2 plasmid may have been caused by the electrophoresis of DNA containing different amounts of salt. Owing to this occurrence, this blot can only be used to confirm the presence of sequences which hybridize to *hph* in the time points from 5 to 52 days (fig. 3.11 lanes B to H). The DNA tracks from all the time points contain one band which strongly hybridizes plus one other slightly weaker band running at a relatively higher molecular weight.

The differences in electrophoresis between lanes may have been caused by varying concentrations of salt in different DNA samples. In retrospect, this problem could have been surmounted by the addition of small amounts of the λHHR size marker to each of the lanes and rehybridizing the membrane to radiolabelled λ DNA. This procedure would have demonstrated whether or not the lanes were running normally.

(iii) *Hin*dIII and *Mbo*I digestion of total genomic DNA (figure 3.12)

The total genomic DNA from the time points of 48 and 52 days were examined in order to detect evidence of possible integration of the p220.2 sequences. The DNA samples are mostly of a high molecular weight, but low molecular weight DNAs such as extrachromosomal vectors are still detected, although weakly. The *hph* probe hybridized to bands in the total genomic DNA from untransfected Raji cells. The *Hin*dIII digestion (fig. 3.12 lane A) shows a hybridizing band at ~3.3kb and the *Mbo*I digestion (fig. 3.12 lane D) may show a band at about 2.5kb. The *Mbo*I digestion is smeared, but reference to a similar track in figure 3.13 (lane F) verifies the presence of a cross-hybridizing band at 2.5kb. These bands are common to the 48 and 52 day samples as well (fig. 3.12 lanes B, C, E and F). One possibility is that the DNA probe hybridization to a sequence in Raji genomic DNA related to the *hph* sequence.

The *Hin*dIII-digested DNA samples from 48 and 52 days contain a 9kb band (fig. 12 lanes B and C) which concurs with the control lane J containing p220.2 linearized with *Hin*dIII (marked by a

Figure 3.12 **Southern blot analysis of total genomic DNA from Raji-p220.2(19.6)**

(a) Ethidium bromide-stained gel of *Mbo*I and *Hin*dIII digested total genomic DNA from Raji cells transfected with p220.2. The DNA samples were taken at the last two time points in the time course of the assay described in section 3.3.1.

(b) The gel was blotted onto Hybond membrane and hybridized to the *hph* DNA probe labelled with ^{32}PdCTP. The bands which correspond to the p220.2 vector sequences in the positive control lanes (J and K) are indicated by arrowheads (9kb), arrows (2.8kb) and double arrowheads (307+364 doublet). The exposure time was 8 hours.

Bands are present in the lanes containing sample DNA (B, C, E and F) which do not correspond to vector DNA bands in the positive control lanes (I, J and K), but these are also present in the lanes containing DNA from untransfected cells. The *hph* DNA probe may have hybridized to related sequences present in the cell line genomic DNA. (Figure 3.13 also shows Raji total genomic DNA hybridized to the *hph* probe. Bands are visible in *Mbo*I and in *Hin*dIII digestions which by comparison with the λHHR size marker conform to the bands **not** highlighted in figure 3.12.)

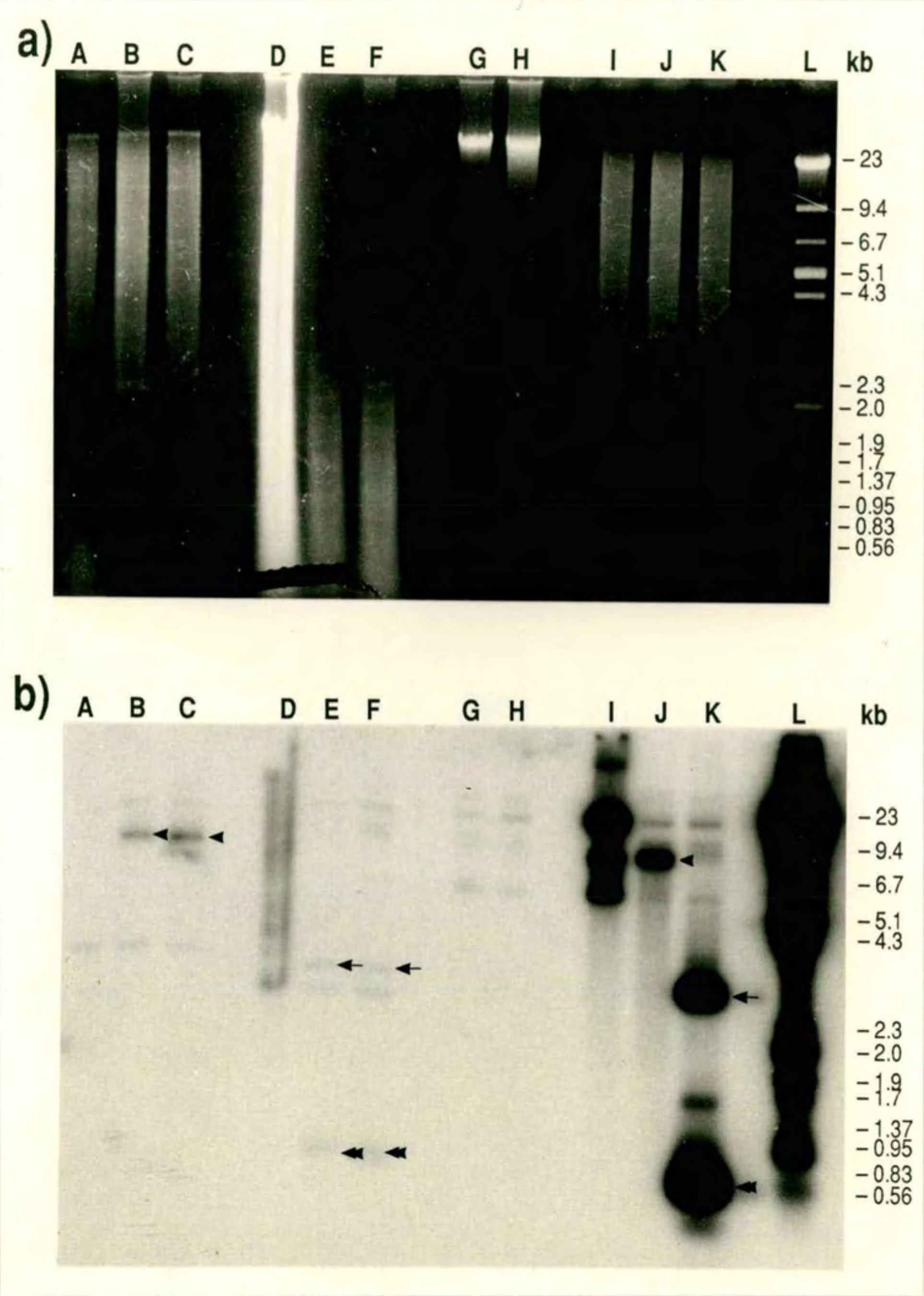
The slight misalignment of the bands was caused by the gel being laid down slightly askew during the blotting procedure.
The fragment sizes of the λHHR DNA size marker are indicated.

Legend
A. 5μg Raji DNA from untransfected cells digested with *Hin*dIII
B. 5μg DNA extracted at 48 days post-electroporation digested with *Hin*dIII
C. 5μg DNA extracted at 52 days post-electroporation digested with *Hin*dIII

D. 10μg Raji DNA from untransfected cells digested with *Mbo*I
E. 5μg DNA extracted at 48 days digested with *Mbo*I
F. 5μg DNA extracted at 52 days digested with *Mbo*I

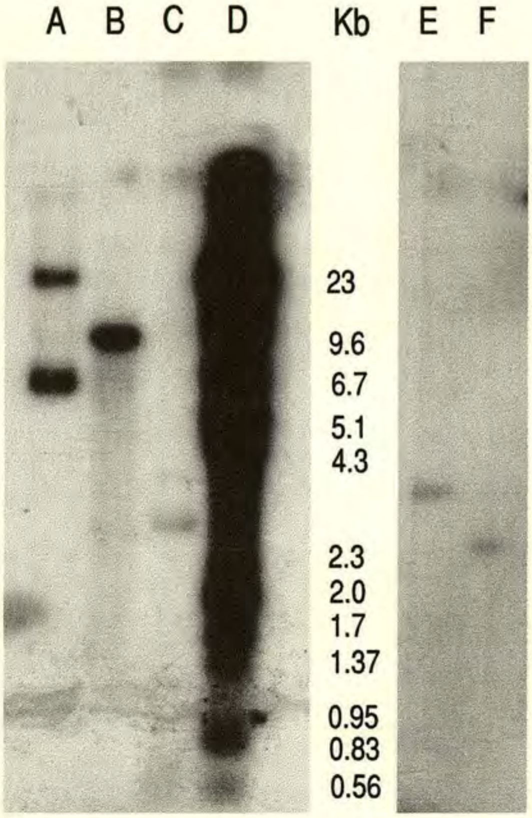
G. 3μg uncut DNA from 48 days
H. 3μg uncut DNA from 52 days

I. p220.2 plasmid uncut loaded with 10μg *Hin*dIII-digested total genomic human DNA
J. p220.2 plasmid digested with *Hin*dIII " " " " "
K. p220.2 plasmid digested with *Dpn*I " " " " "
L. λHHR


a)
A B C D E F G H I J K L kb
- 23
- 9.4
- 6.7
- 5.1
- 4.3
- 2.3
- 2.0
- 1.9
- 1.7
- 1.37
- 0.95
- 0.83
- 0.56
b)
A B C D E F G H I J K L kb
- 23
- 9.4
- 6.7
- 5.1
- 4.3
- 2.3
- 2.0
- 1.9
- 1.7
- 1.37
- 0.95
- 0.83
- 0.56

13 Hybridization of the *hph* DNA probe to Raji total genomic DNA

e shows the autoradiograph resulting from the hybridization of the *hph* probe to Raji total DNA. The probe has hybridized to a 3-3.5kb band in the *Hin*dIII digestion (lane E) and to band in the *Mbo*I digestion (lane F). p220.2 digested with *Dpn*I is present at a low dilution C. Bands which hybridize to *hph* from p220.2 are present at ~2.8kb and, faintly, at <0.5kb in The exposure time was overnight.



Legend

A. Uncut p220.2 plasmid with 10μg *Hin*dIII-digested total genomic human DNA

B. p220.2 linearized with *Hin*dIII " " " "

C. p220.2 digested with *Dpn*I " " " "

D. λHHR size marker

E. 5μg Raji total genomic DNA digested with *Hin*dIII

F. 5μg Raji total genomic DNA digested with *Mbo*I single arrow). The *Mbo*I-digested DNA samples from 48 and 52 days contain a band at <0.5kb which may be the 307+364bp doublet (double arrowhead) and 2.8kb (arrow) bands which are the most strongly hybridizing bands from the control lane K. Uncut samples from the time points were also included on the gel (fig. 3.12 lanes G and H). In these tracks, the high molecular weight genomic DNA is visible at the limiting mobility position of the ethidium bromide-stained gel, greater than 23kb (figure 3.12 a). The hybridization has not produced a strong signal in the uncut genomic DNA band in figure 3.12 b lanes G and H. It may be inferred that the p220.2 vector has not integrated into the genomic DNA and formed tandem arrays. The bands that are visible may correspond to uncut plasmid as seen in the control lane I.

The gel was laid down slightly skewed during the blotting procedure (the line of bands curves down slightly from left-to-right), resulting in the bands on the left of the gel appearing to be higher. In the *Mbo*I digestions (fig. 3.12 lanes D-F),for instance, the cross-hybridizing band present in the untransfected control as well as the 48 and 52 day samples appears to align with the 2.8kb strongly hybridizing fragment from the p220.2 *Mbo*I/*Dpn*I restriction pattern (fig. 3.12 lane K). In order to justify the correlation between the arrowed bands in the lanes containing *Mbo*I-digested DNA from the transfected DNA samples (fig. 3.12 lanes E and F) and the control (fig. 3.12 lane K), another blot of a properly aligned gel has been included. Figure 3.13 shows a hybridization of the *hph* probe to Raji total genomic DNA from untransfected cells. Uncut p220.2 plasmid DNA is included in fig. 3.13 lane A, p220.2 linearized with *Hin*dIII in lane B and p220.2 digested with *Dpn*I in lane C. A weakly hybridizing band at 2.5kb is present in the *Mbo*I-digested DNA (lane F) and a band at ~3.3kb in the *Hin*dIII digestions. By comparison, the larger hybridizing band from the *Mbo*I/*Dpn*I restriction pattern is seen to migrate at the higher position of 2.8kb. Therefore, it may be justifiable to assume that the higher, arrowed bands in lanes E and F (figure 3.12b) correspond to the 2.8kb p220.2 *Dpn*I fragment.

The source of the 1.2kb *Eco*RI fragment of *hph* which was used as a DNA probe was p220.2. The fragment had an overlap of 40-50bp with the 5.5kb *Dra*I-*Hin*dIII fragment of p220.2.

Apparently, this overlap was sufficient in some cases to hybridize and give a visible signal in the autoradiography. The same may apply in the case of the *MboI/DpnI* restriction pattern where the probe hybridized to a 2.8kb fragment in the control and in the test samples. The ***hph*** sequence is cleaved by *MboI/DpnI* into fragments smaller than 400bp according to the restriction site information (Gritz and Davies, 1983) and figure 3.8. Therefore, the *hph* probe may have hybridized to a fragment which extended from *hph* to the next *MboI/DpnI* site in the EBV sequences. The possibility that the *hph* probe cross hybridizes to the EBV sequences may be discounted since signals were not detected in Raji Hirt DNA which is likely to contain EBV sequences.

The data shows that p220.2 DNA sequences were retained in the transfected cells for 52 days (the duration of the assay). The restriction patterns of the hybridizing sequences appear to match those of untransfected plasmid DNA. Transfected DNA apparently became sensitive to *MboI* by ~5 days indicating that the plasmid DNA had been replicated by this time.

In the *MboI*-digested tracks containing Hirt DNA, uncut plasmid DNA appears to be present for the duration of the assay. Therefore it may be concluded that unreplicated plasmid DNA which was *MboI*-resistant was still present in the population. This conclusion is supported by the results of the plasmid rescue experiments described below in section 3.4. This plasmid DNA may be contamination which remained in the tissue culture flask after electroporation. Alternatively, plasmids may have been retained within Raji cells for 52 days without replication despite regular subculturing of the population. Previous research has suggested that the nuclear retention function of EBNA-1 can be separated from its role as a promoter of replication. However, retention of EBV-based plasmids without replication has not been reported to extend longer than ~2 weeks (Reisman *et al.*, 1985; Krysan *et al.*, 1989). Unreplicated plasmids are normally diluted out of the population during growth and subculturing.

The copy number of p220.2 sequences in the transfected cell line Raji-p220.2(19.6) could be estimated. Published observations of the replication of EBV-based plasmids in human cell lines show that the number per cell ranges from 1 to 60 copies, but the average is usually 10 copies per cell. EBV plasmids, like the EBV viral genome, replicate once per cell cycle. However, copies accumulate in some cells owing to mis-segregation (Sugden *et al.*, 1985).

Approximately 10^6 cells were harvested per Hirt extraction. Assuming a 20% transfection efficiency for non-synchronized Raji cells (Teshigarawa and Katsura, 1992), then ($10^6/5 = 20\%$) 2×10^5 cells that contain plasmid. The Hirt extraction results in a 1ml SDS/NaCl supernatant of which 100µl (i.e. 1/10 of the 1ml supernatant) are precipitated to prepare a 60µl DNA sample. Therefore, 2×10^4 (2×10^5 /10) cell equivalents are in the 60µl sample. Since 30µl were used per digestion and per gel lane, 1×10^4 cell equivalents are present per lane.
If the plasmid replicates at a single copy per cell, then 1×10^4 copies are present per lane.

$$\text{amount of DNA per lane} = \frac{\text{no. of copies x molecular weight of 1bp x size of plasmid in bp}}{\text{Avogadro's number}}$$

$=> 1\times10^4 \times 660 \times 9000 / 6.2\times10^{23}$ = amount of DNA per lane = 9.6×10^{-14} g = 0.096pg

or ~0.1pg / lane if single copy

The actual amount of p220.2 DNA present in a lane can be estimated from the lanes of figure 3.9b [the *DraI*/*Hin*dIII digestion of Raji-p220.2(19.6)]. The control lane I contains 25pg of uncut p220.2 plasmid. In the photographic image, the bands in this lane have saturated owing to their high intensity. However, the original autoradiograph allows a comparison between the intensities of the bands in this control lane and of the bands in the 52 day Hirt extract (lane H). (The control bands did not appear to be saturated in the original autoradiograph.) The bands in the control were estimated to be 10-20x more intense than the 52 day bands. Therefore, since lane I contains 25pg DNA, the 52 day sample contains 1.25-2.5pg DNA (25pg /20 and 25pg /10). Dividing this figure by the amount of DNA per lane if single copy (0.1pg DNA which was calculated above) provides the number of copies per cell which is approximately 10-25 copies per cell. The reported copy number of EBV-based plasmids when replicated in mammalian cells is 1-60/cell.

3.3.4 Transfections with the circular vector p220.2TC, and the linear vectors pET and pIT

On the basis of the data from the Southern blots and the stable proliferation of hygromycin resistant cells, p220.2 was judged to be successfully replicating in Raji cells. Transfections of the modified vectors were then carried out. p220.2 was transfected into Raji cells again as a positive control for replication to be compared against p220.2TC (circular), pET and pIT (both linear).

The assay was carried out in the same way as the initial electroporation of p220.2 described in section 3.3.1 above. The cells were harvested at 3 day intervals from the cell populations which had been transfected with each vector.

(i) *Dra*I digestion of Hirt DNA from cells transfected with pET (linear)
(figure 3.14, lanes I-N)

Fig. 3.14 lane A contains Hirt DNA from untransfected cells. Digestion of the pET linear vector with *Dra*I produces two fragments which hybridize to the *hph* probe of about 3.5 and 6kb which are shown in the control lane Q. These fragments are visible in the Hirt DNA samples from transfected cells, but over the time course they fade out from 3 days (lane I) to 12 days (fig. 3.14 lane L) indicating that the vector was not retained within the population after this time.

(ii) *Mbo*I digestion of Hirt DNA from cells transfected with pET (linear)
(figure 3.15)

Lane A contains Hirt DNA from untransfected cells. The control lane I contains uncut pET which runs at a size of 10kb. The vector sequences are not visible by the final time point shown, 15 days (fig.3.15 lane F and uncut in lane H). The pET vector sequences are visible at the time points of 3, 6, 9 and 12 days (fig. 3.15 lanes B, C, D and E respectively), but they remain resistant to digestion with *Mbo*I and stay as a single 10kb band.

Figure 3.14 **The transfection of Raji cells with p220.2TC and pET**

(a) Ethidium bromide-stained gel of *Dra*I-digested Hirt DNA samples extracted over a time course after transfection with (i) p220.2TC, or (ii) pET (see text in section 3.3.4).

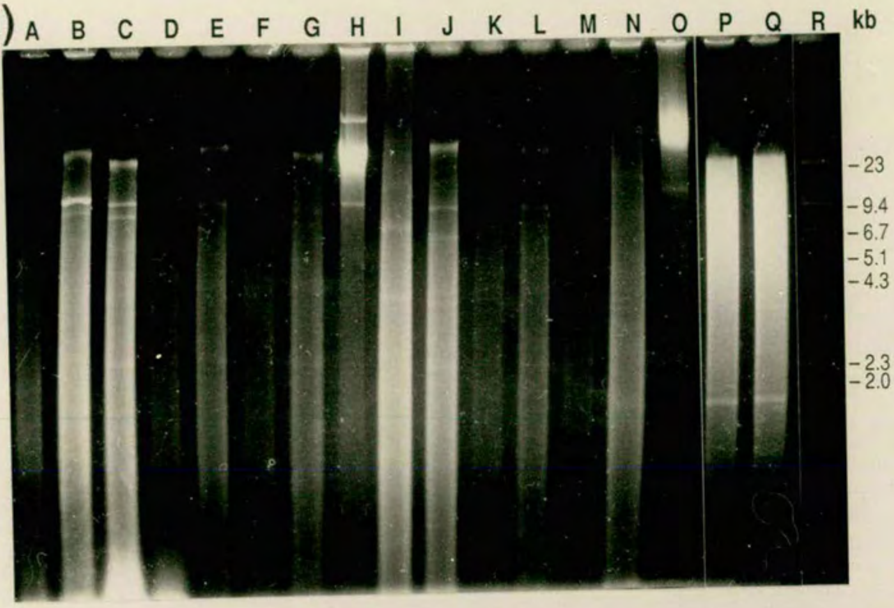
(b) The gel was blotted onto Hybond membrane and hybridized to the *hph* DNA probe labelled with ^{32}PdCTP. The data shows that the two vectors were lost from the cell population by the end of the time course. Cleavage of p220.2TC with *Dra*I prduces a 10.5kb fragment which hybridizes to *hph*. (The bands in lanes C and D appear to be slightly larger, probably owing to variations in the salt in the DNA samples affecting electrophoresis.) *Dra*I cleaves the linear vector pET into two fragments of approximately 3.5 and 6kb, both of which terminate in telomeric sequences at one end. The exposure time was overnight.
The fragment sizes of the λ *Hin*dIII DNA size marker are indicated.

