Analysis of a Gene Amplification Event in Rat Cells

Edith Heard

Thesis presented in partial fulfillment
of the degree of Doctor of Philosophy
to the University of London

Imperial Cancer Research Fund
Laboratories
Lincoln's Inn Fields
London, WC2A 3PX

Department of Biochemistry
Imperial College of Science
and Technology
London, SW7 2AZ

September 1990

For Vincent,

who shared this experience.

ABSTRACT.

The 3B transformed rat cell line was isolated as part of an experiment to try and clone mouse enhancer elements by transfecting immortalized rat cells (the Rat-2 cell line) with a ligation mixture of total restricted mouse DNA and a polyoma virus transforming gene. When the polyoma insert was cloned out of the genomic DNA of the 3B transformant, it was found to be in the form of a large inverted duplication, consisting of both the mouse and polyoma sequences as well as rat DNA. The palindromic sequences were also found to be amplified 20-40 fold in 3B genomic DNA. Thus, the amplified DNA in this cell line existed in the form of large inverted repeats. This has subsequently been found to be a common feature of mammalian amplified DNA sequences. Furthermore, the finding that amplification units are organized as inverted rather than tandem repeats has certain implications on the possible mechanisms by which amplification occurs in mammalian cells. This thesis extends the analysis of the amplified DNA in the 3B cell line with the purpose of elucidating the nature of the processes by which it arose. The structure and the stability of the amplification have been studied, at both a molecular and a karyotypic level. Molecular cloning and long range restriction mapping have shown that the amplified units (each unit being one side of the inverted duplication) seem to be homogeneous, and are approximately 450 kb in size. No rearrangements are detected between the amplified rat sequences and the unamplified alleles of the same region in rat cells, and the structure and copy number of the amplification units appear to be completely stable. Furthermore, the units appear to be clustered in an array of 10-11 Mb. Cytogenetic analysis by fluorescence *in situ* hybridisation of 3B metaphase spreads with a biotinylated probe which detects only the amplified DNA, has shown that all the amplified DNA is stably located at a single chromosomal site. Similar analysis of both Rat-2 and 3B cells, using a rat DNA probe that detects both the amplified and unamplified copies of the region, has revealed the chromosomal origin of the inverted duplication. Karyotypic analysis has shown that the amplified DNA is located on a rearranged chromosome, and that the amplification event probably led to the disruption of the chromosome from which it arose. On the basis of all these observations, the possible mechanism which gave rise to the amplified DNA in 3B cells is considered.

ACKNOWLEDGEMENTS

Many thanks to members of the TVG / EGOE lab., whose solidarity and humour carried me through my time here. Special thanks to Moira Read for her patience, and her expert help with cultured cells; and to Brendan Davies, for being my constant source of advice and competent cells, and for being a tolerant neighbour.

I am also greatly indebted to Sarah Williams for being generous with both her expertise in cytogenetics and her time; and to Zoia Larin and Tony Monaco for their constant encouragement and help with YAC cloning.

Finally, my thanks to Mike, for providing a unique experience, for teaching me about amplifiction, and for initiating me in the ways of science.

ABBREVIATIONS.

ABR	abnormal banded region
AMPD	adenylate deaminase
bp	base pair
CAD	carbamyl phosphate synthetase, aspartate transcarbamylase, dihydroorotase.
DHFR	dihydrofolate reductatase
DM	double minute
DNA	deoxyribonucleic acid
DTT	dithiothreitol
ECR	extended chromosomal region
EDTA	diaminoethane tetra-acetic acid (disodium salt)
FCS	foetal calf serum
hrs	hours
HSR	homogeneously staining region
kb	kilo base
LMP	low melting point
Mb	mega base
MDR	multi-drug resistance
mins	minutes
MTX	methotrexate
O/N	overnight
PALA	N-(phosphonacetyl)-L-aspartate
PCR	polymerase chain reaction
PFGE	pulsed field gel electrophoresis
Py	polyoma virus
rpm	revolutions per minute
SDS	sodium dodecyl sulphate
secs	seconds
Tris	Tris[hydroxymethyl]aminomethane
YAC	yeast artificial chromosome

CONTENTS

		PAGE
CHAPTER 1	GENERAL INTRODUCTION	11
1.1	Developmentally Programmed Gene Amplification in Lower Eukaryotes.	11
1.2	Gene Amplification in Pesticide Resistance.	13
1.3	Gene Amplification in the Germ Line of Mammals.	15
1.4	Gene Amplification in Somatic Cells.	16
1.5	Systems and Strategies Used for the Investigation of Mammalian Gene Amplification.	18
1.6	Karyotypic Hallmarks of Amplification.	21
1.7	Submicroscopic Circular DNA.	22
1.8	Location of Amplified DNA.	23
1.9	Gene Amplification Associated with Chromosomal Rearrangement.	25
1.10	Amplification and Chromosomal Deletion.	28
1.11	The Molecular Structure of Amplified DNA.	29
1.12	Novel Joints and the Organisation of Amplified DNA.	31
1.13	Stability and Evolution of Amplified DNA.	34
1.14	Mechanisms of Amplification in Mammalian Cells.	37
1.15	Models	40
1.16	Questions to be answered.	58
1.17	The Amplified Inverted Duplication in the 3B Rat Cell Line.	59
	OBJECTIVES	64
CHAPTER 2		67
2.1	INTRODUCTION	67
2.2	Isolation of a 15 kb BamHI Rat DNA Clone from the Inverted Duplication.	67
2.3	Isolation of Further Overlapping Sau3A Clones from the Inverted Duplication.	72
2.4	DISCUSSION	76

CHAPTER 3 80
3.1 INTRODUCTION 80
3.2 The Identification of Infrequently Cutting Restriction Enzyme Sites which Lie Within the Inverted Duplication. 82
3.3 The use of 5-azacytidine to Reveal the Presence of Multiple Methylated SalI and XhoI sites in the Amplified Units. 87
3.4 The Use of Infrequently cutting Restriction Enzymes which do not Cleave within the Amplified Rat sequences in 3B Cells in Determining the Size of the Amplification Units. 96
3.5 The Amplified Units may be Clustered in a Single Region of 3B DNA. 100
3.6 Detection of Sites which could be Useful for Chromosome Jumping Along the Inverted Duplication. 108
3.7 DISCUSSION 117

CHAPTER 4 123
4.1 INTRODUCTION 123
4.2 Construction of a YAC Library of 3B EcoRI Partial Fragments Without DNA Size Fractionation. 129
4.3 Stability of a Large Inverted Duplication Present as a YAC in yeast cells. 136
4.4 Attempted Cloning of the 240 kb ClaI Fragment Spanning the Centre of the Inverted Duplication. 141
4.5 Attempted Cloning of the 180 kb XhoI Fragment Spanning the Centre of the Inverted Duplication. 155
4.6 Construction of a Size-Selected YAC library with EcoRI Overlapping Fragments of 3B DNA. 164
4.7 DISCUSSION 177

CHAPTER 5 184
5.1 INTRODUCTION 184
5.2 *In Situ* Hybridisation of 3B Cell Chromosome Spreads with the Amplification-specific λ3B Probe. 186
5.3 *In Situ* Hybridisation of 3B and Rat 2 Chromosome Spreads with the λBam Rat Probe. 192
5.4 Characterisation of 3B and Rat 2 Chromosomes by G-banding 198

Followed by *In Situ* Hybridisation with the λBam probe.
5.5 Karyotypic Analysis of the 3B and Rat 2 Cell Lines. 212
5.6 Investigation of the Translocated Marker Chromosome M1/t(2;3). 237
5.7 DISCUSSION 244

CHAPTER 6 CONCLUSIONS AND FUTURE PROSPECTS 253

CHAPTER 7 MATERIALS AND METHODS 258
Materials 258
Methods 264

REFERENCES 298

APPENDIX 1 308

FIGURES.

		PAGE
1.1	Unequal sister chromatid exchange.	41
1.2	The disproportionate replication model.	45
1.3	The extrachromosomal double rolling circle model.	49
1.4	The chromosomal spiral model.	52
1.5	Restriction map of the centre of the inverted duplication and the λ3B clone.	60
2.2.1	Restriction map of the cloned region of the inverted duplication in 3B cells and a summary of the lambda clones isolated.	68
2.2.2	Southern blotting analysis of 3B and Rat 2 genomic DNA using the 3.9 kb RI-Bam rat DNA probe.	70
2.2.3	Southern blotting analysis of 3B and Rat 2 genomic DNA using a 0.8 kb KpnI-SalI rat DNA probe.	74
3.2.1	Analysis by PFGE of the amplified DNA in 3B cells cut with a range of rare cutter restriction enzymes.	83
3.2.2	Analysis by PFGE of 3B genomic DNA cut with a range of rare cutter restriction enzymes using the Py and RI-Bam probes.	86
3.3.1	Analysis by PFGE of SalI-digested genomic DNA from 3B cells grown in the absence or presence of 5-azacytidine.	88
3.3.2	Analysis by PFGE of SalI and XhoI digested genomic DNA from 3B and Rat 2 cells grown in the presence of different concentrations of 5-azacytidine.	92
3.3.3	Long range restriction map of the inverted duplication in 3B cells and of region of rat genome from which it has been derived.	94
3.4.1	Analysis by PFGE of 3B and Rat 2 genomic DNA cut with the enzymes NarI and EagI.	97
3.5.1	Analysis of the overall organization of the amplification units in 3B cells by PFGE resolution of DNA up to 3 Mb in length.	101
3.5.2	Further analysis of the organization of the amplified units in 3B cells through the resolution of DNA up to approximately 10 Mb in length.	104
3.5.3	The organization of the amplified array in 3B cells and of the native locus in Rat 2 cells.	106
3.6.1	Rare cutter enzymes located within the insert of the λ3B clone.	110

3.6.2 Ability of rare cutter enzyme sites at the centre of the inverted duplication to be cleaved in 3B cellular DNA. 112

3.6.3 Analysis by PFGE of NaeI sites present in the inverted duplication in 3B cells. 115

4.1.1 Strategy for the cloning of large fragments of genomic DNA in yeast artificial chromosomes. 125

4..2.1 Size determination of random YACs from a library of unsize-selected partially cleaved Eco RI fragments of 3B DNA. 132

4.3.1 The pYAC-RC vector. 137

4.3.2 Construction of a YAC containing the complete λ3B 'phage DNA. 139

4.3.3 Analysis by PFGE of the sizes of nine independent YACλ3B isolates. 142

4.3.4 Investigation by southern blotting of the structural integrity of the λ3B DNA insert in independent YAC λ3B isolates. 144

4.4.1 Long range restriction map of ClaI, XhoI and SalI sites within the amplified inverted duplication in 3B cells. 147

4.4.2 Size determination of random YACs from an unsize-selected library of ClaI-cut 3B DNA. 150

4.4.3 Slot blot analysis of size-fractionated, ClaI-cut genomic DNA from 3B cells. 153

4.4.4 Size determination of random YACs from a library of size-selected ClaI-cut 3B DNA fragments. 156

4.5.1 Analysis by PFGE of random YACs from a library of unsize-selected XhoI-cut 3B DNA fragments. 160

4.5.2 Further PFGE analysis of YACs from an unsize-selected library of XhoI-cut 3B DNA fragments. 162

4.6.1 Size comparison of YACs in yeast transformants obtained with or without YAC size selection prior to transformation. 167

4.6.2 Southern analysis of the size-fractionated YAC ligations to be used in the construction of a genomic library of overlapping EcoRI fragments. 169

4.6.3 Analysis of random RI YAC library transformants for the presence of both YAC vector arms. 173

4.6.4 Size determination of random YACs from the size-selected EcoRI YAC library of 3B DNA. 175

5.2.1 Fluorescence *in situ* hybridisation of metaphase spreads from early-passage 3B cells with the λ3B probe. 188
5.2.2 Fluorescence *in situ* hybridisation of metaphase spreads from late-passage 3B cells with the λ3B probe. 190
5.3.1 Fluorescence *in situ* hybridisation of metaphase spreads from 3B cells with the λBam rat probe. 193
5.3.2 Fluorescence *in situ* hybridisation of metaphase spreads from Rat 2 cells with the λBam rat probe. 196
5.4.1 G-banding of Rat 2 metaphase spreads followed by *in situ* hybridisation with the λBam rat probe. 199
5.4.2 G-banding of 3B metaphase spreads followed by *in situ* hybridisation with the λBam rat probe. 202
5.4.3 G-banded karyotype of the Norway rat, *Rattus norvegicus*. 204
5.4.4 G-band nomenclature of rat chromosome 2 (adapted from Levan, 1974). 206
5.4.5 Partial karyotypes of Rat 2 and 3B cells. 208
5.4.6 Diagrammatic representation of G-banding patterns of rat chromosomes 2 and 3, and marker chromosomes M1/t(2;3) and M2. 210
5.5.1 G-banded karyotype of a Rat 2 cell. 215
5.5.2 G-banded karyotype of a Rat 2 cell. 218
5.5.3 G-banded karyotype of a Rat 2 cell. 221
5.5.4 G-banded karyotype of a 3B cell. 225
5.5.5 G-banded karyotype of a 3B cell. 228
5.5.6 G-banded karyotype of a 3B cell. 231
5.5.7 G-banded karyotype of a 3B cell. 234
5.6.1 Fluorescence *in situ* hybridisation of metaphase spreads from Rat 1 and Rat 2 cells with the λBam rat probe. 239
5.6.2 Detection of a fragile site on marker chromosome M1/t(2;3). 242
5.6.3 Detection of the fragile site on chromosome M1/t(2;3) using Giemsa solid staining. 245

CHAPTER 1

GENERAL INTRODUCTION

Certain biological events require the synthesis of high levels of a gene product in a relatively short space of time. When this requirement cannot be met by an increase in the rate of transcription of a particular gene, one solution can be the selective amplification of the DNA sequence containing the gene. This results in an increase in the relative amount of the DNA sequence within a cell. There are examples throughout the evolutionary spectrum of gene amplification being employed as a means of responding to special physiological or developmental demands. In this chapter, a brief overview of some examples of gene amplification that occur in eukaryotes either as part of programmed development, or as a response to toxic agents, will be given together with the possible mechanisms involved in each case. The focus of the chapter will be on gene amplification in mammalian systems, where it is observed only as an abnormal event in cultured cells and in neoplasia. The amplification of proto-oncogenes, and of genes involved in drug resistance, will be described. The cytological and molecular information available for some of these examples of gene amplification will be summarised, and some possible mechanisms and their predictions will be outlined. Finally, the partial characterization of an amplified polyoma virus oncogene, in a rat cell line (3B) isolated by Mike Fried's group will be described.

1.1 Developmentally Programmed Gene Amplification in Lower Eukaryotes.

Selective gene amplification is utilized during the development of some lower eukaryotes, in order to accumulate high levels of a product in specific cells over a short period of time. The amplification of rDNA during amphibian oogenesis is one example of developmentally specific amplification (see Kafatos et al, 1985 for review). This can be explained in terms of the requirement for enormous numbers of ribosomes in a growing oocyte. For sufficient ribosomes to be assembled within a few months, which

is the normal duration of amphibian oogenesis, differential rDNA amplification is essential. The necessary levels of the protein components of ribosomes can be achieved by maximal transcription and translation of the genes, as, unlike rRNAs, the proteins are not direct gene products, and thus can accumulate at much faster rates. Amplification of rDNA occurs at oogenesis in several other animals. In Xenopus, two rounds of amplification (10-40 fold, and then 2500 fold) mean that in the mature oocyte, 70% of the nuclear DNA codes for rRNA. Experimental evidence supports an extrachromosomal rolling circle mechanism for this amplification, involving only a subset of the rRNA genes (Rochaix, 1974).

Some of the best studied examples of developmentally regulated gene amplification are to be found in insects. The selective amplification of chorion genes in the follicular cells of *Drosophila melanogaster*, is one of these (Kafatos et al, 1985). Choriogenesis (the synthesis of the eggshell around the oocyte) is extremely rapid, being almost complete within two hours. In order for the necessary amount of chorion mRNA to be produced in the follicular cells in such a short time, amplification of the two clusters of chorion genes occurs. The mechanism of amplification is thought to involve repeated rounds of local chromosomal replication initiated in or near each gene cluster, such that multiple replication forks spread bidirectionally to variable distances (80-100kb), producing gradients of amplification that are maximal in the gene cluster (Spradling, 1981). Support for this model comes from the observation of multiply branched chromatin fibers encompassing the chorion genes (Osheim and Miller, 1983). In this example of gene amplification, the controls which normally prevent reinitiation of DNA replication within a single S phase have obviously been overcome. However, follicular cells appear to be exceptional in their fate. These cells cease to divide during oogenesis and undergo multiple rounds of endoreduplication thus becoming polyploid. Amplification of the chorion genes begins as endoreduplication ends, and it seems that after this stage in development, the replicons in other chromosomal regions in these cells cannot respond to replication signals. Furthermore, it seems that chorion gene amplification in follicle cells is under the tight control of specific *cis* and *trans*-acting elements (for review see Stark et al, 1990). Thus it seems that disproportionate

replication, which has been favoured as a model for amplification in mammalian cells (see Section 1.15), can only occur in a cell under exceptional circumstances, where the usual control of DNA replication will no longer be required for normal cell division.

Further examples of amplification in dipteran flies include the selective overreplication of several loci in the salivary gland chromosomes of the Sciaraidae family (see Stark et al, 1990). In the late larval stages of development, specific regions of these polytene chromosomes increase in size and open into large transcriptionally active DNA puffs. The genes within these puffs encode abundant proteins which are required for construction of the larval cocoon. The degree of amplification is about 16-fold, and the mechanism is thought to involve disproportionate replication, as in chorion gene amplification in *Drosophila.* Again, these examples involve cells which have undergone endoreduplication, and are not destined to divide further.

1.2 Gene Amplification in Pesticide Resistance.

Gene amplification as a means of generating resistance to toxic agents is a widespread phenomenon in unicellular and multicellular organisms and it can create serious problems, both in medicine and in agriculture.

The protozoan parasites of the genus Leishmania, which cause a spectrum of human disease (Leishmaniasis) ranging from mild to fatal forms, have been found to amplify the dihydrofolate reductase-thymidylate synthase gene in response to the antifolate drug, methotrexate (MTX). Resistance occurs as a result of the high levels of the target protein (DHFR) of methotrexate. Amplification appears to involve the formation of extrachromosomal circular dimers derived from the chromosomal region containing the target gene, which are capable of replicating autonomously. Production of these circular episomes has been reported to be accompanied by deletion of the corresponding chromosomal region in some cases but with its retention in others (Beverley et al, 1988). The same mechanism of amplification has been proposed to occur in some cases of mammalian gene amplification, where episome formation and corresponding chromosomal deletion have been observed (see 1.15).

The development of chloroquine-resistance in the parasite *Plasmodium falciparum* has provided a major setback in the efforts to control malaria globally. It was discovered that gene amplification was the cause for this resistance in some isolates of *P. falciparum* (Foote et al, 1989). However, in these cases it was not a target gene that was amplified, but a homologue of the multidrug-resistance (*mdr*) genes found in mammalian cells (see 1.4). These genes encode the P-glycoprotein which is located on the cell surface and actively pumps toxic agents out of the cell. The efflux of chloroquine in these *P. falciparum* isolates was 40-50 times faster than in the wild-type parasite. The mechanism by which the *pmdr* gene was amplified has not yet been elucidated. However, as the degree of amplification was fairly low (3-5 fold) and the amplified units containing the gene appeared to be arranged in tandem on the same chromosome as the original locus (causing an increase in the size of that chromosome), this would suggest that unequal sister chromatid exchange may have occurred (see 1.15).

Amplification of detoxifying esterase genes is now known to be the frequent cause of high levels of resistance in insects that have been exposed to organophosphorus insecticides. This was first shown in mosquitoes, where a 250-fold increase in copy number of the esterase B1 gene was found in resistant strains (Mouches et al, 1986). The mechanism for this amplification is not known, although the amplified esterase gene was found to be flanked by two repetitive sequences found in other parts of the mosquito genome but not near the unamplified gene (Mouches et al, 1990). This implies that complex DNA rearrangements took place prior to or during the amplification process. The same mechanism of organophosphorus insecticide resistance has been shown to occur in certain aphids (Field et al, 1988).

A surprising, and so far unique example of a *de novo* gene amplification occurring in a human family, is thought to have occurred in response to organophosphorus insecticide exposure (Prody et al, 1989). The amplified gene encoded a defective serum butyrylcholinesterase (BtChoEase). The mutant gene in this family renders cells vulnerable to poisoning by such insecticides. Amplification and thus overexpression of the mutant gene, may compensate for the reduced activity of the detoxifying esterase. It

seems that the 100-fold amplification occurred during early oogenesis, spermatogenesis or embryogenesis in one of the parents of the individual, when the gene is most active and a cell might be put under selective pressure by the presence of insecticides in the environment. The amplified alleles in the resulting individual, were also inherited by his son. The amplified copies are thought to be located at or near the normal site for the BtChoEase gene, but no mechanism of amplification was deduced from this study.

1.3 Gene Amplification in the Germ Line of Mammals.

Examples of unprogrammed, high level gene amplification occurring in the germ cells of higher eukaryotes (such as the one just described) would be expected to be rare. This type of gene amplification resembles that found in mammalian somatic cells (see below), where it is only observed in transformed or neoplastic cells. It has been assumed, therefore, that unprogrammed gene amplification is an abnormal event and reflects the deregulated nature of the cells in which it occurs. In neoplastic cells it can lead to drastic genetic changes, such as chromosomal rearrangement and disruption of gene functions. Similar events in germ line cells might be expected to be potentially catastrophic for the development and survival of the organism. However, the last example in Section 1.2 demonstrates that under the appropriate selective pressure, gene amplification can occur in a mammalian germ cell, and that development can still proceed normally after such an event.

There is, in fact, plenty of evolutionary evidence for gene duplications and amplifications in mammals, in the form of numerous multigene families, including the histones, globins, actins etc. However, in many of these cases, the increase in copy number appears to have been a gradual process, occurring over long periods of evolutionary time by mechanisms such as unequal crossover, gene conversion and slippage replication (Maeda and Smithies, 1986). For high level gene amplification to occur in a single generation, the mechanism involved is probably more akin to the overreplication of DNA that is found in transformed cells under acute selective pressure. There may be many more examples of this type of gene amplification

occurring in higher eukaryotes. One recent report described the presence of multiple copies (35) of large, identical DNA stretches (over 400kb in length) in all strains of the mouse, *Mus musculus*, examined (Koide et al,1990). Another example involves an 18 kb sequence which is amplified approximately 30-fold in a region of human chromosome 17 associated with acute promyelocytic leukaemia (Moore et al, 1989). The amplified sequences are clustered and appear to be organized as large imperfect palindromes. They are postulated to have arisen from a basic ancestral copy which underwent a form of gene amplification to expand the locus.

1.4 Gene Amplification in Mammalian Somatic Cells.

There are no examples to date, of gene amplification occurring as a programmed event during mammalian development. However, gene amplification is frequently observed in abnormal mammalian somatic cells in tissue culture or in tumours, and it is thought to be one symptom of the general loss of growth control found in such cells.

Amplified DNA was first identified in mammalian cells during studies on methotrexate resistance in leukaemia cells (Schimke, 1984 for review). It was found that amplification of the gene encoding the target enzyme of methotrexate, dihydrofolate reductase (DHFR), enabled cells to overcome the inhibitory effects of this drug. Numerous other examples of gene amplification as a means of producing resistance to a variety of drugs in several cell lines have subsequently been discovered and are the subjects of intense study (see Stark et al 1990, for references). Some of these involve amplification of the gene encoding the target enzyme of the inhibitor, as in the case of *dhfr* amplification in response to MTX selection. Amplification of the gene encoding the multifunctional CAD protein is the cause of resistance to the antimetabolite, N-(phosphonacetyl)-L-aspartate (PALA) (Wahl et al, 1979), and amplification of the adenosine deaminase (AMPD) gene produces coformycin resistance (Yeung et al,1983). In other cases, resistance in a cell to a variety of seemingly unrelated cytotoxic drugs (such as colchicine, vinblastine or adriamycin) has been found to be due to amplification of the multidrug resistance gene (*mdr-1*), mentioned above (in 1.2) in the context of parasite resistance. This gene encodes the p170 glycoprotein

which functions as an efflux pump of broad specificity in the plasma membrane of resistant cells (Gerlach et al, 1986).

All these forms of drug resistance in cultured cells, have been used in the study of the mechanisms of gene amplification which will be described below. It was initially thought that amplification might be clinically important as a form of resistance to chemotherapeutic agents in cancer patients. For example, *dhfr* gene amplification following treatment with MTX has been found in cancer patients undergoing chemotherapy (Horns et al 1984, Trent et al 1984, Carman et al 1984). However, very few cases of this have been reported, and when analysed the copy number of the gene has only been found to be elevated 2-4 fold. Multi-drug resistance (MDR) is a major cause of clinical drug resistance. However, it appears that *in vivo* resistance is usually mediated at the transcriptional level by enhancing *mdr-1* expression rather than by MDR gene amplification (Shen et al, 1986, Lemontt et al, 1988). One possibility is that very low levels of resistance are sufficient for tumours to become refractory to treatment, and that these can be achieved by mutation at the level of transcription. In addition, amplification might not be favoured *in vivo* as it usually involves gross chromosomal changes. The types of dosage strategies used for the application of these cytotoxic agents *in vivo* are very different to those used *in vitro*, where high levels of resistance are achieved by multiple selection steps and the preferred mechanism for this seems to be gene amplification.

Although amplification appears to be a rare phenomenon in drug-resistant tumours, there are now several examples where cellular proto-oncogenes have been found amplified in advanced tumours (for reviews see Alitalo and Schwab,1987, and Taya et al,1987). As amplification of oncogenes has generally only been found in cells that have taken at least some of the steps towards neoplastic growth, it is thought to play a role in the later stages of tumour progression rather than initiation. Malignant cells display a wide range of chromosomal aberrations, including deletions, inversions, translocations, as well as DMs or ECRs (the cytological manifestations of amplified DNA, see below). Thus, the overall chromosomal instability in such cells can

precipitate a wide range of genetic damage (eg oncogenic overexpression by amplification, or activation/inactivation of oncogenes by translocation). This genetic instability can generate a great diversity of cells within a tumour population, allowing selection of those with a growth advantage to emerge and promote tumour progression (Bishop, 1987). The amplification of specific proto-oncogenes has been correlated with the progression of particular tumours (for references see Alitalo and Schwab, 1987, and Stark et al 1990). A few examples of this include N-*myc*, which has been found amplified in a high proportion of neuroblastomas (in more than 50% of advanced stage neuroblastomas), and in retinoblastomas; c-*myc* in breast cancer; all the *myc* genes in small-cell lung and other cancers; c-*erb*B-1 in gliomas and breast cancer. There are also numerous cases of proto-oncogene amplification occurring as a sporadic event in a diverse range of advanced tumours and tumour-derived cell lines (Alitalo and Schwab, 1987 for examples).

It seems that DNA amplification is a relatively frequent process in neoplastic cells, and reflects their general genetic instability. When a proto-oncogene becomes amplified in a single cell in a tumour population, and its over-expression provides some growth advantage to the cell (eg the evasion of immune response, metastatic phenotype etc), the amplified DNA will be maintained, as the new mutant cell outgrows the rest of the population. The amplified DNA may itself precipitate additional chromosomal rearrangements, which would accelerate neoplastic progression.

1.5 Systems and Strategies Used for the Investigation of Mammalian Gene Amplification.

Studies on the possible mechanisms of mammalian gene amplification have mostly relied on the investigation of the structure and location of the amplified DNA several generations after the primary event occurred. This is due to the fact that in mammals, amplification of a particular locus is a rare, unscheduled event which takes place in a single cell, and is therefore not amenable to analysis. Despite this limitation, a great deal of information on the molecular nature and the cytogenetic manifestations of amplified DNA has been acquired through studies on cultured cells. Many of these studies have

been on drug resistant cell lines. Although such systems have the technical advantage that the amplification of a gene can be reproducibly selected for using defined drug levels, they may involve mechanisms very different to those responsible for the spontaneous amplification of oncogenes or drug-resistance markers in tumours or tumour-derived cell lines. Many of the agents used to select for amplification actually interfere with DNA metabolism (eg. methotrexate, PALA, hydroxyurea), and can thus directly interfere with DNA replication, recombination or repair processes. The malfunction of such processes could obviously be responsible for the amplification which is seen in such cases. In fact many of these drugs are known to increase the frequency with which gene amplification occurs (Tlsty et al 1982, Brown et al 1983). It is not clear, however, what the range of chromosomal damage produced by each of these agents is. Methotrexate induces thymidine deprivation in a cell, one of the consequences of which is the production of double strand breaks in DNA (Ayusawa et al, 1983). Apart from influencing the nature of the amplification mechanism used in the generation of methotrexate resistant cells, such chromosome damage could induce numerous other chromosomal instabilities and rearrangements (Sager et al 1985, Pallavicini et al 1990), which may or may not involve the amplified DNA and can considerably complicate its analysis. Studies of amplified oncogenes that have arisen in cell lines the absence of such chemical agents, are probably more relevant to the mechanisms of gene amplification *in vivo*. However, the absence of an experimental selection for oncogene amplification means that it is not amenable to detailed temporal analysis such as that attempted in drug resistant cells.

Many studies have involved selection for resistance to drugs by the amplification of endogenous genes, for example *dhfr* amplification (Trask and Hamlin,1990), CAD gene amplification (Saito et al, 1989), or AMPD gene amplification (Hyrien et al, 1988), (for further examples see Stark et al, 1990). However, in order to study the effects of chromosomal position on DNA sequence amplification, amplifications involving transfected genes have been used. The investigation of cloned, transfected genes inserted at various chromosomal locations enables the influence of chromosome position on the frequency, stability and cytogenetic outcome of an amplification to be

assessed. In conducting these studies, it was hoped that chromosomal sites which might initiate or resolve the amplification process, such as origins of replication or recombination-prone sequences, could be identified in this way, and that this would give some insight into the underlying mechanisms. Some of these gene transfection/amplification experiments have involved drug resistance genes. For example, *dhfr* transfection into DHFR$^-$ CHO cells selected for MTX^R, (Kaufman et al 1983, Ruiz and Wahl 1990), or CAD gene transfection into CAD$^-$ CHO cells selected for PALA^R, (Wahl et al 1984, Meinkoth et al 1987). Others have involved oncogenes which were fortuitously found to be amplified following their transfection into cultured cells and selection for a transformed phenotype. For example, following transfection of a v-*src* gene fused to a foreign promoter into mouse cells, one of the resulting transformants was found to exhibit unstable amplification (in the form of DMs) of the integrated oncogene (Glanville, 1985). Other transformants, containing the transfected oncogene did not exhibit this phenotype. Similarly, the transfection of a polyoma virus oncogene fused to a mouse enhancer into immortalised rat cells, led to the production of a transformed cell line (3B) containing a stable amplification of the integrated oncogene (Ford et al 1985). The amplified DNA in the 3B cell line is the subject of this thesis.

Two types of approach have been taken for the investigation of gene amplification in mammalian cells using the systems outlined above. One is to try and analyse the earliest possible amplification events by attempting to identify the initial products of a single amplification step. As mentioned above, this approach is extremely difficult technically, because the frequency of amplification of a particular gene, even in a transformed cell, is extremely low. It is thus not yet possible to induce the amplification of a particular gene in all the cells in a population within a single cell cycle.

Another approach has been to analyse the structure and arrangement of the amplified DNA at both a molecular and a cytological level, in the hope that this will provide insight to the mechanisms by which it arose. The findings described below and the models that have been proposed based on them, are mostly derived from the latter type of approach.

1.6 Karyotypic hallmarks of amplification.

Two of the cytogenetic abnormalities which are characterisitic of DNA amplification in mammalian cells, are double minutes (DMs) and extended chromosomal regions (ECRs). DMs are paired, acentric extrachromosomal chromatin bodies which are mitotically unstable, and are gradually lost in the absence of selection. In metaphase spreads that have been Giemsa-banded, ECRs may be detected as regions that either contain no bands (a homogeneously staining region, HSR) or that have an abnormal banding pattern (abnormal banding region, ABR). As ECRs are linked to a centromere, they segregate properly at mitosis and are usually not lost in the absence of selection. Both of these abnormal structures contain amplified DNA. However, not all amplifications are necessarily detectable at the karyotypic level. Intrachromosomal amplified arrays which consist of only a few repeats (tens rather than hundreds) and/or of a small repeat size (100s rather than 1000s of kilobase pairs), may not visibly disrupt the normal banding pattern of a chromosome (Fougere-Deschatrette et al, 1984). Amplified DNA has also been detected by molecular methods on submicroscopic, extrachromosomal elements termed episomes (see below).

The relationship between DMs and ECRs is still not fully clear. Although the amplified DNA can be located on either of these structures in a single population of cells, they are rarely seen simultaneously in the same cell (Cowell, 1982). In tumour cells that have been analysed after biopsy, DMs are the most common form of amplified DNA. The fact that in cell lines established from these tumours, ECRs are eventually seen to predominate, suggests that the extrachromosomal DMs can integrate into chromosomes to form ECRs (Lin et al, 1985). There appears to be a selection for the presence of DMs *in vivo* and against there presence *in vitro*, (for discussion see Wahl, 1989). Although DMs can give rise to ECRs, it seems that ECRs cannot be excised to form DMs. When a cloned tumour cell line containing an ECR is used to re-establish tumours, the resulting cells do not contain DMs (Cherif et al, 1989). Furthermore, ECRs have not been observed to disaggregate into DMs *in vitro* (Carroll et al 1987).

In addition, DMs are commonly seen in mouse, human and rat cell lines but rarely in hamster cell lines, implying that there are also species differences in the preference for one or other form of amplified DNA. It has been proposed that this is due to rapid reintegration of extrachromosomal DNA which may be highly unstable in hamster cells (Carroll et al, 1988).

1.7 Submicroscopic Circular DNA.

Submicroscopic extrachromosomal elements have been found to contain amplified DNA in a variety of different cell lines (reviewed in Wahl 1989). These "episomes" are self-replicating circular molecules, between 100 and 800kb in size. As episomes can replicate autonomously they must contain functional replication origins. Furthermore, their size range is similar to that expected for mammalian replicons (Vogelstein et al, 1980). It has therefore been suggested that a unit of amplification (an episome) may correspond to a unit of replication. The implications of this for possible amplification mechanisms will be discussed below (Section 1.15).

It has been proposed that episomes are the precursors of DMs, and that these extrachromosomal forms may be initial or early products of an amplification event (Wahl 1989). The first molecular evidence for this came from studies by Carroll et al (1988), who found that episomes with an initial size of about 250 kb gradually enlarged over time, eventually forming DMs, implying that these larger circular elements are mitotically more stable. Subsequent studies on several other gene amplification systems have revealed similar findings (Ruiz et al 1989, Saito et al 1989, Von Hoff et al 1988).

It has also been proposed that episomes and DMs may become chromosomally integrated to form ECRs, both in the case of amplified oncogenes in tumour-derived cell lines, as well as for genes mediating drug resistance (Ruiz and Wahl 1990, also for further refs.). This idea is consistent with the observation that ECRs may be located at many chromosomal locations. Extrachromosomal amplification intermediates which integrate at random, but occasionally integrate at the native locus by homologous recombination, would explain many of the observations concerning the location of

amplified DNA (see Section 1.8).

1.8 Location of amplified DNA.

Karyological analysis of cells containing amplified DNA, as well as *in situ* hybridisation on chromosome spreads using probes for the target gene, have been used to investigate the location of amplified DNA. This has revealed that amplified units tend to be clustered together (sometimes in the form of an ECR or ABR), rather than dispersed at multiple locations. However, clusters of amplified DNA are often found at multiple chromosomal locations within a single cell.

In some cases, the amplified DNA is detected on the same chromosome as that on which the unamplified gene is normally found (eg. Stallings et al 1984, Hyrien et al 1988, Trask and Hamlin 1990). This might suggest that intrachromosomal amplification has occurred at the native site of the gene (ie *in loco* amplification). However, in most cases the amplified DNA is not at the actual site of the original locus on the chromosome. In fact, amplified clusters are sometimes detected at multiple sites along the chromosome (eg. Trask and Hamlin 1990). Furthermore, the chromosome (which carries the native locus) is frequently found to be rearranged in some way (Stallings et al, 1984), and the implications of this are discussed below (see Section 1.9).

There are as many, if not more, examples where amplified DNA is found to be relocated, and lies on a completely different chromosome to that of the native locus. This would suggest that transposition of the amplified DNA and / or chromosomal translocation accompany the amplification process. In a number of tumour-derived cell lines containing an amplified c-*myc* gene, the amplified DNA is detected in ECRs which are located on several different chromosomes other than chromosome 8 (which is the normal site of c-*myc* in human cells) after *in vitro* propagation (eg. Lin et al 1985, Cherif et al, 1989). It seems that the extrachromosomal DMs which initially carried the amplified DNA *in vivo*, reintegrated at random chromosomal locations. The presence of amplified DNA at a different location to that from which it arose supports the idea of extrachromosomal intermediates such as episomes or DMs, which can

reintegrate at random chromosomal locations, or occasionally at the site of an unamplified copy of the region by homologous recombination.

The drawback with most of these studies, is that the chromosomal location of the amplified DNA has been investigated after several amplification events have occurred. Thus multiple chromosomal rearrangements may have resulted in the location of the amplified DNA at a particular site, and conclusions on the mechanism of amplification cannot therefore be drawn.

Amplified DNA has frequently been observed at or close to chromosome telomeres (Kaufman et al 1983 and Trask et al 1990). It has been postulated that telomeric sequences may be prone to the integration of exogenous DNA, such as transfected DNA (Kaufman et al, 1983) or extrachromosomal amplified DNA (Ruiz and Wahl, 1990). Recent evidence concerning the chromosomal rearrangements which immediately follow gene amplification events, suggests that this telomeric location is sometimes a consequence of the chromosomal destabilization precipitated by gene amplification. Chromosome disruption may occur during the amplification process at a number of different stages, and result in dicentric chromosomes (as well as other aberrant stuctures), (see Section 1.9). The breakage of a dicentric chromosome containing amplified DNA may be more likely to occur at or close to the amplified region, as it may have an abnormal chromatin structure and/or be in a region prone to breakage and recombination. The amplified DNA would thus be located at or close to the broken end of one or both of the chromosome fragments. Telomere healing of the broken chromosome ends (without further fusion to other chromosomes), would lead to the apparent relocation of amplified sequences at or close to telomeres.

Chromosomal location (or relocation) is probably also an important factor in determining the mechanism of gene amplification. It is known that the chromosomal position of a (transfected) gene can significantly influence the frequency of its amplification (Wahl et al, 1984), producing 100-fold differences in the rate of amplification at different locations.

As will be discussed below, chromosomal rearrangements may lead to structures or

new sites conducive to the amplification of nearby loci. This idea has been confirmed in a number of different systems. For example, Hahn et al (1987) showed that cells containing X-ray-induced chromosome fragments carrying a *dhfr* gene, had a much higher probability of becoming methotrexate resistant by gene amplification. They proposed that this was through the relocation of the gene and the creation of an amplification-prone region around it (although they pointed out that methotrexate itself enhances the frequency of amplification). Trask and Hamlin (1990) found a striking preference for amplification of an endogenous *dhfr* gene present on a rearranged chromosome and they suggest that the rearrangement may have predisposed this chromosome to DNA amplification events.

1.9 Gene amplification associated with chromosomal rearrangement.

Amplified DNA may lie on the same or a different chromosome as that from which it was derived. However, it is clear that in either case the chromosome involved is often found to be rearranged.

There are numerous examples of amplified genes located on translocated chromosomes. In some cases the amplified DNA is at the site of translocation as in one example, where c-*abl* sequences were found amplified 4-8 fold in human leukaemia cells containing a translocated Philadelphia chromosome. The reciprocal translocated chromosome also carried amplified sequences, implying that the translocation event was closely associated with the amplification that occurred at the exchange site (Collins et al 1983). In another report, Biedler et al (1988) studied the location of the amplified DNA in several independently isolated MDR Chinese hamster cell lines, and found that the amplified DNA was always located on a translocated chromosome which was sometimes (but not always) derived from part of the chromosome on which the native locus is normally found.

There are a number of ways in which chromosome translocation might be linked with DNA sequence amplification:

i) A translocation event itself might precipitate amplification, either through the juxtaposition of a gene with a region that favours recombination and amplification, or

through the formation of some structure, such as an inverted duplication, which is prone to amplification (see 1.8).

ii) A chromosome translocation could be one of the consequences of an amplification event. If the initial stages of amplification cause disruption of the chromosome carrying the target gene, the resulting chromosome fragments might be translocated to other locations. If the amplification occurred *in loco*, the site of the amplification might be close to or at the translocation junction. Alternatively, if amplification proceeds via extrachromosomal intermediates, then the excision of an episome containing the gene to be amplified might precipitate chromosome disruption resulting in chromosome fragments with broken ends. Extrachromosomally amplified DNA might then preferentially recombine with these fragments. This would sometimes result in its relocation at or close to the chromosomal site from which it arose, and the chromosome would usually be translocated, as proposed by Biedler et al (1988).

iii) Recombination-prone regions are often the sites of chromosome translocations (Laird et al, 1987 for review). Such regions might also be the preferred integration sites of extrachromosomally amplified DNA.

iv) Reintegration of extrachromosomal, amplified DNA can itself precipitate chromosomal destabilisation, and result in the location of the amplified sequences on rearranged chromosomes, as discussed in Section 1.13.

It has been proposed that chromosome aberrations such as aneuploidy, deletions, inversions and translocations, which are often observed in mammalian cells containing amplified genes, may reflect a common fundamental defect in, for example, DNA recombination and repair mechanisms (for review see Meuth 1989). Consistent with such a hypothesis is the observation that agents which cause chromosome disruption and rearrangement also stimulate gene amplification.

The frequency of amplification of a gene is not significantly affected by mutagens,

such as ethylmethane sulphonate, which produce point mutations, whereas treatment with agents that inhibit DNA synthesis or damage DNA can have profound effects on this frequency. For example, cell-irradiation has been shown to enhance the rate of gene amplification (eg. Tlsty et al, 1985), presumably by causing breaks in DNA which can result in translocation and amplification events such as those mentioned above. Drugs that interfere with DNA metabolism (eg. methotrexate, hydroxyurea) and therefore with DNA replication, repair and recombination, also increase the frequency of chromosome aberrations and gene amplification. For example, thymidine deprivation (by methotrexate) stimulates recombination in bacteria and yeast, presumably because uracil misincorporation followed by excision repair leads to nicks and gaps in the DNA (Ayusawa et al 1983). Many of these agents that interfere with DNA metabolism (eg. MTX, PALA, coformycin) are used in the selection of resistance by gene amplification, as mentioned above. The amplification events studied in these systems may therefore be induced rather than spontaneous, and would not necessarily be expected to reflect the processes that underlie gene amplification *in vivo*.

The integration of exogenous DNA has also been shown to stimulate DNA amplification events in a variety of different systems. The exogenous DNA may be a viral genome, transfected DNA, or extrachromosomal amplified DNA. Chromosome destabilisation that is precipitated by integration events may result in translocations and/or DNA amplification. In the case of viral DNA, EBV infection has been correlated with *myc* gene rearrangement and amplification in Burkitt's Lymphoma cells in culture (Lacy et al, 1987). Integration of hepatitis B virus (HBV) is frequently accompanied by formation of large inverted duplications (Tokino et al 1987, for two examples and for further refs.), which are structures that have been intimately associated with the initial steps of gene amplification (Passananti et al, 1987; see below). There are also several examples (Popescu et al,1989) where human papilloma virus DNA (HPV16) that was transfected into cultured human keratinocytes was found to produce a range of chromosome alterations, including DMs and HSRs, and extensive chromosome fragmentation. Gene transfer studies have also revealed that integrated DNA is

frequently found at or close to sites of chromosomal rearrangement (Robins et al 1980), and can be amplified at high frequency (Wahl et al 1984, Glanville 1985, Ford and Fried 1986, Ruiz and Wahl 1988). The integration of exogenous DNA (of whatever nature) can thus sometimes result in the formation of amplification-prone structures as well as in chromosomal translocations.

The integration of extrachromosomal amplified DNA is thought to be another cause of chromosome destabilization and rearrangement. Recent evidence in support of this has come from the studies of Biedler et al (1988), and from Ruiz and Wahl (1990) who propose that the reintegration of extrachromosomal intermediates of amplification (ie episomes or DMs) can cause chromosome disruption and lead to rearrangements. Aberrant structures such as dicentrics, acentrics and ring chromosomes are often observed in cells where the amplified DNA is still evolving in structure and location (eg. Kaufman et al 1983, Trask and Hamlin, 1990, Ruiz and Wahl 1990). As was originally proposed by McClintock (1951), these structures represent the intermediates of breakage-fusion-bridge (B-F-B) cycles, which are characteristic of instability due to telomeric loss. Cycles of B-F-B are triggered by chromosome disruption, (such as the integration of exogenous DNA), which can generate acentric fragments and chromosomes without telomeres. Fusion between the broken ends of sister chromatids, or of these chromosome fragments with other chromosomes, results in dicentrics or translocations. Dicentric chromosomes will be broken at anaphase, generating two new chromosomes lacking telomeres and further B-F-B cycles can result. Eventual stabilization results from acquisition of telomere function, and this may sometimes be achieved by chromosome translocation.

1.10 Amplification and chromosomal deletion.

There is strong evidence that gene amplification may be linked to chromosomal deletion. However, the presence of other alleles of a locus which can mask the deletion of one allele, can make this difficult to ascertain. Nevertheless, using cell lines containing transfected genes, Carroll et al (1988) and Ruiz and Wahl (1990) provide evidence that the initial event in amplification involves the excision and corresponding

chromosomal deletion of a locus.

Conclusive evidence for chromosomal deletion accompanying gene amplification *in vivo,* has been provided by some elegant experiments conducted by Hunt et al (1990), involving amplified N-*myc* in human neuroblastomas. By making use of somatic cell hybrids, they were able to separate and analyze the two homologues of chromosomes 2, which is the native locus of the N-*myc* gene, from a neuroblastoma cell line containing in vivo amplified N-*myc*. In this way they were able to show that the locus had been deleted from one of the chromosomes. This suggests that excision may be the first step in gene amplification in a naturally occurring system. Thus, it appears that the same recombination events that bring about the excision of a locus, may lead to deletion and allele loss on the one hand, or deletion and allele over-expression through amplification, on the other. Some of the models proposed for mechanisms of gene amplification, incorporate the deletion of the parental locus and the simultaneous formation of a circular episome which can be amplified, and these are described in Section 1.15.

1.11 The molecular structure of amplified DNA.

Karyotypic analysis has established that intrachromosomal amplified DNA exists as arrays of repeats. Estimates of the degree of amplification have also been made, based on cytogenetic measurements, and these vary from a few copies up to thousands (2,500 copies of *dhfr* in one MTX^R cell line, Yeung et al 1983). As a complement to such karyotypic analysis, and in order to investigate the basic organization of amplified DNA, a number of techniques have been used for its molecular characterisation.

One of the approaches used for the isolation and molecular analysis of amplified DNA has been the random cloning of amplified sequences, either from purified DMs and HSRs, or by differential screening of genomic and cDNA libraries, or by in-gel renaturation techniques (for specific examples and other references, see Stark, 1986). Another approach has been to chromosome walk along amplification units (in lambda or cosmid vectors). By using numerous unlinked probes from the amplified DNA of a

particular cell line, estimates have been made for the size of the amplification unit. These range from a few hundred up to as many as 10,000 kb (Giulotto et al,1986).

The large size of amplified units prevents their resolution by standard gel electrophoretic techniques. However, the advent of pulsed field gel electrophoresis (PFGE) has facilitated long range mapping of the units, and more accurate size estimates have now been made in several cases. This technique is also being used to try and define the structural arrangement of the amplified units. The largest units (analysed by PFGE) reported so far are at least 750 kb and 800 kb, in the case of two independent cell lines containing amplified *mdr* genes (Borst et al 1987); and 650 kb in the case of a cell line containing an amplified *dhfr* gene (Looney et al, 1988). Thus, in all mammalian systems analysed it appears that the amount of DNA coamplified with a gene is far larger than the gene itself.

The probes isolated from the amplified DNA in one cell line have been used to investigate independent amplification events of the same locus, in other cell lines. This has enabled the analysis of the structure of the DNA coamplified with various genes such as the CAD (Ardeshir et al 1983), DHFR (Looney et al 1988), and adenylate deaminase (Debatisse et al 1988) genes. It appears that the extent of coamplified DNA varies between independent amplification events, such that the sequences closest to the selected locus are amplified in all cases, but more distant sequences are sometimes not. In fact, different sequences can be amplified to different degrees in the same cell line. This seems to be due to the fact that most cell lines analysed contain amplified DNA produced by more than one amplification event, especially in cell lines undergoing drug selection. A different amount of flanking DNA is coamplified with the target gene at each amplification event, and the resulting units can thus have an extremely complex structure.

Based on such observations, it has been suggested that initial amplification products are large and homogeneous and that shorter, more heterogeneous repeats are formed in subsequent rounds of amplification. This is supported by data from Borst et al (1987), and others (see Stark and Wahl, 1984 for other examples), who have found that the heterogeneity in the size of the amplification units is greatest when the copy number of

the amplified genes is high. The reasons for this progressive trimming of units as the degree of amplification increases, are most probably to do with the stability of the amplified DNA, and this is discussed in Section 1.13.

1.12 Novel joints and the organisation of amplified DNA.

During amplification, DNA sequences from different parts of the genome are brought together to form novel joints. It would obviously be desirable to investigate the novel structures present immediately after an initial amplification event, in order to understand the exact nature of the recombination processes involved, but this can be difficult for technical reasons (as mentioned above). However, sequence analysis of novel joints identified in amplified regions, that were created during the formation and evolution of the amplified DNA, has provided some insight into the possible mechanisms of gene amplification. Some novel joints are formed in the initial step of amplification and are closely implicated in the molecular mechanisms involved. Other amplified novel joints appear to be the result of fortuitous amplification of a rearrangement (the joining of two non-adjacent genomic regions) created during an earlier step, and have no mechanistic significance.

It had been assumed, before the characterisation of any novel joints, that mammalian amplified DNA would be arranged in the form of tandem (head-to-tail) repeats, as is found in bacteria and yeasts. However, the first such junction to be characterized (Ford et al,1985) was found to lie at the centre of an amplified inverted duplication. The amplification in this rat cell line (3B) contained a polyoma oncogene linked to a mouse enhancer element, and the selection for overexpression of the oncogene involved loss of growth control of the cells. The 3B cell line is described in more detail at the end of this chapter. As the cloned inversion joint was found to be present between all the copies of the amplifed DNA, it was deduced that it must have been formed either at the time of, or immediately prior to the amplification event. Following the identification of this unusual structure in the amplified DNA of one cell line, Ford and Fried (1986) then investigated other cell lines containing well-characterized arrays of amplified CAD

genes and *myc* genes for the presence of inverted repeats, as opposed to tandem repeats. To do this, they used the "snapback" technique which enables the selective isolation of inverted repeats from genomic DNA. In each of the cell lines examined, some of the amplified DNA was found to be in the form of inverted repeats. Based on these findings, Passananti et al (1987) proposed that the formation of inverted duplications is a requisite part of amplification.

Subsequently, several other examples of inverted repeat structures in amplified DNA have been reported, involving regions containing the *aprt* (Nalbantoglu and Meuth, 1986), AMPD (Hyrien et al, 1988), CAD (Saito and Stark 1986) and DHFR (Looney and Hamlin 1987, Ma et al 1988, Heartlein and Latt 1989) loci. In each of these cases there is strong evidence that the formation of an inverted duplication is an initial step in the amplification process, and that this structure tends to be preferentially amplified. Nalbantoglu and Meuth (1986) found that simultaneous formation and (3-4 fold) amplification of an inverted joint took place, with a concommitant deletion of the *aprt* structural gene. This amplification occurred in the absence of any selection, and this is consistent with the proposal made by Passananti et al (1987), that formation and amplification of an inverted duplication are mechanistically coupled, and that this is a primary event in the amplification process. In the case of the AMPD locus, the formation and 40-fold amplification of the inverted duplication occurred after primary selection (for coformycin resistance), suggesting once again that formation of this structure was accompanied by its immediate amplification. This inversion joint was then preferentially reamplified 150-fold after secondary selection, where it was thought to be present in most if not all of the amplified DNA in the cell line (Hyrien et al, 1988). In the case of the DHFR locus, inverted repeat structures were found at 85-90% of the novel joints in one amplified gene array (Ma et al, 1988), suggesting that the inverted joints were formed during the early amplification events, and were thus present between most copies of the amplified repeats, even after multiple selection steps. Further support for this comes from the results of Ruiz and Wahl (1988), who isolated a cell line that exhibited high frequency amplification of a transfected *dhfr* gene, and found that the donated sequences, flanked by host cellular DNA sequences, were in the

form of a large inverted duplication. This structure existed in an extrachromosomal episomal form and was amplified under MTX selection. However, even in the absence of selection this structure was found to be sporadically amplified up to 7-fold, implying that it was somehow prone to amplification, in a similar manner to the *aprt* and *ampd* inversion joints described above.

All the amplified inversion joints analysed so far, have the same general structure, consisting of two long palindromic arms (hundreds of kilobases long) which flank a small, non-inverted central region of 200-1000 bp. A number of these joints and their unrearranged counterparts have been sequenced, and their fine structures are all similar. In 3B rat cells, the palindromic arms flank a non-inverted region of less than 1 kb. No unusual sequences or structures, such as transposable elements, were detected in the unrearranged region from which the inversion was derived. The sequence of the two breakpoints suggests that a simple illegitimate recombination event occurred during the formation of the inversion, with no other rearrangements. One of these breakpoints occurred within a region which contained dyad symmetry that could have been used to form hairpin structures, and the other occurred in an AT-rich region, (AT-rich DNA is known to favour recombination). The inversion in the *aprt* system consisted of palindromic arms over 100 kb in length, separated by a 672 bp non-inverted central region. The breakpoints used to form the junction contained no unusual sequences, apart from some short inverted repeats. The structure of the inverted duplication in the AMPD system is the same as the others described, with two palindromic arms of over 120 kb flanking a non-inverted 862 bp region, and the junction was again the product of a simple illegitimate recombination event. However, unlike any of the others cloned so far, this amplified inversion joint was found to be derived from a region containing a recombination hotspot, which had been shown to be involved in the formation of numerous other rearrangements in independent events (Hyrien et al, 1987). This region is AT-rich and contains stretches of Alu-like repeats with dyad symmetry capable of forming significant stem loop structures. Both breakpoints map within one of the potential stem loop structures. Legouy et al (1989) isolated four novel joints from a highly PALA-resistant Syrian hamster cell line, one of which was found to be inverted.

This joint contained a central non-inverted region of 157 bp, and had arisen from a region rich in AT-sequences, as well as imperfect palindromes (26-144 bp long) which might have been capable of hairpin structure formation. Another inversion joint cloned by Latt et al (1986), was found to have a similar structure to those described above, with no sequence peculiarities apart from an AT-rich stretch surrounding the breakpoint. Overall, it appears that all these amplified inverted duplications were initially formed by an illegitimate recombination event, for which there are no obvious sequence or structural requirements, although AT-rich regions do seem to be preferentially involved at the sites of recombination.

Numerous amplified, non-inverted joints have also been detected in amplified arrays. However, many of these are found in cells which have undergone several steps of drug selection, and may be the by-products of multiple amplification events or other rearrangements, as described below. For example, three of the four novel joints isolated by Legouy et al (1989) mentioned above, were straightforward recombinations between two regions of wild-type DNA. The formation of non-inverted novel joints does not appear to be a fundamental part of the mechanism of amplification. Moreover, these joints are sometimes lost in subsequent amplification steps, whilst inverted duplications tend to be preferentially reamplified (eg. Hyrien et al,1988).

1.13 Stability and evolution of amplified DNA.

Analysis of cells isolated as close to primary amplification events as possible, or that have undergone multiple amplification steps, or that have been maintained in culture in the absence of any selection, have been used to investigate the factors that affect the stability of amplified DNA. From such studies, it seems that the early products of amplification have a tendency to be unstable.

The generation of extrachromosomal episomes is thought to be an initial event in many examples of amplification (see below), and such acentric elements can either rapidly evolve under increased selection, or they can be lost in the absence of selection. Unexpectedly, however, it has also been observed that intrachromosomal amplified

arrays detected at the early stages of drug selection, can be lost when selection is removed (Gudkov et al 1987, Saito et al 1989). Stable, amplified CAD gene arrays are only found after further steps of drug selection and amplification. This might be due to a requirement for chromosome rearrangement to occur, so that the amplified DNA can become stably integrated at a suitable chromosomal location. The chromosome disruption which leads to such rearrangement might be induced by the increasing concentrations of PALA used.

Support for the idea that chromosomal relocation of amplified DNA is necessary for its stability, comes from studies on the differences in the kinetics of *dhfr* gene amplification in tumourigenic and non-tumourigenic Chinese hamster cells (Sager et al 1985). It seems that in transformed cell lines, amplified arrays became stabilised at one or a few intrachromosomal locations, on rearranged chromosomes, soon after methotrexate selection. However, in the untransformed, cells numerous karyotypic rearrangements accompany the acquisition of resistance. In addition, the frequency and level of amplification is found to be much lower in these cells. These results are also consistent with the idea proposed in Section 1.8, that relocation of a locus may enable it to undergo more rapid amplification. Thus if strong selective pressure is applied the copy number of the gene can be rapidly increased. Such relocation is more likely to occur in transformed cells where chromosomal rearrangements are a common feature. A striking example of an early amplification or recombination (eg translocation) event resulting in a rearrangement which is unstable and prone to amplification, was described by Debatisse et al (1987). They found that in one case of an AMPD gene amplification in transformed hamster cells, a single copy of a complex rearrangement was produced in a mutant after primary drug selection. This was then found to be preferentially reamplified in all of the independently derived mutants isolated after futher drug selection. This rearrangement somehow promoted the further evolution of the amplified DNA.