Legend
A. DNA from untransfected cells digested with *Dra*I
B. DNA extracted 3 days after transfection with p220.2TC (*Dra*I digest)
C. " " 6 days " " "
D. " " 9 days " " "
E. " " 12 days " " "
F. " " 15 days " " "
G. " " 18 days " " "
H. " " 3 days " " " uncut

I. DNA extracted 3 days after transfection with pET (*Dra*I digest)
J. " " 6 days " " "
K. " " 9 days " " "
L. " " 12 days " " "
M. " " 15 days " " "
N. " " 18 days " " "
O. " " 3 days " uncut

P. p220.2TC plasmid cut with *Dra*I loaded with 10µg *Hin*dIII-digested total genomic human DNA
Q. p220.2TC plasmid cut with *Not*I + *Dra*I " " " "
R. λ *Hin*dIII

a)


A B C D E F G H I J K L M N O P Q R
kb
-23
-9.4
-6.7
-5.1
-4.3
-2.3
-2.0

b)

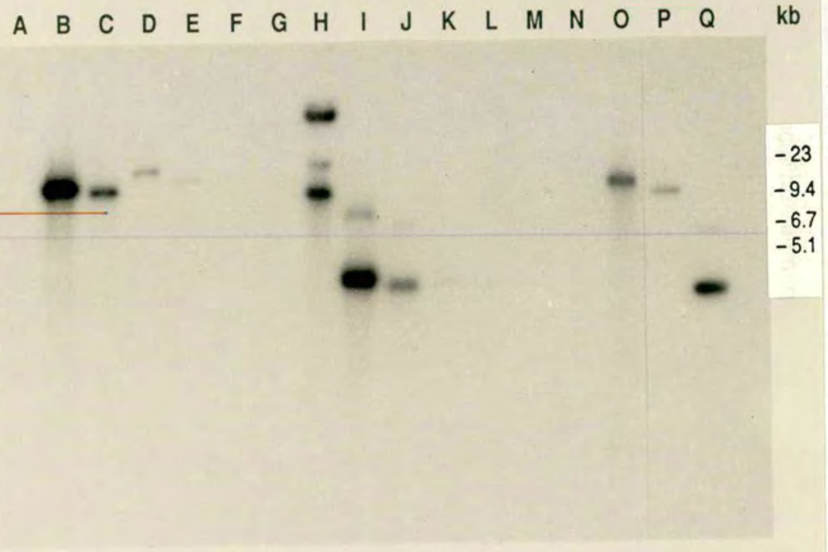
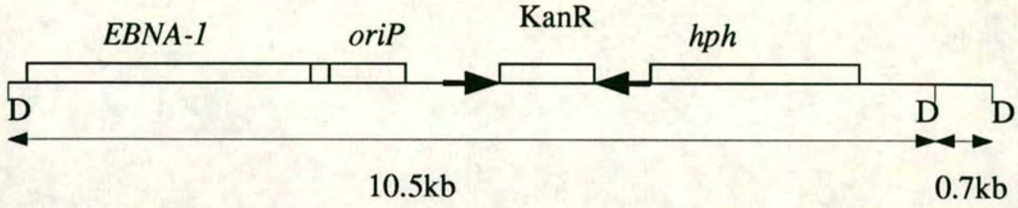

A B C D E F G H I J K L M N O P Q
kb
-23
-9.4
-6.7
-5.1

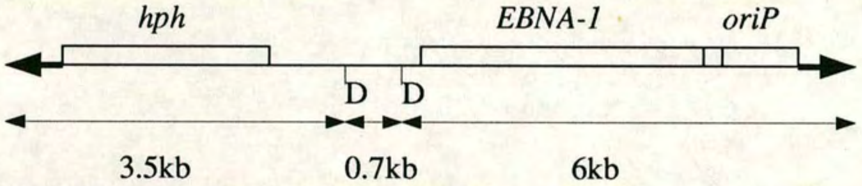
Figure 3.14c The linear *Dra*I restriction maps of p220.2TC and pET

p220.2TC



Only the large 10.5kb *Dra*I (D) fragment of p220.2TC is detected by the *hph* DNA probe.

pET



The *hph* DNA probe was able to detect the 3.5kb probe strongly, but also weakly detected the larger 6kb fragment.

Figure 3.15 Hirt-extracted DNA from Raji cells transfected with pET

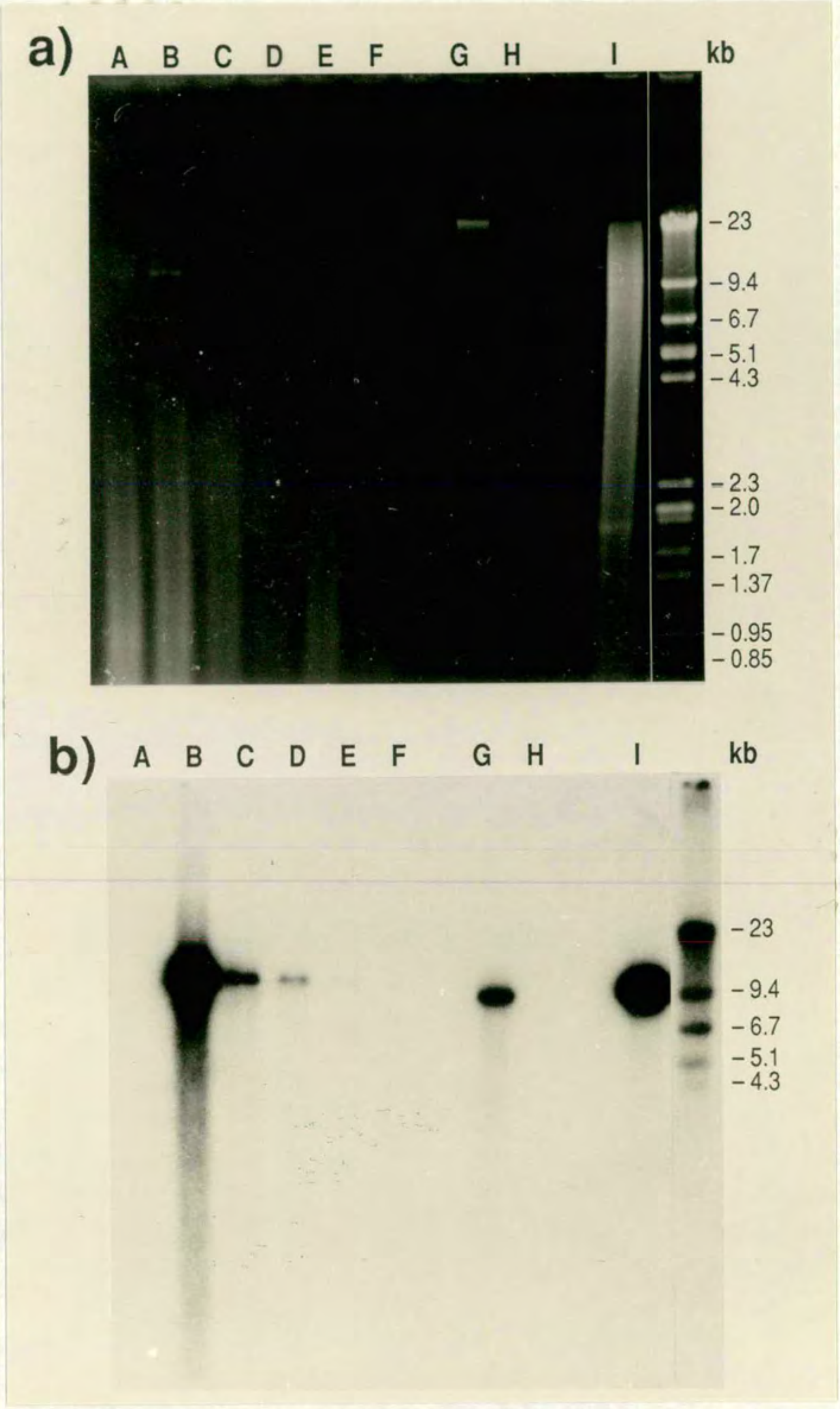
(a) Ethidium bromide-stained gel of *Mbo*I-digested Hirt DNA samples from the time course described in section 3.3.4 and figure 3.14.

(b) The gel was blotted onto Hybond membrane and hybridized to the *hph* DNA probe labelled with ^{32}PdCTP. pET vector DNA which was *Mbo*I-sensitive was not detected for the duration of the assay indicating that the DNA was not replicated by the mammalian cell (see text of section 3.3.4). The exposure time was 4 hours.
The fragment sizes of the λHHR DNA size marker are indicated.

Legend
A. DNA from untransfected cells digested with *Mbo*I enzyme
B. DNA extracted at 3 days post-electroporation "
C. " " 6 days " " " "
D. " " 9 days " " " "
E. " " 12 days " " " "
F. " " 15 days " " " "

G. " " 3 days uncut
H. " " 15 days uncut

I. p220.2TC plasmid cut with *Dra*I loaded with 10μg *Hin*dIII-digested total genomic human DNA
J. λHHR size marker


a)
A B C D E F G H I
kb
23
9.4
6.7
5.1
4.3
2.3
2.0
1.7
1.37
0.95
0.85
b)
A B C D E F G H I
kb
23
9.4
6.7
5.1
4.3

(iii) *Dra*I digestion of Hirt DNA from cells transfected with p220.2TC (circle) (figure 3.14, lanes B-H)

Fig. 3.14 lane A contains Hirt DNA from untransfected cells. Digestion of the plasmid p220.2TC with *Dra*I produces a single band of 10.5kb which hybridizes to *hph (*included in the control lane P). The hybridizing sequences are visible in the 3, 6, 9 and 12 day time points (fig.3.14 lanes B, C, D, and E respectively), but are no longer visible at 15 days (fig. 3.14 lane F). According to this Southern blot, the plasmid was lost from the population 15 days after transfection.

The vector sequences were present in the early samples at a high copy number because the hybridizing bands in fig. 3.14 lanes B, C and H apparently match the bright bands in the ethidium-stained gel. A strong hybridization to these bands was obtained which resulted in a relatively short exposure time for the autoradiograph. It could be argued that a longer exposure time might reveal vector sequences present at a lower copy number in the later time points. However, after 18 days, selection with hygromycin was applied to the transfected cells which were subsequently killed. This indicated that the *hph* sequence was no longer present in the cells in either an episomal or an integrated form.

(iv) *Mbo*I digestion of Hirt DNA from cells transfected with p220.2TC (figure 3.16)

Fig. 3.16 lane A contains Hirt DNA from untransfected cells. Lane H contains an undigested sample of Hirt DNA containing p220.2TC which appears as an uncut plasmid. The time course covers the time points 3, 6, 9, 12, 15 and 18 days (fig. 3.16 lanes B, C, D, E, F and G respectively). Digestion of the Hirt DNA samples with *Mbo*I showed that the plasmid remained uncut and, therefore unreplicated, until it disappeared from the population.

(v) *Dra*I and *Mbo*I digestion of Hirt DNA from cells transfected with pIT (linear with internal telomeric sequences) (figure 3.17)

DNA from untransfected Raji cells is present in fig. 3.17 lanes A and G. Lane P contains the *Mbo*I/*Dpn*I restriction pattern of p220.2. Lane O contains p220.2TC cleaved with *Dra*I

Figure 3.16 <u>Hirt-extracted DNA from Raji cells transfected with p220.2TC</u>

(a) Ethidium bromide-stained gel of *Mbo*I-digested Hirt DNA samples from the time course described in section 3.3.4 and figure 3.14.

(b) The gel was blotted onto Hybond membrane and hybridized to the *hph* DNA probe labelled with ^{32}PdCTP. The p220.2TC plasmid did not become *Mbo*I-sensitive before the end of the time course. The exposure time was overnight.
The fragment sizes of the λHHR DNA size marker are indicated.

Legend
A. DNA from untransfected cells digested with *Mbo*I enzyme
B. DNA extracted at 3 days post-electroporation "
C. " 6 days " "
D. " 9 days " "
E. " 15 days " "
G. " 18 days " "
H. " 3 days uncut

I. p220.2TC plasmid cut with *Dra*I loaded with 10µg of *Hin*dIII-digested total genomic human DNA
J. λHHR size marker

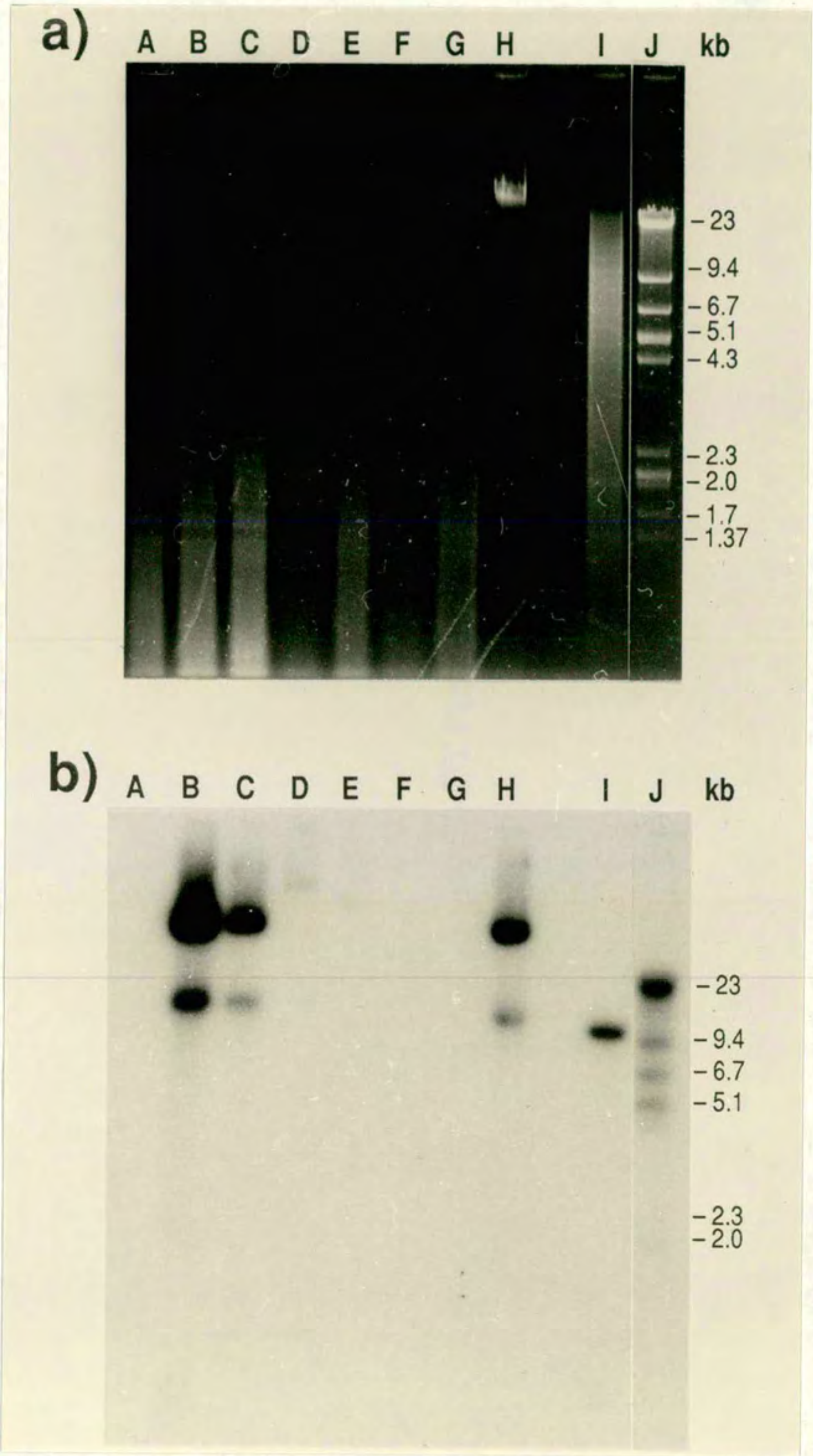

a)
A B C D E F G H I J kb
- 23
- 9.4
- 6.7
- 5.1
- 4.3
- 2.3
- 2.0
- 1.7
- 1.37
b)
A B C D E F G H I J kb
- 23
- 9.4
- 6.7
- 5.1
- 2.3
- 2.0

Figure 3.17 **Hirt-extracted DNA from Raji cells transfected with pIT**

(a) Ethidium bromide-stained gel of *Dra*I-digested and *Mbo*I-digested Hirt DNA samples from the time course described in section 3.3.4.

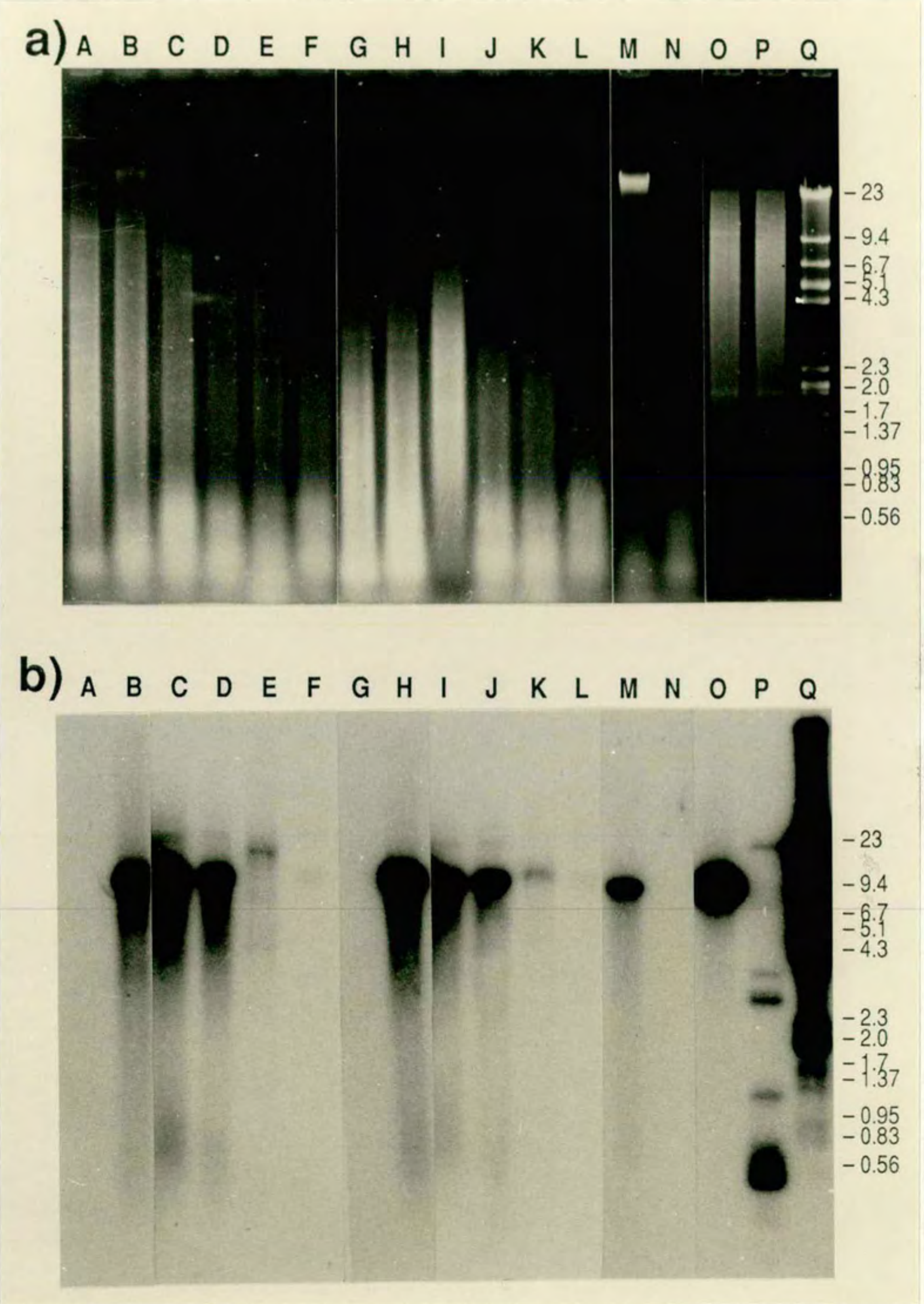
(b) The gel was blotted onto Hybond membrane and hybridized to the *hph* DNA probe labelled with ^{32}PdCTP. *Dra*I which leaves pIT intact was used to digest the background low molecular weight Raji chromosomal DNA to show that the DNA samples were fully cleavable by a restriction enzyme which was not affected by methylation. pIT was lost from the population by the end of the time course and did not become sensitive to *Mbo*I. The autoradiograph is a composite of films exposed to the membrane for different times: lanes A, B, G, H and M-Q were exposed overnight; lanes C-F and I-L were exposed for 2 days.
The fragment sizes of the λHHR DNA size marker are indicated.

Legend
A. DNA from untransfected cells digested with *Dra*I
B. DNA extracted 3 days post-electroporation "
C. " " 6 days " "
D. " " 9 days " "
E. " " 12 days " "
F. " " 15 days " "

G. DNA from untransfected cells digested with *Mbo*I
H. DNA extracted 3 days post-electroporation "
I. " " 6 days " "
J. " " 9 days " "
K. " " 12 days " "
L. " " 15 days " "

M. DNA extracted at 3 days uncut
N. " " at 15 days uncut

O. p220.2TC plasmid cut with *Dra*I loaded with 10μg of *Hin*dIII-digested total genomic human DNA
P. p2202. plasmid cut with *Dpn*I " " " "
Q. λHHR size marker


a)
A B C D E F G H I J K L M N O P Q
23
9.4
6.7
5.1
4.3
2.3
2.0
1.7
1.37
0.95
0.83
0.56
b)
A B C D E F G H I J K L M N O P Q
23
9.4
6.7
5.1
4.3
2.3
2.0
1.7
1.37
0.95
0.83
0.56

producing a 10.5kb band which is equivalent to uncut pIT. Hirt DNA samples from the time points 3, 6, 9, 12 and 15 days are included in fig. 3.17 lanes B, C, D, E and F (*Dra*I digestion), and fig. 3.17 H, I, J, K and L (*Mbo*I digestion). *Dra*I does not cleave within pIT, so only the contaminating cellular DNA has digested. The time course shown in lanes B-F shows that the hybridizing band which corresponds to the linear pIT molecule has almost faded by the 15 day time point showing that it was not retained within the population. The *Mbo*I digestions of the same DNA samples (lanes H-L) show that the sequences did not become sensitive to the enzyme (the 10.5kb band does not change in size) and, on this basis, were not replicated.

(vi) *Mbo*I digestion of Hirt DNA from cells transfected with p220.2 (positive control) (figure 3.18)

Fig. 3.18 lane A contains Hirt DNA from untransfected cells. The *Mbo*I/*Dpn*I restriction pattern of p220.2 is included in the control lane J. Lane H contains uncut p220.2 plasmid and lane I contains linearized p220.2 plasmid as single 9kb band. Lane G contains undigested Hirt DNA from the 3 day time point and demonstrates the appearance of uncut plasmid. The DNA samples are from the 3, 6, 9, 12 and 15 day time points (fig. 3.18 lanes B, C, D, E and F respectively). The hybridizing bands correspond to uncut plasmid showing that on this occasion, p220.2 sequences were not replicated by the mammalian cells and did not become sensitive to *Mbo*I enzyme. Additionally, hybridizing bands are not visible after 12 days indicating that the plasmid had been lost from the population without being replicated.

The Southern blot data shows that none of the four vectors, including the positive control p220.2, were retained or replicated in the cells over a 18 day period. This is a marked contrast to the previously described experiment where p220.2 had replicated by 10 days. After selection was applied to the cells at 18 days post-electroporation, no resistant cells appeared in the populations.

The two linear molecules, pET (which possessed exposed telomeric DNA sequences) and pIT (no exposed telomeric DNA) were lost from the transfected populations at about the same time and did not become sensitive to *Mbo*I.

Figure 3.18

Hirt-extracted DNA from Raji cells transfected with p220.2

(a) Ethidium bromide-stained gel of *Mbo*I-digested Hirt DNA samples from the time course described in section 3.3.4.

(b) The gel was blotted onto Hybond membrane and hybridized to the *hph* DNA probe labelled with ^{32}PdCTP. The p220.2 vector sequences were not detectable, or *Mbo*I-sensitive by the end of the assay. Unlike the results for Raji-p220.2(19.6) described in figures 3.10, 3.11 and 3.12, p220.2 sequences were not replicated in Raji cells on this occasion. The exposure time was overnight.

The fragment sizes of the λHHR DNA size marker are indicated.

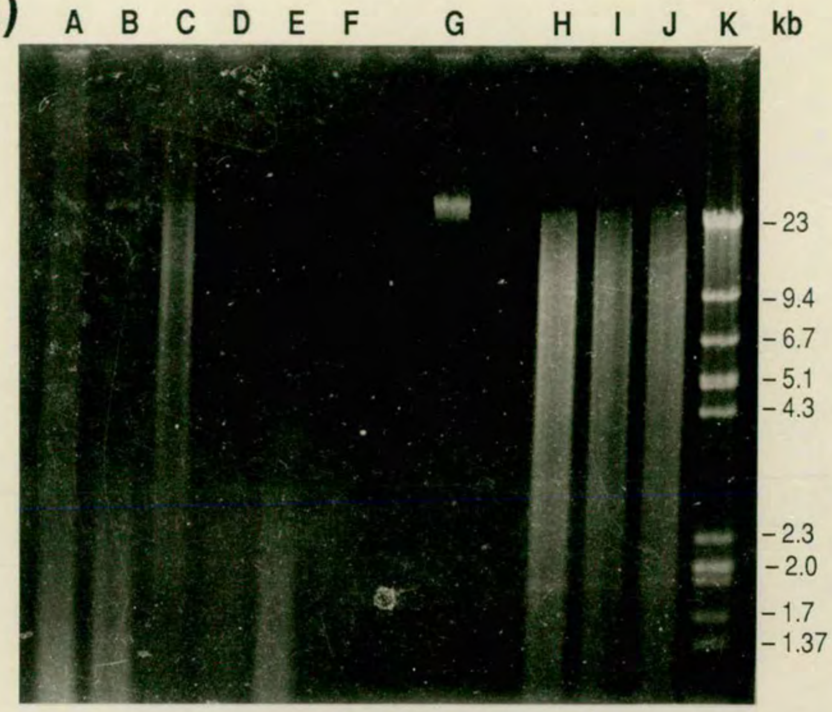
Legend

A. DNA from untransfected cells digested with *Mbo*I enzyme
B. DNA extracted at 3 days post-electroporation "
C. " 6 days " "
D. " 9 days " "
E. " 12 days " "
F. " 15 days " "

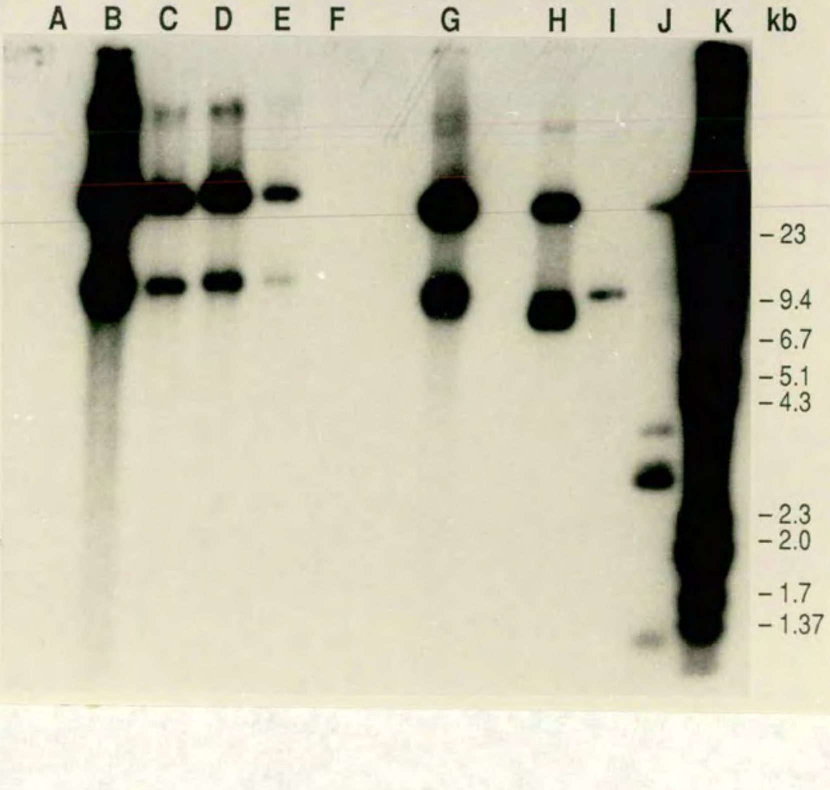
G. " 3 days uncut

H. p220.2 uncut plasmid loaded with 10μg of *Hin*dIII-digested total genomic human DNA
I. p220.2 plasmid cut with *Hin*dIII " " "
J. p220.2 plasmid cut with *Dpn*I " " "
K. λHHR size marker

a)


A B C D E F G H I J K kb
- 23
- 9.4
- 6.7
- 5.1
- 4.3
- 2.3
- 2.0
- 1.7
- 1.37

b)


A B C D E F G H I J K kb
- 23
- 9.4
- 6.7
- 5.1
- 4.3
- 2.3
- 2.0
- 1.7
- 1.37

The experiment was repeated 13 times, but on these subsequent occasions, the cells stopped proliferating by ~15 days post-electroporation with or without hygromycin selection. Different conditions of electroporation were tried, for instance lower pulse voltages and cells synchronized at G2/M phase (Goldstein *et al.*, 1989), but the modifications had no effect upon improving the health of the cells. Untransfected Raji cells continued to proliferate normally which discounted the possibility that the general growth conditions were deleterious to the cells. Control cells which were electroporated in the absence of DNA also died, so the assumption must be made that the electroporation itself was having deleterious effects upon the cells.

3.3.5 Summary

The concentration of the sheared chromosomal DNA present in the Hirt DNA varied between different samples. This variation was visible in ethidium bromide stained agarose gels. The variation was not removed from the samples because there was no correlation between the concentrations of the visible sheared chromosomal DNA and the transfected DNA. Measurement of the DNA concentration, for instance, via an absorbance reading at 260nm would provide a net value for the mixture of sheared chromosomal and vector DNA. Therefore, equalizing the "net concentrations" of different Hirt DNA samples might have introduced new variation into the concentrations of vector DNA alone. The lack of correlation between the concentration of total visible DNA and the vector DNA is best illustrated by figure 3.18. In the image of the autoradiograph (fig. 3.18b), strong vector hybridization signals are present in tracks where very little DNA is visible in the corresponding agarose gel (fig. 3.18a), notably lanes B-E.

The Southern blot data provided evidence that p220.2 could successfully transfect Raji cells on the following criteria: (i) the vector sequences were replicated and became sensitive to *Mbo*I enzyme; (ii) the vector sequences were retained within the population for a long term period; (iii) transfected cells became resistant to hygromycin B; (iv) the copy number of hybridizing sequences was 10-25/cell (within the expected range). It was apparent that if the p220.2 plasmid was replicating, the DNA was detectable by Southern blot analysis for at least 30 days without the use of selection. This observation corroborates published evidence that, in short term assays, replicating EBV plasmids are not rapidly diluted out of a population of proliferating cells.

The inconsistency of the transfection assay may be due to the complexity of the interaction between the EBV replication sequences and the host cell. There was no guarantee that after transfection the EBNA-1 protein would bind the vector DNA leading to (1) replication and (2) retention of the DNA. It may be hypothesized that slight variations in the culture conditions which did not affect growth of untransfected cells may have affected the ability of electroporated cells to (1) to continue proliferation and (2) allow the EBNA-1-*oriP* interaction.

In the final attempt at the experiment, the long term maintenance of p220.2 was apparently obtained. The cells which had been transfected with p220.2 continued to proliferate steadily after electroporation and showed resistance to 200µg/ml HmB. This cell line was designated Raji-p220.2(12.1), but insufficient time remained in the project to analyse it by Southern blots. The cells which had been transfected with p220.2TC, pET and pIT at the same time stopped dividing and died after 17 days. (This transfection was carried out using asynchronous cells which were electroporated using a pulse of 0.25kV at 500µF.)

3.4 RESCUE OF TRANSFECTED PLASMIDS

3.4.1 Introduction

Southern blot analysis of transfected cell lines can usually show if the vector sequences have been retained in the cells. However, it is not always clear if the vector sequences (1) are still extrachromosomal, (2) are free of rearrangements, (3) have been replicated by the mammalian cell.

The rescue of the transfected plasmids back into bacteria can usually answer these questions. Hirt-extracted DNA is enriched for low-molecular weight and extrachromosomal DNA. Plasmids which are present in Hirt-extracted DNA can be successfully retrieved by transformation of competent *E.coli*. The restriction enzyme, *Dpn*I, recognises the site GATC but only cleaves it if the adenine residue is methylated. When DNA is replicated in dam^+ bacteria this site is methylated and, therefore, sensitive to *Dpn*I, but when a transfected plasmid is replicated in a mammalian cell, it loses the bacterial methylation pattern. The GATC restriction site becomes unmethylated and, therefore, insensitive to *Dpn*I (although the site becomes sensitive to the isoschizomer *Mbo*I which cannot cleave the methylated GATC site). Pre-digestion of Hirt-extracted DNA with *Dpn*I degrades any bacterially replicated plasmid DNA and leaves intact only the DNA which was replicated by the mammalian cell. *Dpn*I-resistant plasmid DNA should be able to transform *E.coli* and produce colonies. The plasmid DNA can then be extracted from the bacteria, analysed by restriction digestion and compared against a sample of untransfected plasmid DNA.

In section 3.3.3, the copy number of replicating p220.2 in Raji cells was estimated to be 10-25 copies/cell. A 60µl Hirt DNA sample was calculated to contain 2×10^4 cell equivalents, therefore 1µl contained $2\times10^4 / 60 = 333$ cell equivalents /µl. If the copy number per cell is assumed to be 20, the number of copies per µl is ~6666. The amount of plasmid DNA in 1µl can therefore be calculated using the same formula as in section 3.3.3:

$6666 \times 660 \times 9000 / 6.2\times10^{23} = 6.4\times10^{-14}$g = **0.064pg p220.2 DNA/µl**

Transformations of *E.coli* were carried out using untreated and, for some samples, *Dpn*I-treated Hirt-extracted DNA. 1µl (0.064pg) of untreated Hirt-extracted DNA was used as a positive control for transformations.

*Dpn*I digestions were carried out using 10µl of the Hirt-extracted DNA (0.64pg) in a 30µl reaction volume. 10ng of pUC4KN were added to the reaction as a control for the efficiency of digestion with *Dpn*I. When pUC4KN was included in the reaction, the resulting transformation was plated out onto media containing kanamycin as well as ampicillin in order to check that no bacterially replicated plasmids were surviving the *Dpn*I digestion intact. (After digestion with *Dpn*I, no colonies should be obtained on media containing kanamycin as well as ampicillin.) After digestion, the reaction mix was dialysed using Millipore 0.025µm filters and 3µl (0.064pg of p220.2 + 1ng of pUC4KN) were used to transform competent *E.coli* by electroporation.

3.4.2 Rescue of p220.2 from Raji-p220.2(19.6)

Southern blot analysis of Raji cells transfected with p220.2 (figures 3.8, 3.9 and 3.10) showed that vector sequences which were *Mbo*I-sensitive were present up to 52 days post-electroporation. However, variation in the running of DNA in different lanes necessitated additional evidence that the plasmid DNA had not integrated or rearranged.

Despite some indications from the Southern blot data described in section 3.3.3 that p220.2 had replicated episomally, *Dpn*I-resistant plasmids could not be retrieved from the Hirt DNA samples.

The Hirt-extracted DNA from the time points of 3, 10, 25 48 and 52 days were used to transform *E.coli* (table 3.3). Bacterial colonies were obtained from transformations with untreated Hirt DNAs from all these time points.

The possibility existed that although plasmids were present in the Hirt-extracted DNA, they may have contained deletions or other rearrangements. "Mini-prep" DNA was obtained from the colonies and analysed by restriction digestion (figure 3.19). In all cases, the DNA proved to have the same restriction pattern as p220.2.

The subsequent retransfection of p220.2 into Raji cells which created Raji-p220.2(12.1) was judged to be successful on the basis of hygromycin resistance and transformations of *E.coli* to rescue plasmids. However, bacterial colonies were not obtained from *Dpn*I-digested Hirt DNA at 41 days after electroporation (table 3.5).

The efficiency of transformation in the presence of Hirt DNA could be calculated from the plating efficiency of pUC4KN. 1ng of pUC4KN was used in a transformation. 1/100 of the actual transformation was plated onto selective media, therefore the colonies were produced by 1ng/100 = 0.01ng = 10pg DNA. 10pg of pUC4KN produced a mean colony number of 623 (from 640, 367 and 861) which is equivalent to an efficiency of 62 colonies/pg. The presence of the sheared genomic DNA present in Hirt DNA samples may have reduced the overall transformation efficiency. The amount of replicated p220.2 plasmid DNA per µl was 0.064pg, therefore it may be assumed that the assay was not sensitive enough to rescue replicated p220.2 plasmid.

Table 3.3* Hirt DNA from Raji-p220.2(19.6)

Days post-electroporation	**untreated DNA**
3	44
10	3
25	10
48	6
52	52

Table 3.4 Hirt DNA from Raji-p220.2(19.6)

Days post-electroporation	**untreated DNA (on Ap media)**	**untreated DNA + pUC4KN (on Ap & Kan)**	***Dpn*I-digested + pUC4KN (on Ap media)**	***Dpn*I-digested + pUC4KN (on Ap & Kan media)**
25	10	640*	0	0
52	11	367*	0	0

Table 3.5 Hirt DNA from Raji-p220.2(12.1)

Days post-electroporation	**untreated DNA (on Ap media)**	**untreated DNA + pUC4KN (on Ap & Kan)**	***Dpn*I-digested + pUC4KN (on Ap media)**	***Dpn*I-digested + pUC4KN (on Ap & Kan media)**
41	6	861*	0	0

Table 3.6 p220.2TC (unreplicated)

Days post-electroporation	**untreated**	***Dpn*I-digested**
6	211	0
12	16	0
18	1	0

Table 3.7 p220.2 (unreplicated)

Days post-electroporation	**untreated**	***Dpn*I-digested**
6	374	0
12	44	0
18	2	0

1/10 of the incubated SOC culture was plated out except where marked *.

*1/100 of the incubated SOC culture was plated out.

Figure 3.19 **The rescue of transfected plasmids into *E.coli***

Hirt-extracted DNA samples from transfected cells were used to transform competent *E.coli* and plated onto selective media. DNA was extracted from the colonies and digested with appropriate restriction enzymes. The restriction digests were electrophoresed through agarose gels stained with ethidium bromide to confirm that the plasmids did not contain deletions or rearrangements.
The fragment sizes of the λHHR DNA size marker are indicated.

(a) Rescue of p220.2 plasmids from transfected cell line Raji-p220.2(19.6) (25 days post-electroporation)

12 colonies were picked from transformants resulting from bacteria transformed with *Dpn*I-digested Hirt DNA. A background of contaminating bacterial chromosomal DNA is present but the *Eco*RI restriction pattern of p220.2 (lanes 7 and 15) is visible. (The smallest band of <0.5kb is only faintly visible in most of the lanes although it was visible in the agarose gel.)

1 -6 and 9 -14: DNA from individual transformed colonies digested with *Eco*RI
7 and 15: p220.2 plasmid DNA digested with *Eco*RI
8 and 16: λHHR size marker

(b) Rescue of p220.2 plasmids from Raji-p220.2(19.6) (52 days post-electroporation)

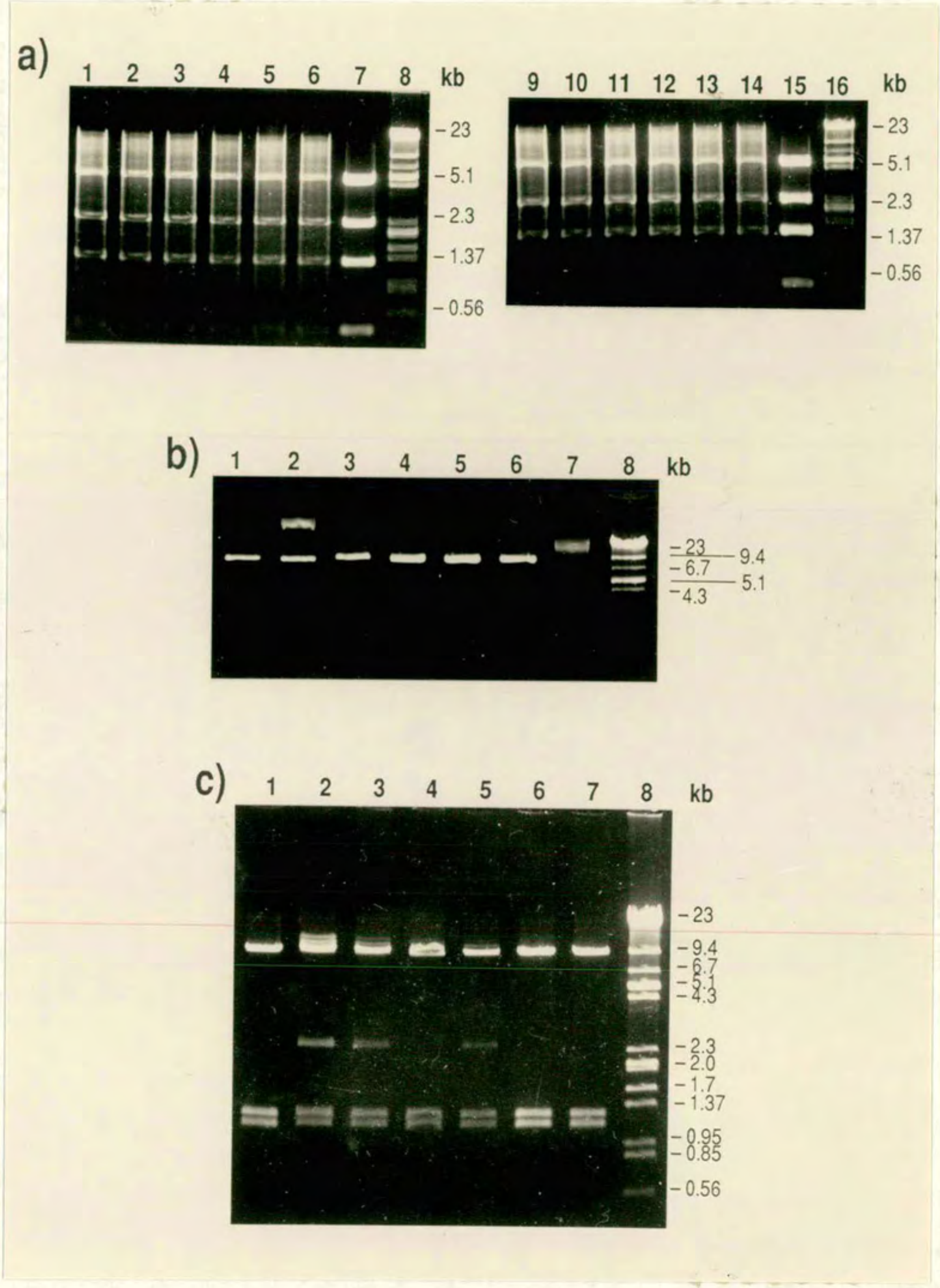
5 colonies were picked from transformants resulting from bacteria transformed with Hirt DNA. Digestion of the DNA with *Hin*dIII produced a single 9kb fragment which matches linearized p220.2 (lane 6).