Novel joints between inverted amplified repeats (inversion joints) also tend to be preferentially reamplified as mentioned in Section 1.12. These structures are usually found to be stably propagated even after multiple amplification steps, whereas other

amplified novel joints are gradually lost as the molecular structure of the amplified DNA evolves.

Gene transfer studies have revealed that not only are some regions of the genome more prone to amplification than others (eg. for transfected *CAD*, Wahl et al 1984; transfected v-*src*, Glanville 1985; transfected *dhfr*, Ruiz and Wahl 1988) but also that different nuclear environments seem to favour distinct cytogenetic outcomes for amplification (eg Meinkoth et al, 1987). It appears that hamster nuclei favour stable, intrachromosomal amplified arrays, whereas mouse nuclei exhibit both intrachromosmal and extrachromosomal DNA. This is consistent with the observation that DMs are hardly ever seen in hamster cells, although they are often found in transformed mouse, rat and human cells. The currently favoured explanation for this, is that extrachromosomal elements can be produced in hamster cells, but that they tend to reintegrate rapidly. This phenomenon is probably as a result of different growth characteristics in different species (see Ruiz and Wahl 1990, for discussion).

Molecular analysis of the structure and arrangement of amplified units after multiple steps of amplification (using cloning and PFGE techniques), has revealed that a large proportion of the flanking DNA that is initially coamplified with a selected gene is preferentially lost at each subsequent amplification event. In some highly drug resistant cell lines, only a minimal amount of DNA around the selected gene is retained (see Stark et al 1990, for references). One possibile reason for selective loss of co-amplified DNA would be the presence of deleterious sequences within it, such as genes which become toxic when overexpressed. A deletion or rearrangement of sequences in the amplified DNA might reduce the expression of the deleterious gene. On the other hand preferential reamplification of sequences, excluding the gene, might occur.

Another factor that influences the retention or loss of flanking DNA, is the amount of DNA involved at each amplification step. Initially, amplification units can be several hundred or even thousand kilobases long. As the copy number increases at each amplification step, the huge amount of DNA involved might impose too great a burden on a cell's replication functions. There will thus be a counter-selection for smaller units,

through the loss of some previously co-amplified DNA, in later amplification steps (eg. Borst et al 1987, Ma et al 1988, Saito et al 1989). As a consequence of this progressive loss of flanking DNA, numerous rearrangements are usually found within unstable and/or highly amplified arrays. Multiple novel joints are created in such cases and the units are very heterogeneous in size. This can greatly complicate the analysis of the amplified DNA.

In summary, studies on the stability of amplified DNA show that early amplification events lead to simple repeated structures, involving large stretches of DNA. These amplification units are initially quite uniform in structure. However, a number of factors can affect the stability of these early products, such as the form in which they exist (extrachromosomal or intrachromosomal), or their chromosomal location, or the nature of their nuclear environment (both in terms of species and state of transformation). Different outcomes of initial amplification events will influence both the stability of the amplified DNA, as well as the mechanism of subsequent events. Additional factors, such as the nature and extent of the DNA co-amplified with a selected gene, will determine the degree of molecular heterogeneity seen within an amplified array after multiple amplification events.

1.14 Mechanisms of Amplification in Mammalian Cells.

Based on findings such as those described above concerning the nature of amplified DNA in mammalian systems, a number of models for the mechanisms of mammalian gene amplification have been proposed. The wide range of structures and chromosomal manifestations of amplified DNA observed in different cell types, or sometimes for different loci in the same cell, clearly suggest that more than one mechanism may be possible, and also that the same mechanism may have several different outcomes. In fact, there is now considerable evidence in support of the idea that different steps of amplification proceed by different mechanisms. The generation of a particular structure at an early step using one mechanism, may enable another mechanism to act in subsequent steps. The examples of developmentally regulated gene amplification, and

gene amplification produced in response to toxic agents, described earlier, demonstrate the wide range of amplification mechanisms that seem to be used by eukaryotes in nature. It might therefore be expected that transformed and neoplastic mammalian cells, whose inherent genetic instability enables numerous types of gross chromosomal rearrangements to occur, would also employ a variety of different mechanisms to amplify a DNA sequence.

A summary of some of the observations that must be accounted for by any model for amplification is presented:

a) The amount of DNA amplified with a gene is hundreds or even thousands of kilobases long; much larger than the gene under selection.

b) The amplified units are clustered in linear arrays rather than being dispersed as single entities at multiple locations.

c) The multiple amplified copies can either be integrated chromosomally (sometimes detected as ECRs or ABRs), or they can be located extrachromosomally in the form of episomes or DMs. (DMs and ECRs containing the same sequences are rarely found to co-exist in the same cells).

d) Amplification is often associated with the appearance of the amplified DNA at a new chromosomal location. This location is often on a different chromosome to that on which the unamplified allele is normally found. However in some cases the amplified DNA may be on the same chromosome, but is seen at multiple locations, other than the site of the native locus.

e) Intrachromosomal amplified DNA is often detected on rearranged chromosomes and aberrant chromosomal structures such as dicentrics, acentrics, and ring chromosomes. Furthermore amplified DNA is frequently found at or close to sites of chromosomal

translocation, although the temporal connection between amplification and translocation is unclear.

f) Amplification is sometimes found to be associated with excision and chromosomal deletion of the sequences involved. Evidence for this is difficult to obtain, as the deletion of a locus on one homologue may be masked by the unamplified copy on the other homologue(s), as well as all the amplified copies. It has only recently been possible to address this question, by using highly sensitive techniques such as the polymerase chain reaction (PCR) and fluorescence *in situ* hybridisation (for a recent example of this type of thorough analysis, see Hunt et al, 1990).

g) Extrachromosomal episomes carring amplified DNA are often detected early on in amplification. In two cases generation of such episomes has been linked with chromosomal deletion of a transfected gene (Carroll et al 1987, Ruiz and Wahl 1988), suggesting that excision of the sequences in the form of an episome may be an initial event in amplification.

h) Amplified units are frequently found to be arranged as inverted repeats (ie in a head-to-head or tail-to-tail orientation), rather than as tandem repeats (ie in a head-to-tail orientation). The formation and amplification of inverted duplications appear to be mechanistically coupled.

i) The novel joints which lie between such inverted repeats are usually present between all the amplified units. Thus, the ends of the amplification units around inversion joints are precise. However, an entire amplification unit has never been isolated in order to determine whether both the ends of all the units produced by a single amplification event are homogeneous.

j) Different sequences may be co-amplified with a selected gene in different clones derived from the same parental cell line. In addition, the extent of the flanking DNA co-

amplified usually decreases with each subsequent amplification event.

k) Rapid accumulation of amplified units can occur in a single or a few cell generations. Although it is difficult to isolate a single cell immediately after it has undergone amplification, studies on drug resistant cell populations selected in what is presumed to be a single step (see Stark and Wahl 1984, for review) suggest that around 10-20 copies are usually seen at a single chromosomal site.

1.15 Models.

Before much of the information described above concerning the molecular structure and the organisation of amplification units had been obtained, the two most favoured models for gene amplification in mammalian cells were unequal sister chromatid exchange (USCE) and disproportionate replication (the 'onion skin' model). Both of these models had been shown to mediate gene amplification in other organisms, and were therefore extended to try and explain the process in mammals.

Unequal sister chromatid exchange (USCE).

This model (shown in Fig.1.1) is known to be the mechanism of amplification of rRNA genes in Drosophila (Endow and Atwood 1988). In its simplest form, this model predicts that the amplified copies would be located on the same chromosome as the selected gene, and that the gene will be preserved intact at its native location. The size of the amplified units would be the distances between suitable sites of recombination.

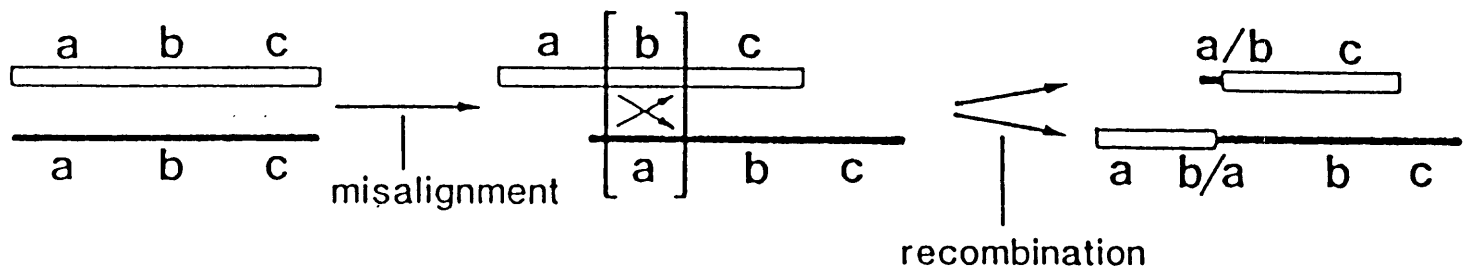
However, USCE only yields a two-fold increase in copy number in a single cell division. In order to generate the longer amplified arrays observed in mammalian cells, several USCE events would have to take place over a number of cell divisions. This is not consistent with the rapid generation of intrachromosomal amplified DNA that is usually observed in mammalian cells. The chromosomal deletion of a locus which is amplified also cannot be explained by this model. Although it was originally suggested that circular extrachromosomal elements such as episomes and DMs might be generated

FIGURE 1.1

Unequal Sister Chromatid Exchange.

The drawing is adapted from Stark and Wahl (1984). Homologous portions on two chromatids of a chromosome are shown, prior to mitosis. Misalignment between the two chromatids can result in an unequal recombination event. This generates one chromatid with a tandem duplication, and the other with a deletion. After mitosis and another round of DNA replication, one cell is deleted for the region between **a** and **b** on this chromosome, whilst the other cell now has two copies of the region on this chromosome. Note that this can only lead to a two fold increase in copy number of the region involved within a single cell cycle.

The initial unequal cross-over event may occur by homologous recombination between tandemly repeated elements, such as satellite DNA sequences, or other short repeats, or, in fact, larger amplification units already present.


a b c
a b c
misalignment
a b c
a b c
a/b c
a b/a b c
recombination

by intrachromosomal homologous recombination, from the amplified arrays produced by USCE (Hamlin et al 1984), this order of events is now thought not to occur (see Section 1.6). Furthermore, this would not explain why in several cases unstable episomes and DMs are seen immediately after the presumed amplification event, but not in later cell generations when the amplified DNA is found to be stably integrated. Thus, USCE does not easily account for the movement of amplified DNA to new chromosomal locations.

At the molecular level, USCE would lead to the formation of multiple heterogeneous novel joints (ie as different recombination events occur for each 2-fold amplification) rather than amplified (homogeneous) novel joints, which are present between all the units and which seem to be preferentially reamplified. Most importantly, USCE predicts that amplified units will be arranged in head-to-tail tandem arrays rather than in the arrays of inverted repeats commonly observed.

Although USCE cannot account for the observations made in many cases, it does fit the data of some instances of mammalian gene amplification, where low copy number intrachromosomal amplification may have led to the generation of an abnormally long chromosome in which the amplified copies are seperated by a long distance (eg Giulotto et al 1986). Trask and Hamlin (1990) used high resolution fluorescence *in situ* hybridisation to reveal that both the native copies of the (DHFR) gene were intact at their original locations, and that in some (but not all) cases, amplified clusters were located on the same chromosome, >50 Mb away from the native locus. These data could mostly be explained by USCE, although this alone will not explain why clusters of units rather than single copy units should be found along the length of the chromosome; and the relocation of the amplified DNA in some cells can only be explained by invoking the formation of dicentrics and other intermediates of breakage-fusion-bridge cycles.

It is worth noting that all of the possible examples of USCE as a mechanism of amplification in mammalian cells, have been in cells selected for resistance to drugs that are known to cause chromosome breaks and promote mitotic crossover.

Disproportionate replication.

As described in Part 1.1, this mechanism is known to occur for chorion gene amplification in Drosophila follicular cells. However, it has already been pointed out that these cells are terminally differentiated and that following the amplification events, they are no longer required to divide. This is an important consideration as the model (shown in Fig.1.2) requires that a single origin of replication is used multiple times during a single cell cycle. Eukaryotic origins of replication are strictly only used once during an S-phase in normal, dividing cells. Although the cells in which gene amplification occurs in mammals are transformed and hence genetically unstable, it seems unlikely that the degree of deregulation required for spontaneous unscheduled rereplication to occur would not be lethal for actively dividing cells, even of a transformed type. The model was originally proposed to occur in mammalian systems by Bullock and Botchan (1982) during the excision and amplification of SV40 DNA integrated into a mammalian chromosome. However during this process the SV40 origin of replication is used and the T-antigen is required.

Nevertheless, disproportionate replication is still frequently proposed as a mechanism of mammalian gene amplification (Schimke 1986), as it predicts the rapid generation of multiple amplified copies in a single generation, and it is flexible enough to account for most types of molecular product. The generation of long tandem arrays by such a mechanism requires that multiple recombination events must occur after rereplication, in order to resolve the onionskin structure. Although the structure could conceivably be resolved into head-to-head tandem arrays, there is no prerequisite for these as opposed to head-to-tail arrays. Furthermore, the multiple recombination events would produce an equal number of different (heterogeneous) novel joints. The only way a highly amplified (homogeneous) novel joint could arise in a single amplification step (as seen by Ford et al 1985, Nalbantoglu and Meuth 1986, Hyrien et al 1988) by this mechanism would be if recombination occurred before replication was complete. However, in this case also there is no prerequisite for such a joint to be inverted, (as found in many cases cited above, in 1.12. Although the onionskin structure could also be resolved to form extrachromosomal episomes, which might then lead to relocation of

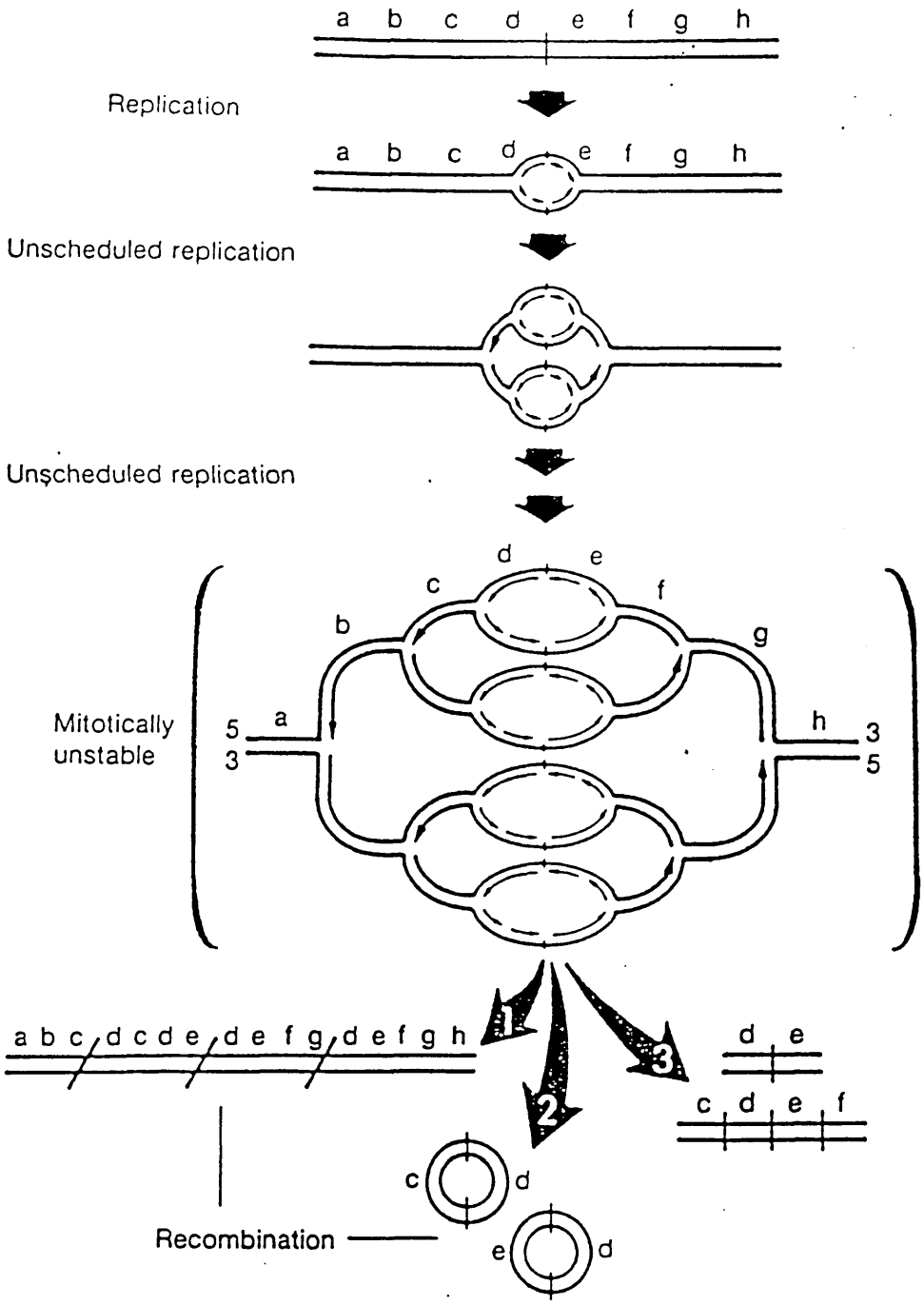
FIGURE 1.2

The Disproportionate Replication Model.

The diagram shown is adapted from Stark and Wahl (1984). Bidirectional replication at an origin (between d and e) generates a bubble that can undergo further rounds of unscheduled replication, resulting in an "onion-skin" of partially replicated duplexes. This relies on multiple reinitiations at the same origin of replication within a single S phase in a cell cycle.

Note that there are two contiguous chromosomal strands. It is postulated that this structure could either be resolved into an intrachromosomal array (pathway 1) by multiple recombination events among the different duplexes; or, that recombination within the same duplex could generate extrachromosomal circles (pathway 2); or, that linear duplex DNA could become detached from the structure if two replication forks approach one another very closely (pathway 3).

Note also that each of the possible outcomes involves multiple recombination events, which will lead to an equivalent number of different novel joints between each pair of repeat units.


a b c d e f g h
Replication
a b c d e f g h
Unscheduled replication
Unscheduled replication
d e
c f
b g
Mitotically unstable
5
3
a
h
3
5
a b c d c d e d e f g d e f g h
1
2
3
d e
c d e f
c d
e d
Recombination

the amplified DNA, it would not predict a chromosomal deletion to be associated with episome formation.

The only example in mammalian cells that seems to fit all the criteria required by this model, including the gradient of decreasing copy number seen with increasing distance from the proposed origin, is one involving the amplification of transfected APRT-TK genes (Roberts and Axel 1982, Roberts et al 1983). However, in this case it appears that the replication origin repeatedly used for the amplification, actually lies within plasmid rather than mammalian genomic sequences. Furthermore the average size of the units (40 - >200kb) was much smaller than that observed in most other instances of mammalian gene amplification, suggesting that this example was exceptional in many ways.

The above models cannot adequately explain some of the recent data concerning mammalian gene amplification. Most significantly, the finding that the formation of inverted duplications seemed to be an obligatory part of the amplification process (Ford and Fried 1986), and that excision of episomal circles (shown to contain inverted duplications in one case, Ruiz and Wahl 1988) accompanied by deletion may be an early event in amplification (Carroll et al 1987), have led to the proposal of a number of different models.

The extrachromosomal double rolling circle model.

This model was proposed by Passananti et al (1987) to explain the formation of inverted as opposed to tandem repeats in mammalian amplified DNA. This had first been noted for the amplified DNA in the 3B rat cell line (Ford et al 1985, see Section 1.17) and subsequently in other cell lines (Ford and Fried 1986). The model (shown in Fig. 1.3) is based on one originally proposed by Futcher (1986) for the amplification of the 2μ plasmid in the yeast *Saccharomyces cerevisiae*. In the Futcher model, high level amplification of the 2μ plasmid can occur in a single S phase without the need for multiple initiations at the replication origin. Furthermore, the plasmid contains an

inverted duplication which is a prerequisite for its amplification. Futcher postulates that following initiation of replication at the single origin on the 2μ plasmid, a homologous recombination event occurs between a newly replicated copy of the inverted sequence and an as yet unreplicated copy, resulting in the inversion of one of the replication forks. The two forks which had previously been proceeding towards each other, now follow each other indefinitely (in the form of a double rolling circle) leading to the generation of a long array of 2μ plasmid copies. Another recombination event which reinverts one of the replication forks, means that replication can terminate when the two forks meet. Single 2μ circles are then excised by homologous recombination from this array. The 2μ plasmid encodes a site-specific recombinase, and this acts within the inverted sequence to mediate the recombination event responsible for flipping the replication forks, which is essential for amplification to occur (Volkert and Broach 1986).

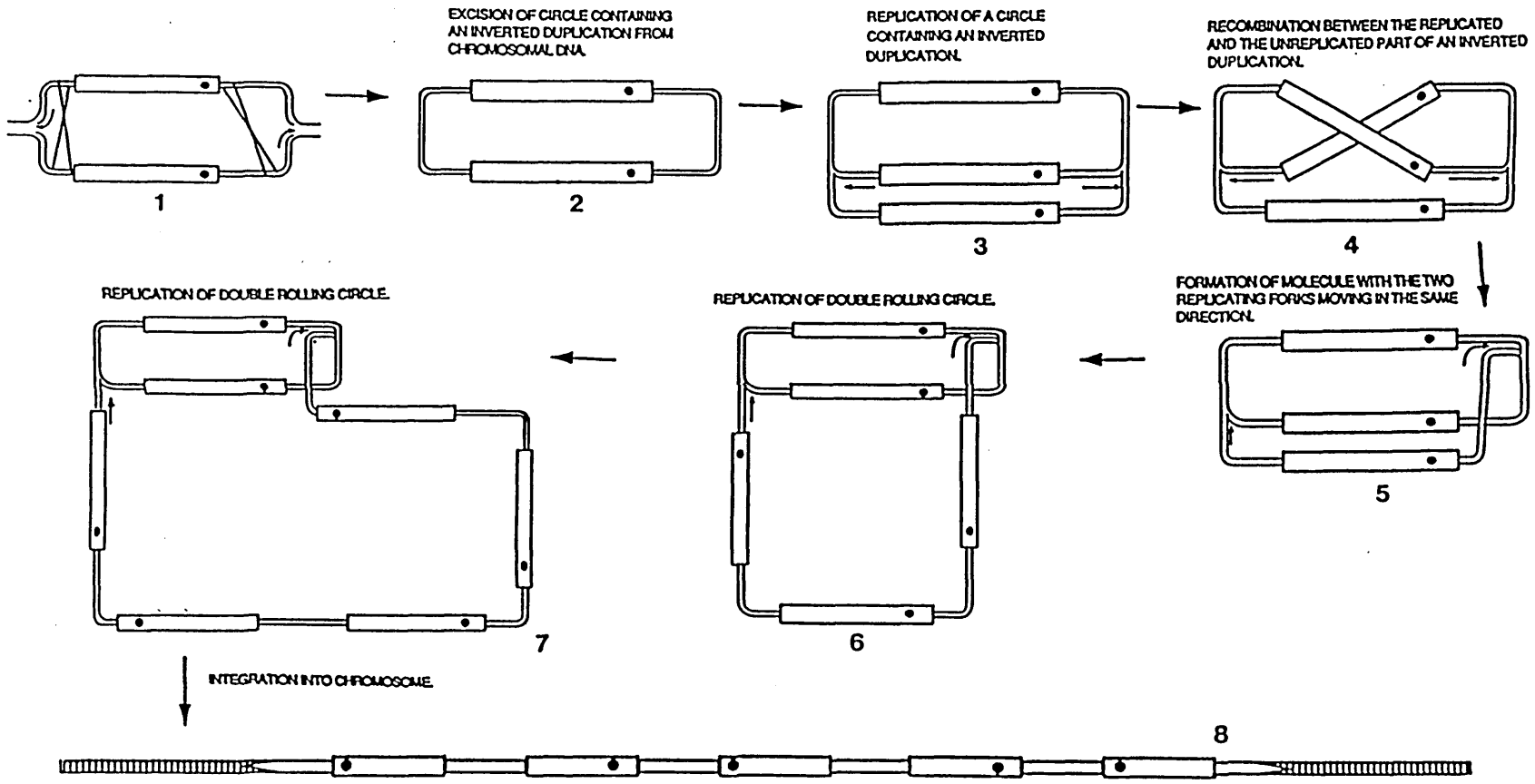
Passananti et al (1986) suggested that a circle containing a functional origin of replication, could be excised from a chromosome in the form of a large inverted duplication by two illegitimate recombination events (see Fig.1.3). The large size of inverted repeats would be necessary to provide an adequate target area for a (non site-specific) homologous recombination event to occur in the absence of a site-specific recombinase. This will enable amplification of a locus to occur in the form of a double rolling circle within a single S phase, without the need for multiple reinitiation at an origin of replication. The result will be an amplified array of large inverted duplications, which can either remain extrachromosomal or become integrated at some chromosomal location.

This model neatly explains the presence of highly amplified novel inversion joints detected in many amplified arrays of both oncogenes and drug resistance genes. It also shows how such structures might be preferentially reamplified in subsequent events, as seen, for example, by Hyrien et al (1988). Furthermore the predicted large sizes of the inverted sequences, are consistent with the size estimates made for various different amplification units. It also links the formation of inverted duplications with the production of episomes, which are thought to be the precursors of DMs (Carroll et al,

FIGURE 1.3

The Extrachromosomal Double Rolling Circle Model.

The drawing is adapted from Passananti et al (1987). This model is based on the Futcher model (Futcher, 1986) for amplification of the 2μ circle of yeast. A circular molecule (2) containing an inverted duplication (rectangle) harbouring the gene to be amplified (dot) is excised from the chromosome possibly by illegitimate recombination between newly replicated DNA strands (1). This excision would result in a deletion of the region at its chromosomal location. Replication is initiated from a sequence within the circle that acts as an origin. Following a single replication of the target gene (3), homologous recombination occurs between a newly replicated copy of the inverted sequence, and the copy of the inverted sequence that has yet to be replicated (4). As a result of this the replication forks flip, so that they are no longer converging, but instead are chasing each other in the form of a double rolling circle (5). Further replication results in the generation of an amplified array of inverted duplications (6,7), derived from a single initiation of replication within a single S phase. The amplified array can either remain extrachromosomal (in the form of episomes and/or DMs) (7), or it could reintegrate into a chromosome (8).


EXCISION OF CIRCLE CONTAINING AN INVERTED DUPLICATION FROM CHROMOSOMAL DNA.
REPLICATION OF A CIRCLE CONTAINING AN INVERTED DUPLICATION.
RECOMBINATION BETWEEN THE REPLICATED AND THE UNREPLICATED PART OF AN INVERTED DUPLICATION.
1
2
3
4
FORMATION OF MOLECULE WITH THE TWO REPLICATING FORKS MOVING IN THE SAME DIRECTION.
REPLICATION OF DOUBLE ROLLING CIRCLE.
REPLICATION OF DOUBLE ROLLING CIRCLE.
5
6
7
INTEGRATION INTO CHROMOSOME.
8

1987). The random integration of the extrachromosomally amplified DNA would account for the presence of amplified sequences at new chromosomal locations. The large size of the units might also be expected to lead to occasional homologous recombination, into a chromosome containing a native unamplified copy of the locus. The model predicts the probable deletion of the actual copy of a locus which has been excised and amplified by this mechanism. Such a deletion, or indeed the chromosomal reintegration of the amplified DNA, could lead to chromosome disruption and the frequent association of amplified DNA with aberrant structures and with translocation as has already been discussed in Section 1.9.

The examples of mammalian gene amplification where the generation of large inverted duplications is coupled with immediate amplification, can all be explained by such a model (eg. Ford and Fried 1986, Saito and Stark 1986, Nalbantoglu and Meuth 1986, Looney and Hamlin 1987, Ruiz and Wahl 1988, Heartlein and Latt 1989). It accounts for both the karyotypic and molecular data of such cases. One prediction of this model is that the other ends of the amplified units will also be precise, and that another amplified novel joint should exist between them. However, this has not so far been tested, as it requires the isolation of a complete amplification unit (which is beyond the size range of most cloning vectors), in order to demonstrate that the novel joints at each end are truly adjacent.

The chromosomal-spiral model.