1 - 5: DNA from individual transformed colonies digested with *Hin*dIII
6: p220.2 plasmid DNA digested with *Hin*dIII
7: uncut DNA from 1.
8: λHHR

(c) Rescue of unreplicated p220.2TC plasmids from Raji-p220.2TC(22.9)

Hirt DNA samples from 3 time points which had not been digested with *Dpn*I were transformed into bacteria. The restriction patterns produced by digestion with *Hin*dIII match that of untransfected p220.2TC. *Hin*dIII digestion of p220.2TC produces three bands of 1.0, 1.2 and 9.0kb. The 9kb fragment is the whole of the p220.2 sequence. The 1.0 and 1.2kb fragments are the two fragments of the telomere cassette which are produced by *Hin*dIII digestion (see figure 3.4). The band seen at ~2.3kb in lanes 2, 3 and 5 is a partial digestion product and represents the entire telomere DNA cassette.

1 - 3: colonies from Hirt DNA extracted at 6 days
4 - 5: colonies from Hirt DNA extracted at 12 days
6: colony from Hirt DNA extracted at 18 days
7: p220.2TC plasmid DNA digested with *Hin*dIII
8: λHHR size marker


a)
1 2 3 4 5 6 7 8 kb
- 23
- 5.1
- 2.3
- 1.37
- 0.56
9 10 11 12 13 14 15 16 kb
- 23
- 5.1
- 2.3
- 1.37
- 0.56
b)
1 2 3 4 5 6 7 8 kb
- 23
9.4
- 6.7
5.1
- 4.3
c)
1 2 3 4 5 6 7 8 kb
- 23
- 9.4
- 6.7
- 5.1
- 4.3
- 2.3
- 2.0
- 1.7
- 1.37
- 0.95
- 0.85
- 0.56

3.4.3 Rescue of p220.2TC from Raji cells

The Southern blot data of the transfection of p220.2TC and the retransfection of p220.2 in figures 3.14, 3.16 and 3.18 show that in this experiment neither plasmid was retained or replicated in Raji cells. This was verified by transformations of bacteria with Hirt DNA from 6, 12 and 18 days post-electroporation (tables 3.6 and 3.7).

No colonies were obtained from *Dpn*I-digested DNA and the transformation frequency is reduced proportionally with time. The plasmids were still present at 18 days post-electroporation although sub-culturing of the cells was gradually diluting them out of the population. Restriction digestion of the p220.2TC plasmid extracted from the bacteria showed that the DNA had not been rearranged by the transfection procedure (figure 3.19).

3.4.4 Conclusions

The numbers of colonies obtained were the result of the transformation of Hirt DNA, only a fraction of which is plasmid DNA. Much of the ~20ng of DNA used in the transformations consisted of sheared cellular DNA. Therefore, although the efficiency appears at first sight to be low, it should be noted that these transformations are not directly comparable to those done using a similar amount of pure plasmid DNA.

The plasmid, p220.2, could not be recovered from Hirt-extracted DNA from transfected Raji cells after digestion with *Dpn*I enzyme. Unreplicated plasmid could, however, be retrieved from the Hirt DNA in unrearranged forms from all the time points which were tested (figure 3.18). Since it is unlikely that the plasmid rescue assay was sensitive enough to detect replicated p220.2, the assay was not able to prove or disprove the presence of episomally replicated p220.2 plasmids in the Raji cells. However, unreplicated episomal plasmids were present in the cell population for over 7 weeks. This apparent discrepancy will be discussed in the following chapter.

The plasmid rescue data substantiated the Southern blot data in section 3.3.4 and figures 3.14 and 3.16 concerning the loss without replication of p220.2 TC from populations of transfected Raji cells.

According to the data in table 3.3, *Dpn*I-sensitive p220.2 persisted in the transfected cells for the duration of the 52 day time course. It might have been expected that in the long term, the only episomal plasmid DNA that remained in the cell population after regular subculturing, if any, would be endogenously replicated and *Dpn*I-resistant. If this were the case, then the same number of bacterial colonies would be obtained from *Dpn*I-digested and undigested Hirt DNA. However, the data shows that colonies could only be retrieved from undigested DNA. It is not clear whether the bacterially replicated plasmid DNA was actually inside the cells or simply contaminating the surfaces of the flask vessel.

It may be argued that the persistance of even an EBV-based plasmid inside cells for over 50 days without being replicated is unlikely and that any remaining bacterially replicated plasmids must be external to the cells. It is valid, though, to point out that episomal p220.2 was not detected in any form where the transfection of p220.2 was completely unsuccessful (see section 3.3.4 and figure 3.18). Examination of the Hirt DNA samples from a transfected cell line where p220.2 had been lost 18 days after electroporation, showed that (1) the vector DNA sequences did not become sensitive to *Mbo*I enzyme, (2) the vector sequences could not be detected by hybridization to the *hph* DNA probe after 15 days (figure 3.18) and (3) the number of bacterial colonies that were obtained by transformation with Hirt DNA decreased with successive time-points (table 3.7).

CHAPTER 4 DISCUSSION

DISCUSSION

4.1 Summary

An existing EBV-based bacterial-mammalian cell shuttle vector was modified by the insertion of a telomeric DNA cassette. This circular modified shuttle vector was converted into a linear vector by the excision of a stuffer fragment which exposed the telomeric sequences. The linear and circular modified vectors were transfected into mammalian cells in order to test their ability to replicate compared with the unmodified vector. None of the modified EBV vectors were episomally replicated by the human cells. Additionally, no increase in the size of the linear vector via elongation of the telomeric sequences was detected. The unmodified vector, p220.2, did appear to replicate in the mammalian cells on the basis of data from Southern blots and the conferral of resistance to hygromycin. Plasmids which had been replicated episomally were not rescued from the transfected cells, but this may be a result of lack of sensitivity in the assay. These observations are discussed below in relation to the published literature.

4.2 Vectors

Several strategies were attempted in order to insert tracts of telomeric DNA repeats into p220.2. This was accomplished by the simultaneous ligation into p220.2 of two telomeric DNA fragments together with a stuffer fragment carrying kanamycin resistance. The correct orientations of the fragments were ensured by the utilization of cohesive termini (see section 3.1.3, The construction of p220.2TC). The resultant plasmid, p220.2TC, was converted into a linear vector by the excision of the stuffer fragment using a pair of rare-cutter restriction sites.

A crucial drawback in the use of p220.2 for this work was not discovered until the project was well underway. Gahn and Schildkraut (1989) reported that the EBV *oriP* element contains both the initiation and termination sites of DNA replication. The termination site is situated in the block of direct repeats where the EBNA-1 protein binds (figure 1.1) and was identified by 2-dimensional gel electrophoresis as a replication fork barrier. It has been suggested that bound molecules of

EBNA-1 protein at *oriP* might provide an actual physical barrier to the fork (Dhar and Schildkraut, 1991). In circular DNA molecules, such as the EBV genome and the plasmid p220.2, replication is carried out by the unblocked fork moving in the opposite direction which completes the circle. If, however, *oriP* is present on a linear vector, replication of the entire molecule cannot be completed if the fork barrier is present. Therefore, the linear pET vector could not be fully replicated.

The experiments of Krysan and Calos (1991) and Caddle and Calos (1992) suggested that putative ARS elements such as that within pLIB16 promote the initiation of replication not only within the human genomic DNA fragment, but also within the surrounding bacterial vector sequences. I had hoped that the putative human ARS in pLIB16 could bypass the fork barrier in *oriP*. The 14kb genomic DNA fragment is adjacent to *oriP* in the plasmid, therefore it is theoretically possible that the ARS element could initiate replication bubbles on both sides of *oriP* enabling the complete replication of a linearized version of pLIB16. However, this experiment could not be executed because the telomere cassette could not be inserted into pLIB16 without substantial deletions occurring.

The plasmid p220.2TC may be of use in the construction of linear vectors for other eukaryotic systems as it was for the work in *Podospora anserina* (Javerzat *et al.*, 1993). The telomeric DNA cassette, while extremely difficult to construct, proved to be stable in p220.2TC. A great deal of effort would be saved in the construction of future vectors if this plasmid was used as a source of cloned telomeric tracts in an opposing orientation which can easily be excized as a single fragment.

4.3 Transfections

Early data from Southern blots had shown that the p220.2 vector sequences had been maintained in the transfected cells over a time course and that the plasmid had lost its original bacterial methylation indicating replication by the mammalian cell. Calculations of the copy number of p220.2 DNA sequences in the 52 day Hirt sample were later carried out and found to be 10-25/cell which agrees with published literature (Sugden *et al.*, 1985; Krysan *et al.*, 1989). On this basis, further experiments were carried out using the modified circular vector, p220.2TC and the linear vectors, pET and pIT. These modified vectors were not replicated or retained by Raji cells. Results concerning these modified vectors could not be obtained from a set of transfections in which p220.2 was seen to replicate on the same criteria as described above despite many attempts. The results of plasmid rescue tests and Southern blotting analysis showed that the DNA was successfully introduced into the cells by electroporation. Entry into the cells did not prove to be a guarantee that the DNA would be replicated.

Plasmid rescue experiments were carried out in order to verify that the cell line Raji-p220.2(19.6) was truly supporting extrachromosomal replication of the p220.2 EBV plasmid. Although unreplicated plasmids could be rescued into *E.coli* bacteria by transformation with Hirt-extracted DNA using samples taken from the final time point of the assay, extrachromosomal replicated plasmids could not be retrieved after pre-digestion of the DNA with *Dpn*I enzyme. From estimates of the copy number per cell, in conjunction with estimates of the concentration of cell equivalents in the Hirt extracted DNA samples, the concentration of the plasmid DNA in the Hirt extracts was found (~0.064pg/µl). Calculations of the transfromation efficiency which was obtained in the plasmid rescue experiments showed that under the conditions used, the assay was probably not sensitive enough to retrieve p220.2 plasmids which might have been present.

Southern blots of the Hirt-extracted and total genomic DNA of Raji-p220.2(19.6) had seemed to indicate that p220.2 DNA sequences were present in an unrearranged and replicated state. It may be argued that a few copies of the plasmid had integrated into the host chromosomes without rearrangement, thereby enabling replication by the host cells and conferring stable resistance to hygromycin. This kind of integration which gives the impression of autonomous replication has been reported previously in research into putative murine ARS elements (Gilbert and Cohen, and Grummt, 1989). Although integration of the vector sequence is a logical explanation, the analysis of the total genomic DNA of Raji-p220.2(19.6) (figure 3.10) does not provide good evidence for integration. The bands that hybridized to the *hph* DNA probe in the uncut samples of DNA may correspond to the uncut plasmid bands in the positive control, but, more importantly, did not localize, as was found in the work of Gilbert and Cohen, and of Grummt, to the high molecular weight band running at the limiting mobility of the gel. The work on murine ARSs also found that the integrated vectors were present in the chromosomes in tandem arrays of high copy number which is not the case in Raji-p220.2(19.6). Such an array would have produced a very strong signal in uncut total genomic DNA at the position of limiting mobility as well as signals in the background chromosomal smear in digested samples. The only hybridizing signals are in accordance with residual limited copy number plasmids. The higher molecular weight bands present in the *Mbo*I digestions correspond to uncut plasmid DNA bands, but it was already shown from study of the Hirt-extracted DNA that unrearranged *Mbo*I-resistant plasmids were present in the DNA samples.

The presence of apparently unreplicated, unrearranged episomal copies of p220.2 at 53 days after transfection remains as an anomaly. Insufficient time remained in the project to investigate whether the plasmid was retained within the cells by an interaction with the EBNA-1 protein, or if the DNA was simply contaminating flask surfaces. A control set of successive Hirt extractions where DNA had simply been mixed with non-electroporated DNA, however, showed that contaminating extracellular DNA was lost relatively quickly.

The inconsistency of the replicative capability of p220.2 may be the result of unknown variations in the growth conditions, e.g. different batches of RPMI 1640 medium, coupled with the complexity of the interaction between the EBNA-1 protein and its binding sites in *oriP* and possible binding sites in the host nucleus. The molecular mechanism by which the EBNA-1 protein initiates replication and acts in nuclear retention of linked DNA is still unknown.

Owing to the inability of the positive control to replicate consistently, the modified vectors, p220.2TC and pET, could not be assayed in a truly valid context. Ideally, the fate of the vectors in Raji cells should have been studied in an assay where (1) p220.2 was replicating in cells from the same population and (2) all populations of transfected cells proliferated for the duration of the assay in the absence of selection.

On the basis of the available data, it is not clear whether the failure of p220.2TC to replicate was caused by the introduction of tracts of telomeric DNA repeats adjacent to the *oriP* element. Since the *oriP* element also contains repeated DNA, the telomeric DNA may have interfered with its function. However, when the absence of replication of p220.2 is considered, it is possible that p220.2TC is competent to replicate, but at a still lower frequency. The plasmid could be retrieved into bacteria from transfected cells for 2 weeks after electroporation without suffering mutations or rearrangements.

There is no obvious reason why the telomeric sequences on pET were not detectably elongated by telomerase. Farr *et al.* (1991) showed that as little as 500bp of terminal (TTAGGG)n repeats could seed telomere formation in mammalian cells. However, this tract of repeats was acting as a telomere on a broken chromosome as opposed to an episomal vector. Unlike the linear vector, pET, the broken chromosome was stably maintained in the cells for an adequately long period for telomerase activity to be detected. Presumably, Raji cells contain active telomerase like other immortalized cell lines. The telomeres of immortalized cell lines have been observed to be set at a constant length (Counter *et al.*, 1992). The telomerase activity is probably regulated, both positively and negatively, otherwise the telomeres would show steady growth. If Raji cells are assumed to regulate telomerase in the same manner, maybe the vector pET was not maintained in the cells long enough for telomerase to show any appreciable activity. A hypothetical regulation mechanism might recognise that all the endogenous telomeres were already at the "standard length", so telomerase activity was not induced.

It may be argued that the work of Farr *et al.* (1991), who introduced a new telomere onto a chromosome by breakage, negated the necessity of this experiment. However, the context of the end of a full-sized chromosome and a 10kb linear vector are different. No published work is yet available which shows that human telomerase can elongate telomeric sequences which are present on a non-chromosomal vector inside a cell. It was thought that the results of this type of experiment might be difficult to interpret because a linear vector might simply integrate into the chromosomes as a response to antibiotic selection. However, the EBV-based linear vector, pET, did not appear to integrate which implies that the terminal telomeric DNA sequences were still exposed to any active telomerase in the cell. (When, according to the Southern blot data, the pET DNA sequences had been lost from the Hirt-extracted DNA, hygromycin resistant cells did not arise in the population.) These observations must be balanced against inconsistent replication at the EBV *oriP* element and, implicitly, function of the EBNA-1 protein as a nuclear retention factor.

4.4 The viral-based approach to linear vectors

The inability of the unmodified p220.2 plasmid to replicate consistently was disappointing, but not wholly unusual. Vectors based on bovine papillomavirus-1 have been developed to act as episomes in mammalian cells (Sambrook *et al.*, 1989), but BPV vectors have been found to be inconsistent in their episomal stability and often integrate into the host chromosomes (Bostock and Allshire, 1986).

Any functional linear vector which was developed using EBV-based vectors would be restricted in its uses. EBV-based vectors may not be suitable for applications in, for instance, gene therapy owing to the requirements for expression of the *EBNA-1* gene. EBNA-1 negative cells may become transformed into cancer cells by the integration of EBV sequences leading to rearrangements, or as yet unknown properties of the EBNA-1 protein. In addition, EBV vectors do not replicate episomally in murine cells (Yates *et al.*, 1985) which provide the major model system for genetic study in mammals.

Attempts at creating linear vectors from viral-based plasmids have been made by other researchers. Linear episomes that terminate in human telomeric repeats were constructed which utilized the SV40 origin of replication and neomycin or histidinol resistance selectable markers. The only drug resistant transfectants that were obtained had integrated copies of the selectable marker indicating that the telomeric repeats were not sufficient to prevent integration (John Hanish and Titia De Lange, meeting abstract).

4.5 Other approaches to linear vectors

The creation of artificial chromosomes or linear vectors can take several approaches. In the introduction (sections 1.5.3 and 1.5.5) the idea of constructing *E.coli*: yeast:mammalian shuttle vectors was raised. This work is being carried out by other researchers and the resultant vectors are in the early stages of testing in mammalian cells. Existing YAC arms have been modified by the introduction of human telomeric sequences into the ends of both arms. The neomycin resistance selectable marker has been inserted into one of the arms. In order to obtain a mammalian ARS element, these arms were used to create a YAC library with human genomic DNA inserts of 100-280kb. It is probable that genomic DNA fragments of this size contain origins of replication (Huberman and Riggs, 1968). Although larger vectors can be designed in yeast, the technical difficulty exists of purifying the YAC vector prior to transfection into mammalian cells.

The work described in this project involves the "building up" of a vector from separate elements. Alternatively, an existing chromosome could be broken to create a stable minichromosome carrying an integrated selectable marker.

One such idea utilizes the "half-YAC" cloning procedure which was used to isolate a human telomere to create a mouse minichromosome (Brown, 1989; Cross *et al.*, 1989). Mouse chromosomes are telocentric, i.e. the centromere is adjacent to one of the telomeres. Therefore, it is feasible to create a murine minichromosome by ligating a single YAC arm carrying both yeast and mammalian selectable markers to large fragments of genomic DNA in order to "catch" a telomere and centromere together on one fragment. Transformation into yeast of the resulting library would be followed by colony hybridizations that used (a) TTAGGG repeats and (b) mouse minor satellite repeats as probes for mouse telomeres and centromeres respectively to identify a YAC-mouse minichromosome.

The telomere-associated chromosome fragmentation (TACF) event described by Farr *et al.* (1991) has opened up new avenues in the creation of minichromosomes by breakage. The new telomere was formed close to the site of the pre-existing chromosomal telomere in the human X-chromosome of the rodent-human somatic cell hybrid which they used. Recent work (Farr *et al.*, 1992) describes the creation of a set of nested terminal deletions on the q arm extending to the centromere. These were made using positive selection for histidinol resistance (integrated into the p arm) and negative selection against *HPRT* (situated on the distal portion of the q arm). It is theoretically possible by using the same techniques to make truncated chromosomes which contain little more than a centromere, reintroduced telomeres and selectable markers. The chromosomes of the mouse may be more realistic candidates for this approach than human chromosomes, since the centromere is already adjacent to one telomere.

4.6 Concluding remarks

The construction of linear vectors for mammalian cells is a challenging area of research. After human telomeres were cloned in YAC vectors, linear vectors were thought to be the next step. However, it is now apparent that the successful linear vectors or artificial chromosomes for mammalian cells will not be easily constructed. To date, artificial chromosomes have only been constructed for one of the least complex eukaryotes, *Saccharomyces cerevisiae*. Research is still underway to develop an artificial chromosome-based cloning system for the fission yeast, *Schizosaccharomyces pombe*.

A simplistic approach to constructing a linear vector might involve the following plan. A cosmid or plasmid containing a tract of centromeric DNA, e.g. α-satellite, and a mammalian selectable marker is linearized and telomeric DNA sequences are ligated onto the termini. The ligation product is then directly transfected into mammalian cells. This type of experiment might not provide any useful findings. Haaf *et al.* (1992) transfected cosmid clones of α-satellite DNA into immortalized African Green Monkey cells which randomly integrated into the endogenous chromosomes. The introduced satellite DNA sequences then rearranged and amplified into large arrays. The same result may await the theoretical linear vector described above.

Research into chromosomal biology is a necessary prerequisite to the construction of artificial chromosomes. A circular problem is present in that "basic chromosomal elements", e.g. ARSs and centromeric sequences, are required to construct artificial chromosomes, but some form of artificial chromosome is required to assay the same elements for function. Viral-based linear vectors may only be restricted to assays of function of, for instance, putative centromeric sequences and not suitable for the future construction of artificial chromosomes, but they may be important in this area. An example is the testing of putative human ARS elements in EBV vectors (Krysan *et al.*, 1989; Heinzel *et al.*, 1991; Krysan and Calos, 1991; Caddle and Calos, 1992). However, obtaining consistent replication of a chosen viral-based vector is necessary before valid work can be carried out.

REFERENCES

REFERENCES

Agard, D.A. and Sedat, J.W. (1983) Three dimensional architecture of a polytene nucleus. *Nature* **302**: 676-681.

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 of *T.thermophila* cross hybridizes with human telomeres. *Nature* **332**: 656-659.

Allshire, R.C., Dempster, M. and Hastie, N.D. (1989) Human telomeres contain at least three types of G-rich repeat distributed randomly. *Nucl. Acids Res.* **17**: 4611-4627.

Ariga, H., Imamura, Y. and Iguchi-Ariga, S.M.M. (1989) DNA replication origin and transcriptional enhancer in c-*myc* gene share the c-*myc* protein binding sequences. *EMBO Journal* **8**: 4273-4279.

Bell, S.P. and Stillman, B. (1992) ATP-dependant recognition of eukaryotic origins of replication by a multiprotein complex. *Nature* **357**: 128-134.

Bernards, A., Michels, P.A.M., Lincke, C.R. and Borst, P. (1983) Growth of chromosomal ends in multiplying trypanosomes. *Nature* **303**: 592-597.

Bernat, R.L., Borisey, G.G., Rothfield, N.F. and Earnshaw, W.C. (1990) Injection of anti-centromere antibodies in interphase disrupts events required for chromosome movement at mitosis. *Journal of Cell Biology* **111**: 1519-1533.

Biamonti, G., Della Valle, G., Talarco, D., Cobianchi, F., Riva, S. and Falaschia, A. (1985) Fate of exogenous recombinant plasmids introduced into mouse and human cells. *Nucl. Acids Res.* **13**: 5545-5561

Biessmann, H., Carter, S.B., and Mason, J.M. (1990) Chromosomes in *Drosophila* without telomeric sequences. *Proc. Natl. Acad. Sci. USA* **87**: 1758-1761.

Biessmann, H., Valgeirsdottir, K., Lofsky, A., Chin, C., Ginther, B., Levis, R.W. and Pardue, M.-L. (1992) HeT-A, a transposable element specifically involved in "healing" broken chromosome ends in *Drosophila melanogaster*. *Mol. Cell. Biol.* **12**: 3910-3918.

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

Blackburn, E.H. and Gall, J.G. (1978) A tandemly repeated sequence at the termini of the extrachromosomal rRNA genes in *Tetrahymena*. *Journal of Molecular Biology* **120**: 33-53.

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 Szostak, J.W. (1984) The molecular structure of centromeres and telomeres. *Ann. Rev. Biochem.* **53**: 163-194.

Blackburn, E.H. and Karrer, K.M. (1986) Genomic organization in ciliated protozoans. *Ann. Rev. Genetics* **20**: 501-521.

Blochlinger, K. and Diggelman, H. (1984) Hygromycin B phosphotransferase as a selectable marker for DNA transfer experiments with higher eukaryotic cells. *Mol. Cell. Biol.* **4**: 2929-2931.

Bloom, K., Fitzgerald-Hayes, M. and Carbon, J. (1983) Structural analysis and organization of yeast centromeres. *Cold Spring Harbor Symp. Quant. Biol.* **47**: 1175-1185.

Borowiec, J.A., Dean, F.B., Bullock, P.A. and Hurwitz, J. (1990) Binding and unwinding - how T antigen engages the SV40 origin of DNA replication. *Cell* **60**: 181-184.

Bostock, C.J. and Allshire, R.C. (1986) Comparison of methods for introducing vectors based on bovine papillomavirus-1 DNA into mammalian cells. *Somatic Cell and Molecular Genetics* **12**: 357-366.

Brown, W.R.A. (1989) Molecular cloning of human telomeres in yeast. *Nature* **338**: 774-776.

Brewer, B.J. and Fangman, W.L. (1987)The localization of replication origins on ARSs in *S.cerevisiae*. *Cell* **51**: 463-471.

Bullock, W.O., Fernandez, J.M. and Short, J.M. (1987) XL-1 Blue - a high-efficiency plasmid transforming recA *Escherichia coli* strain with beta-galactosidase selection. *Biotechniques* **5**: 376-379.

Burhans, W.C., Vassilev, L.T., Caddle, M.S., Heintz, N.H. and DePamphilis, M.L. (1990) Identification of an origin of bidirectional DNA replication in mammalian chromosomes. *Cell* **62**: 955-965.

Burke, D.T., Carle, G.F. and Olson, M.V. (1987) Cloning of large segments of exogenous DNA in yeast by means of artificial chromosome vectors. *Science* **236**: 806-812.

Caddle, M.S. and Calos, M.P. (1992) Analysis of the replication behaviour in human cells of the dihydrofolate reductase putative chromosomal origin of replication. *Nucl. Acids Res.* **20**: 5971-5978.

Carson, M.J. and Hartwell, L. (1985) *CDC*17: an essential gene that prevents telomere elongation in yeast. *Cell* **42**: 249-257.

Carson, M.J. (1987) *CDC17*, the structural gene for DNA polymerase 1 of yeast: mitotic hyperecombination and effects on telomere metabolism. Ph.D. thesis, University of Washington, Seattle, WA.

Chan, C.S.M. and Tye, B.-K. (1980) Autonomously replicating sequences in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* **77**: 6329-6333.

Chan, C.S.M. and Tye, B.-K. (1983a) Organization of DNA sequences and replication origins at yeast telomeres. *Cell* **33**: 563-573.

Chan, C.S.M. and Tye, B.-K. (1983b) A family of *Saccharomyces cerevisiae* repetitive autonomously replicating sequences that have very similar genomic environments. *Journal of Molecular Biology* **168**: 505-523.

Chikashige, Y., Kinoshita, N., Nakaseko, Y., Matsumoto, T., Murakami, S., Niwa, O. and Yanagida, M. (1989) Composite motifs and repeat symmetry in *S.pombe* centromeres: direct analysis by integration of *Not*I restriction sites. *Cell* **57**: 739-751.

Clarke, L. and Carbon, J. (1980) Isolation of a yeast centromere and construction of functional small circular chromosomes. *Nature* **287**: 504-509.

Clarke, L. and Carbon, J. (1983) Genomic substitutions of centromeres in *Saccharomyces cerevisiae*. *Nature* **305**: 23-28.

Clarke, L. (1990) Centromeres of budding and fission yeasts. *Trends in Genetics* **6**: 150-154.

Clarke, L. and Baum, M. (1990) Functional analysis of a centromere from fission yeast: a role for centromere-specific repeated DNA sequences. *Mol. Cell. Biol.* **10**: 1863-1872.

Comings, D.E. and Okada, T.A. (1971) Fine structure of the kinetochore in Indian muntjac. *Exp. Cell Res.* **67**: 97-110.

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, C.A., Heck, M.S. and Earnshaw, W.C. (1987) The inner centromere protein (INCENP) antigens: movement from inner centromere to midbody during mitosis. *Journal of Cell Biology* **105**: 2053-2067.

Cooke, H.J., Brown, W.R.A. and Rappold, G.A. (1985) Hypervariable telomeric sequences from the human sex chromosomes are pseudoautosomal. *Nature* **317**: 687-692.

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.

Coren, J.S., Epstein, E.M. and Vogt, V.M. (1991) Characterization of a telomere-binding protein from *Physarum polycephalum*. *Mol. Cell. Biol.* **11**: 2282-2290.

Counter, C.M., Avilion, A.A., LeFeuvre, C.E., Stewart, N.G., Greider, C.W. Harley, C.B. and Bacchetti, S. (1992) Telomere shortening associated with chromosome instability is arrested in immortal cells which express telomerase activity. *EMBO Journal* **11**: 1921-1929.

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.H., Lindsey, J., Fantes, J., McKay, S.J., McGill, N.I. and Cooke, H.J. (1990) The structure of a subterminal repeated sequence present on many human chromosomes. *Nucl. Acids Res.* **18**: 6649-6657.

De Lange, T., Shiue, L., Myers, R., 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.

De Lange, T. (1992) Human telomeres are attached to the nuclear matrix. *EMBO Journal* **11**: 717-724.

DePamphilis, M.L. (1988) Transcriptional elements as components of DNA replication. *Cell* **52**: 635-638.

Dhar, V. and Schildkraut, C.L. (1991) Role of EBNA-1 in arresting replication forks at the Epstein-Barr virus *oriP* family of tandem repeats. *Mol. Cell. Biol.* **11**: 6268-6278.

Diffley, J.F.X. and Stillman, B. (1990) The initiation of chromosomal DNA replication in eukaryotes. *Trends in Genetics* **6**: 427-432.

Diffley, J.F.X. and Cocker, J.H. (1992) Protein-DNA interactions at a yeast replication origin. *Nature* **357**: 169-172.

Ding, R., McDonald, K.L. and McIntosh, J.R. (1993) Three-dimensional reconstruction and analysis of mitotic spindles from the yeast, *Schizosaccharomyces pombe*. *Journal of Cell Biology* **120**: 141-151.

DuBridge, R.B., Tang, P., Hsia, H.C., Leong, P.-M., Miller, J.H. and Calos, M.P. (1987) Analysis of mutation in human cells by using an Epstein-Barr virus shuttle vector. *Mol. Cell. Biol.* **7**: 379-387.

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.

Earnshaw, W.C., Sullivan, K.F., Machlin, P.S., Cooke, C.A., Kaiser, D.A., Pollard, T.D., Rothfield, N.F. and Cleveland, D.W. (1987) Molecular cloning of cDNA for CENP-B, the major human centromere antigen. *Journal of Cell Biology* **104**: 817-829.

Earnshaw, W.C., Ratrie, H. and Stetten, G. (1989) Visualization of centromeric proteins CENP-B and CENP-C on a stable dicentric chromosome in cytological spreads. *Chromosoma* **98**: 1-12.

Fangman, W.L. and Brewer, B.J. (1991) Activation of replication origins within yeast chromosomes. *Ann. Rev. Cell Biol.* **7**: 375-402.

Farr, C.J., Goodfellow, P. and Cooke, H.J. (1991) Functional reintroduction of human telomeres into mammalian cells. *Proc. Natl. Acad. Sci. USA* **88**: 7006-7010.

Farr, C.J., Stevanovic, M., Thomson, E.J., Goodfellow, P.N. and Cooke, H.J. (1992) Telomere-associated chromosome fragmentation: applications in genome manipulation and analysis. *Nature Genetics* **2**: 275-282.

Feinberg, A.P. and Vogelstein, B. (1983) A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **132**: 6-13.

Feinberg, A.P. and Vogelstein, B. (1984) Addendum: a technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **137**: 266-267.

Fishel, B., Amstutz, H., Baum, M., Carbon, J. and Clarke, L. (1988) Structural organization and functional analysis of centromeric DNA in the fission yeast *Schizosaccharomyces pombe*. *Mol. Cell. Biol.* **8**: 754-763.

Gahn, T.A. and Schildkraut, C.L. (1989) The Epstein-Barr virus origin of plasmid replication, *oriP*, contains both the initiation and termination sites of DNA replication. *Cell* **58**: 527-535.

Gilbert, D. and Cohen, S., and Grummt, F. (1989) Autonomous replication in mouse cells: a correction. *Cell* **56**: 143-144. (Letters to the editor.)

Gluzman, Y. (1981) SV40-transformed simian cells support the replication of early SV40 mutants. *Cell* **23**: 175-182.

Goldstein, S., Fordis, C.M. and Howard, B.H. (1989) Enhanced transfection efficiency and improved cell survival after electroporation of G2/M synchronized cells and sodium butyrate. *Nuc. Acids. Res.* **17**: 3959-3971.

Gottschling, D.E. and Cech, T.R. (1984) Chromatin structure of the molecular ends of *Oxytrichia* macronuclear DNA. *Cell* **38**: 501-510.

Gottschling, D.E. and Zakian, V.A. (1986) Telomere proteins: specific recognition and protection of the natural termini of *Oxytrichia* macronuclear DNA. *Cell* **47**: 195-205.

Grady, D.L., Ratcliff, R.L., Robinson, D.L., McCanlies, E.C., Meyne, J. and Moyzis, R.K. (1992) Highly conserved repetitive DNA sequences are present at human centromeres. *Proc. Natl. Acad. Sci. USA* **89**: 1695-1699.

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 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**: 331-336.

Greider, C.W. (1990) Telomeres, telomerase and senescence. *BioEssays* **12**: 363-369.

Greider, C.W. (1991) Telomerase is processive. *Mol. Cell. Biol.* **11**: 4572-4580.

Gritz, L. and Davies, J. (1983) Plasmid encoded hygromycin B resistance: the sequence of hygromycin B phosphotransferase gene and its expression in *Escherichia coli* and *Saccharomyces cerevisiae*. *Gene* **25**: 179-188.

Haaf, T., Warburton, P.E. and Willard, H.F. (1992) Integration of human α-satellite DNA into simian chromosomes: centromere protein binding and disruption of normal chromosome segregation. *Cell* **70**: 681-696.

Haase, S.B. and Calos, M.P. (1991) Replication control of autonomosly replicating human sequences. *Nuc. Acids. Res.* **19**: 5053-5058.

Hanahan, D. (1983) Studies on transformation of *Escherichia coli* with plasmids. *Journal of Molecular Biology* **166**: 557-580.

Hahnenberger, K.M., Baum, M.P., Polizzi, C.M. Carbon, J. and Clarke, L. (1989) Construction of functional artificial minichromosomes in the fission yeast *Schizosaccharomyces pombe*. *Proc. Natl. Acad. Sci. USA* **86**: 577-581

Harley, C.B., Futcher, A.B. and Greider, C.W. (1990) Telomeres shorten during ageing of human fibroblasts. *Nature* **345**: 458-460.

Harrington, L.A. and Greider, C.W. (1991) Telomerase primer specificity and chromosome healing. *Nature* **353**: 451-454.

Hartwell, L. (1973) Three additional genes required for deoxyribonucleic acid synthesis in *Saccharomyces cerevisiae*. *J. Bacteriology* **115**: 966-974.

Hastie, N.D.. Dempster, M., Dunlop, M.G., Thompson, A.M., Green, D.K. and Allshire, R.C. (1990) Telomere reduction in human colon carcinoma and with ageing. *Nature* **346**: 866-868.

Hayden, J.H., Bowser, S.S. and Rieder, C.L. (1990) Kinetochores capture astral microtubules during chromosome attachment to the mitotic spindle: direct visualization in live newt lung cells. *Journal of Cell Biology* **111**: 1039-1045.

Heery, D.M., Gannon, F. and Powell, R. (1990) A simple method for subcloning DNA fragments from gel slices. *Trends in Genetics* **6**: 173.

Heinzel, S.S., Krysan, P.J. Tran, C.T. and Calos, M.P. (1991) Autonomous DNA replication in human cells is affected by the size and the source of the DNA. *Mol. Cell. Biol.* **11**: 2263-2272.

Henderson, E.R. and Blackburn, E.H. (1989) An overhanging 3' terminus is a conserved feature of telomeres. *Mol. Cell. Biol.* **9**: 345-348.

Henikoff, S. (1990) Position-effect variegation after 60 years. *Trends in Genetics* **6**: 422-426.

Hirt, B. (1967) Selective extraction of polyoma DNA from infected mouse cell cultures. *Journal of Molecular Biology* **26**: 365-369.

Hofmann, J.F.X. and Gasser, S. (1991) Identification and purification of a protein that binds the yeast ARS consensus sequence. *Cell* **64**: 951-960.

Huberman, J.A. and Riggs, A.D. (1968) On the mechanism of DNA replication in mammalian chromosomes. *Journal of Molecular Biology* **32**: 327-341.

Huxley, C. and Gnirke, A. (1991) Transfer of yeast artificial chromosomes to mammalian cells. *BioEssays* **13**: 545-550.

Hyman, A.A. and Mitchison, T.J. (1991) Two different microtubule-based motor activities with opposite polarities in kinetochores. *Nature* **351**: 206-211.

Hyman, A.A., Middleton, K., Centola, M., Mitchison, T.J. and Carbon, J. (1992) Microtubule-motor activity of a yeast centromere protein complex. *Nature* **359**: 533-536.

Iguchi-Ariga, S.M.M., Itani, T., Kiji, Y. and Ariga, H. (1987) Possible function of the c-*myc* product: promotion of cellular DNA replication. *EMBO Journal* **6**: 2365-2371.

Iguchi-Ariga, S.M.M., Okazaki, T., Itani, T., Ogata, M., Sato, Y. and Ariga, H. (1988) An initiation site of DNA replication with transcriptional enhancer activity present upstream of the c-*myc* gene. *EMBO Journal* **7**: 3135-3142.

Jankelevich, S., Kolman, J.L., Bodnar, J.W. and Miller, G. (1992) A nuclear matrix attachment region organizes the Epstein-Barr viral plasmid into a single DNA domain. *EMBO Journal* **11**: 1165-1176.

Javerzat, J.-P., Bhattacherjee, V. and Barreau, C. (1993) Isolation of telomeric DNA from the filamentous fungus *Podospora anserina* and construction of a self-replicating linear plasmid showing high transformation efficiency. *Nuc. Acids. Res.* **21**: 497-504.

Kipling, D.G. and Cooke, H.J. (1990) Hypervariable ultra-long telomeres in mice. *Nature* **347**: 400-402.

Klein, H.L. (1988) different types of recombination events are controlled by the *RAD1* and *RAD52* genes of *Saccharomyces cerevisiae*. *Genetics* **120**: 367-377.

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.

Krysan, P.J., Haase, S.B. and Calos, M.P. (1989) Isolation of human sequences that replicate autonomously in human cells. *Mol. Cell. Biol.* **9**: 1026-1033.

Krysan, P.J. and Calos, M.P. (1991) Replication initiates at multiple locations on an autonomously replicating plasmid in human cells. *Mol. Cell. Biol.* **11**: 1464-1472.

Kuno, K, Murakami,S. and Kuno, S. (1991) Single-stranded-binding factor(s) which interact with *ARS1* of *Saccharomyces cerevisiae*. *Gene* **95**: 73-77.

Larsen, D.D., Spangler, E.A. and Blackburn, E.H. (1987) Dynamics of telomere variation in *Tetrahymena thermophila*. *Cell* **50**: 477-483.

Leach, D.R.F. and Stahl, F.W. (1983) Viability of λ phages carrying a perfect palindrome in the absence of recombination nucleases. *Nature* **305**: 448-451.

Lechner, J. and Carbon, J. (1991) A 240kd multisubunit protein complex, CBF3, is a major component of the budding yeast centromere. *Cell* **64**: 717-725.

Levis, R.W. (1989) Viable deletions of a telomere from a *Drosophila* chromosome. *Cell* **58**: 791-801.

Lindahl, T., Adams, A., Bjursall, G., Bornkamm, G.W., Kascheka-Dierach, C. and Jehn, U. (1976) Covalently closed circular duplex DNA of Epstein-Barr virus in a human lymphoid cell line. *Journal of Molecular Biology* **102**: 511-530.

Lindsey, J., McGill, N.I., Lindsey, L.A., Green, D.K. and Cooke, H.J. (1991) In vivo loss of telomeric repeats with age in humans. *Mutation research* **256**: 45-48.

Linskens, M.H.K. and Huberman, J.A. (1990) The two faces of higher eukaryotic DNA replication origins. *Cell* **62**: 845-847.

Liu, Z. and Tye, B.-K. (1991) A yeast protein that binds to vertebrate telomeres and conserved yeast telomeric junctions. *Genes & Development* **5**: 49-59.

Lundblad, V. and Szostak, J.W. (1989) A mutant with a defect in telomere elongation leads to senescence in yeast. *Cell* **57**: 633-643.

Lundblad, V. and Blackburn, E.H. (1990) RNA-dependant polymerase motifs in Est1: Tentative identification of a protein component of an essential yeast telomerase. *Cell* **60**: 529-530.

Lustig, A.J. and Petes, T.D. (1986) Identification of yeast mutants with altered telomere structure. *Proc. Natl. Acad. Sci. USA* **83**: 1398-1402.