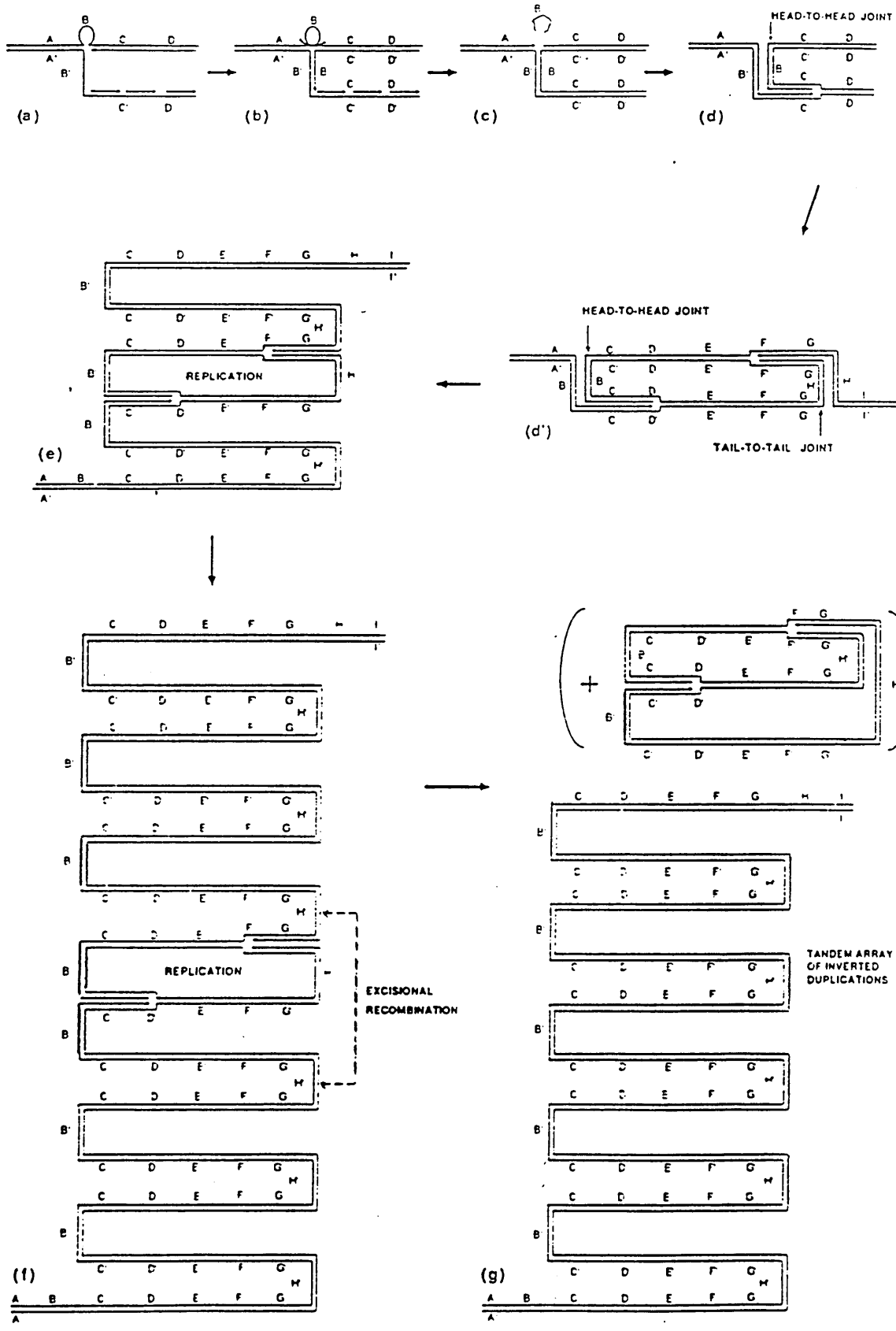
This model was proposed by Hyrien et al (1988) to account for the presence of an amplified array of inverted repeats containing the AMPD gene, at the original chromosomal locus of the gene. The mechanism of amplification (see Fig.1.4) and predictions are very similar to the double rolling circle model, except that here amplification occurs intrachromosomally and does not involve a chromosome break. Hyrien et al (1988) proposed that the presence of a region rich in short palindromic sequences might promote strand switching during replication, which would result in the formation of an inverted duplication with a small loop of single stranded DNA at its

FIGURE 1.4

The Chromosomal Spiral Model.

The drawing is adapted from Hyrien et al (1988). This model accounts for the formation of an inverted duplication and intrachromosomal (*in loco*) amplification. In steps (a) to (d), a head-to-head joint is formed (C'/BC) by copy-choice recombination, which may, for example, be precipitated by the formation and excision of a hairpin loop (see steps a to c). In step (d'), the same process is postulated to occur at the other end of the replication bubble shown, resulting in a tail-to tail novel joint (G/H'G') in addition to the head-to-head joint (C'/BC). In steps (d') to (f), the two replication forks at the two triple junctions engage into the inverted duplicated DNA and follow each other in an indefinite spiral (steps e and f), which results in an amplified array of inverted duplications in the chromosome, without relocation. In step (g) a possible way of generating extrachromosomal circles is shown, by excisional recombination within the amplified array.

Note that, in this model the formation of extrachromosomal amplified elements is not a prerequisite, and the formation of such elements would not lead to deletion of the original locus, unless excision of the whole array were to occur.


HEAD-TO-HEAD JOINT
(a)
(b)
(c)
(d)
HEAD-TO-HEAD JOINT
(d')
TAIL-TO-TAIL JOINT
REPLICATION
(e)
REPLICATION
EXCISIONAL RECOMBINATION
(f)
TANDEM ARRAY OF INVERTED DUPLICATIONS
(g)

centre, if the parental strands were slightly misaligned at the site of switching. Excision of this loop might prime repair replication to occur, so that a fork will proceed around the previously replicated DNA. The model requires that two such events should occur in a single cell cycle, one at each end of a replicon around a single origin, so that the replication forks will follow each other around a "spiral" yielding a tandem array of inverted duplications.

This model accounts for the appearance of highly amplified inversion joints in a single step. It also predicts an identical structure and organization of the amplified DNA as the previous model. In addition, Hyrien et al suggest that episomes might be generated by homologous recombination from the intrachromosomally amplified array, and that these could either lead to DM formation or to the relocation of the amplified DNA. However, chromosomal deletion and simultaneous episome formation are not prerequisites in this model, and these events appear to precede amplification in a number of cases.

Although this model was originally proposed to account for the presence of the amplified AMPD array at the site of its native locus, no other cases of an amplified cluster located at its site of origin have been conclusively demonstrated. It seems that amplified DNA is most frequently relocated at random chromosomal locations, as predicted by the double rolling circle model, or the deletion/episome model described below. The double rolling circle model in fact easily accomodates the results from the AMPD system, when it is considered that the "native site" of the locus from which the initial amplified inversion joint arose already contained multiple copies (7-10) of the region (Debatisse et al 1986). Homologous recombination of the extrachromosomally amplified DNA generated from one of these copies with the other copies remaining at the chromosomal site, could easily be envisioned to occur preferentially over random non-homologous reintegration in this situation.

Deletion and unequal segregation of episomes.

This mechanism focuses on the formation of episomes coincident with chromosomal deletion, as has been shown to occur in some cases of mammalian gene amplification

(Carroll et al 1988, Ruiz and Wahl 1988), as well as in the production of R region episomes in *Leishmania* (Beverley et al 1988). An episome containing an origin of replication, can be excised from its chromosomal location in the form of an inverted duplication by recombination across a replication loop, essentially as described for the first step of the double rolling circle mechanism (see Fig.1.3). However, the formation of an inverted duplication is not then used as the basis for rapid rereplication in a single cell cycle, as is the case for the double rolling circle model. Instead, this model proposes that unequal segregation of these acentric episomes at mitosis leads to amplification, thus avoiding the need for a homologous recombination event to invert the replication forks and lead to a multimeric array of inverted duplications. The model predicts deletion of the chromosomal region contained in the episome by recombination across a replication loop. On the other hand, conservative mechanisms such as disproportionate replication and the chromosome spiral model would predict the preservation of the locus excised.

This model accounts for the early appearance of submicroscopic, autonomously replicating, circular molecules containing amplified genes that are seen in some cases (see Wahl 1990, for review). The sizes of these episomes (usually about 500 kb), are consistent with the size range proposed for eukaryotic replicons (see Section 1.7). It is suggested that these episomes may oligomerize to form larger DMs and/or to reintegrate at random locations and form ECRs. However, it is not clear why arrays of amplified units should always be seen in cells containing chromosomally located amplified genes, whereas single copies of amplified sequences that might be expected to occur by chromosomal integration of monomeric episomes are rarely (if ever) seen. Wahl (1989) and others have proposed that the episomes enlarge in order to become mitotically more stable. Another unsatisfactory aspect of such a model, is that the copy number of the episome carrying the selected gene can only increase over several cell generations, as in the unequal sister chromatid exchange model. The rapid generation of an array of inverted repeats commonly observed, would not be expected to occur via the gradual unequal segregation of episomes.

Models involving chromosome destabilization.

A number of other models involving unequal segregation of chromosomes or chromosome fragments as a method of amplification have been proposed (eg Hahn et al 1987) for drug resistant cell lines where chromosome breaks and rearrangements are common. However, these do not attempt to explain many of the molecular structural and organizational features found in amplified arrays. Another form of chromosomal amplification, involving double strand-breaks and subsequent breakage-bridge-fusion cycles, precipitated by the integration of exogenous DNA was originally proposed by Kaufman et al (1983). Integration of a transfected *dhfr* gene was proposed to precipitate its amplification. The creation of inverted duplications by fusion of broken sister chromatid ends at the site of integration, and subsequent dicentric chromosome breakage at mitosis leading to one chromosome with two copies of the gene, were proposed to be initial events in the gradual amplification by B-F-B cycles. This model was extended by Ruiz and Wahl (1990) to explain the presence of similar B-F-B intermediates containing amplified DNA, in cells in which extrachromosomal episomes with the amplified sequences had previously been detected. They proposed that the chromosomal integration of these episomes precipitated similar chromosome destabilization and amplification events to those seen with transfected and viral DNA molecules. Such chromosomal destabilization would lead to numerous structures which could undergo further recombination and amplification. Again, this range of aberrations has only been seen in mammalian cells subjected to drug selection, and never in neoplastic cells containing amplified proto-oncogenes.

Combinations of mechanisms.

It is likely that amplification proceeds by many different mechanisms in mammalian cells. Early events frequently seem to involve the formation and simultaneous amplification of inverted duplications (eg by the double rolling circle or chromosomal spiral models). These events are often (perhaps always) coupled to the excision of episomes, which may contain inverted duplications as predicted by the double rolling

circle model. The two nonconservative models described (ie the double rolling circle and deletion/episome models) predict that deletion accompanies the initial excision step before amplification. In a recent report of *in vivo* amplified N-*myc* genes in neuroblastoma cells, Hunt et al (1990) showed definitively that excision of an N-*myc* gene had occurred in cells containing DMs carrying the amplified copies of the gene. This strongly suggests that excision is a requisite part of the amplification process. More studies such as these are required, especially in cells isolated as soon after the primary amplification event as possible, to determine whether deletion and episome formation is the most frequent first step of mammalian gene amplification.

Amplification of the primary structures described above would be expected to produce arrays of homogeneous amplification units with precise ends, and amplified novel joints (one at each end of a unit). Recent data from PFGE studies in cell lines containing non-complex amplified DNA support this idea. However, until a complete unit has been characterised, heterogeneously sized units produced by other mechanisms, such as disproportionate replication or USCE cannot be ruled out. These arrays of units may expand extrachromosomally to form unstable DMs, or may form ECRs either by reintegration (double rolling circle and deletion/episome models) or by intrachromosomal amplification (chromosome spiral model). The reintegration of episomes or DMs may lead to destabilization at the site of chromosome integration, which will precipitate the formation of aberrant structures and possible further amplification by recombination mechanisms or by unequal segregation, as suggested by Kaufman et al (1983) and more recently by Ruiz and Wahl (1990).

These early amplification events can sometimes lead to the presence of multiple chromosomal sites containing amplified arrays (following random integration of episomes or DMs), and these may favour different amplification processes. For example, expansion of ECRs has been shown to occur (cytologically) by unequal sister chromatid exchange (Holden et al 1985). Large stretches of tandem repeats (such as tandem arrays of inverted duplications) found in amplified arrays would be expected to greatly increase the probability of USCE. The presence of clusters of amplified units extending along the length of chromosomes (often seen as HSRs) in cells containing

high levels of amplification (eg. Trask and Hamlin 1990) would be consistent with this sequence of events. In addition, inter- and intrachromosomal homologous recombination between amplified DNA present at different locations could also be another way of precipitating chromosome destabilization and further amplification events, as described above and by Ruiz and Wahl (1990).

Every amplification event leads to the formation of novel rearrangements in the genome. These rearrangements may or may not subsequently be amplified. A few of these novel joints are sometimes found to be preferentially reamplified, as discussed in Section 1.12. Furthermore, in cells containing high levels of amplification, multiple recombination events, expansions of some amplified sequences and deletion of others will also occur. These points underline the importance of analysing the structures generated after discrete amplification events, rather than after multiple events, in order to draw any conclusions concerning the mechanism involved at each step.

1.16 Questions to be answered.

In order to verify or eliminate some of the amplification mechanisms described above, and gain a deeper insight into the nature and order of events that take place, certain crucial issues still need to be clarified:

-Are amplification units produced by a single amplification event homogeneous or heterogeneous, ie do the amplified arrays contain amplified novel joints, or multiple different joints?

-Do early amplified arrays consist of only inverted repeats, or a combination of inverted and tandem repeats?

-Is the chromosomal location of the selected gene preserved or deleted, ie are conservative or non conservative processes involved?

-Can amplification occur *in loco* or is movement obligatory?

-If relocation does occur, is it by transposition or translocation, and does the amplification event precede or follow this relocation?

-If it is extrachromosomal, does amplified DNA integrate at single or multiple

chromosomal locations after a single amplification event?

-If amplified DNA is located on a translocated chromosome, does it often reside exactly at the exchange site of a translocation? (ie what, if any, could be the mechanistic association between amplification and translocation?)

-Can a single gene be amplified by different mechanisms depending on the chromosmal region surrounding it? (ie how does the chromosomal environment of a relocated gene influence its amplification?)

-What are the mechanistic differences between gene amplification *in vivo* and *in vitro*?

-Are there fundamental differences in the amplification processes behind oncogene amplification and amplification of drug-resistance genes?

The ability to analyse cells as soon after an amplication event as possible, and the development of techniques for long range molecular analysis and for sensitive karyological analysis should yield some answers to these questions. Long range restriction mapping of amplified units using PFGE, as well as cloning techniques such as chromosome jumping and YACs should enable the structure and organization of amplified units to be accurately determined. High resolution fluorescence *in situ* hybridisation allows the detection of single copy DNA sequences on chromosomes, which had not previously been possible.

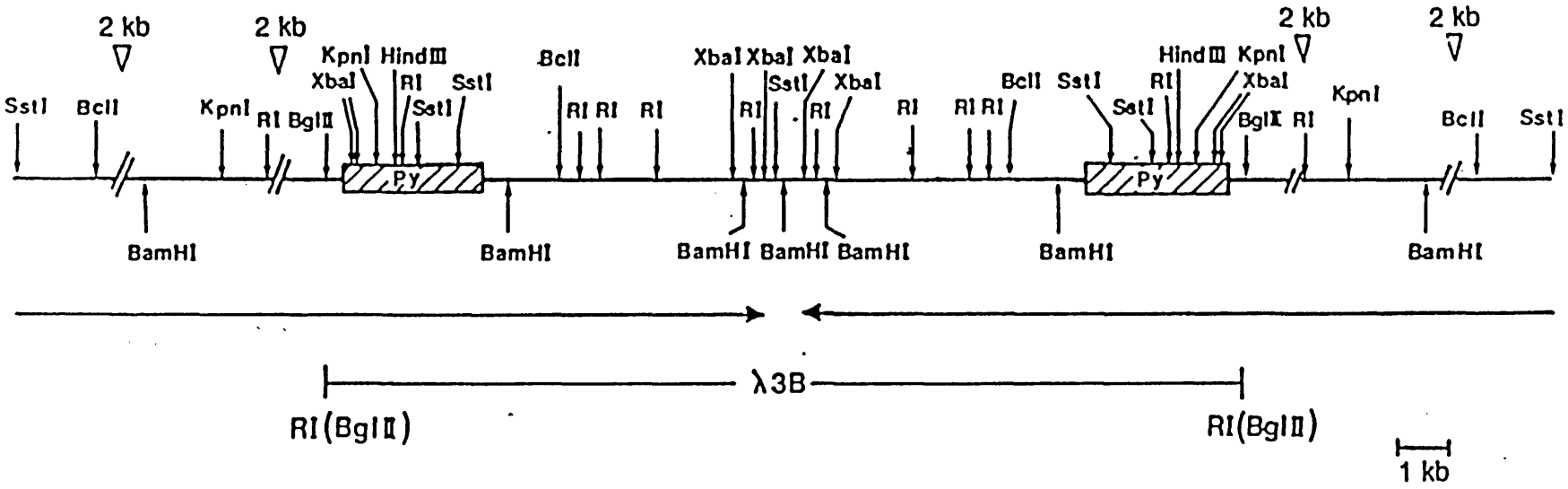
1.17 The Amplified Inverted Duplication in the 3B Rat Cell Line.

The transformed cell line, 3B, was isolated after transfection of Rat-2 cells with a ligated mixture of restricted enhancerless polyoma virus (Py) transforming region DNA, and restricted mouse cellular DNA (Ford et al 1985). This was done as part of the "expression selection strategy" developed in Mike Fried's laboratory (Fried et al 1983) in an attempt to isolate mouse cellular enhancer elements. The presence of such an element would activate the Py oncogene and lead to transformation of the immortalized Rat-2 cells. The transformed 3B cellular clone was found to contain a single type of Py DNA insert, which was molecularly cloned as an 18 kb BglII fragment (λ3B, shown in Fig.1.5). A weak cellular enhancer was identified upstream

FIGURE 1.5

Restriction Map of the Centre of the Inverted Duplication and the λ3B Clone.

A restriction map of the 3B genomic DNA from which the λ3B clone was derived is shown (taken from Ford et al, 1985). The 18 kb BglII insert of λ3B (which symmetrically spans the centre of the inversion) is indicated below the map. The region is in the form of an inverted duplication, with identical palindromic sequences (indicated by the arrows below the map) flanking a central non-inverted region of < 1 kb in size. The 2.8 kb polyoma virus inserts in the palindromic arms are indicated by the hatched boxes (Py). The inverted duplicated sequences were shown to be at least 22 kb long, by restriction analysis and southern blotting using a polyoma virus probe (Ford et al, 1985). The genomic DNA between the Py inserts (ie flanking their 5'ends), was later shown to be of mouse origin, and the genomic sequences flanking the 3' ends of the Py inserts in the inverted duplication, were shown to be of rat origin (Passananti et al, 1987).


2 kb
2 kb
2 kb
2 kb
SstI
BclI
KpnI
RI
BglII
XbaI
HindIII
RI
SstI
Py
BclI
RI RI RI
XbaI XbaI XbaI
RI
SstI
XbaI
RI
RI
RI RI
BclI
SstI
SstI
RI
HindIII KpnI
XbaI
BglII
RI
KpnI
BclI
SstI
Py
BamHI
BamHI
BamHI BamHI BamHI
BamHI
BamHI
λ3B
RI(BglII)
RI(BglII)
1 kb

from the Py sequences. This enhancer and the Py insert were found to be in the form of a large inverted duplication. This structure was also shown to be present in 3B genomic DNA by using a "snapback" technique which selectively enriched for inverted duplicated sequences (Ford et al 1985). Southern blot analysis of restricted genomic DNA from 3B cells and from 7AXT cells (a well-characterized Py-transformed cell line known to contain only a single insert of Py sequences) revealed that the Py and enhancer sequences within the inverted duplication in 3B cells were amplified 20-40 fold. It was suggested by Ford et al (1985) that the amplification of the weakly expressed Py oncogene and cellular enhancer sequences may have been responsible for the strongly transformed phenotype observed in 3B cells. In case the amplification had been induced by the presence of the viral origin in the Py insert it was shown that the insert was in fact inactive for replication (as part of the large T-antigen gene necessary for viral replication, had been deleted).

This was the first report of amplified DNA containing inverted rather than tandem duplications in mammalian cells. In order to verify that this was not a peculiarity of the 3B cell line, Ford and Fried (1986) went on to investigate a number of better characterized amplification events for the presence of inverted repeats in the amplified DNA, again using the snapback technique. Cell lines containing amplified CAD and c-*myc* genes were investigated, and the amplified DNA was found to contain inverted repeats in every case.

In order to understand the processes behind the formation of the inverted duplication and its amplification in 3B cells, its molecular characterization was continued by Passananti et al (1987). As the inversion joint was found to be amplified 10-20 fold its formation must have preceeded the amplification event. Unique DNA probes derived from the central region around the joint (ie upstream of the Py insert) were all found to be of mouse origin. In order to investigate the origins of the cellular DNA downstream of the Py insert, an 11.5 kb BCll fragment (λBclI, shown in Fig. 2.2.1) was cloned from 3B genomic DNA (Passananti et al 1987). Unique probes derived from this clone revealed that the DNA sequences downstream of Py were all of rat origin, and

furthermore that the size and organization of these sequences were identical to that of the parental Rat-2 cell DNA, and extended at least 19.5 kb from the Py DNA. The presence of rat DNA within the inverted duplication showed that the inversion present in 3B cells must have taken place after integration of the transfected DNA into a rat chromosome.

DNA sequencing was used to investigate the nature of the rearrangements that led to the formation of the inversion. The palindromic sequences between the Py inserts in the duplication, were found to be separated by a non-inverted central region of about 1 kb. The amplified inversion joint appeared to have arisen by a simple illegitimate recombination event, which involved two sites, 1 kb apart, in the transfected mouse DNA. Apart from a single nucleotide at the point of inversion, all the other nucleotides could be accounted for by the parental DNA strands involved in the recombination event. The sequences around the inversion joint did not contain any unusual features (such as transposable elements) which could have participated in its formation, apart from some dyad symmetries which might have formed hairpin structures at one of the sites of recombination, and an AT-rich region around the other site of recombination.

The structure of this inversion joint, in relation to others that have subsequently been cloned and analysed, was discussed previously. On the basis of these findings in 3B cells, the double rolling circle model of amplification was proposed (Passananti et al 1987). As described in Section 1.15, this model incorporated the formation of inverted duplications as an obligatory part of the amplification process, as none of the existing models at that time (disproportionate replication, unequal sister chromatid exchange) could adequately explain this association.

The results presented in this thesis extend the molecular and karyotypic analysis of the amplified DNA in the 3B cell line.

Objectives.

As described above, the large inverted duplication found amplified in 3B cells represented the first report of amplified DNA arranged in the form of inverted, as opposed to tandem, repeats. This led to the proposal of an alternative model for gene amplification (the double rolling circle model) which incorporated the formation of inverted repeats as a prerequisite in the amplification process (Passananti et al, 1987). In order to gain more insight in to the mechanism by which the amplification in 3B cells arose and to test the predictions of the double rolling circle and the other models (presented in Part 1.15), a knowlege of the molecular structure of the amplified DNA as well as of its karyotypic manifestation was essential.

At the molecular level, the aim was to try and analyse the structure of a complete amplification unit by chromosome walking along its length, ultimately reaching the other end(s) of the inverted duplication. In this way the homogeneity of the units could be investigated, and the integrity of the DNA in the amplification units in 3B could be compared to the original (native) region in parental Rat-2 cellular DNA, to see whether any deletions or other rearrangements had occurred. The amplified novel joint at one end of the inverted duplication had been cloned and sequenced (Passanati et al, 1987). Further novel joints found in the amplified DNA would be examined for any unusual sequences or structures. The initiation of such a walk using lambda 'phage vectors is described in Chapter 2. The probes isolated with this approach would be useful in the attempts to extend the cloned region (in Chapter 4), for further structural analysis of the amplified DNA at the molecular level (as described in Chapter 3), as well as for a cytogenetic analysis (as described in Chapter 5).

The size of amplification units in mammalian cells is known to be in the order of hundreds of kilobases (as described in Part 1.11), which is well beyond the limit of resolution of standard agarose gel electrophoresis. Therefore pulsed field gel electrophoresis (PFGE) has been used to conduct a long range analysis of the amplified DNA in Chapter 3. The size and homogeneity of the amplification units could be investigated in this way, and it could also be predicted whether the units were all

clustered in an array or dispersed at several locations. The palindromic arms of the amplified inverted duplication were known to be precise (homogeneous) around one inversion joint (cloned and analysed by Ford et al, 1985 and Passananti et al, 1987). If the amplified units are found to be homogeneous and clustered together by this long range PFGE mapping, this would be strong evidence that another amplified novel joint exists between the other ends of adjacent inverted duplications. From the cloning of the other end of the inverted duplication (attempted in Chapters 2 and 4) it was hoped that it could be seen whether the amplified DNA is organised as an array of head-to-head and tail-to tail repeats (as predicted by the double rolling circle model, Passananti et al, 1987). Heterogeneity in the size of the units however, would make the cloning and analysis of the amplified DNA a much more complicated matter. The best approach for the molecular cloning of the amplified DNA in 3B cells was also investigated by this long range analysis. The large size of the amplification units and the paucity of rare cutter enzyme sites within the cloned region of the inverted duplication, which became apparent during the PFGE analysis in Chapter 3, led to the choice of yeast artificial chromosomes (YACs) as the most appropriate cloning vectors in the case of the 3B amplification (Chapter 4).

In Chapter 5, a cytogenetic analysis of the chromosomal origins and manifestation of the 3B amplification is described. The highly sensitive technique of fluorescence *in situ* hybridisation has been used, which allows the efficient detection of single copy sequences on metaphase chromosome spreads. In this way, the amplified DNA has been located and furthermore, the region from which the amplified DNA was derived has been identified in the parental cell line, Rat-2. It has been determined whether the amplified DNA exists intra- or extrachromosomally, and whether it is located in a single place or scattered at various locations. This cytogenetic analysis provides an essential complement to the molecular analysis of amplified DNA, as it can confirm whether or not amplification units are clustered together, and therefore the likelihood of another amplified novel joint being present, between the tandem inverted duplications.

The karyotypic consequences of the 3B amplification event have been investigated in Chapter 5, by comparison of 3B and Rat-2 chromosome karyotypes. In addition, the

questions are addressed of whether the amplification in 3B cells proceeded via an extrachromosomal intermediate (ie transposition) and whether the amplification event was accompanied by deletion of the original locus, or whether it occurred intrachromosomally; and also the association between chromosome translocation and amplification.

CHAPTER 2

2.1 INTRODUCTION.

In order to examine the cellular sequences downstream of Py, an 11.5 kb BCll fragment (see Fig. 2.2.1) had been cloned in λBCll (Passananti et al, 1987) and probes derived from the region downstream of Py were all found to be of rat origin. Southern blot analysis using these probes showed that the size and organization of restriction fragments in the region were identical in 3B and Rat-2 cellular DNA (Passananti et al, 1987), as described in Section 1.15.

This chapter describes a chromosome walk conducted in order to extend the cloned region of the rat DNA in the inverted duplication in 3B cells. Chromosome walking involves the use of probes from cloned fragments to screen a library, in order to obtain overlapping clones from which probes for further steps can be derived. Both directional cloning (ie the cloning of a restriction fragment of suitable length which has been identified by Southern blot analysis) and cloning of overlapping fragments of partially digested genomic DNA, were used to extend the map of the inverted duplication, and to isolate unique DNA fragments for use as probes.

RESULTS:

2.2 Isolation of a 15 kb BamHI rat DNA clone from the inverted duplication.

A 15 kb BamHI fragment was detected on a Southern blot of 3B DNA using a unique 2.5 kb BglII probe which had been subcloned from λBCll (Passananti et al,1987), as shown in Fig.2.2.1. This 15 kb BamHI fragment would extend the cloned region of rat DNA by 12 kb, and therefore its cloning was used as the first step of the chromosome walk. A genomic library enriched for this fragment was constructed (as described in Materials and Methods) using 3B genomic DNA digested with BamHI which had been size-fractionated by sedimentation in a sucrose density gradient. The fraction containing DNA fragments of 15 kb was cloned in the lambda 'phage vector EMBL4 (Frischauf et al, 1983). There were approximately 60 positive signals when a library of 120-150,000

FIGURE 2.2.1

Restriction map of the cloned region of the inverted duplication in 3B cells and a summary of the lambda clones isolated.

The map in the upper part of the figure represents the extent of the cloned region of the inverted duplication (76 kb) in 3B genomic DNA. The 18 kb BglII insert of the λ3B clone isloated by Ford et al (1985) is shown. Hatched boxes represent the 2.8 kb polyoma virus (Py) DNA inserts. The arrows above the map indicate the two palindromic arms of the inverted duplication.