Lustig, A.J., Kurtz, S. and Shore, D. (1990) Involvement of the silencer and UAS binding protein RAP1 in regulation of telomere length. *Science* **250**: 549-553.

McClintock, B. (1941) The stability of broken ends of chromosomes in *Zea mays*. *Genetics*. **26**: 234-282.

McClintock, B. (1942) The fusion of broken ends of chromosomes following nuclear fusion. *Proc. Natl. Acad. Sci. USA* **28**: 458-463.

McKnight, S. and Tijan, R. 91986) Transcriptional selectivity of viral genes in mammalian cells. *Cell* **46**: 795-805.

Mann, C. and Davis, R.W. (1983) Instability of dicentric plasmids in yeast. *Proc. Natl. Acad. Sci. USA* **80**: 228-232.

Masumoto, H., Masukata, H., Muro, Y., Nozaki, N. and Okazaki, T. (1989) A human centromere antigen (CENP-B) interacts with a short specific sequence in alphoid DNA, a human centromeric satellite. *Journal of Cell Biology* **109**: 1963-1973.

Masse, M.J.O., Karlin, S., Schachtel, G.A. and Mocarski, E.S. (1992) Human cytomegalovirus origin of DNA replication (oriLyt) resides within a highly complex repetitive region. *Proc. Natl. Acad. Sci. USA* **89**: 5246-5250.

Mellon, P., Parker, V., Gluzman, Y. and Maniatis, T. (1981) Identification of DNA sequences required for transcription of the human α1-globin gene in a new SV40 vector-host system. *Cell* **27**: 279-288.

Meyne, J., Ratliff, R.L. and Moyzis, R.K. (1989) Conservation of the human telomere sequence, $TTAGGG_n$, among vertebrates. *Proc. Natl. Acad. Sci. USA* **86**: 7049-7053.

Mitchison, T.J. and Kirschner, M.W. (1985) Properties of the kinetochore *in vitro*. II. Microtubule capture and ATP-dependent translocation. *Journal of Cell Biology* **101**: 766-777.

Morin, G.B. and Cech, T.R. (1986) The telomeres of the linear mitochondrial DNA of *Tetrahymena thermophila* consist of 53bp tandem repeats. *Cell* **46**: 873-883.

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 chromosome truncation site associated with α-thalassaemia by human telomerase. *Nature* **353**: 454-456.

Moroi, Y., Peebles, C., Fritzler, M.J., Seigerwald, J. and Tan, E.M. (1980) Autoantibody to the centromere (kinetochore) in scleradoma sera. *Proc. Natl. Acad. Sci. USA* **17**: 1627-1631.

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.

Muller, H.J. (1938) The remaking of chromosomes. *Collect. Net.* **13**: 181-198.

Muller, H.J. and Herskowitz, I.H. (1954) Concerning the healing of chromosome ends produced by breakage in *Drosophila melanogaster*. *Am. Nat.* **88**: 177-208.

Murray, A.W. and Szostak, J.W. (1983) Construction of artificial chromosomes in yeast. *Nature* **305**: 189-193.

Murray, A.W., Claus, T.E. and Szostak, J.W. (1988) characterization of two telomeric DNA processing reactions in *Saccharomyces cerevisiae. Mol. Cell. Biol.* **8**: 4642-4650.

Neil, D.L., Villasante, A., Fisher, R.B., Vetrie, D., Cox, B. and Tyler-Smith, C. (1990) Structural instability of human tandemly repeated DNA sequences cloned in yeast artificial chromosome vectors. *Nucl. Acids Res.* **18**: 1421-1428.

Newlon, C.S. (1988) Yeast chromosome replication and segregation. *Microbiol. Rev.* **52**: 568-601.

Olovnikov, A.M. (1973) A theory of marginotomy. *Journal of Theoretical Biology* **41**: 181-190.

Palmer, D.K., O'Day, K. and Margolis, R.L. (1990) The centromere specific histone CENP-A is selectively retained in discreet foci in mammalian sperm nuclei. *Chromosoma* **100**: 32-36.

Pluta, A.F. and Zakian, V.A. (1989) Recombination occurs during telomere formation in yeast. *Nature* **337**: 429-433.

Pluta, A.F., Saitoh, N., Goldberg, I. and Earnshaw, W.C. (1992) Identification of a subdomain of CENP-B that is necessary and sufficient for localization to the human centromere. *Journal of Cell Biology* **116**: 1081-1093.

Potter, H., Weir, L. and Leder, P. (1984) Enhancer-dependent expression of human κ immunoglobulin genes introduced into mouse pre-B lymphocytes by electroporation. *Proc. Natl. Acad. Sci. USA* **81**: 7161-7165.

Price, C.M. and Cech, T.R. (1989) Properties of the telomeric DNA-binding protein from *Oxytrichia nova. Biochemistry* **28**: 769-774.

Price, C.M. (1990) Telomere structure in Euplotes crassus: characterization of DNA-protein interactions and isolation of a telomere-binding protein. *Mol. Cell. Biol.* **10**: 3421-3431.

Prosser, J., Frommer, M., Paul, C. and Vincent, P.C. (1986) Sequence relationships of three human satellite DNAs. *Journal of Molecular Biology* **187**: 145-155.

Pulvertaft, R.J.V. (1965) A study of malignant tumours in Nigeria by short-term tissue culture. *J. Clin. Pathol.* **18**: 261-271.

Rawlins, D.R., Milnan, G., Hayward, S.D. and Wayward, G.S. (1985) Sequence-specific DNA binding of the Epstein-Barr virus nuclear antigen (EBNA-1) to clustered sites in the plasmid maintenance region. *Cell* **42**: 859-868.

Reisman, D., Yates, J. and Sugden, B. (1985) A putative origin of replication of plasmids derived from Epstein-Barr virus is composed of two *cis*-acting components. *Mol. Cell. Biol.* **5**: 1822-1832.

Richardson, C.C. (1971) Polynucleotide kinase from *Escherichia coli* infected with bacteriophage T4. *Procedures in nucleic acid research* **2**: 815.

Riethman, H.C., Moyzis, R.K., Meyne, J., Burke, D.T. and Olson, 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.

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.

Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. and Erlich, H.A. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**: 487-489.

Saitoh, H., Tomkiel, J., Cooke, C.A., Ratrie, H., Maurer, M., Rothfield, N.F. and Earnshaw, W.C. (1992) CENP-C, an autoantigen in scleroderma, is a component of the human inner kinetochore plate. *Cell* **70**: 115-125.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular cloning: a laboratory manual. 2nd edition. Cold Spring Harbour Laboratory Press.

Sanger, F. and Coulson, A.R. (1978) The use of thin acrylamide gels for DNA sequencing. *FEBS Letters* **87**: 107-110.

Sanger, F., Micklens, S. and Coulson, A.R. (1977) DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**: 5463-5467.

Schlessinger, D. (1990) Yeast artificial chromosomes: tools for mapping and analysis of complex genomes. *Trends in Genetics* **6**: 248-258.

Schmidt, A.M.A., Herterich, S.U. and Krauss, G. (1991) A single-stranded DNA binding protein from *S.cerevisiae* specifically recognises the T-rich strand of the core sequence of ARS elements and discriminates against mutant sequences. *EMBO Journal* **10**: 981-985.

Shampay, J., Szostak, J.W. and Blackburn, E.H.(1984) DNA sequences of telomeres maintained in yeast. *Nature* **310**:154-157.

Shizuya, H., Birren, B., Kim, U.-J., Mancino, V., Slepak, T., Tachira, Y. and Simon, M. (1992) cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc. Natl. Acad. Sci. USA* **89**: 8794-8797.

Shore, D. and Nasmyth, K.A. (1987) Purification and cloning of a DNA binding from yeast that binds to both silencer and activator elements. *Cell* **51**: 721-732.

Southern, E.M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology* **98**: 503-517.

Sugawara, N. and Szostak, J. (1986) Telomeres of *Schizosaccharomyces pombe*. *Yeast* 2: Suppl. 373 (abstract).

Sugden, B., Marsh, K. and Yates, J. (1985) A vector that replicates as a plasmid and can be efficiently selected in B-lymphocytes transformed by Epstein-Barr virus. *Mol. Cell. Biol.* **5**: 410-413.

Starling, J.A., Maule, J. Hastie, N.D. and Allshire R.C. (1990) Extensive telomere repeat arrays in mouse are hypervariable. *Nucleic Acids Res.* **18**: 6881-6888.

Stinchcomb, D.T., Struhl, K. and Davis, R.W. (1979) Isolation and characterization of a yeast chromosome replicator. *Nature* **282**: 39-43.

Szostak, J.W. and Blackburn, E.H. (1982) Cloning yeast telomeres on linear plasmids. *Cell* **29**: 245-255.

Takahashi, K., Murakami, S., Chikashige, Y., Niwa, O. and Yanagida, M. (1991) A large number of tRNA genes are symmetrically located in fission yeast centromeres. *Journal of Molecular Biology* **218**: 13-17.

Teshigarawa, K. and Katsura, Y. (1992) A simple and efficient mammalian gene expression system using an EBV-based vector transfected by electroporation in G2/M phase. *Nuc. Acids. Res.* **20**: 2607.

Tijan, R. (1978) The binding site on SV40 DNA for a T antigen-related protein. *Cell* **13**: 165-179.

Tyler-Smith, C. and Brown, W.R.A. (1987) Structure of the major block of alphoid satellite DNA on the human Y chromosome. *Journal of Molecular Biology* **195**: 457-470.

Umek, R.M., Linskens, M.H.K., Kowalski, D. and Huberman, J.A. (1989) New beginnings in studies of eukaryotic DNA replication. *Biochem. Biophys. Acta* **1007**: 1-14.

Vassilev, L. and Johnson, E.M. (1990) An initiation zone of chromosomal DNA replication located upstream of the c-*myc* gene in proliferating HeLa cells. *Mol. Cell. Biol.* **10**: 4899-4904.

Vaughn, J.P., Dijkwel, P.A. and Hamlin, J.L. (1990) Replication initiates in a broad zone in the amplified CHO dihydrofolate reductase domain. *Cell* **61**: 1075-1087.

Walmsley, R.M., 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.

Wang, S.-S. and Zackian, V.A. (1990) Telomere-telomere recombination provides an express pathway for telomere acquisition. *Nature* **345**: 456-458.

Warburton, P.E. and Willard, H.F. (1990) Genomic analysis of sequence variation in tandemly repeated DNA: evidence for localized homogeneous sequence domains within arrays of α-satellite DNA. *Journal of Molecular Biology* **216**: 3-16.

Watson, J.D. (1972) Origin of concatameric T4 DNA. *Nature New Biology* **239**: 197-201.

Weiss, B., Jacquemin-Sablon, A., Live, T.R., Fareed, G.C. and Richardson, C.C. (1968) Enzymatic breaking and joining of deoxyribonucleic acid. *J. Biol. Chem.* **243**: 4543-4555.

Weiss, M.C. and Green, H. (1967) Human-mouse hybrid cell lines containing partial complements of human chromosomes and functioning human genes. *Proc. Natl. Acad. Sci. USA* **58**: 1104-1111.

Wellinger, R.J. and Zakian, V.A. (1989) Lack of positional requirements for autonomously replicating sequence elements on artificial yeast chromosomes. *Proc. Natl. Acad. Sci. USA* **86**: 973-977.

Willard, H.F. and Waye, J.S. (1987) Hierarchical order in chromosome-specific human alpha-satellite DNA. *Trends in Genetics* **3**: 192-198.

Willard, H.F. (1990) Centromeres of mammalian chromosomes. *Trends in Genetics* **6**: 410-416.

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.

Wright, J.H., Gottschling, D.E. and Zakian, V.A. (1992) *Saccharomyces* telomeres assume a non-nucleosomal chromatin structure. *Genes & Development* **6**: 197-210.

Wright, W.E. and Shay, J.W. (1992) Telomere positional effects and the regulation of cellular senescence. *Trends in Genetics* **8**: 193-197.

Yao, M.-C. (1981) Ribosomal RNA gene amplification in *Tetrahymena* may be associated with chromosome breakage and DNA elimination. *Cell* **24**: 765-774.

Yasuda, S. and Hirota, Y. (1977) Cloning and mapping of the replication origin of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **74**: 5458-5462.

Yates, J., Warren, N., Reisman, D. and Sugden, B. (1984) A *cis*-acting element from the Epstein-Barr viral genome that permits stable replication of recombinant plasmids in latently infected cells. *Proc. Natl. Acad. Sci. USA* **81**: 3806-3810.

Yates, J., Warren, N. and Sugden, B. (1985) Stable replication of plasmids derived from Epstein-Barr virus in various mammalian cells. *Nature* **313**: 812-815.

Yen, T.J., Compton, D.A., Wise, D., Zinkowski, R.P., Brinkley, B.R., Earnshaw, W.C. and Cleveland, D.W. (1991) CENP-E, a novel human centromere-associated protein required for progression from metaphase to anaphase. *EMBO Journal* **10**: 1245-1254.

Yen, T.J., Li, G., Schaar, B.T., Szilak, I. and Cleveland, D. (1992) CENP-E is a putative kinetochore motor that accumulates just before mitosis. *Nature* **359**: 536-539.

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 telomerase RNAs. *Nature* **344**: 126-132.

Yu, G.-L. and Blackburn, E.H. (1991) Developmentally programmed healing of chromosomes by telomerase in *Tetrahymena*. *Cell* **67**: 823-832.

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.

Zakian, V.A. (1989) Structure and function of telomeres. *Ann. Rev. Genetics*. **23**: 579-604.

Zhong, Z., Shiue, L., Kaplan, S. and De Lange, T. (1992) A mammalian factor that binds telomeric TTAGGG repeats *in vitro*. *Mol. Cell. Biol.* **12**: 4834-4843.

Zinkowski, R.P., Meyne, J. and Brinkley, B.R. (1991) The centromere-kinetochore complex: a repeat subunit model. *Journal of Cell Biology*. **113**: 1091-1110.

© 1993 Oxford University Press Nucleic Acids Research, 1993, Vol. 21, No. 3 497–504

Isolation of telomeric DNA from the filamentous fungus *Podospora anserina* and construction of a self-replicating linear plasmid showing high transformation frequency

Jean-Paul Javerzat, Vasker Bhattacherjee[1] and Christian Barreau*
Laboratoire de Génétique, UA CNRS 542, Université de Bordeaux II, Avenue des Facultés, 33405 Talence Cedex, France and [1]MRC Human Genetics Unit, Western General Hospital, Crewe Road, Edinburgh EH4 2XU, UK

Received October 26, 1992; Revised and Accepted December 31, 1992

ABSTRACT

It has been previously shown that linear plasmids bearing *Tetrahymena* telomeric sequences are able to replicate autonomously in the filamentous fungus *Podospora anserina* (1). However, autonomous replication occurs in only 50–70% of the transformants, suggesting a defect in the recognition of the *Tetrahymena* telomeric template by the putative *P. anserina* telomerase so that only a fraction of entering DNA is stabilized into linear extrachromosomal molecules. We have cloned DNA sequences added to the *Tetrahymena* $(T_2G_4)_n$ ends of the linear plasmid. Nucleotide sequencing showed that these sequences are exclusively composed of T_2AG_3 repeat units. Hybridization experiments of *Bal*31 treated DNA showed that T_2AG_3 repeats are confined within 200 bp in chromosomal *P. anserina* telomeres. A new plasmid has been constructed so that after linearization, the terminal sequences contain T_2AG_3 repeats. This linear molecule transforms *P. anserina* with a high frequency (up to 1.75×10^4 transformants/μg), autonomous replication occurs in 100% of the transformants and the plasmid copy number is about 2–3 per nucleus. These results underscore the importance of the telomeric repeat nucleotide sequence for efficient recognition as functional telomeric DNA *in vivo* and provide the first step toward the development of an artificial chromosome cloning system for filamentous fungi.

INTRODUCTION

Telomeres are the specialized structures that form the termini of linear eukaryotic chromosomes. They are responsible for the complete replication of the extreme ends of chromosomal molecules and contribute to chromosome stability in protecting them from exonucleolytic degradation and end to end fusion events (2). Telomeric DNAs have been isolated from numerous evolutionary distant species ranging from yeast to human. Telomere function seems to be fully defined within a few hundred base pairs in ciliates or yeasts and thousands base pairs in vertebrates. The overall structure of all known eukaryotic telomeres consists of tandemly repeated short sequences of G- and C- rich complementary strands with an orientation specificity so that the G rich telomeric DNA runs 5′ to 3′ towards the terminus of a chromosome (3). Telomeric DNA sequences can be composed of precise repeats such as T_2G_4 in *Tetrahymena* (4) and T_4G_4 in *Oxytricha* (5). In other species such as *S.cerevisiae* (6), *S.pombe* (7) or *Dictyostelium* (8), the repeated units were found to be more heterogeneous. In ciliates and humans, at least, the 3′ end of the G rich strand is elongated by a ribonucleoprotein called telomerase (9, 10, 11, 12, 13). The conservation of the overall telomere structure among widely divergent eukaryotes is reflected in the observation that telomeric sequences from ciliates (6, 14), fission yeast (15) or humans (16, 17) can function as telomeres on linear plasmids in *S.cerevisiae*. However, heterologous telomeric DNAs exhibit a broad spectrum of activity in their ability to be recognized as functional telomeres *in vivo*. For instance, it has been reported (15) that after propagation in *S.cerevisiae*, YAC vectors are rarely replicated as linear minichromosomes when transferred into *S. pombe* cells. In roughly all the transformants, minichromosomes have undergone rearrangements such as circularization events often accompanied by deletions of internal sequences. In the same way, it has been reported that transformation of *S.pombe* cells with a linear plasmid bearing *Tetrahymena* telomeric sequences gives rise to only a very small fraction of transformants containing a plasmid with the correct structure (7). These observations suggest that neither *S.cerevisiae* nor *Tetrahymena* telomeric repeats are fully functional in *S.pombe*.

Similar features can be observed with the fate of linear plasmids bearing *Tetrahymena* telomeres in *P. anserina* : autonomous linear replication occurs in only 50–70% of the transformants, the remaining resulting from integration events into chromosomal DNA (1). As circular plasmids seem unable to replicate autonomously in *P. anserina* (18), transformation can only be achieved either by self replication of linear molecules or by integration in chromosomal DNA. The most likely explanation

---

* To whom correspondence should be addressed

for the observed high number of integrative transformants is that *Tetrahymena* telomeric sequences are not perfectly recognized as telomeric templates in *P. anserina*. Therefore, a competition takes place between mechanisms involved in telomere recognition/elongation and those involved in recombination events giving rise to the final ratio of the two classes of transformants. This interpretation was supported by the observation that after propagation in *P. anserina*, the linear plasmid does not integrate anymore into chromosomal DNA when reintroduced in *P. anserina* protoplasts (18). Therefore, the addition of *P. anserina* telomere-like sequences onto *Tetrahymena* templates seems to fully inhibit recombination between the transforming plasmid and chromosomal DNA. Taken together, these observations have led us to investigate what could be the effect of homologous telomeric sequences on *P. anserina* linear vectors. We have cloned *P. anserina* telomeric repeats and shown that a sequence containing homologous telomeric repeats fully prevents integration of the transforming vector into chromosomal DNA.

MATERIALS AND METHODS

Strains

The characteristics of *P. anserina* were reviewed by Esser (19). The isolation and characterization of a *ura5-6* mutant deficient for OMPppase activity have been described elsewhere (20). The bacterial strain *Escherichia coli* DH5α (F^- *endA1 hsdR17 supE44 thi-1 recA1 gyrA96 relA1* (*ϕ80d lacZΔM15*); Bethesda Research Laboratories) was used for bacterial transformation and plasmid propagation. Bacterial plasmids used in this study were constructed by standard recombinant DNA techniques.

Plasmids

The pPYACRC5 vector was contructed as follow: The 1.55 kbp *Eco*RI fragment carrying the *P. anserina ura5* gene (21) was blunt ended with the klenow fragment of DNA polymerase I and cloned into the filled-in *Sal*I site of pYAC4 (22). The resulting plasmid was named pPYAC4. The final construction, pPYACRC5, was created by replacing the *Bam*HI-*Sma*I fragment containing *TRP1*, *ARS1* and *CEN4* in pPYAC4 with the *BamHI*-*Sma*I fragment carrying *TRP1*, *ARS1*, *CEN4* and the cloning sites from pYAC-RC (23). The DNA fragment containing *Tetrahymena* telomeric repeats in pPYACRC5 was deleted by digestion with *Xho*I followed by religation to create the plasmid pPRC5. The pPATRC5 plasmid was constructed as follows. The plasmid pHUTEL-2-end (24) was digested with *Eco*RI and *Hae*III. The *Eco*RI-*Hae*III fragment of 0.5 kbp was cloned into the *Eco*RI-*Sma*I sites of pGEM7 creating the plasmid pEND2. The 0.5 kbp *Hind*III-*Xho*I fragment containing the human telomeric repeats (HTEL) was ligated to the 1.3 kbp *Sal*I fragment containing the kanamycin resistance gene (kan^r) from pUC4KN (a derivative of pUC4K (25, 26) in which *Not*I sites have been inserted into the polylinker flanking the gene). The ligated fragments were inserted in one step into the *Hind*III site of the plasmid p220.2 (27) to create p220.2TC. This plasmid contains a 'telomeric cassette' of two 0.5 kbp telomeric DNA fragments in opposing orientations flanking the kan^r gene which acts as a stuffer fragment and selectable marker. The entire 'telomeric cassette' has been removed by *Asu*II digestion and inserted into the *Cla*I site of pPRC5 to create pPATRC5.

Transformation of *P. anserina* and determination of the mitotic stability of the transformants

Protoplasts from the *ura5-6* strain were prepared and transformed as previously described (28, 29). The transformants selected on minimal medium were grown for 4 days on solid medium with uridine (100 mg/liter). From each transformant, small pieces of mycelium were picked at the thallus periphery and grown on minimal medium with or without uridine. The number of nuclei in each mycelial fragment was not known precisely but was higher than 1,000. If no regeneration of mycelium was observed on the minimal medium without uridine, it was concluded that the nuclei had lost the ura^+ transforming gene during growth on non-selective medium.

DNA isolation and Southern blotting

Genomic DNA of *P.anserina* was isolated from the mycelium. The mycelium was lyophilized, ground, and lysed in 100 mM Tris hydrochloride (pH 9.0)−10 mM EDTA−1% sarcosyl−200 μg/ml proteinase K for 2 h at 65°C. After centrifugation (10,000×g, 10 min), the supernatant was extracted twice with phenol, once with chloroform, and nucleic acids were precipitated with ethanol. The DNA fiber was collected, washed several times with 70% ethanol, and dissolved in 10 mM Tris hydrochloride (pH 8.0)−1 mM EDTA. Digestions with restriction enzymes were performed overnight under the conditions recommended by the supplier (Bethesda Research Laboratories). DNA samples were electrophoresed on agarose gels and transferred to nylon membranes (Hybond N+, Amersham). Filters were prehybridized at 65°C for 1 h in 5×SSC (1×SSC is 0.15 M NaCl plus 15 mM sodium citrate)−40 mM Tris hydrochloride (pH 8.0)−5×Denhardt's solution−0.5% SDS−100 μg of denatured salmon sperm DNA per ml. Double-strand DNA fragments were labeled with (α-^{32}P) dCTP (Amersham) using a Random Primed Labeling kit (Boehringer Mannheim). Hybridizations were carried out at 42°C for 16 h in 50% formamide−5×SSC−40 mM Tris hydrochloride (pH 8.0)−5×Denhardt's solution−0.5% SDS−100 μg of denatured salmon sperm DNA per ml. The filters were then washed twice at 65°C in 0.1×SSC−0.1% SDS for 30 min and subjected to autoradiography at −80°C with an intensifying screen. The synthetic probe $(T_2AG_3)_4$ was produced on an Applied Biosystem DNA synthesizer and labeled at the 5′ end with ($\gamma^{32}P$) ATP by using T4 polynucleotide kinase under the conditions recommended by the enzyme supplier (Bethesda Research Laboratories). Hybridizations were performed at 52°C for 16 h in 5×SSC−40 mM Tris hydrochloride (pH 8.0)−5×Denhardt's solution−0.5% SDS−100 μg of denatured salmon sperm DNA per ml. Filters were washed twice with 2×SSC−0.1% SDS at 65°C and subjected to autoradiography.

*Bal*31 digestion

Genomic DNA from the *ura5-6* strain (480 μg) was mixed with 60 ng of *Hind*III digested λ DNA in 2.4 ml of 1×*Bal*31 buffer (20 mM Tris hydrochloride pH 7.2−600 mM NaCl−12.5 mM $MgCl_2$ −12.5 mM $CaCl_2$ −1 mM EDTA) and equilibrated at 30°C for 10 min. After addition of 24 U of *Bal*31 (Boehringer Mannheim), 400 μl samples were removed at 0.5, 2, 7.5, 15, and 60 min. Reactions were terminated by adding EDTA to a final concentration of 50 mM and quick freezing. DNA samples were phenol extracted and precipitated with ethanol.

Cloning and sequencing of *P. anserina* telomeric repeats

The cloning procedure of Wang and Zakian (30) was followed. Total DNA prepared from a transformant containing the pPYACRC5 linear plasmid was digested with *Cla*I and extracted with phenol-chloroform. The digested DNA (5 μg) was treated with T4 DNA polymerase (final concentration, 67 mU/ml) at 15°C for 15 min in the presence of a high concentration of the four deoxynucleotides (final concentration of each deoxynucleotide, 100 μM), extracted with phenol-chloroform, precipitated with ethanol, and self-ligated with T4 ligase at a low DNA concentration (10 ng/μl) to circularize plasmid DNA. The ligation products were concentrated by ethanol precipitation and used to transform *E.coli* DH5α competent cells (Bethesda Research Laboratories) to ampicillin resistance. Sequencing was performed on double strand plasmid DNA by Sanger's method (31) using the synthetic primer shown in Fig. 1.

RESULTS

Cloning of *P. anserina* telomeric DNA

In order to obtain a YAC library of *P. anserina* chromosomal DNA we have constructed a modified YAC vector by the addition of the *P. anserina ura5* gene (21) to the pYACRC vector (23). The resulting plasmid, pPYACRC5 (Fig. 1), carries all sequences from pYACRC and is able to transform yeast AB1380 spheroplasts as other YAC vectors. Transformation of the *P. anserina ura5-6* strain either with the previously described pPATura1 linear vector (1) or with *BamH*I cut pPYACRC5 gives rise to similar results: transformation frequency is in the range of 200−1000 transformants/μg of DNA, 50−70% of the transformants are highly unstable through vegetative growth and the linear plasmid is transmitted with a very low frequency through meiosis. This suggests that the yeast sequences *ARS1* and *CEN4* are not functional in *P. anserina*. Fig. 2 shows molecular analysis of unstable *P. anserina* transformants carrying the pPYACRC5 linear plasmid. Total DNA was cut with *Pst*I, run on an agarose gel and probed with radiolabeled pBR322. This probe detect weakly the 1.8 kbp fragment which contains only 30% of pBR322 sequences. However, the hybridization pattern is in good agreement with the *Pst*I restriction map of the linear plasmid shown in Fig. 1B. But the comparison of restriction patterns between plasmid DNA extracted from *E.coli* (Fig. 2, lane 2 and 6) and plasmid DNA in *P. anserina* transformants (Fig. 2, lanes 3 to 5) reveals that the linear plasmid has been modified during propagation in *P. anserina*. Internal *Pst*I fragments detected by the probe have exactly the same mobility as in the control lane whereas terminal fragments migrate slightly slower. As it has been previously shown in *S.cerevisiae* (6) and *S.pombe* (7) one might expect that *Tetrahymena* telomeric sequences have been healed *in vivo* by addition of host specific telomeric sequences. It can be estimated from this experiment that at least 100 to 150 bp of DNA have been added to the ends of the linear plasmid. This is in agrement with the 200 bp increase of telomeric ends we previously reported for the linear vector (1).

To isolate these sequences we used the enzyme T_4 DNA polymerase which has been shown to enable the isolation of telomeres with virtually no loss of telomeric DNA (30). Genomic DNA from an unstable transformant was digested with *Cla*I which cuts only once in the plasmid (Fig. 1), treated with T_4 DNA polymerase and self ligated to circularize plasmid DNA. Transformation of highly competent *E.coli* cells with the ligation mixture gave rise to one Ap^R clone. Restriction analysis of the rescued plasmid revealed that it had the expected structure. The fragment containing putative *P. anserina* telomeric DNA was sequenced by using the synthetic primer shown in Fig. 1B. Close to the primer, the sequence starts with a 25 units track of the repeat T_2AG_3, followed by the *Tetrahymena* T_2G_4 telomeric sequence. Thus, the *in vivo* processing of *Tetrahymena* telomeric templates consists of the addition of a homogeneous track of the repeated sequence T_2AG_3. Restriction analysis of the rescued plasmid showed that the *Xho*I-*Sal*I fragment containing telomeric repeats is 800 bp in length. Assuming that the *Xho*I-*BamH*I fragment containing *Tetrahymena* telomeric repeats is 700 bp in pPYACRC5, about 50 bp of *Tetrahymena* DNA have been removed by the healing process.

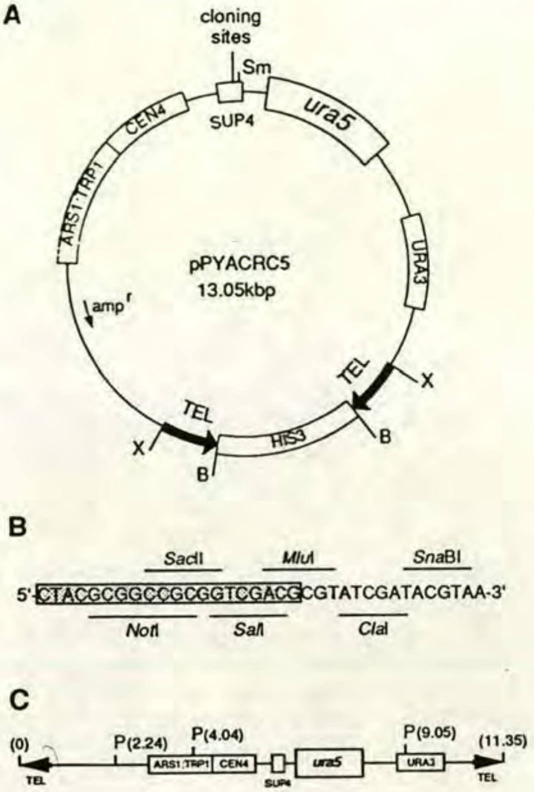


Figure 1. (A) Map of the pPYACRC5 vector. *SUP4*, *TRP1*, *HIS3* are yeast genes. *ARS1* and *CEN4* are sequences from yeast's chromosome IV: *ARS1* is an autonomous replication sequence while *CEN4* provides centromere function. The TEL sequences are derived from the termini of *Tetrahymena* macronuclear rDNA molecules. The *ura5* gene is from *P. anserina*. pBR322 derived sequences are shown as thin lines. **(B)** nucleotide sequence of the pPYACRC5 polylinker. The shaded box corresponds to the nucleotide sequence of the primer used for sequencing. **(C)** *Pst*I restriction map of *Bam*HI linearized pPYACRC5. The sizes (kbp) are indicated from the left end of the linear molecule. Restrictions sites are: B, *Bam*HI; P, *Pst*I; Sm, *Sma*I; X, *Xho*I.

T_2AG_3 repeats are the component of chromosomal *P. anserina* telomeres

To demonstrate that chromosomal *P. anserina* telomeres are also composed of T_2AG_3 repeats we set up hybridization experiments using the synthetic probe $(T_2AG_3)_4$. Genomic DNA from the

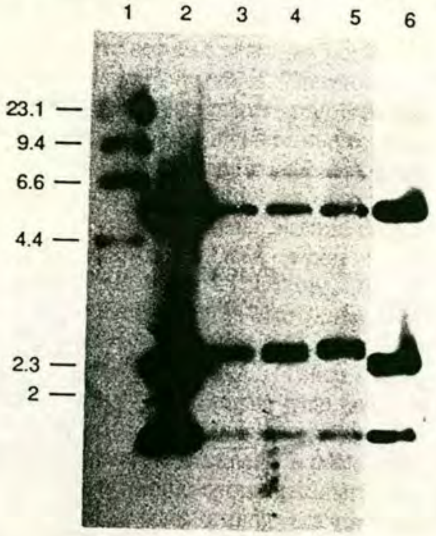


Figure 2. Southern blot analysis of the linear plasmid after propagation in *P. anserina* transformants. DNA samples were electrophoresed on a 0.8% agarose gel, transferred to nylon membrane, and probed with a mixture of ^{32}P-labeled pBR322 DNA (2.5 ng/ml; 10^9 cpm/μg) and ^{32}P-labeled λ DNA (0.5 ng/ml; 10^8 cpm/μg). Lane 1, *Hind*III digested λ DNA (0.2 μg); lane 2 and 6, *Bam*HI-*Pst*I digested pPYACRC5 DNA (0.5 ng), (lane 6 corresponds to a 10 times shorter exposure to autoradiography); lanes 3 to 5, *Pst*I digested DNA (1 μg) from three unstable pPYACRC5 transformants. Numbers to the left refer to DNA size standards (kilobases).

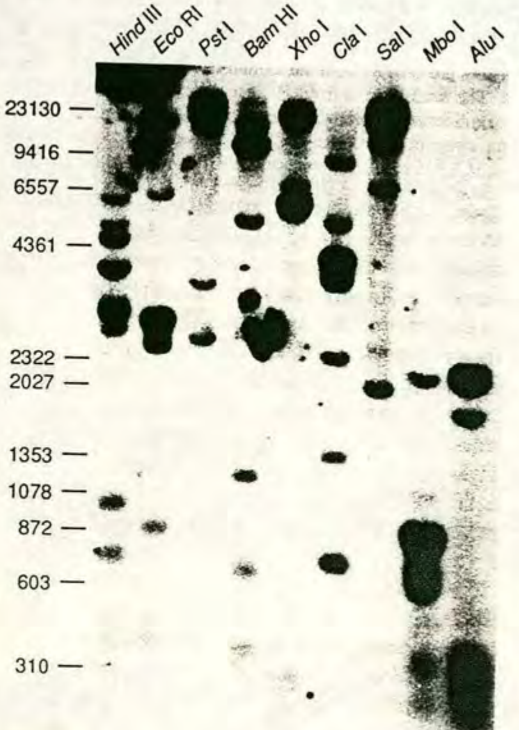


Figure 3. Southern blot analysis of *P. anserina* genomic DNA with the telomeric probe $(T_2AG_3)_4$. Genomic DNA (10 μg) from the *ura5-6* strain was digested with different restriction enzymes, electrophoresed on a 0.8% agarose gel, transferred to nylon membrane, and probed with ^{32}P-labeled $(T_2AG_3)_4$. Numbers to the left refer to DNA size standards (base pairs).

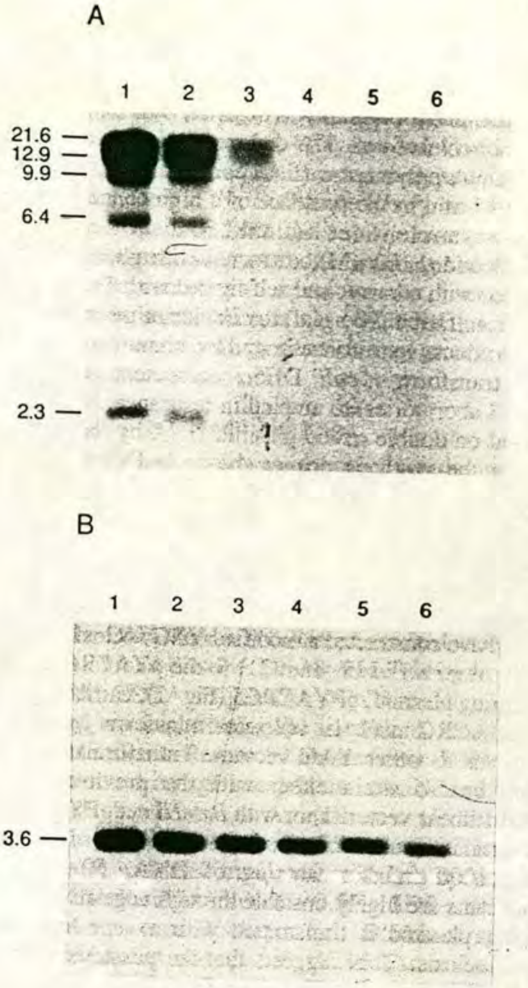


Figure 4. $(T_2AG_3)_4$ hybridization of *Bal*31-treated *P. anserina* DNA. Genomic DNA from the *ura5-6* strain was treated with *Bal*31 nuclease for 0 (lane 1), 0.5 (lane 2), 2 (lane 3), 7.5 (lane 4), 15 (lane 5), or 60 min (lane 6) and subsequently digested with *Sal*I. **(A)** DNA samples (10 μg) were electrophoresed on a 0.8% agarose gel, transferred to nylon membrane, and hybridized with ^{32}P-labeled $(T_2AG_3)_4$. **(B)** The membrane shown in (A) was reprobed with the ^{32}P-labeled 1.55 kbp *Eco*RI fragment containing the *ura5* gene. Size (kilobases) and position of detected fragments are indicated to the left of each panel.

ura5-6 strain was digested with different restriction enzymes, run on an agarose gel and probed with the radiolabeled oligonucleotide $(T_2AG_3)_4$ (Fig. 3). Several genomic DNA fragments hybridize with the probe at high stringency, demonstrating that some chromosomal DNA fragments do contain T_2AG_3 sequences. In the *Hind*III digested DNA, eight bands can be discerned with some bands being doublets or triplets. It can be estimated that at least twelve DNA fragments are detected by the probe. Assuming that *P. anserina* genome contains seven chromosomes (32, 33), this result is in good agreement with the expected number of telomeres. The considerable variation in the intensities of the fragments detected may also traduce some variation in the number of telomeric repeats present at each chromosome end. This can be observed particularly for some of the small fragments which give a very weak signal suggesting they contain very few repeats. In addition, the bands detected are rather broad and heterogenous especially for the smallest fragments from *Mbo*I or *Alu*I digests. This may

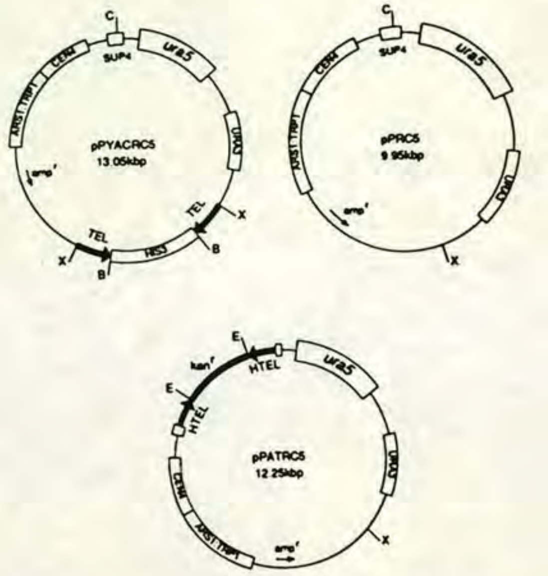


Figure 5. Maps of the linear vectors with different ends. The symbols used are as in Fig.1. In pYACRC5, the ends are *Tetrahymena* telomeres. In pPRC5, the telomeres have been deleted. pPATRC5 contains human telomeres (HTEL) with T_2AG_3 telomeric repeats, separated by the kanamycin resistance gene (kan^r). Restriction sites are: B, *Bam*HI; C, *Cla*I; E, *Eco*RI; X, *Xho*I.

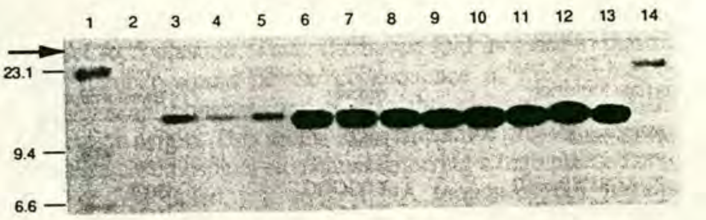


Figure 6. Southern blot analysis of DNA from *P. anserina* transformants. Total DNA (1 μg) was electrophoresed on a 0.8% agarose gel, transferred to nylon membrane, and hybridized with a mixture of ^{32}P-labeled pBR322 DNA (2.5 ng/ml; 10^9 cpm/μg) and ^{32}P-labeled λ DNA (0.5 ng/ml; 10^8 cpm/μg). Lane 1 and 14, *Hind*III digested λ DNA (0.2 μg); lane 2, DNA from the *ura5-6* strain; lanes 3 to 5, DNA from three unstable pPYACRC5 transformants; lanes 6 to 13, DNA from eight pPATRC5 transformants. Numbers to the left refer to DNA size standards (kilobases). The position of the ethidium bromide stained genomic DNA is indicated by the arrow.

reflect some heterogeneity in size as generally observed for DNA fragments ending with telomeres (34). Most of the fragments detected by the $(T_2AG_3)_4$ oligonucleotide are very small in the *Mbo*I and *Alu*I digested DNAs suggesting that the telomeric repeats can be as short as 300 bp or even shorter.

In order to demonstrate that T_2AG_3 sequences have a telomeric location, we showed that the DNA fragments detected by the probe are lost after brief digestion of intact DNA with the *Bal*31 exonuclease (Fig.4). After 7.5 min of *Bal*31 treatment, all the DNA fragments homologous to the $(T_2AG_3)_4$ probe are completely deleted whereas the same blot probed with the *ura5* gene shows that internal chromosomal sequences remain undigested even after 60 min of *Bal*31 treatment. As *Hind*III digested λ DNA was added to *P. anserina* genomic DNA before *Bal*31 treatment, the kinetic of exonuclease activity could be easily assessed by probing the same blot with λ DNA (data not shown). About 200 bp were removed after 7.5 min of digestion, suggesting that T_2AG_3 repeats are confined within the terminal 200 bp of *P. anserina* chromosomes. Because *Bal*31 can remove telomeric repeats very quickly, this value may be largely underestimated, but it is in agreement with the small size observed above for some of the telomeric fragments.