Presented below is a detailed restriction map of one arm (38 kb) of the inverted duplication, at twice the size scale of the map above. Solid bars above this map represent unique sequence DNA probes and the checked bar represents a repetitive sequence DNA probe. The inserts of the different lambda clones (described in Parts 2.2 and 2.3) isolated for the construction of this map, are shown below it.

MAP OF CLONED REGION OF INVERTED DUPLICATION IN 3B CELLS.

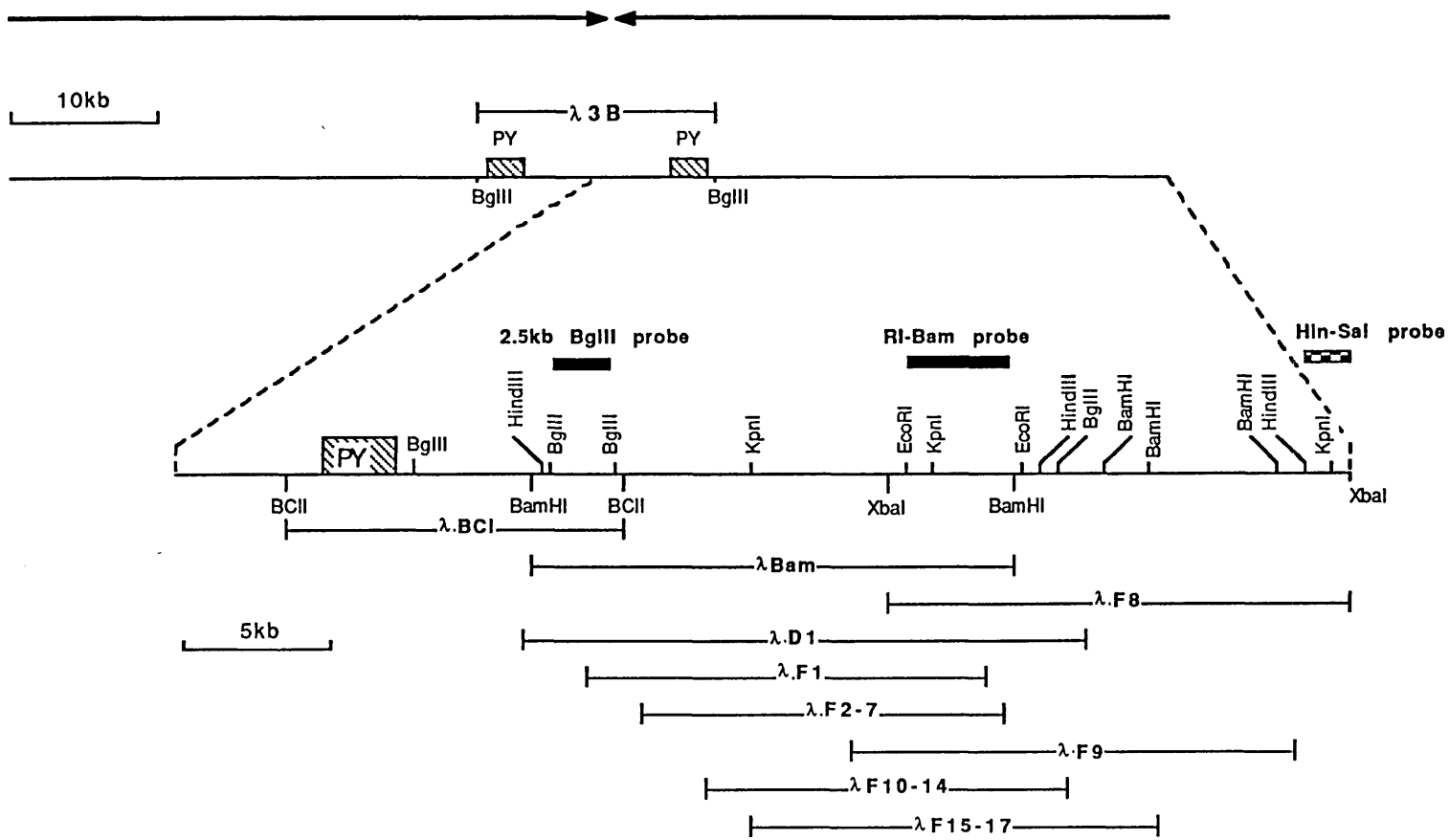
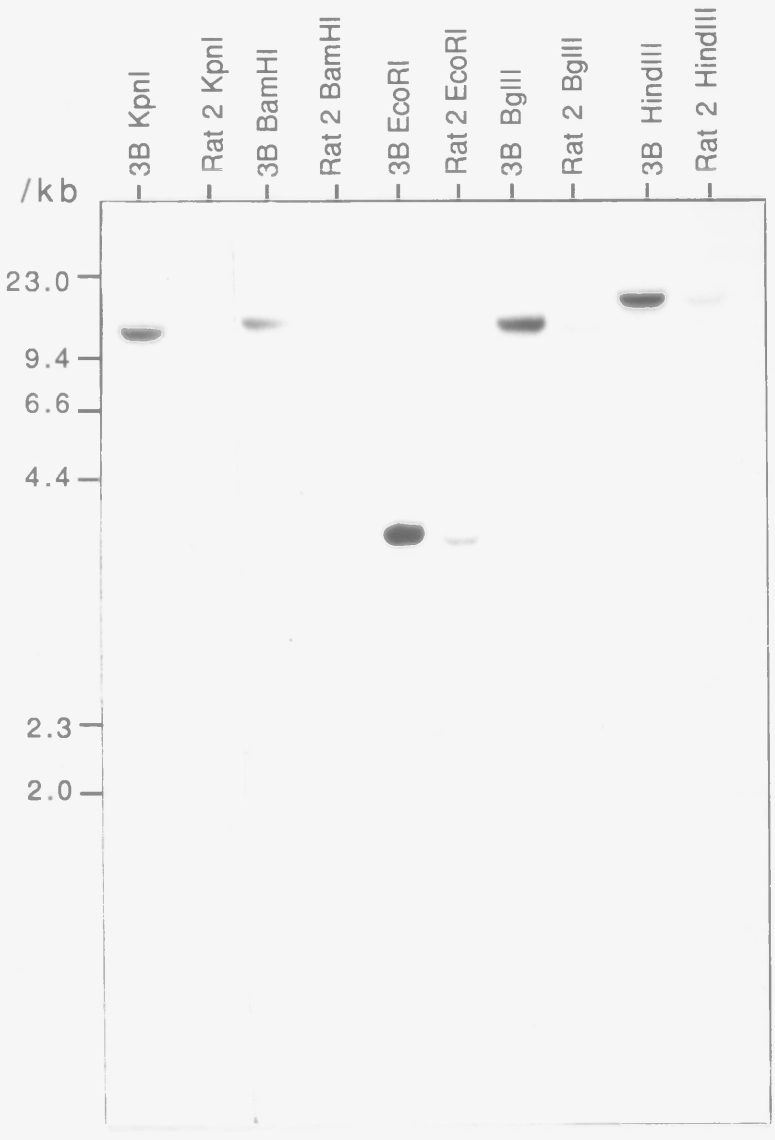


FIGURE 2.2.2

Southern blotting analysis of 3B and Rat-2 genomic DNA using the 3.9 kb RI-Bam rat DNA probe.

The Southern blot presented is of rat genomic DNA from either 3B or Rat-2 cells, digested with a variety of restriction enzymes, as indicated above each lane. All the tracks contained an equal amount (10 μg) of genomic DNA. The probe hybridised to this blot was the 3.9 kb RI-Bam fragment derived from the rat DNA within the inverted duplication in 3B cells, and shown in Fig. 2.2.1. Note the increased intensity of the (amplified) 3B bands when compared to the corresponding Rat-2 bands. The sizes indicated in kilobases, were derived from lambda HindIII markers.



3.9kb RI-Bam Rat probe

plaques was screened with the 2.5 kb BglII probe, which was about the number expected when taking into account the 20 fold enrichment of size-fractionated BamHI fragments, together with the 20-40 fold amplification of this region in 3B genomic DNA. Twelve of these positives were plaque purified and found to contain identical inserts of 15 kb. A partial restriction map of this clone is shown in the map of the inverted duplication in Fig.2.2.1. Hybridisation of a southern blot of the 15 kb insert of λBam cut with a variety of restriction enzymes with radiolabelled total rat cellular DNA was used to reveal any fragments containing highly repetitive DNA. It was thus possible to identify a number of unique fragments which could potentially be used as probes for further cloning, and also to construct a restriction map of the newly cloned region (Fig.2.2.1). End-labelling of the BamHI digested lambda clone DNA, followed by EcoRI digestion, revealed a unique 3.9 kb EcoRI-BamHI fragment (RI-Bam) which was located at the end of the clone farthest from the 2.5 kb BglII probe. This RI-Bam fragment was therefore an ideal candidate for use as a probe to extend the walk and it was subcloned in the plasmid Pxf3.

When the RI-Bam probe was hybridised with a southern blot of genomic DNA from 3B and Rat-2 cells cut with a variety of restriction enzymes, the bands detected were of identical sizes (Fig. 2.2.2). The intensity of the bands in 3B was consistent with a 20-40 fold amplification of this region compared to Rat-2 DNA. These results indicated that there were no rearrangements of the rat DNA in the amplification within the new region cloned and that this region was amplified to the same extent as the rest of the cloned duplication.

2.3 Isolation of further overlapping rat Sau3A DNA clones from the inverted duplication.

As seen in Fig. 2.2.2, the RI-Bam probe did not detect any large fragments which would have been convenient for directional cloning. A 12 kb KpnI fragment (one end lying within the probe region) was the only fragment which extended the map by a useful distance for cloning (see Fig.2.2.1). However, it was decided to construct a library of overlapping fragments from partially digested 3B DNA in order to maximise

the size of the next step of the chromosome walk. Such a library could also be used indefinitely for any subsequent steps that could not be taken by directional cloning

Two libraries of overlapping Sau3A fragments were constructed using 3B DNA partially cut with Sau3A, and the lambda vectors λFIX and λDASH (Stratagene). The λDASH vector had BamHI ends whilst the λFIX vector contained XhoI ends which needed to be filled in to generate Sau3A compatible overhangs (as described under Materials and Methods). In both cases the partial conditions were chosen such that they would generate fragments of an average size over 20 kb (the maximum cloning capacity of both vectors being 22 kb).

In the case of the λDASH library the partially cut Sau3A fragments were size selected on a sucrose density gradient, and only the fractions containing fragments of average size over 20 kb were used for cloning into the vector. This was done as an additional measure to maximise the size of the clones. Part of this library was plated out and 80,000 plaques screened using the RI-Bam probe. Only two positive plaques were detected in total, and these were picked for plaque purification. One of these (λD1) had an insert of 18kb, but it was also found to hybridise to the 2.5 kb BglII probe (from λBcl, see Fig.2.1) and contained the complete BamHI fragment cloned previously (λBam). A partial restriction map of λD1 can be seen in Fig.2.2.1 which shows that it only extended the cloned region by 2.5 kb. The second positive clone (λD2) had an insert of 16kb and seemed to contain an extensive amount of new sequence beyond the RI-Bam probe fragment. However, although λD2 contained the 3.9kb RI-BamHI fragment and immediately adjacent restriction fragments intact, the rest of the restriction fragments (on both sides of the probe) did not correspond at all to the map of the region expected from southern blotting. Thus, λD2 appeared to be highly rearranged and although it was attempted to salvage a fragment for use as a probe (if it could be identified on the predicted map of 3B by southern analysis) most of the new sequences in the insert were found to be highly repetitive. Therefore, λD2 could not be used to extend the map of the inverted duplication.

In the case of the λFIX library, 420,000 plaques in total were screened with the RI-

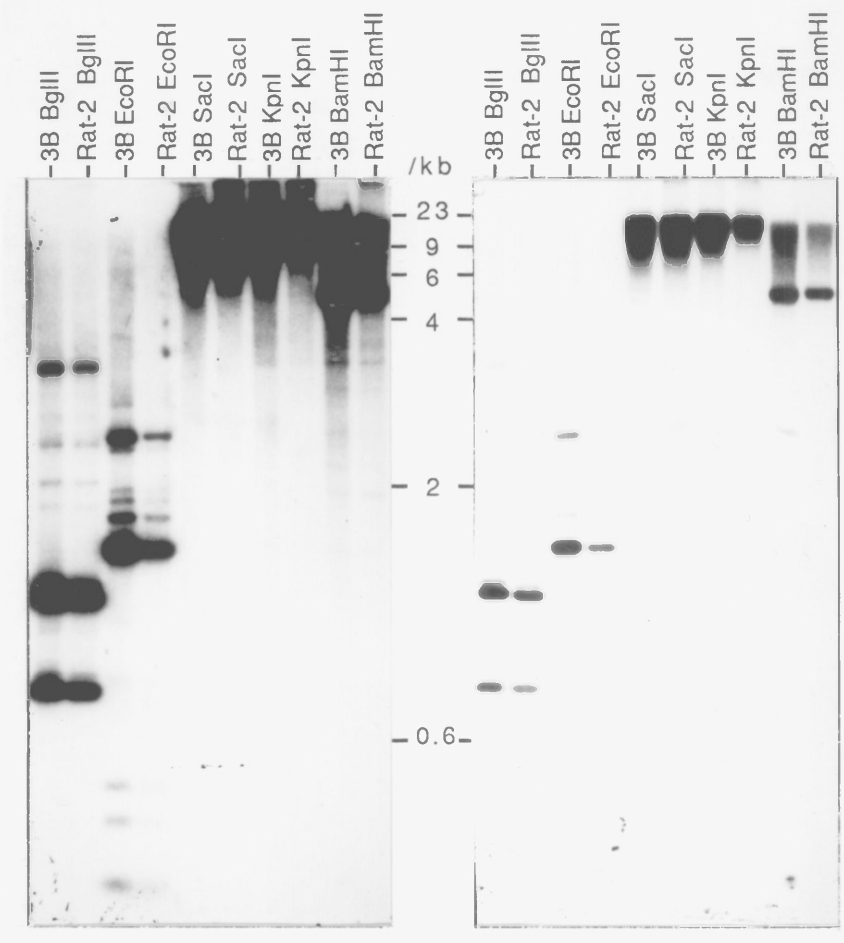
FIGURE 2.2.3

Southern blotting analysis of 3B and Rat-2 genomic DNA using a 0.8 kb KpnI-SalI rat DNA probe.

The Southern blot shown on both sides of the figure, is of 3B and Rat-2 genomic DNA cut with a variety of restriction enzymes, as indicated above each lane. The probe used was a 0.8 kb KpnI-SalI rat DNA fragment, derived from the 1.5 kb Hin-Sal fragment whose position within the inverted duplication is shown in Fig. 2.2.1. On the left hand side, the probe was pre-annealed with rat competitor DNA under the conditions described in Materials and Methods, prior to hybridisation. On the right hand side the probe was instead pre-annealed with four fold more competitor DNA prior to hybridisation. Each track contained approximately equal amounts (5 μg) of 3B or Rat-2 genomic DNA. Note that the intensity of bands is approximately the same for 3B and Rat-2 DNA cut with a particular enzyme. This differs from the situation shown in Fig. 2.2.2, where the unique RI-Bam probe detected amplified DNA sequences in 3B, but not Rat-2, cells. Furthermore, the bands detected in the blot on the right hand side correspond to satellite DNA bands, which were also seen on the ethidium bromide stained gel.

The sizes indicated were deduced from lambda HindIII size markers.

The bands on the right hand blot shown here correspond to the rat satellite bands detected on the ethidium bromide stained gel from which the blot was derived. The KpnI-SalI probe detects only repetitive rat DNA sequences.



Probe pre-annealed with
1 x rat competitor

Probe pre-annealed with
4 x rat competitor

Bam probe and 30 positives were detected, of which 17 were picked and plaque purified. The average size of the inserts was only about 12 kb despite the choice of partial conditions, although in this case the genomic DNA had not been size fractionated. Partial restriction maps of these clones (λF1-17) are shown in Fig. 2.2.1. Out of these seventeen clones, fifteen were found to extend the mapped region by less than 5 kb and only two, λF8 and λF9, extended the map a sufficient distance to appear useful. λF8 contained the largest insert (15.5 kb) and extended the cloned region by the longest distance (10.5 kb) including the Kpn I site detected by the RI-Bam probe on the Southern blot of 3B and Rat-2 DNA (Fig. 2.2.2). Therefore λF8 was analysed further for the presence of suitable (unique) probe fragments, as previously described for the isolation of the RI-Bam probe from λBam (Part 2.2). However, the 10.5 kb of new sequence appeared to consist largely of repetitive DNA, as all the fragments generated by a variety of restriction enzymes hybridised strongly with total rat DNA. Nevertheless, the terminal 1.5 kb HindIII-SalI fragment (furthest away from the probe) was subcloned in Pxf3 (Hin-Sal probe, see Fig.2.2.1). When this fragment (isolated from the Pxf3 clone) was used as a probe on a genomic southern blot it was found to contain highly repetitive sequences which were impossible to compete away totally (see Materials and Methods) even when four times the suggested amount of sonicated total rat DNA was used in the pre-annealing step. Even smaller fragments of this Hin-Sal plasmid clone were isolated in an attempt to find a non-repetitive probe. A 0.8 kb KpnI-SalI fragment at the furthest end of λF8 appeared to be non-repetitive, when the 1.5 kb Hin-Sal was restricted and probed with total rat genomic DNA after Southern blotting. However, when this 0.8 kb KpnI-SalI fragment was used as a probe on genomic DNA, it was also found to be repetitive (see Fig. 2.3.1). Again, even when excess competitor rat DNA was used, only repetitive DNA fragments hybridised to it. These repetitive bands corresponded to satellite rat DNA bands, and could also be seen on the ethidium bromide stained gel.

2.4 DISCUSSION.

The procedure of chromosome walking in lambda vectors described in this chapter

can be fairly time-consuming. It can only proceed in steps of less (and often much less) than the average length of an insert, as was found with the λFIX and λDASH libraries screened. However, this problem can be avoided if a suitably large restriction fragment is found in the region which can be cloned after enrichment by size fractionation, as for the cloning of λBam. Other inherent problems in chromosome walking include the possibility of encountering regions of the genome which are almost impossible to cross, either because of long stretches of repetitive sequences or because of the presence of a mammalian sequence which may be difficult to clone in *E.coli* (Wyman et al,1985).

The chromosome walk along the inverted duplication in 3B was undertaken despite the possible difficulties for two main reasons: Firstly, there was no alternative method for cloning large stretches of DNA available at that time, and without any knowledge of the size and structure of this amplification unit it would not have been possible to try and elucidate the process by which this amplification arose. Secondly, the region being cloned had the advantage that it was amplified 20-40 fold which meant that the 3B genomic libraries constructed should be highly enriched for these sequences, thus increasing the probability of being able to take a large step at each round of screening with a new probe. However this did not avoid the problems of highly repetitive DNA (short sequences of less than 1kb which are highly abundent in mammalian genomes) and unclonable regions.

The first step of the chromosome walk which consisted of the cloning of λBam was straightforward, and yielded a unique probe (the 3.9kb RI-Bam probe) which was useful not only for taking the next step of the walk, but also in the later analysis of 3B, (see Chapters 3,4 and 5).

When the λFIX and λDASH libraries of overlapping fragments were constructed the steps became more problematic. In the case of the λDASH library two positives out of a total of 80,000 plaques screened were detected. This was approximately the number expected given that the average size of DNA fragments in this library was over 15 kb and taking the 20-40 fold amplification of this region into account. However, of the

two positives, one was highly rearranged, while the other mainly covered sequences already cloned and only provided 2.5 kb of new sequence. This suggested that the region beyond the RI-Bam probe might be difficult to clone. If certain sequences within this region could not be tolerated by *E.coli*, there would be a selection against any Lambda recombinants containing these sequences, making extension of the cloned region difficult. This suspicion was reinforced when the λFIX library was screened. The number of positives obtained in this case (30 out of 420,000 plaques screened) was close to that expected. However in this case the library had not been strongly size selected for DNA over 20kb (as was the case with λDASH) so that the average insert size was only 12-15 kb, and fifteen out of the seventeen positives analysed did not extend beyond about 5 kb into the uncloned region. Again this suggested that the region might be difficult to clone, but two positives were eventually isolated which did extend for more than 5 kb, and did not appear to be rearranged as far as could be detected by restriction analysis. The region was therefore not impossible to clone, although its presence in lambda seemed to provide some instability. This region (about 10 kb beyond the probe) was found to be full of highly repetitive sequences, making it impossible to isolate further probes which could be used to continue the chromosome walk.

At the time that the chromosome walk described in this chapter was initiated, the size of the amplification unit was unknown. It was subsequently discovered (as described in Chapter 3) that its size was actually hundreds, rather than tens, of kilobases long. It appears that in mammalian gene amplification, although the amplified gene may be small (in this case the Py gene is 2.8 kb), the amplification units themselves are hundreds and even thousands of kilobases in length, and the degree of amplification can be many hundred-fold. In the case of 3B, the pulsed field gel electrophoresis analysis described in Chapter 3 revealed that the size of each side of the inverted duplication was in the order of 400-450 kb. It therefore became clear that trying to clone this region and reach the other end of the inverted duplication using lambda vectors would be futile, especially in the light of the difficulties in chromosome walking which were encountered during the short walk described here. Alternative strategies

therefore had to be considered to try and overcome both the problems of the size and the unusual structure of the region to be cloned (an inverted duplication). Initial attempts using cosmids (which have a capacity of approximately 40kb) were made but when the large size of the amplification unit was discovered, it was felt that other approaches for the analysis and cloning of large regions of DNA should be investigated. Also, none of the problems found with cloning in lambda (such as possible unclonable mammalian sequences in a bacterial host, and repetitive regions) were avoided by the use of cosmids. The particular approaches chosen for the analysis of the amplified DNA in 3B cells are described in Chapters 3 and 4.

The work described in this chapter extended the map of one side of the inverted duplication in 3B to 38kb away from its centre, (the end of λF8) - ie the map of the complete inversion was already 76kb. This mapping yielded some useful rat DNA probes: the unique RI-Bam probe which was used to show that the amplified DNA in 3B was still intact when compared to its parental line, Rat-2, and which was later useful in the long range mapping of the amplified DNA, and the λBam clone which was used as a probe for *in situ* hybridisation onto 3B and Rat-2 chromosomes, as described in Chapter 5. This work also illustrated the various difficulties found with chromosome walking, which resulted in the need to adopt alternative methods that would be more suited to the nature of the particular project.

CHAPTER 3

3.1 INTRODUCTION.

The results of the previous chapter extended the mapped region of the inverted duplication in 3B to 76 kb. This size was already well beyond the limit of resolution of standard gel electrophoresis techniques. Fortunately at that time the technique of pulsed field gel electrophoresis (PFGE), initially described by Schwartz and Cantor (1984), and later modified by Carle et al (1986) and Chu et al (1986) had begun to be widely used for long range DNA analysis. PFGE allows the separation of DNA fragments up to several thousand kilobase pairs (for review see Anand, 1986) when used in conjunction with restriction enzymes which cut genomic DNA infrequently. Enzymes which cut the mammalian genome infrequently either have long recognition sites (8bp for NotI and SfiI) and/or contain one or more CpG dinucleotides in their recognition sequence. This dinucleotide is underrepresented in mammalian DNA (Razin and Riggs, 1980) and is also frequently found to be methylated at the cytosine residue (Brown et al, 1986). Hence such enzymes generate fragments of mammalian DNA which are several hundred or thousand kilobases in length, and can be resolved by PFGE. This technique can thus be of great use in defining the topography of large stretches of genomic DNA, such as those involved in mammalian gene amplification.

Long range analysis of the amplified DNA in 3B was undertaken with a number of specific objectives in mind. Using a variety of enzymes and different size resolution conditions, the aim was to try and predict the size of the amplification units, and in addition to try and determine whether the units were homogeneous or heterogeneous in size. The DNA from 3B cells and their parental Rat-2 cells could also be compared for any rearrangements in the amplified DNA. Furthermore, using PFGE under very high molecular weight resolution conditions it would be possible to determine whether the amplified units were clustered together as an array, or whether they were dispersed at a number of different locations. This information would be extremely useful in the understanding of the mechanism by which the amplification arose.

Analysis of the DNA already cloned from 3B in conjunction with the data from the pulsed field gel (PFG) analysis would indicate the most efficient way to go about cloning the amplified inverted duplication. In particular the possibilities of using various long range cloning techniques such as chromosome jumping (Collins and Weissman1984, Poustka and Lehrach, 1986) or Yeast Artificial Chromosomes (Burke et al, 1987) could be explored, once the size and organisation of the amplified DNA had begun to be analysed. A knowledge of the available infrequently cutting restriction enzyme sites in the region would be essential for all such techniques.

It is quite a common phenomenon to detect multiple bands of genomic DNA restricted with infrequently cutting enzymes with a single copy probe on Southern blots of pulsed field gels. This is usually due to incomplete digestion by the enzyme owing to methylation of the cytosine in the CpG dinucleotides which are found in the cleavage sites of all these infrequently cutting enzymes. Methylation can inhibit the cleavage of a site by an enzyme to varying extents, and to complicate matters further there appears to be heterogeneity in the methylation patterns of genomic DNA, even between alleles of the same region, see discussion (Part 3.7). The resulting complex patterns of restriction fragments detected by a probe need not be too problematic when dealing with a single copy sequence. However, where there are multiple copies of a sequence as is the case with a gene amplification, partial cutting of heterogeneously methylated DNA would make the interpretation of results very difficult, as heterogeneity in the size of the units would also manifest itself as a complex pattern of bands. Thus, an understanding of the organisation of the amplified DNA could be very difficult if the cleavage of some restriction sites is inhibitted due to methylation when there is also the possibility of amplification unit size heterogeneity. In fact this has turned out to be a notorious problem in the analysis of other cases of gene amplification, especially if the degree of amplification is high which often means that there is a great deal of heterogeneity in the size of the units (see Chapter 1). In this chapter an approach which allows incomplete restriction site digestion to be distinguished from unit heterogeneity is described.

RESULTS:

3.2 The Identification of Infrequently Cutting Restriction Enzyme Sites which lie within the Inverted Duplication.

In order to examine the size and structure of the amplification units and compare the DNA in the inverted duplication with the corresponding unamplified rat DNA, pulsed field gel electrophoresis (PFGE) was conducted under a variety of conditions (voltage, pulsing interval, buffer temperature and run time) which resolved different size ranges of large DNA fragments generated using infrequently cutting enzymes. Both orthogonal field alternating gel electrophoresis, OFAGE, (based on the system described by Schwartz and Cantor, 1984 and Carle and Olson, 1985) and contour-clamped-homogeneous-electrophoresis, CHEF, (based on the system described by Chu et al, 1987) were used during the course of this work. DNA from 3B or Rat-2 cells embedded in LMP agarose blocks (prepared as described in Materials and Methods) was used throughout this long range mapping analysis in order to minimize DNA shearing.

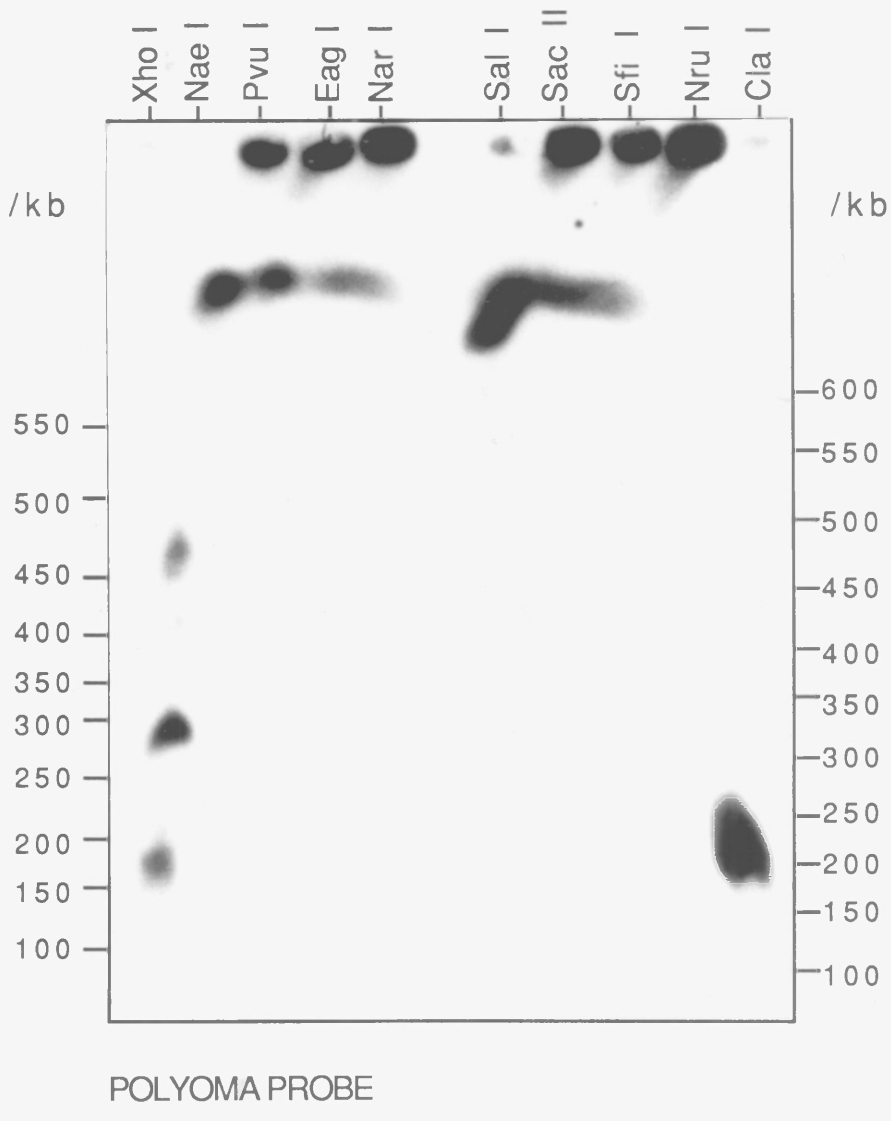
Southern blot analysis of rat DNA resolved on PFGs was conducted using either of two probes derived from the amplified units: the Polyoma (Py) probe, which can detect only the amplification specific DNA; and the RI-Bam probe (see Chapter 2) which detects both amplified copies and the native copy of the rat DNA.