Theses results demonstrate that *P. anserina* telomeres are composed of short tracks of T_2AG_3 repeats and therefore, confirm the expected telomeric nature of DNA sequences added to the termini of linear plasmids after replication in *P. anserina*.

Effect of telomeric sequences containing T_2AG_3 repeats on autonomous replication of a linear plasmid in *P. anserina*

In order to assess the effect of T_2AG_3 telomeric repeats, we constructed the plasmids shown in Fig. 5. The three plasmids carry the *ura5* selective marker but differ by their termini after linearization with the appropriate restriction enzyme. In pPYACRC5 *Bam*HI digested DNA, the termini are composed of *Tetrahymena* telomeric sequences whereas there are no telomeres in *Xho*I or *Cla*I linearized pPRC5. As human telomeres have been shown to be also composed of T_2AG_3 repeats (16, 17), we used human telomeric sequences (HTEL) to construct the plasmid pPATRC5. The human telomeric DNA has been previously isolated by complementation of a deficient yeast artificial chromosome (16) and has been shown to induce the formation of functional telomeres when reintroduced in mammalian cells (35). From the *Eco*RI sites in pPATRC5, the telomeric sequence corresponding to an *Eco*RI-*Hae*III DNA fragment issued from pHUTEL-2-end (24) starts with 12 bp of yeast telomeric repeats, followed by about 500 bp of T_2AG_3 repeats and a few related variants such as $GTAG_3$ and $TCAG_3$.

As shown in table 1, either *Cla*I or *Xho*I linearized pPRC5 DNA transforms the *P. anserina ura5-6* strain with a low frequency and the ura^+ phenotype is stable through vegetative growth. Transformation occurs by integration of the plasmid in chromosomal DNA as we reported previously for linear plasmids without telomeric ends (1, 20). Transformation frequency shows a 3 fold increase when protoplasts are transformed with the *Bam*HI linearized pPYACRC5 plasmid, bearing *Tetrahymena* telomeric repeats. About 70% of the transformants are unstable and loose the ura^+ phenotype upon growing on non selective medium whereas the remaining 30% are stable through vegetative growth. Transformation frequency of the linear plasmid bearing T_2AG_3 repeats (*Eco*RI linearized pPATRC5) shows nearly a 60 fold increase over the transformation frequency observed with *Bam*HI cut pPYACRC5 and reaches 1.75×10^4 transformants/μg DNA. All the pPATRC5 transformants tested were unstable when grown on non selective medium but unlik pPYACRC5 transformants, the loss of the ura^+ phenotype wa not complete. After 4–5 days culture upon non selectiv conditions, pieces of mycelium picked at the edge of the thall were able to regenerate a mycelium with a lag time of 48h afte transfer on minimal medium without uridine. The delay observe for growth recovery on selective medium indicates that som nuclei still contained the ura^+ gene after propagation on no selective medium. This could arise either by stabilization event such as integration in chromosomal DNA or by persistance o a few nuclei containing the linear replicating plasmid. To examin these two possibilities, transformants were subcultured sever times on minimal medium supplemented with uridine until lo of the ura^+ phenotype. All the transformants lost the *ura*

Table 1. Transformation frequency of the linear plasmid bearing human telomeric DNA and mitotic stability of the transformants

Type of DNA used in transformation	Type of telomeric repeat	Total number of transformants[a]	Number of transformants tested	% of viable transformants[c]	% of unstable transformants[d]
pPRC5 *Xho*I	none	120	20	100	0
pPRC5 *Cla*I	none	100	20	100	0
pPYACRC5*Bam*HI	TTGGGG	307[b]	48	91	66
pPATRC5 *Eco*RI	TTAGGG	1.75×10^{4}[b]	100	100	100

[a] Protoplasts were prepared from the *ura5-6* mutant. Equal number of protoplasts (5×10^7) were transformed with 1 μg of plasmid DNA preparation.

[b] Calculated from two independant transformation experiments.

[c] Transformants which were able to grow after transfer from transformation plates to minimal medium.

[d] Transformants were grown on minimal medium supplemented with uridine (0.1 mg/ml). After 4–5 days, four small pieces of mycelium were picked at the edge of the thalli and grown on minimal medium with or without uridine to determine the ura$^+$ or ura$^-$ phenotypes

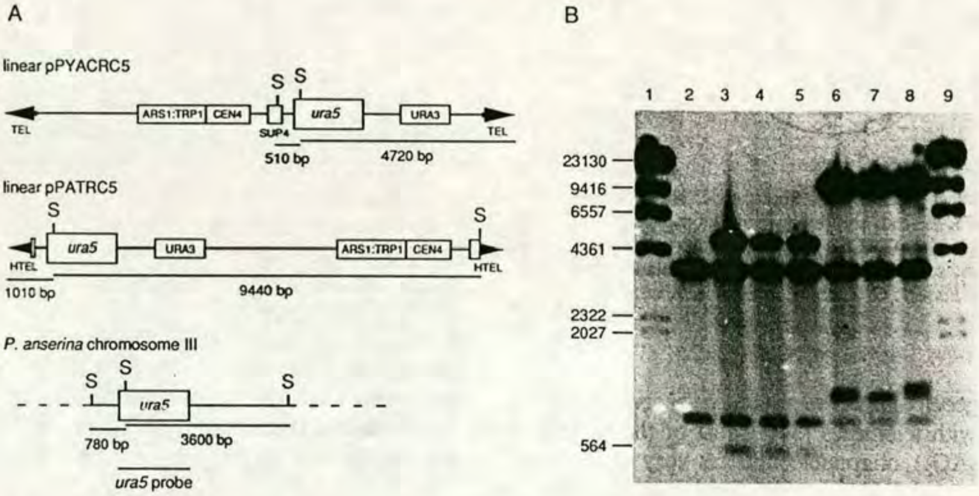


Figure 7. Determination of the plasmid copy number in pPYARC5 and pPATCR5 transformants. **(A)** Location of *Sal*I sites (S) in the linear plasmids and in the DNA region encompassing the chromosomal *ura5* gene. Thin lines represent the expected size of the DNA fragments detected by the *ura5* probe. The symbols used in the drawing are as in Fig. 1 and 5. **(B)** Southern blot analysis of DNA from pPYACRC5 and pPATRC5 transformants. Total DNA (1 μg) was digested with *Sal*I, electrophoresed on a 0.8% agarose gel, transferred to nylon membrane, and probed with a mixture of the ^{32}P-labeled 1.55 kbp *Eco*RI fragment carrying the *ura5* gene (2.5 ng/ml; 10^9 cpm/μg) and ^{32}P-labeled λ DNA (0.5 ng/ml; 10^8 cpm/μg). Lanes 1 and 9, *Hind*III digested λ DNA (0.2 μg); lane 2, DNA from the *ura5* recipient strain; lanes 3 to 5, DNA from three unstable pPYACRC5 transformants; lanes 6 to 8, DNA from three pPATRC5 transformants. The autoradiogram was exposed for 72 h. Relative intensities were determined by using a scanning densitometer on X-ray films exposed for 2, 16 and 48 h. Numbers to the left refer to DNA size standards (kilobases).

phenotype after 2 or 3 passages on non selective medium suggesting that the observed increased stability of the transformants does not result from integration events.

To determine whether autonomous linear replication occured in pPATRC5 transformants, the presence of pBR322 sequences on extrachromosomal molecules was examined by analysis of DNA extracted from transformed strains. Total undigested DNA was run on an agarose gel and probed with radiolabeled pBR322 (Fig. 6). A single band of the correct size is detected for each transformant and no hybridization is observed with chromosomal DNA. This means that the linear plasmid replicates autonomously without integration of pBR322 sequences into chromosomal DNA. The comparison of band intensities between DNAs from pPYACRC5 and pPATRC5 transformants clearly indicates a higher plasmid copy number in pPATRC5 transformants.

Using the single copy chromosomal *ura5* gene as a reference, we determined the amount of plasmid DNA in pPYACRC5 and pPATRC5 transformants. Total DNA was cut with *Sal*I, run on an agarose gel and probed with the 1.55 kbp *Eco*RI fragment carrying the *ura5* gene (21). Fig. 7A shows *Sal*I restriction maps of the linear plasmids and the location of *Sal*I sites encompassing the chromosomal *ura5* gene. As the size of *Sal*I DNA fragments detected by the probe is different between chromosomal and plasmid DNA, the comparison of bands intensity in each lane gives an accurate measurement of the plasmid copy number in each transformant. From the quantitation of the autoradiogram shown in Fig. 7B we calculated an average plasmid copy number of 0.4 and 2.7 per nucleus for pPYACRC5 and pPATRC5 transformants respectively. This result provides evidence that the replacement of *Tetrahymena* telomeric templates by human telomeric repeats leads to a seven fold increase in plasmid copy number and may explain the increased mitotic stability of pPATRC5 transformants.

In order to know if the plasmid copy number has an effect on the meiotic stability of the linear plasmid, pPATRC5 and pPYACRC5 transformants were crossed with the *ura5-6* strain. With pPYACRC5 transformants, the frequency of *ura*$^+$ strains in the progeny was very low (below 10^{-2}), as previously reported for linear plasmids bearing *Tetrahymena* telomeric repeats in *P. anserina* (1). The frequency of transmission of the linear pPATRC5 plasmid through meiosis was in the same order of magnitude, indicating that the increased plasmid copy number in pPATRC5 transformants does not improve the frequency of transmission of the plasmid through meiosis.

DISCUSSION

Telomeric DNA and structure are similar among otherwise widely divergent eukaryotes. Telomere repeats from the ciliate *Tetrahymena* can seed telomere formation on linear plasmids in *S.cerevisiae* (36), *S.pombe* (7), and *P. anserina* (1). The telomere terminal transferase activity in *Tetrahymena* cell extracts recognizes any G-rich telomere strand as a primer (9, 10). Such a process could explain how the putative yeasts and *P. anserina* telomerases recognize heterologous telomeric repeats. However, linear plasmids bearing *Tetrahymena* telomeric sequences transform *P. anserina* with a low frequency and autonomous replication occurs in only a fraction of the transformants. This observation suggests that the primary defect is in heterologous telomeric DNA recognition by the putative *P. anserina* telomerase. We have shown here that the healing of *Tetrahymena* telomeric templates results in the addition of an homogeneous track of T_2AG_3 repeats. Hybridization experiments demonstrated that T_2AG_3 repeats are the component of chromosomal telomeres and are confined within the terminal 200 bp of each *P. anserina* chromosome. As there is only a single base change between telomeric repeats from *Tetrahymena* and *P. anserina*, this could explain why *Tetrahymena* telomeric DNA can seed telomere formation in *P. anserina*. Alternatively, this single base difference may be deleterious for efficient recognition and/or elongation by the putative *P. anserina* telomerase. To test the effect of homologous telomeric sequences, we constructed the plasmid pPATRC5 in which *Tetrahymena* repeats from pPYACRC5 have been replaced by a human telomeric DNA fragment containing T_2AG_3 repeats. The change in telomeric templates results in a 60 fold increase in transformation frequency, autonomous replication occurs in 100% of the transformants and the plasmid copy number is shifted from 0.4 to 2.7 per nucleus. The most likely explanation for these results is that homologous T_2AG_3 repeats are more efficiently recognized as telomeric DNA than *Tetrahymena* repeats. However, the shift in plasmid copy number remains an intriguing result. After propagation in the transformants, it is likely that both pPYACRC5 and pPATRC5 plasmids have acquired functional *P. anserina* telomeres so that the main difference between the two linear plasmids concerns internal sequences immediately adjacent to the termini. It is to notice that we could never demonstrate that the linear plasmid with *Tetrahymena* telomeres was bearing a specific sequence functioning as a replication origin in *P. anserina* and the replication is thought to initiate at non specific sequences. Therefore, the seven fold increase in plasmid copy number could be mediated by a replication origin lying on human derived telomeric sequences. It has been shown that the X and Y elements in *S.cerevisiae* subtelomere domains can function as autonomous replicating sequences in yeast cells (36). However, the human telomeric DNA fragment used in this study corresponds to the terminal repeats of pHUTEL-2end (24) and does not contain any subtelomeric sequence. If no particular sequences are responsible for the increased plasmid copy number in pPATRC5 transformants, this suggests that *Tetrahymena* sequences have a deleterious effect on the replication or on the maintenance of the pPYACRC5 plasmid. Thus, the seven fold increase in plasmid copy number would be the consequence of the removal of *Tetrahymena* sequences rather than the addition of a putative replication origin. This seems to be true since previous attempts to isolate chromosomal elements conferring a high plasmid copy number had failed : if some DNA sequences can increase transformation frequency, none have been found to have an effect on the plasmid copy number (18). The 700 bp *Tetrahymena* telomeric fragment contains approximately 400 bp of unique AT-rich DNA followed by 300 bp of T_2G_4 repeats (37). It is not clear whether the low copy number of the pPYACRC5 plasmid could be due to the presence of the T_2G_4 track or to internal sequences. Further experiments are needed to confirm the unexpected effect of *Tetrahymena* sequences on the maintenance of linear plasmids in *P. anserina*.

In *Fusarium oxysporum*, linear self replicating plasmids are produced by the *in vivo* rearrangement of normally integrative circular vectors. The termini of the linear plasmids are composed of T_2AG_3 repeats and the plasmid copy number is approximately 10 per haploid genome (38). As telomere formation seems to be sufficient to allow autonomous replication in *F. oxysporum* and *P. anserina*, the existence of replication origins as specific DNA sequences remains to be investigated.

The replacement of *Tetrahymena* telomeric template by human telomeric sequences results in a net increase on the mitotic stability of the linear plasmid but does not improve its transmission through meiosis. Therefore, the observed increase in mitotic stability can simply be explained as a consequence of a more efficient replication of the linear plasmid rather than an improved fidelity of transmission through mitosis. However, it has been recently demonstrated that human telomeres are tightly associated with the nuclear matrix and it seems likely that the T_2AG_3 repeats mediate the anchorage of the telomere (39). As telomeric repeats are identical between *P. anserina* and human, the same attachment may occur in *P. anserina* . Thus, a single base difference in the sequence of the telomeric repeat may destabilise the anchorage to the nuclear matrix and explain the difference in stability between the two linear plasmids. It is not presently known if the attachment to the nuclear matrix is mediated by a nucleoprotein complex forming at the telomere but it is demonstrated that some proteins bind specifically to the telomere in ciliates (40, 41) and in yeast (42, 43). In this last case, binding of the protein RAP1 to the telomeric end has been shown to have a direct effect on telomere lengh and on chromosome stability. If such proteins exist in *P. anserina*, a single nucleotide change in the sequence of the telomeric repeat of the linear plasmid may also affect the stability of the molecule.

In *S.cerevisiae* and *S.pombe*, centromeric DNAs have been shown to allow artificial minichromosomes to be segregated properly through nuclear divisions (44, 15). Experiments are now in progress to use the plasmid pPATRC5 as a cloning vector to isolate a functional *P. anserina* centromere.

ACKNOWLEDGEMENTS

We thank Joel Bégueret and Michel Perrot for helpful discussions and critical reading of the manuscript. This work was supported by the Centre National de la Recherche Scientifique and the Université de Bordeaux II. J.P.J. was supported by a grant from the Ministère de la Recherche et de la Technologie.

REFERENCES

1. Perrot, M., Barreau, C. and Bégueret, J. (1987). Mol. Cell. Biol. **7,** 1725- 1730.
2. Blackburn, E. H., and Szostak, J. W. (1984). Ann. Rev. Biochem. **53,** 163–194
3. Blackburn, E. H. (1991). Nature (London) **350,** 569–573.
4. Blackburn, E. H., and Gall, J. G. (1978). J. Mol. Biol. **120,** 33–53.
5. Klobutcher, L. A., Swanton, M. T., Donini, P. and Prescott, D. M. (1981). Proc. Natl. Acad. Sci. USA **78,** 3015–3019.
6. Shampay, J., Szostak, J. W. and Blackburn, E. H. (1984). Nature (London) **310,** 154–157.
7. Sugawara, N., and Szostak, J. W. (1986). Yeast **2** (suppl.), 373 (abstr.).
8. Emery, S.H., and Weiner, A. M. (1981). Cell **26**: 411–419
9. Greider, C. W. and Blackburn, E. H. (1985). Cell **43,** 405–413.
10. Greider, C. W. and Blackburn, E. H. (1987). Cell **51,** 887–898.
11. Greider, C. W. and Blackburn, E. H. (1989). Nature (London) **337,** 331–337.
12. Morin, G. B. (1989). Cell **59,** 521–529.
13. Shippen-Lentz D., and Blackburn, E. H. (1990). Science **247,** 546–552.
14. Pluta, A. F., Dani, G. M., Spear, B. B. and Zakian, V. A. (1984). Proc. Natl. Acad. Sci. USA **81,** 1475–1479.
15. Hahnenberger, K. M., Baum, M. P., Polizzi, C. M., Carbon, J. and Clarke,L. (1989). Proc. Natl. Acad. Sci. USA **86,** 577–581
16. Cross, S. H., Allshire, R. C., McKay, S. J., McGill, N. I. and Cooke, H.J. (1989). Nature (London) **338,** 771–774.
17. Brown, W. R. A. (1989). Nature (London) **338,** 774–776.
18. Bégueret, J., Turcq, B., Razanamparany, V., Perrière, M.,Denayrolles, M., Berges, T., Perrot, M., Javerzat, J. P. and Barreau, C. (1989). Proceedings of the EMBO-Alko workshop on Molecular Biology of Filamentous Fungi. Ed. by H. Nevalainen & M. Penttilä. Foundation for Biotechnical and Industrial Fermentation Research. **6,** 41–49
19. Esser, K. 1974. *Podospora anserina*. In R. C. King (ed.), Handbook of genetics. Plenum Publishing Corp., New York. p. 513.
20. Razanamparany, V., and Bégueret, J. (1986). Curr. Genet. **10,** 811–817.
21. Turcq, B., and Bégueret, J. (1987). Gene **53,** 201–209.
22. Burke, D. T., Carle, G.F. and Olson, M. V. (1987). Science **236,** 806–812.
23. Marchuk, D., and Collins, F. S. (1988). Nucleic Acids Res. **16**: 7743.
24. Cross, S., Lindsey, J., Fantes, J., McKay, S. J., McGill, N. and Cooke, H. (1990). Nucleic Acids Res. **18,** 6649–6657.
25. Viera, J., and Messing, J. (1982). Gene **19,** 259–268.
26. Oka, A., Sugisaki, H. and Takanami, M. (1981). J. Mol. Biol. **147,** 217 226.
27. Dubridge, R. B., Tang, P., Hsia, H. C., Leong, P. M., Miller, J. H. and Calos, M. P. (1987). Mol. Cell. Biol. **7,** 379–387.
28. Bégueret, J., Razanamparany, V., Perrot, M. and Barreau, C. (1984). Gene **32,** 487–492.
29. Berges, T. and Barreau C. (1989). J. Gen. Microbiol. **135,** 601–604.
30. Wang, S. S. and Zakian, V. A. (1990). Mol. Cell. Biol. **10,** 4415–4419.
31. Sanger, F., Nicklen, S. and Coulson, A. R. (1977). Proc. Natl. Acad. Sci. USA **74,** 5463–5467.
32. Marcou, D., Piccard-Bennoun, M. and Simonet, J. M. (1984). Genetic map of *Podospora anserina*. In Genetic Maps, S., J. O'Brian (ed.) Cold Spring Harbor Laboratory Publications, Cold Spring Harbor, NY., **3,** 252–261.
33. Simonet, J. M., and Zickler, D. (1979). Curr. Genet. **175,** 359–367.
34. Szostak, J. W., and Blackburn, E. H. (1982). Cell **29,** 245–255.
35. Farr, C., Fantes, J., Goodfellow, P. and Cooke, H. (1991). Proc. Natl. Acad. Sci. USA **88,**7006–7010.
36. Horowitz, H. and Haber, J. E. (1985). Mol. Cell. Biol. **5,** 2369–2380.
37. Kiss, G. B., Amin, A. A. and Pearlman, R. E. (1981). Mol. Cell. Biol. **1,** 535–543.
38. Powell, W. A. and Kistler, H. C. (1990). J. Bacteriol. **172,** 3163–3171.
39. De Lange, T. (1992) EMBO J., **11,** 717–724.
40. Price, C.M. (1990) Mol. Cel. Biol., **10,** 3421–3431.
41. Price, C.M. and Cech, T.R. (1989) Biochemistry, **28,** 769–774.
42. Conrad, M.N., Wright, J.H., Wolf, A.J. and Zakian, V.A. (1990) Cell, **63,** 739–750.
43. Lustig, A.J., Kurtz S. and Shore D. (1990) Science, **250,** 549–553.
44. Murray, A. W. and Szostak, J. W. (1983). Nature (London) **305,** 189–193.