DNA from 3B was digested with a wide variety of rare cutter enzymes (see Materials and Methods for conditions) in order to establish which of these enzymes would be most informative for further analysis of the amplified DNA. In Fig. 3.2.1 a Southern blot of a PFG of 3B DNA probed with a Polyoma virus (Py) probe is shown. Most of the enzymes used did not cut the amplified DNA into fragments of less than 600 kb (the limit of resolution of this gel), so that the probe hybridizes to the unresolved compression zone and to DNA which has remained in the slots. However the probe does detect a ClaI fragment of 240 kb, which implies that there is a ClaI site 120 kb away from the centre of the inverted duplication (as no ClaI sites were found in the region between the two Py sequences in the λ3B clone, as will be described in Part

FIGURE 3.2.1

Analysis by PFGE of the amplified DNA in 3B cells cut with a range of rare cutting restriction enzymes.

The blot of the pulsed field gel presented has been hybridised with a polyoma virus (Py) probe, which detects only the amplified DNA in 3B cells. Each track contains about 10 μg of 3B genomic DNA digested, in an agarose block, with the enzyme indicated (see Materials and Methods for block preparation and digestion conditions). The gel was run in an OFAGE apparatus for 16 hours, at 350 V, with a pulse interval of 30 secs, in 0.25 x TBE running buffer. The limit of resolution under these conditions was about 600 kb. The sizes on each side indicate the positions of concatemerized bacteriophage lambda markers. Note that the non-homogeneous electric field in the OFAGE apparatus leads to a different degree of curvature in the migration pathway of the DNA in each track, so that the fragment sizes are not parallel across the width of the gel.



POLYOMA PROBE

3.3). As the ClaI band is discrete this shows that the amplified palindromic arms must be homogeneous for at least 120 kb. In Fig.3.2.2 a single band of the same size (240 kb) is detected using the RI-Bam rat probe. This implies that the native (unamplified) rat ClaI fragment lies in the same size range as the amplified 240 kb fragment and thus cannot be distinguished.

The Py probe detects a ladder of XhoI fragments in Fig.3.2.1, the three major bands being 180 kb, 320 kb and 460 kb in length. Such a ladder could either be the result of a number of XhoI sites along the length of the inverted duplication, some of which were cut more efficiently than others due to differential methylation of the sites. If this is the case then the locations of the major XhoI sites would be 90 kb and 230 kb away from the centre (no XhoI sites exist in the region previously cloned, see part 3.3). On the other hand, a ladder of fragments could be an indication of heterogeneity in the amplification units, different units having different XhoI sites (although the ClaI data suggests that the units are identical for at least 120 kb).

In the case of SalI a faint band of approximately 550-600 kb, and another band just below the compression zone can be seen, but the rest of the hybridising fragments remain unresolved. The SalI bands are resolved more clearly in the PFG blot shown in Fig.3.2.2, where the limit of resolution is 800 kb. This blot was hybridised with both the (amplification specific) Py probe and the rat RI-Bam probe (which detects both the amplified and unamplified copies of the DNA). In the case of SalI-cut 3B DNA, a major band of 600 kb, a faint band of 550 kb, an even fainter band of 700 kb and a 750 kb band, as well as some unresolved fragments were detected by both the probes. However, the RI-Bam probe detected an extra band (or two closely migrating bands) of about 350 kb. These additional bands might represent the (unamplified) native copy of the rat DNA, and this is later confirmed by hybridisation of the same probe (RI-Bam) on SalI cut Rat-2 DNA (see Fig. 3.3.2). As with the complex pattern of XhoI fragments, the multiple bands of the amplified DNA could either be due to incomplete digestion of a number of adjacent SalI sites, or else heterogeneity in the size of the different copies of the amplification unit.

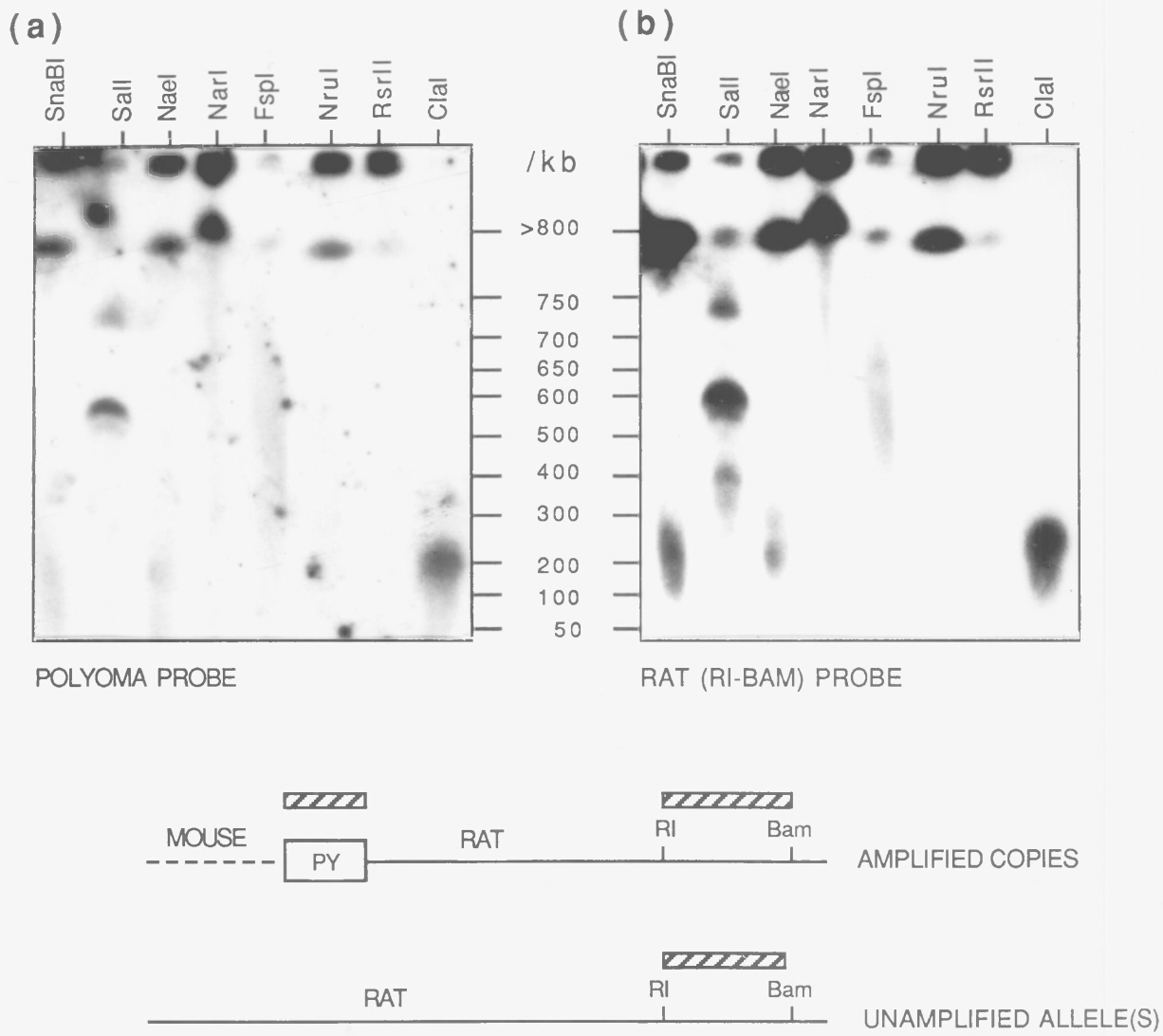
FIGURE 3.2.2

Analysis by PFGE of 3B genomic DNA cut with a range of rarely cutting restriction enzymes using the Py and RI-Bam Probes.

The same pulsed field gel blot of 3B DNA, was probed with (**a**) Polyoma virus DNA and (**b**) the 3.9 kb RI-Bam rat fragment. The same hybridisation and washing conditions were used in each case (see Materials and Methods).

Each track contains about 10 μg of 3B DNA cut with the enzyme indicated. The gel was run in a CHEF apparatus at 170 V for 24 hrs, in 0.25 x TBE running buffer using a pulse interval of 70 secs. The limit of resolution is 800 kb under these conditions. The sizes indicated between the blots were deduced from lambda concatemer markers.

Below, the upper map shows the presence and relative positions of the Py insert and the RI-Bam fragment in the amplified DNA in 3B cells; the lower map indicates the additional presence of the RI-Bam fragment on the unamplified allele(s) of this region of rat DNA in 3B cells. The hatched bars above each map represent the DNA fragments used as probes.


(a)
(b)
SnaBI
SalI
NaeI
NarI
FspI
NruI
RsrII
ClaI
/kb
>800
750
700
650
600
500
400
300
200
100
50
POLYOMA PROBE
RAT (RI-BAM) PROBE
MOUSE
PY
RAT
RI
Bam
AMPLIFIED COPIES
RAT
RI
Bam
UNAMPLIFIED ALLELE(S)

3.3 The Use of 5-azacytidine to Reveal the Presence of Multiple Methylated SalI and XhoI Sites in the Amplification Units.

In figures 3.2.1 and 3.2.2 it can be seen that the majority of infrequently cutting enzymes that were tested did not cut within the amplification units. Given the limited number of enzymes which did cleave within the amplification units (ClaI, XhoI, SalI) and which appeared to be informative at this stage in the analysis, it was important to clarify the exact nature of this information. It was therefore decided to try and determine whether the multiple XhoI and SalI fragments detected in the amplified DNA were a consequence of methylated restriction sites within the amplification units rather than heterogeneity in the sequences of the units. Methylation of cytosines in the sites of infrequently cutting enzymes inhibits cleavage which leads to the generation of partial digestion products. The agent 5-azacytidine (5-azaC) is a cytidine analogue which, when incorporated into mammalian DNA, greatly inhibits methylation of cytosine residues (Jones and Taylor, 1980) and leads to hypomethylation of the DNA. This effect is attributed to the impaired methylation of the cytosine in CpG dinucleotides, when a proportion of them have been substituted with 5-azaC. Analysis of the DNA derived from 3B cells grown with or without 5-azaC would reveal the presence of methylated restriction sites if a change in the pattern of bands was detected in the hypomethylated DNA.

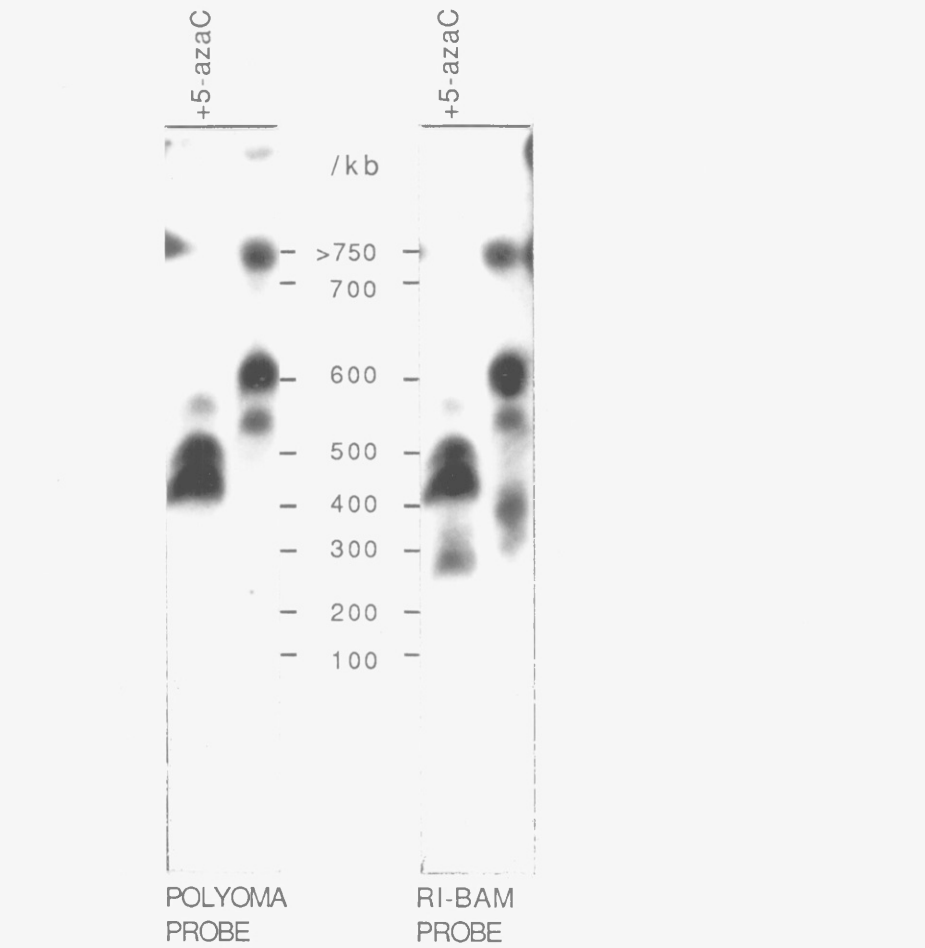
Both 3B and Rat-2 cultured cells were grown in different concentrations of 5-azaC (0.5 μM, 2 μM and 5 μM, as described under Materials and Methods) for 24 hours, after which the cells were immediately harvested and their DNA prepared in LMP agarose blocks. The effect of 5-azaC substitution and the resulting presence of unmethylated restriction sites was first examined in 3B DNA digested with SalI. This is presented in Fig.3.3.1 where a PFG blot of SalI-cut DNA from 3B cells grown in the absence or presence of (5 μM) 5-azaC was hybridised with both the RI-Bam and Py probes. A striking shift in the pattern of bands can be seen in the 5-azaC treated DNA compared to the untreated DNA. The same pattern as that in Fig.3.2.2 is found with untreated DNA, the amplified DNA being detected (by the RI-Bam and Py probes) in a

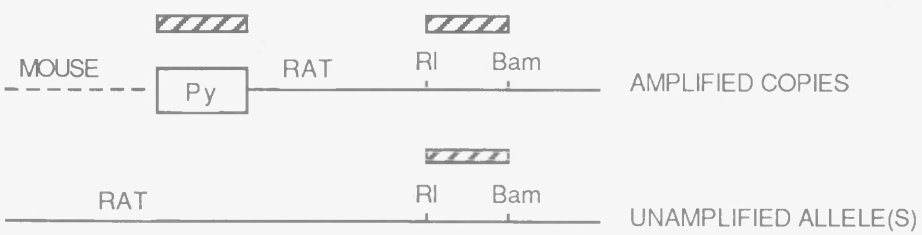
FIGURE 3.3.1

Analysis by PFGE of SalI-digested genomic DNA from 3B cells grown in the absence or presence of 5-azacytidine.

The two panels shown above represent the same southern blot hybridised with the polyoma virus probe (left), and the rat RI-Bam probe (right). The track labelled "+ 5-azaC" in this blot, contains 10 μg of SalI-cut genomic DNA derived from 3B cells which had been grown in the presence of 5 μM 5-azacytidine. The other (unlabelled) track contains 10 μg of SalI-cut genomic DNA from 3B cells grown without 5-azaC. The SalI restriction digestion conditions were identical for the two DNA samples. The gel was run in a CHEF apparatus at 170 V, for 18 hrs, in 0.25 x TBE running buffer, using a pulse interval of 60 secs. The limit of resolution is 750 kb under these conditions. The sizes indicated were deduced from lambda concatemers. Note that equivalent bands differ slightly in position between the two tracks, owing to the faster migration rate of the bulk of the SalI-cut genomic DNA from 5-azaC treated cells (as it has a lower average molecular weight).

Below are shown two maps, the upper of which indicates the presence, and relative positions in the amplified DNA of the Py insert and RI-Bam fragment, used as probes. The lower map illustrates the additional presence of the RI-Bam fragment in any unamplified alleles of this region of rat DNA in 3B cells.


+5-azaC
+5-azaC
/kb
>750
700
600
500
400
300
200
100
POLYOMA PROBE
RI-BAM PROBE


MOUSE
Py
RAT
RI
Bam
AMPLIFIED COPIES
RAT
RI
Bam
UNAMPLIFIED ALLELE(S)

600 kb band, a fainter 550 kb band and the rest in the unresolved region (DNA >750 kb). In the 5-azaC DNA the majority of the amplified signal is seen in a 500 kb band not previously detected, and to lesser extents it is detected in bands of 550 kb and 600 kb, decreasing in intensity in that order. In this case all the amplified DNA resides in these three fragments with none remaining unresolved. These patterns of bands were thus probably due to differential methylation of a number of adjacent SalI sites, because in hypomethylated DNA their relative intensity shifted from higher molecular weight to lower molecular weight bands. Assuming that this was the case, the locations of these SalI sites would be predicted to lie in the order shown in Fig.3.3.3. There appears to be a highly methylated SalI site (Sal^a) 250 kb away from the centre of the inverted duplication which generated a 500 kb band (but only in 5-azaC treated DNA). In addition, there must be a SalI site (Sal^b) 50 kb away from Sal^a (ie 300 kb away from the centre of the inverted duplication) which is cut much more efficiently (perhaps being less frequently methylated) generating a band of 600 kb. Thus in methylated DNA, site Sal^a is cut very infrequently, generating only the faint 550 kb band in conjunction with the Sal^b site on the other side of the inverted duplication. On the other hand, in 5-azaC-treated DNA most of the Sal^a sites were cut, but not all, so that the 550 kb and 600 kb bands are still visible. This indicates that although Sal^a is cut more frequently in 5-azaC treated DNA, it was still not cut 100% of the time. This was either because some copies of it were still methylated even after treatment with 5 μM 5-azaC, or because of site preferences displayed by the enzyme.

There is probably at least one more SalI site in the amplified DNA, Sal^c, about 150 kb away from Sal^b. The putative Sal^c site sometimes generates a very faint 700 kb partial product with the Sal^a site on the other side of the inverted duplication, and a stronger 750 kb band with the more efficiently cut Sal^b site (see Fig.3.2.2). Thus in methylated DNA only the 750 kb band is present (but unresolved in Fig.3.3.1), whilst in hypomethylated DNA the 700 kb can just be seen.

In Fig.3.3.1, the extra bands of 350 kb and 400 kb detected by the RI-Bam rat probe, but not the amplification specific Py probe, can also be seen to shift in relative intensity after 5-azaC treatment, the 350 kb band becoming the more intense of the two in

hypomethylated DNA. These bands probably represent the partial digestion products from more than one SalI site in the native (unamplified) copy of the region in rat DNA. The bands differ by 50 kb, with one site obviously more methylated than the other, which is consistent with them corresponding to the Sal^a and Sal^b sites in the amplified DNA. As these sites are 250 kb and 300 kb away from the joint at the centre of the inverted duplication (ie the point of rearrangement), not all of the sequences in the 350 kb and 400 kb native rat SalI fragments could be present in the corresponding region of the amplified inverted duplication (see map in Fig.3.3.3).

In order to confirm these results and to look at the effect of similar treatment on the XhoI ladder of bands, this analysis was extended using DNA from 3B and Rat-2 cells treated with a variety of concentrations of 5-azaC. The results shown in the PFG blot in Fig.3.3.2 are consistent with those previously obtained for the SalI-cut DNA, with an increasing dose of 5-azaC causing a corresponding shift in the relative intensities of the bands. The lowest molecular weight fragment predominates as the proportion of unmethylated SalI sites increases. The two (unamplified) native rat bands detected only by the RI-Bam probe are shifted in intensity with 5-azaC treatment as before. Moreover, their origin is confirmed, as in Rat-2 DNA the RI-Bam probe detects two bands, which are of the same size, and which display an identical shift in intensity upon demethylation as those in 3B DNA.

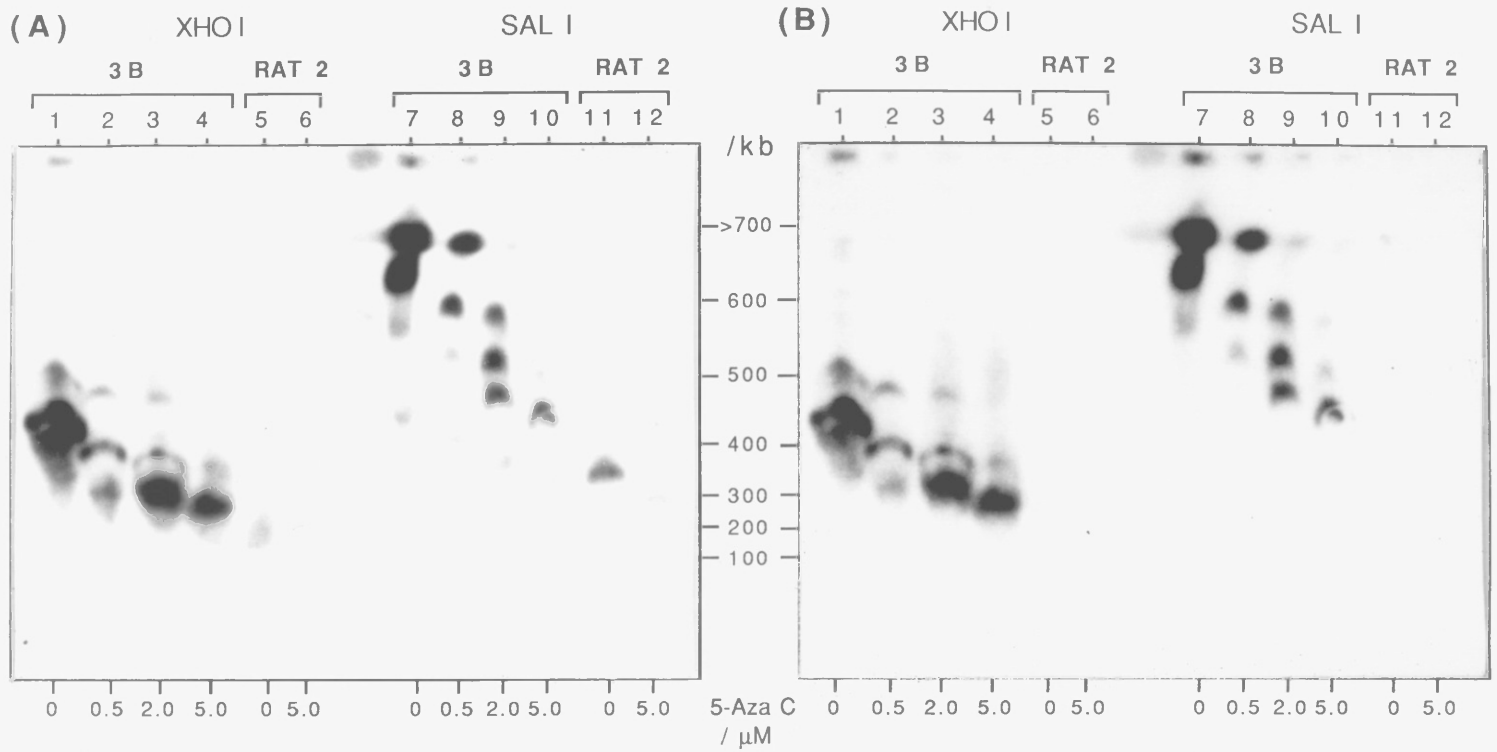
Increasing concentrations of 5-azaC had a very similar effect on the XhoI ladder of bands. There appear to be a number of XhoI sites along each arm of the inverted duplication which are all cut with different efficiencies probably as a consequence of methylation as well as the location of sites. As with SalI, demethylation of the DNA markedly increased the efficiency with which the XhoI sites were cut, with most of the high molecular weight bands diappearing and the lowest molecular weight bands becoming the most intense as the 5-azaC dose increased. The predicted map of these XhoI sites is also shown in Fig.3.3.3. The existence of a number of methylated XhoI sites in this region of the rat genome was confirmed in Rat-2 DNA where the RI-Bam probe detected two bands of approximately 120 kb and 150 kb, which were reduced to

FIGURE 3.3.2

Analysis by PFGE of SalI and XhoI digested genomic DNA, from 3B and Rat-2 Cells grown in the presence of different concentrations of 5-azacytidine.

The two panels at the top show the same southern blot, hybridised in (B) with the polyoma virus probe and in (A) with the rat RI-Bam probe. As indicated above the lanes, **tracks 1-4** each contain 10 µg of 3B genomic DNA cut with XhoI; **tracks 7-10** each contain 10 µg of 3B genomic DNA cut with SalI. **Tracks 5 and 6** contain 10 µg and 5 µg respectively, of Rat-2 DNA cut with XhoI; **tracks 11 and 12** each contain 10 µg of Rat-2 DNA cut with SalI. The genomic DNA used was isolated from 3B or Rat-2 cells that had been grown either in the presence of 0.5, 2.0 or 5.0 µM 5-azaC, or in its absence. The particular 5-azaC concentration used is indicated below each track. For each restriction enzyme, the digestion conditions for all the samples were identical. The gel was run at 170 V, for 21 hrs, in 0.5 x TBE running buffer, using a pulse interval of 50 secs. The limit of resolution under these conditions was 700 kb. The sizes indicated between the panels were deduced from lambda DNA concatemer markers.

Below, the upper of the two maps shown, indicates the presence and relative positions of the Py insert and the RI-Bam rat fragment in the amplified DNA in 3B cells. The lower map shows the presence of the RI-Bam fragment in the single copy alleles of the same region which exist in both Rat-2 and 3B cells. The hatched bars illustrate that the Py probe will detect amplified DNA, only present in 3B cells; whereas the RI-Bam rat probe will detect both amplified and unamplified DNA in 3B cells, as well as the single copy DNA in Rat-2 cells. In this context, note the extra bands present in (A) but not (B), which represent unamplified DNA fragments detected by the rat probe.


(A)
XHO I
SAL I
3B
RAT 2
1 2 3 4 5 6 7 8 9 10 11 12
/kb
>700
600
500
400
300
200
100
0 0.5 2.0 5.0 0 5.0 0 0.5 2.0 5.0 0 5.0
5-Aza C / μM
RAT (RI-BAM) PROBE
(B)
XHO I
SAL I
3B
RAT 2
1 2 3 4 5 6 7 8 9 10 11 12
0 0.5 2.0 5.0 0 5.0 0 0.5 2.0 5.0 0 5.0
POLYOMA (PY) PROBE

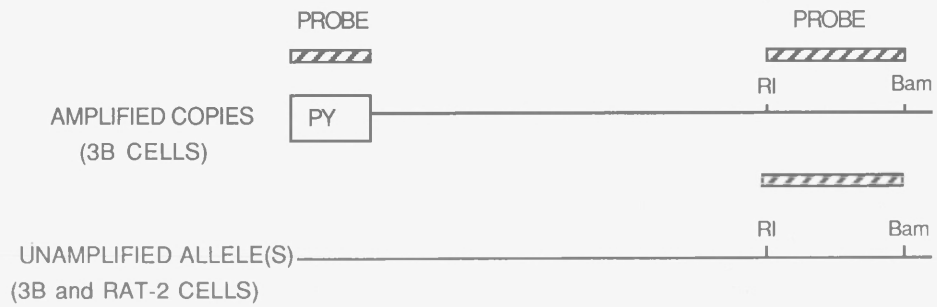
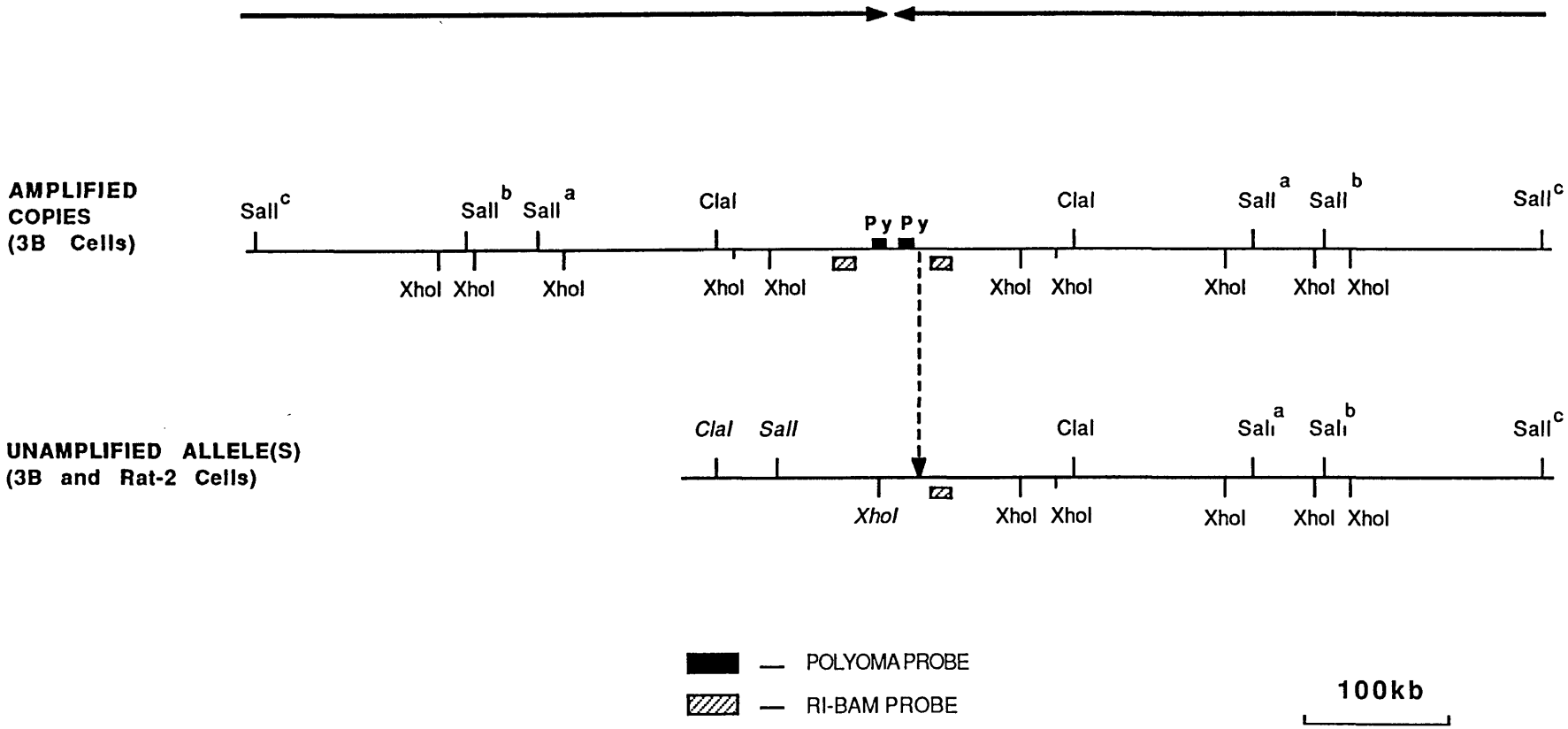

PROBE
PROBE
RI
Bam
AMPLIFIED COPIES
(3B CELLS)
PY
RI
Bam
UNAMPLIFIED ALLELE(S)
(3B and RAT-2 CELLS)

FIGURE 3.3.3

Long range restriction map of the inverted duplication in 3B cells and of the region of rat genome from which it has been derived.

The upper map represents the complete inverted duplication found amplified in 3B cells. The two arrows above it indicate the palindromic arms. The sites shown for the rare cutting enzymes ClaI, SalI and XhoI were deduced by PFGE analysis (Parts 3.2 and 3.3). The multiple XhoI and SalI sites were deduced by southern blot analysis of the sizes of partial restriction products generated by these enzymes. Both the Py and the rat RI-Bam probes (shown as solid and hatched bars respectively) hybridised to the same amplified fragments. The size of each arm of the inverted duplication was predicted to be at least 450 kb from this analysis. The minimum size of the complete inverted duplication is thus 900 kb.

Below, a map of the equivalent unamplified region of rat DNA is shown, which is present in Rat-2 cells as well as 3B cells. The arrow extending from the upper to the lower map indicates the point in the rat DNA at which rearrangement with the polyoma insert occurred during the formation of the inverted duplication in 3B cells. The sites shown in italics, to the left of this point, are therefore not present in the amplified inverted duplication. These single copy sites were mapped by southern blot analysis of Rat-2 DNA and 3B DNA, using the RI-Bam probe (indicated by the hatched bar).


AMPLIFIED COPIES (3B Cells)
SalI c
SalI b
SalI a
ClaI
Py Py
ClaI
SalI a
SalI b
SalI c
XhoI XhoI XhoI XhoI XhoI XhoI XhoI XhoI XhoI XhoI
UNAMPLIFIED ALLELE(S) (3B and Rat-2 Cells)
ClaI SalI
ClaI
SalI a
SalI b
SalI c
XhoI XhoI XhoI XhoI XhoI XhoI XhoI
POLYOMA PROBE
RI-BAM PROBE
100kb

a single 120 kb band by 5-azaC treatment. (These bands cannot be easily distinguished in 3B DNA due to the close migration of a number of the amplification-specific bands in the 100-300 kb region). The size of the smallest unamplified rat fragment (120 kb) is consistent with the first XhoI site being 90 kb away from the centre of the inverted duplication, as the formation of the amplified joint must have disrupted the native 120 kb XhoI fragment.

The use of hypomethylated 3B and Rat-2 cellular DNAs in this PFG analysis provided strong evidence that the multiple SalI and XhoI fragments which had been observed were a result of incomplete restriction site cleavage. This led to the conclusion that the amplification units are probably homogeneous in size and contain multiple SalI and XhoI sites. From the analysis described in this section, the predicted map of these sites (see Fig.3.3.3) enabled the minimum size of each amplification unit to be estimated at 375-400 kb.

3.4 The Use of Infrequently Cutting Restriction Enzymes Which Do Not Cleave Within the Amplified Rat Sequences in 3B cells in Determining the Size of the Amplification Units.

The analyses in Parts 3.2 and 3.3 made use of enzymes which were found to cut within the amplified units to create a map of the inverted duplication and estimate its minimum size. However, the majority of rare cutting restriction enzymes tested did not appear to cleave within the inverted duplication. The aim here was to make use of some of these enzymes in conjunction with pulsed field gel electrophoresis using conditions that gave resolution of larger DNA fragments than before in order to estimate the upper size limit of the amplification unit.

In Fig. 3.4.1 a blot of a PFG which resolved DNA up to 1 Mb in length can be seen. Based on the published sequence of the polyoma virus genome (Tooze, 1981), the viral DNA within the amplified inverted duplication was known to contain an NarI site. Unfortunately, this site was extremely inefficiently cut, (as will be described in part 3.4) and the enzyme NarI usually never generated any clearly resolved fragments, even after 5-azaC treatment (see track on the far left in both (a) and (b) in Fig.3.4.1).

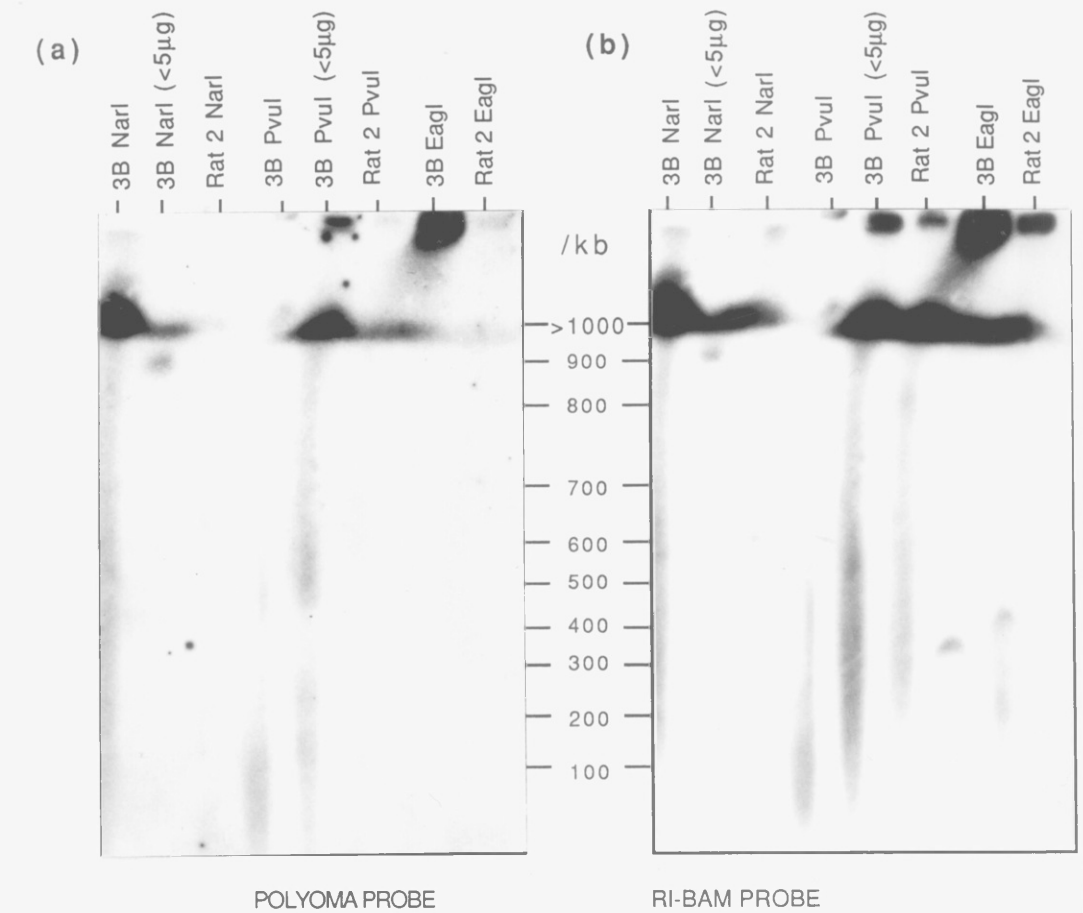
FIGURE 3.4.1

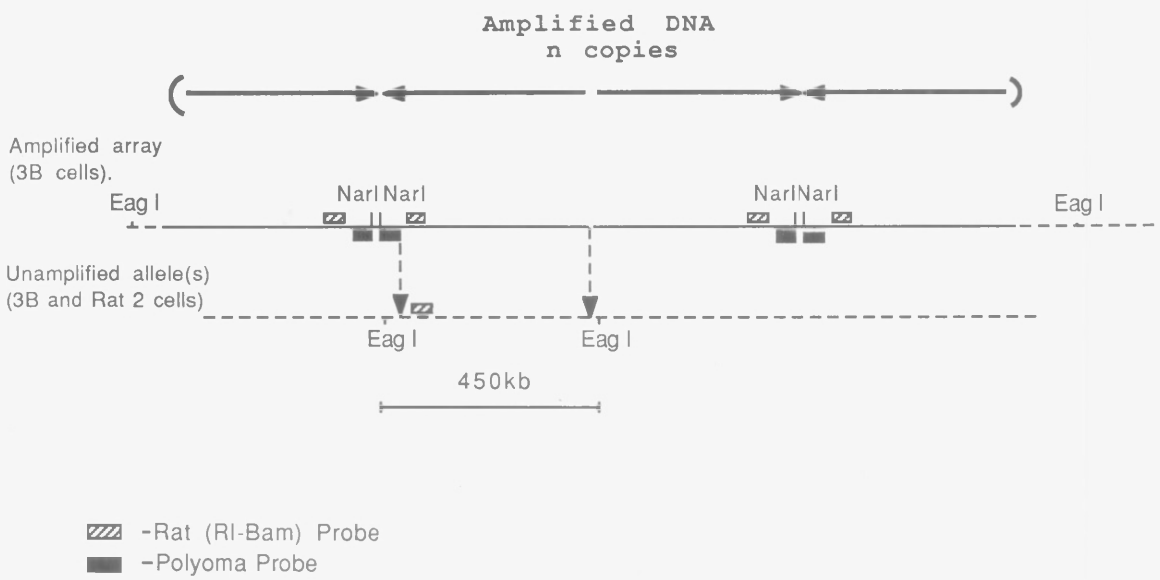
Analysis by PFGE of 3B and Rat-2 genomic DNA cut with the enzymes NarI and EagI.

The two panels above show the same southern blot hybridised in (**a**) with the Py probe and in (**b**) with the RI-Bam rat probe. Each track contains 3B or Rat-2 genomic DNA digested with NarI or EagI, as indicated above each lane. The second track from the left contains less than 5 μg of DNA, whereas the rest contain approximately 10 μg of DNA each. The gel was run in a CHEF apparatus at 120 V, for 44 hrs, in 0.5 x TBE running buffer using a pulse interval of 110 secs. The limit of resolution was 1 Mb under these conditions. The sizes indicated were deduced from lambda DNA concatemers and yeast (*S. cerevisiae*) chromosomes.

The upper of the two maps shown in the diagram below, is a hypothetical representation of the amplified DNA as an array of inverted duplications at a position on a rat chromosome. Only two copies of the complete inverted duplication from the array are shown. The dashed lines correspond to unamplified rat DNA, and solid lines to amplified DNA. The NarI sites that lies within the polyoma inserts in the amplified DNA are indicated. The polyoma probe is represented by a black bar, and the RI-Bam rat probe by a hatched bar. Two hypothetical EagI sites are shown in this map, located outside the amplified array, which would generate unresolved fragments under these PFGE conditions. These two sites are not necessarily the same as those shown in the map below

The lower map shows the native rat locus from which the amplified DNA was derived. This unamplified locus is found in Rat-2 cells as well as 3B cells. The arrows extending from the upper map onto the lower one, indicate the extent of the rat DNA present in the inverted duplication, and the presence of the RI-Bam fragment in this region is shown. The two EagI sites that generate a single copy 450 kb fragment which is detected by the RI-Bam probe [see last two tracks, panel (**b**)], are shown lying just outside this region (about 450 kb apart), and are thus not present in the amplified array in 3B cells.


(a)
(b)
3B NarI
3B NarI (<5µg)
Rat 2 NarI
3B PvuI
3B PvuI (<5µg)
Rat 2 PvuI
3B EagI
Rat 2 EagI
/kb
>1000
900
800
700
600
500
400
300
200
100
POLYOMA PROBE
RI-BAM PROBE


Amplified DNA
n copies
Amplified array
(3B cells).
Eag I
NarI NarI
NarI NarI
Eag I
Unamplified allele(s)
(3B and Rat 2 cells)
Eag I
Eag I
450kb
-Rat (RI-Bam) Probe
-Polyoma Probe

However, using an excess of enzyme on a minimal amount of DNA (<5 μg), it was possible to observe some cleavage of the amplified DNA (see second track from the left, in (a) and (b), Fig. 3.4.1). In this case, a discrete band of about 900 kb was detected by both the RI-Bam and Py probes. The 900 kb band of amplified DNA detected in 3B cells could either be due to cleavage at the NarI sites within the polyoma DNA in adjacent inverted duplications; or, to cleavage at a Py NarI site and a rat NarI site within a unit. The latter explanation seems unlikely because when Rat-2 DNA was cut with NarI (see third lane from left in (b), Fig.3.4.1) it produced a ladder of very faint bands with the RI-Bam probe, with most of the probe hybridising to unresolved fragments. This indicates that all the NarI sites detected in the Rat-2 DNA are cut very inefficiently in the **native** copy of the region, and would therefore not be expected to be cut efficiently in the corresponding **amplifed** copies of the region. It seems more likely that the faint 900 kb band was a product of NarI cleavage in the polyoma inserts of adjacent inverted duplication centres (see map, Fig.3.4.1). From this result it may be predicted that if the centres of adjacent inverted duplications are 900 kb apart, then each amplification unit must be 450 kb in length. This is consistent with the minimum size estimate of 375-400 kb for each side of the inverted duplication made with the SalI map in Part 3.3.

The enzyme EagI was also very useful in this analysis, as it never appeared to cut the amplified DNA in 3B (detected with the Py probe) into fragments that could be resolved by PFGE. However, it can be seen (in the two tracks on the far right in (b), Fig.3.4.1) that a band of 450 kb representing the native unamplified copy of the region is detected in both 3B and Rat-2 DNA cut with EagI when hybridised with the rat RI-Bam probe. This band is not detected with the Py probe (same tracks in (a), Fig.3.4.1). Most of the RI-Bam probe signal and all of the Py probe signal is found in very high molecular weight DNA (in the slot and the compression zone) in 3B DNA (track second from right in (a) and (b), Fig.3.4.1). This suggests that EagI has no sites within the amplified DNA in 3B and the probes can therefore only detect extremely large fragments which are derived from sites outside the amplification units (see map in Fig.3.4.1). On the other hand, the inability of EagI to cleave the amplified DNA in 3B

cells, could be a consequence of differences in the methylation of EagI sites between the amplified copies and the unamplified native copy of the rat DNA. This seems unlikely as no sign of even partial EagI cleavage was ever detected with 3B DNA. Assuming, therefore, that the native region of the rat DNA contains a 450 kb EagI fragment but that neither of the sites which generate it are present in the inverted duplication, this implies that the DNA comprising each palindromic arm lies within this 450 kb fragment. In other words, each side of the inverted duplication must be less than 450 kb in length.

3.5 The Amplification Units May Be Clustered in a Single Region of 3B DNA.

In order to assess whether the amplified DNA was clustered as an array of units in a single position or dispersed at a number of locations, a number of other rare cutting restriction enzymes which did not appear to cut within the amplification units were used. PFGE of 3B genomic DNA cut with such enzymes, under conditions that resolved fragments which were several megabases in size, would indicate the distribution of the amplification units.

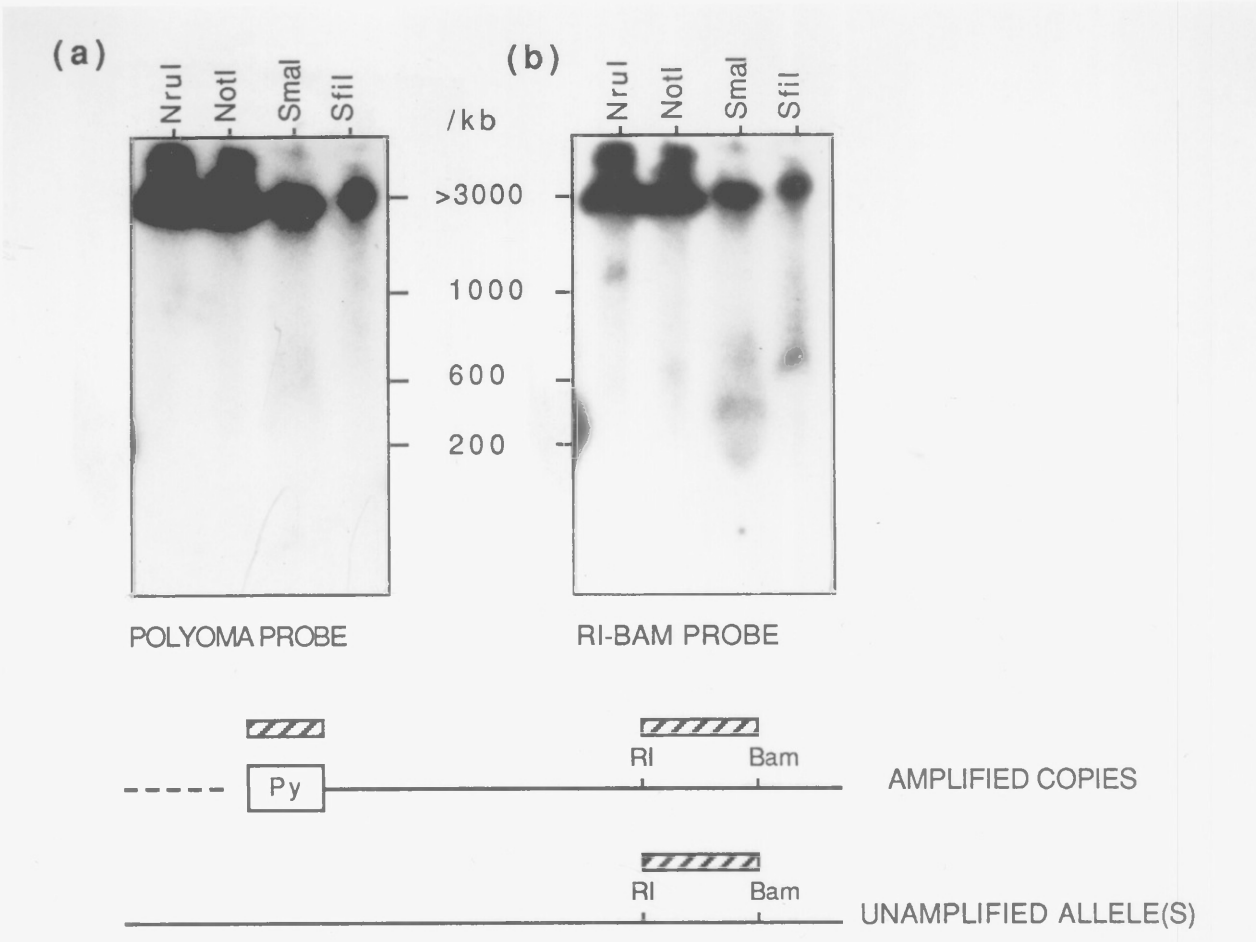
In Fig.3.5.1 a blot of a PFG is shown, on which DNA fragments of up to 3 Mb in length have been resolved, hybridised with both Py and RI-Bam probes. Using the (amplification specific) Py probe all the DNA detected is in the unresolved region ie over 3 Mb, [see Fig.3.5.1 (a)]. This indicates either that all the amplfied DNA lies in a fragment(s) of over 3 Mb in length, or that the enzymes had not cleaved the DNA to completion. The same blot hybridised with the RI-Bam probe [see Fig.3.5.1 (b)] provided a very useful control, as the single copy bands due to the native rat region could be seen in each track, in addition to all the amplified copies over 3 Mb (see map in Fig.3.5.3). It could therefore be assumed that the DNA had been digested to completion by all the restriction enzymes, as a single copy sequence could easily be detected in each track. It was thus concluded that all the amplified DNA must lie within fragments of >3Mb, as the amplification-specific Py probe detected no bands (single copy or otherwise) below this size. It may be noted, however, that in the case of SmaI,

FIGURE 3.5.1

Analysis of the overall organization of the amplified units in 3B cells by PFGE resolution of DNA up to 3 Mb in length.

The same southern blot was hybridised with the polyoma virus (Py) probe in (**a**), and the RI-Bam rat probe in (**b**). Each track contains 10 μg of 3B genomic DNA digested with NruI, NotI, SmaI or SfiI as indicated. The gel was run in a CHEF apparatus at 60 V for 5 days, in 0.5 x TBE running buffer using a pulse interval of 20 mins. The sizes shown were deduced from yeast (both *S.pombe* and *S.cerevisiae*) chromosome size markers. Note that although the majority of DNA fragments hybridising to polyoma appear to be unresolved in (**a**), this is not due to incomplete enzyme digestion, as single copy bands are detected in the same tracks by the RI-Bam probe in (**b**).

Below, two maps are shown which illustrate the fact that the polyoma virus probe will only detect restriction fragments from the amplified DNA in 3B cells; whereas the rat RI-Bam probe will detect both the amplified DNA and the unamplified copy (or copies) of this region of rat DNA.


(a)
NruI
NotI
SmaI
SfiI
/kb
>3000
1000
600
200
POLYOMA PROBE
(b)
NruI
NotI
SmaI
SfiI
RI-BAM PROBE
Py
RI
Bam
AMPLIFIED COPIES
RI
Bam
UNAMPLIFIED ALLELE(S)

an approximately 300 kb unamplified rat DNA fragment is detected with the RI-Bam probe [see Fig.3.5.1(b)]. This suggests that there should be a SmaI site within the amplification unit, if the size estimate of 450 kb is correct. However, the Py probe does not hybridise to any SmaI fragments below 3 Mb. This implies that there might be differences in the methylation of SmaI sites between the amplified and unamplified copies of the same region of rat DNA. In fact, Rat-2 DNA cut with SmaI also produces this 300 kb fragment (data not shown). However, 5-azaC treatment never had any effect on SmaI digestion of 3B DNA.

In order to see whether it was possible to resolve the amplified DNA in to more than one fragment of >3 Mbp with any of the enzymes tested above, a PFG of 3B DNA cut with each of them was run under conditions resolving even higher molecular weight DNA. A blot of this hybridised with the two probes is shown in Fig.3.5.2. As no size markers were available for DNA over 5.7 Mb (the largest *S.pombe* chromosome) the limit of resolution of this gel could only be estimated to be about 10 Mb. The DNA fragments in this size range have usually begun to shear, so the tracks appear more smeared than in other PFG blots. Furthermore, these PFGE conditions resolve lower molecular weight fragments (in the range of hundreds of kilobases) much more poorly, so that the single copy rat fragments detected by only the RI-Bam probe are less clear. However, it can still be seen that with the Py probe all the signal is again in the unresolved region of the gel [see Fig. 3.5.2 (a)]. The RI-Bam probe detected some unamplified rat DNA of lower molecular weight, but it mainly hybridised to unresolved fragments (>10 Mbp) which presumabely were derived from the amplified DNA [see Fig. 3.5.2 (b)]. Hence, the amplified DNA appears to be clustered together in fragments of over about 10 Mb. Resolution of higher molecular weight DNA than this would not produce any meaningful data owing to the degradation of such large fragments.

A summary of the enzymes which do not cleave within the amplification units, but have sites located outside the proposed amplified array, can be seen in the map in Fig.3.5.3.

FIGURE 3.5.2

Further analysis of the organisation of the amplified units in 3B cells through the resolution of DNA up to approximately 10 Mb in length.

The same southern blot is shown hybridised with the polyoma virus probe in (**a**) and the RI-Bam rat probe in (**b**). Each track contains genomic DNA from 3B cells cut with either NotI, SacII or NruI. In order to try and maximise the DNA cleavage by these enzymes, a number of different approaches were used. For the NotI and NruI digestions, DNA from cells grown in the presence of 2 μM 5-azaC as well as from untreated cells was used, as indicated above the lanes. Two SacII digestions were carried out, one involving 10 μg and the other 5 μg of DNA (as indicated) with the same amount of enzyme for each.

The gel was run at 75 V, for 4 days, in 0.5 x TBE, using a pulse interval of 65 mins. The limit of resolution under these conditions was estimated to be about 10 Mb, although the longest of the *S.pombe* chromosomes, used as size markers here, is only 5.7 Mb. The smear of hybridisation that can be seen in each track is probably due to partial shearing of the very high molecular weight DNA being resolved.

The maps shown below, illustrate the fact that the polyoma probe can only detect the amplified DNA in 3B cells, whilst the rat RI-Bam probe detects both the amplified DNA and the unamplified copy (copies) of the same region of rat DNA.

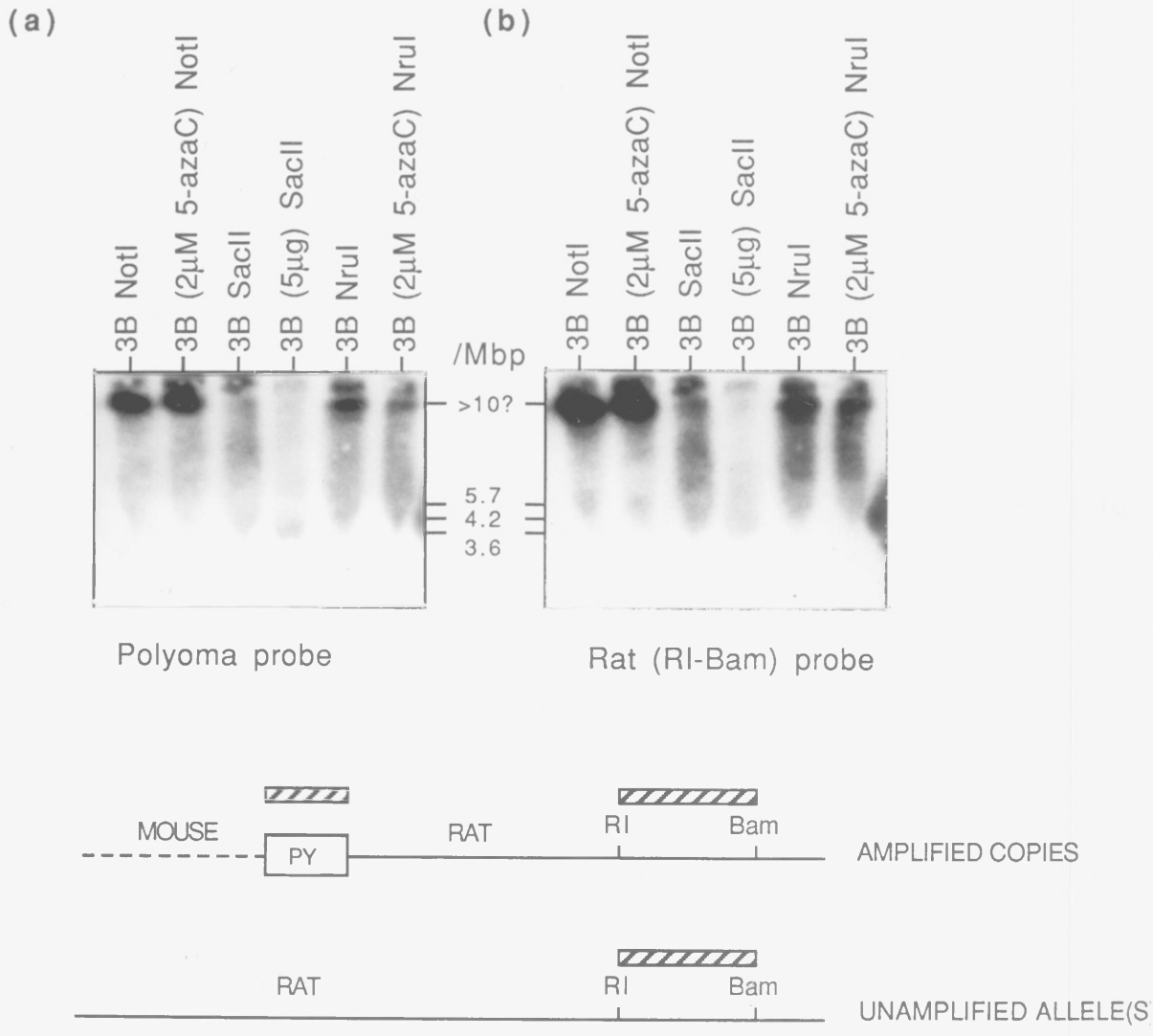

(a)
(b)
3B NotI
3B (2μM 5-azaC) NotI
3B SacII
3B (5μg) SacII
3B NruI
3B (2μM 5-azaC) NruI
/Mbp
>10?
5.7
4.2
3.6
Polyoma probe
Rat (RI-Bam) probe
MOUSE
PY
RAT
RI
Bam
AMPLIFIED COPIES
RAT
RI
Bam
UNAMPLIFIED ALLELE(S

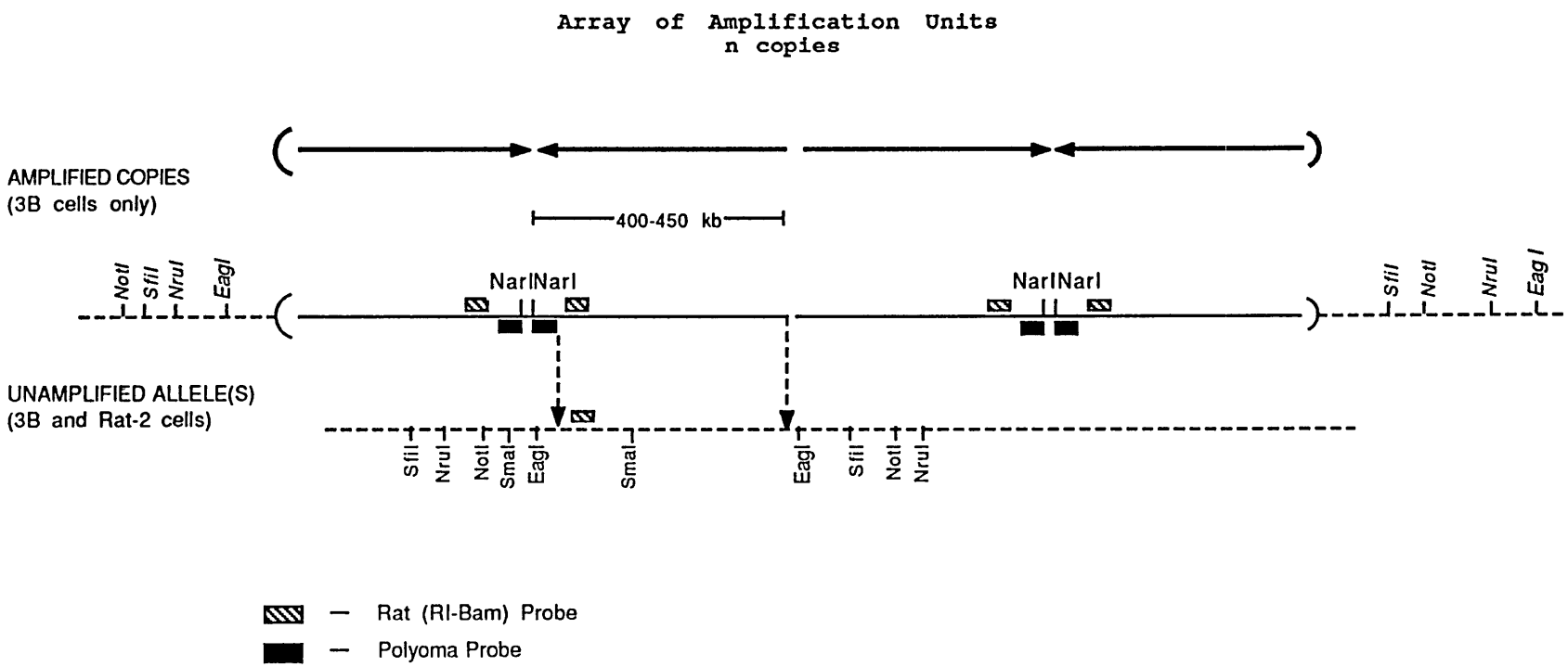
FIGURE 3.5.3

The organization of the amplified array in 3B cells, and of the native locus in Rat-2 cells.

The maps presented here shows the restriction sites of only those enzymes that do not cleave the rat DNA within the amplified inverted duplication, and were thus useful for the investigation of the overall organisation of the amplification units.

The upper map is a hypothetical representation of the amplified DNA in 3B cells as an array of inverted duplications at some location on a rat chromosome. Each amplification unit is 400-450 kb long as indicated. Only four out of the twenty to forty amplification units are shown. The arrows above the upper map indicate the palindromic orientations of the units. Restriction sites for NarI within the polyoma inserts but not the rat DNA are shown. The dashed lines represent the unamplified rat DNA that is presumed to flank the amplified array, and possible sites (in italics) are shown for the enzymes which do not cut within the amplified DNA, although these sites have not been mapped as they generate fragments of over 10 Mb in length (Part 3.5). The locations of the Py (solid bar) and the RI-Bam (hatched bar) probes are indicated.

The map below represents the unamplified rat allele of the region from which the inverted duplication was generated. The arrows from the map above indicate the extent of this region lying within the amplification unit. A SmaI site is shown within this region only in the unamplified allele. As described in Part 3.5, this site is either uncleavable or absent in the amplified DNA. Sites are also shown for the enzymes which cleave outside the region within the amplification unit in 3B cells. These sites generate single copy DNA fragments that are detected by only the RI-Bam probe (hatched bar) of ≥ 400-450 kb. The sites are not necessarily in the order shown.


Array of Amplification Units
n copies
AMPLIFIED COPIES
(3B cells only)
400-450 kb
NotI
SfiI
NruI
EagI
NarINarI
NarINarI
SfiI
NotI
NruI
EagI
UNAMPLIFIED ALLELE(S)
(3B and Rat-2 cells)
SfiI
NruI
NotI
SmaI
EagI
SmaI
EagI
SfiI
NotI
NruI
Rat (RI-Bam) Probe
Polyoma Probe

3.6 Detection of Sites which could be useful for Chromosome Jumping along the Inverted Duplication.

The results presented so far in this chapter, indicated that the size of the inverted duplication was several hundred kilobases long and, in addition, that the units of amplification were homogeneous as far as could be mapped using PFGE. It therefore seemed reasonable to try and clone the other end of the inverted duplication if a suitable long range cloning strategy could be found, with the aim of analysing the arrangement of the next novel joint in the amplified DNA.

When various possible cloning strategies were being considered, chromosome jumping had just begun to be used with some success by a number of laboratories. Chromosome jumping (Poustka and Lehrach 1986, for review) is the cloning of two ends of a large DNA fragment, without the intervening sequences between them. This is accomplished by circularizing large DNA fragments generated either by complete digestion of infrequently cutting enzymes or by partial digestion with frequently cutting enzymes, around a selectable marker and cloning the junction fragments thus generated in *E.coli*. Although the length of these ligated junction fragments is limited by the capacity of the cloning vector, the distance between the two fragments can be hundreds of kilobases. Furthermore, only the ends of the large DNA fragment need to be clonable and non-repetitive. This method is thus supposedly less sensitive to sequences which are either unclonable or too repetitive to be crossed by chromosome walking. It therefore has a number of advantages over chromosome walking in cosmid or lambda vectors, although it does mean that intervening DNA between jumping points is not cloned and therefore cannot be analysed in detail. The most straightforward type of jump in terms of the number of manipulations required and the amount of starting material necessary to generate adequate numbers of jumping clones, is one in which rare cutter enzymes are used to generate the large "jump" fragments. For this approach a rare cutter site to jump from as well as one to jump to are required.

As will be discussed later in this chapter, many groups working on amplified cellular

genes have taken advantage of the unmethylated rare cutter enzyme sites often found upstream of the expressed gene in an HTF island, to analyse the long range structure of the amplified DNA. The PFGE analysis described previously had not reveal any HTF islands (Bird et al, 1985) ie clusters of rare cutter enzyme sites, within the inverted duplication which could have provided useful jumping points.

A detailed search of λ3B was made in the hope of finding possible rare cutter enzyme sites which were unmethylated for use in a jump. The known sequence of the polyoma genome (Tooze, 1981) revealed the presence of a single NarI site. To test for the presence of this site and any others in the adjacent mouse sequences upstream of the polyoma virus oncogene in 3B cellular DNA, the cloned λ3B DNA was cut with a range of infrequently cutting enzymes. Two blots of the digested 'phage DNA hybridised with the Py probe are shown in Fig. 3.6.1. NarI cuts the inverted duplication within the two polyoma oncogenes as expected. Two SmaI sites were also detected in the mouse DNA within λ3B. Double digestion with a number of other enzymes predicted the position of one SmaI site on each side of the centre of the inverted duplication and confirmed the position of the NarI site in each of the Polyoma regions (data not shown); a map of these sites is shown in Fig.3.4.1. In order for any of these sites to be useful for cloning and mapping, it had to be determined whether they could be cut in genomic DNA, as CpG dinucleotides are methylated in mammalian cells (unlike *E.coli* in which lambda DNA is propagated). The enzyme NaeI had previously been observed to produce some faint bands containing Py sequences in 3B genomic DNA (data not shown), thus it too was tested even though it did not appear to cut λ3B DNA. Fig.3.6.2 shows a blot of 3B genomic DNA cut with each of the potentially useful rare cutting enzymes hybridised with a Py probe. In each case the enzyme was used alone and in combination with KpnI. KpnI is known to cut within the Py sequences and also at a site 5 kb downstream in the rat DNA (Passananti et al,1987), as shown in the map. A 5 kb KpnI band is thus detected in each of the double digestions. NarI generates a 2 kb fragment with KpnI as was expected, but the digestion appears incomplete and a large excess of enzyme was required to produce this result. SmaI would have been expected to produce a band of 3 kb in combination with

FIGURE 3.6.1

Rare cutter enzymes sites located within the insert of the λ3B clone.

Two southern blots are shown of λ3B cloned DNA cut with a range of rare cutter enzymes. Each track contains 1 μg of λ3B DNA cleaved with the enzyme indicated. The sizes indicated were deduced from lambda HindIII size markers.

Both blots were probed with polyoma virus DNA. The positions of the polyoma sequences which are detected by this probe, are indicated (PY) in the map of λ3B shown below. The inverted duplicated sequences extend towards the centre of the λ3B insert as far as the points denoted by the two arrowheads. Only the enzymes NarI and SmaI were found to cleave the λ3B insert from this southern blot analysis, and their positions are shown. Additional SmaI sites present in the lambda arms (shown in italics) were useful in the mapping of the SmaI sites in the mouse DNA.

Note that the enzyme NaeI does not cut the λ3B insert (see lower blot, third track from left). Sites were subsequently detected for NaeI in genomic DNA from 3B cells (see Fig. 3.6.2). The λ3B digestion was repeated using excess enzyme, as well as spermidine, but no cleavage was ever detected.

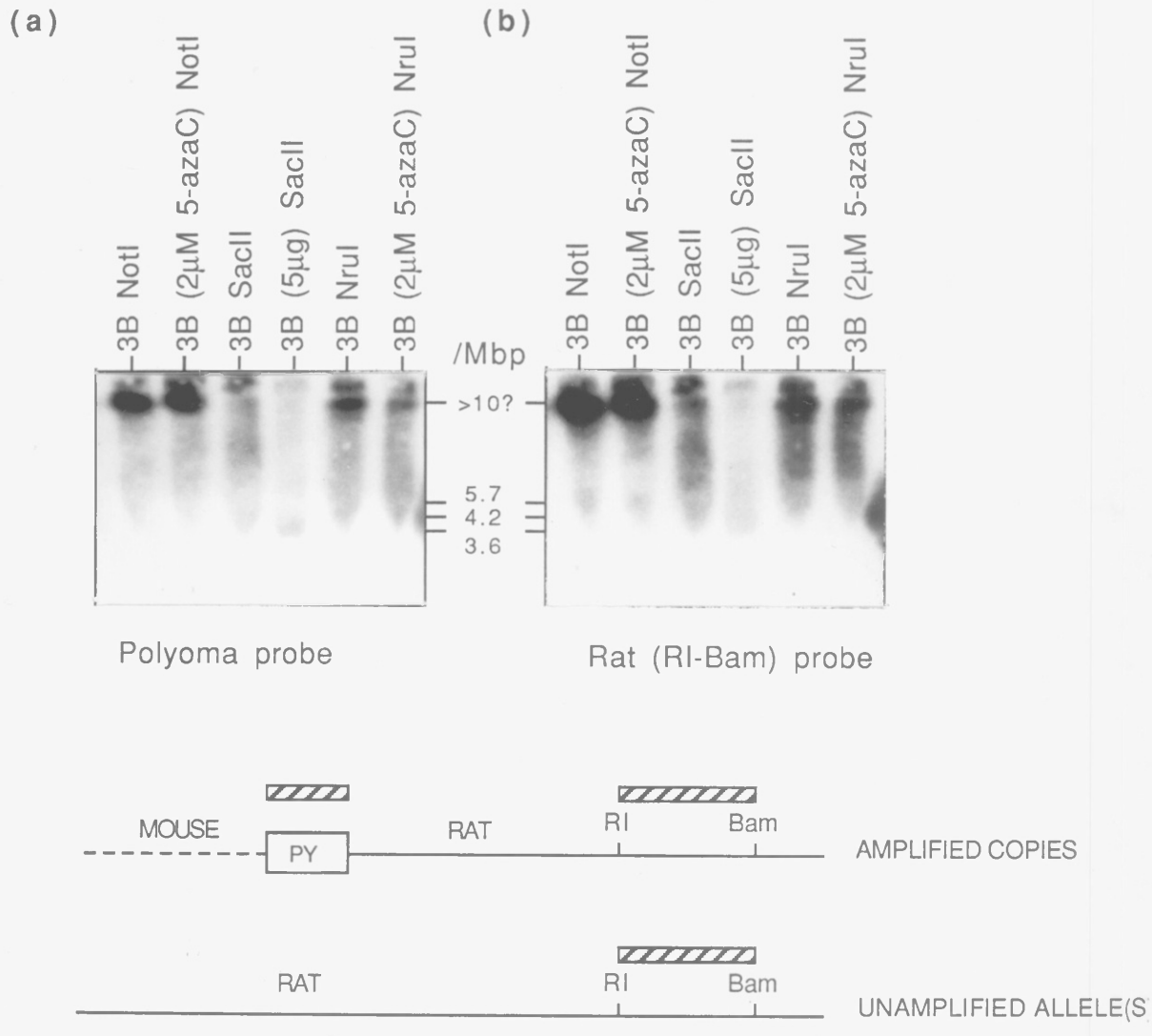
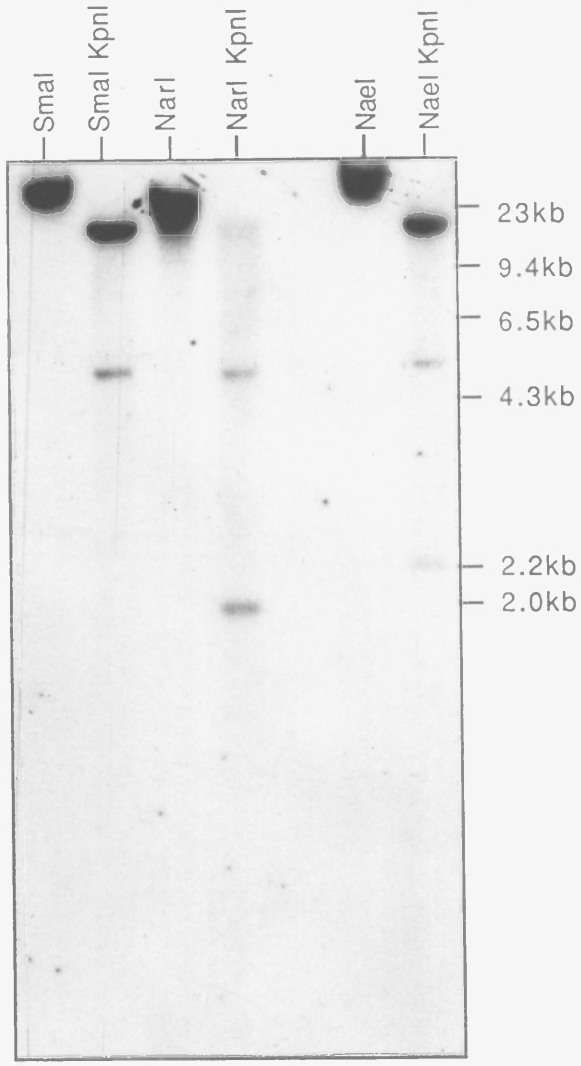

(a)
(b)
3B NotI
3B (2μM 5-azaC) NotI
3B SacII
3B (5μg) SacII
3B NruI
3B (2μM 5-azaC) NruI
/Mbp
>10?
5.7
4.2
3.6
Polyoma probe
Rat (RI-Bam) probe
MOUSE
PY
RAT
RI
Bam
AMPLIFIED COPIES
RAT
RI
Bam
UNAMPLIFIED ALLELE(S

FIGURE 3.6.2

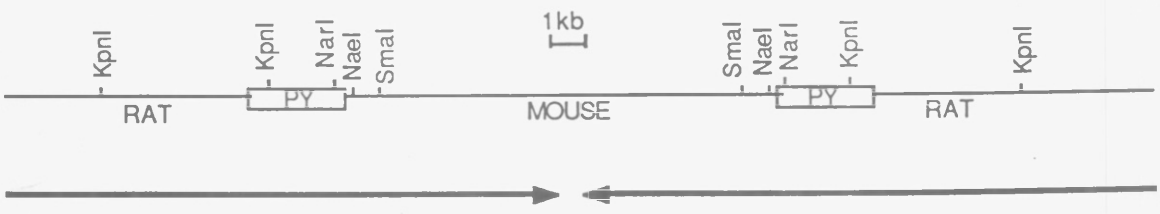
Ability of rare cutter enzyme sites at the centre of the inverted duplication to be cleaved in 3B cellular DNA.

A southern blot of 3B genomic DNA cut with rare cutter enzymes alone or in combination with KpnI is shown. Each track contains 7 μg of genomic DNA cut with the enzyme(s) indicated. For the double digestion reactions, the DNA sample was first cleaved with KpnI and then with the rare cutter enzyme (having adjusted the salt concentration appropriately). In all cases, rare cutter enzyme was used in 10-20 fold excess, in order to maximise cleavage. The sizes indicated were deduced from lambda HindII size markers.

The blot was probed with polyoma virus DNA, denoted as PY in the map below. This shows the central region of the inverted duplication, and the palindromic arms are indicated by the two arrows below the map. The positions of the sites for the enzymes used in the southern blot above are indicated. The NarI and SmaI sites had already been mapped (see Fig. 3.6.1), as had the positions of the KpnI sites (Ford et al, 1985), and their positions were confirmed by the blot above. The enzyme NaeI was found to cleave 3B genomic DNA (see above) at the positions shown on the map, but these sites could never be cleaved in cloned λ3B DNA (see Fig.3.6.1).



POLYOMA PROBE



KpnI. However, no such band is visible indicating that this site is methylated in 3B genomic DNA and is thus not efficiently cleaved. On the other hand NaeI (which had not cut λ3B DNA) did produce a faint band of 2.2 kb, in combination with KpnI. The position of this NaeI site is shown in the map in Fig.3.6.2. However, NaeI appeared to cut even genomic DNA inefficiently as a large excess of enzyme was required to produce this result. This seemed to suggest that the cleavability of this NaeI site within the mouse DNA of the inverted duplication, was dependant on the origin/condition of the DNA being used in the digestion.

Although none of the enzymes appeared to cut the known sites in 3B DNA very efficiently, these were the only sites in the cloned region which could provide feasible starting points from which to conduct a chromosome jump. SmaI is an enzyme which generates fragments with blunt ends and so would not be a reasonable choice for use in cloning. In any case, it did not appear to cut the necessary site in genomic DNA. The other two enzymes however, NarI and NaeI, which both cut to some extent, were potentially useful as cloning enzymes. For this purpose, it had to be determined whether there were any NarI or NaeI sites to jump to, further along the sequences of the inverted duplication. Such sites had to generate fragments of an appropriate size (a few hundred kilobases) for a jump to be technically possible and this was investigated by PFGE analysis. As was mentioned in Part 3.2, these enzymes did not appear to generate any resolved fragments of amplified DNA under the standard digestion conditions. However, NarI did cut the amplified DNA when excess enzyme and minimum DNA were used, as was shown in Fig.3.4.1 to generate a band of about 900kb. A jump of such a large size would be difficult technically. In any case, the results of Part 3.4 suggest that the two NarI sites involved in producing this fragment are actually at equivalent positions in adjacent inverted duplications. This jump would therefore miss the other end of the inverted dulication.

NaeI is an enzyme that is known to be particularly sensitive to the nature of the sequences around its sites (New England Biolabs Catalogue, 1989). It has been reported that spermidine (which affects the conformation of DNA) can have an effect on

FIGURE 3.6.3

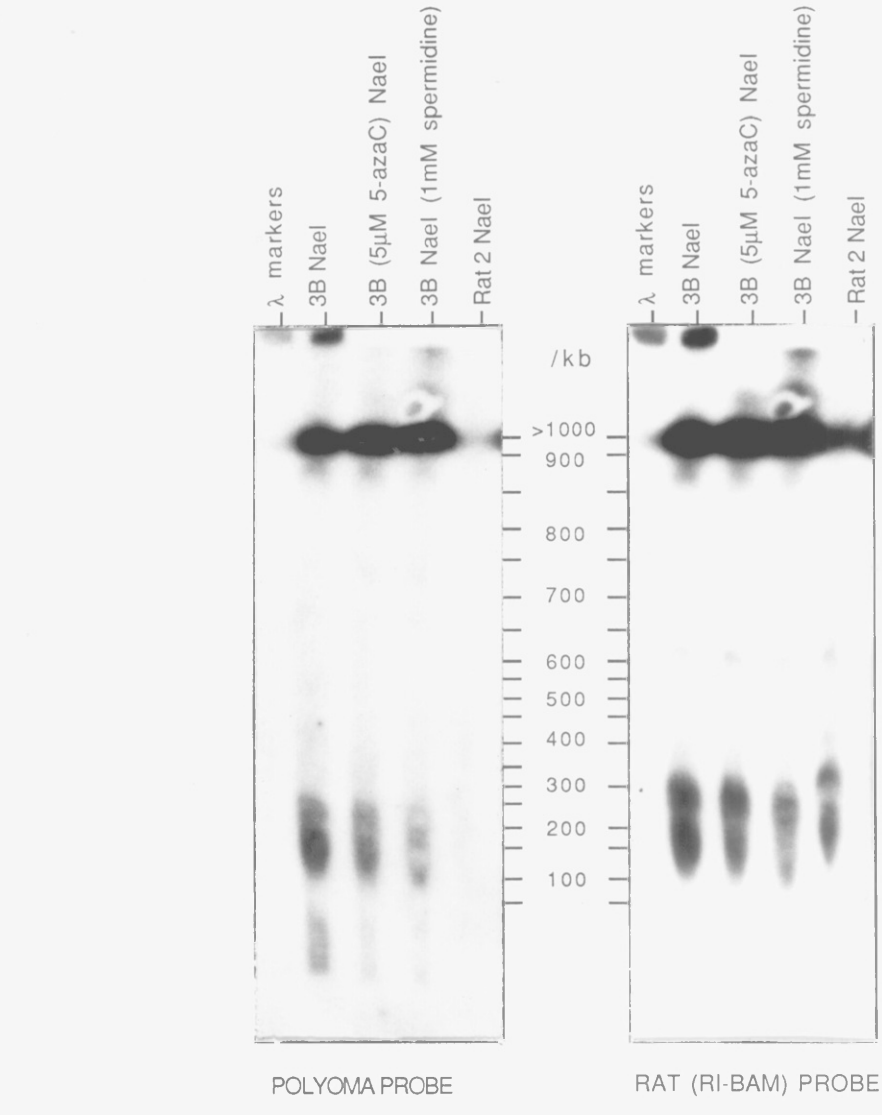
Analysis by PFGE of the NaeI sites present in the inverted duplication in 3B cells.

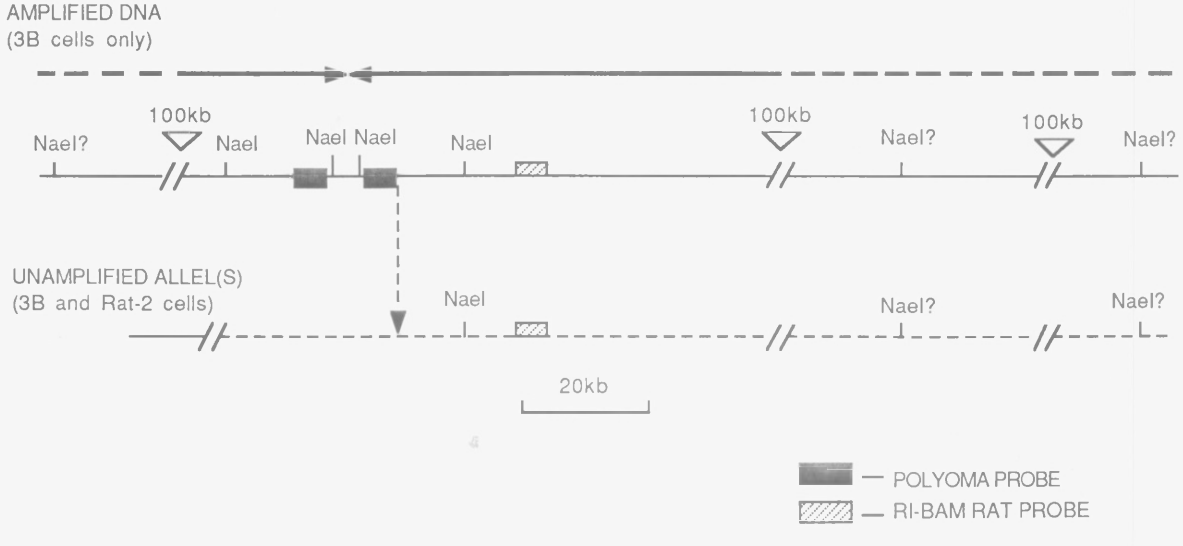
The southern blot of a pulsed field gel shown, was probed with polyoma virus DNA (left side), and with the RI-Bam rat fragment (right side). Each track contains 10 μg of NaeI-cut genomic DNA, from either 3B or Rat-2 cells as indicated. In the third track the DNA was derived from 3B cells that were grown in the presence of 5 μM 5-azacytidine. Identical NaeI digestion conditions were used for all samples, except for one, which was carried out in the presence of 1 mM spermidine, as indicated above the lane. The sizes shown were deduced from lambda concatemer size markers (see first track). The gel was run in a CHEF apparatus at 120 V for 44 hrs, in 0.5 x TBE running buffer, and using a pulse interval of 110 secs.

The maps below summarise the probable locations of NaeI sites deduced from the blots shown above, as well as in Fig. 3.6.2. The upper map represents part of one inverted duplication in 3B cells, as indicated by the arrows above. The lower map represents the unamplified allele of the region. The arrow extending from the above map indicates the point at which rearrangement occurred during formation of the inverted duplication. The NaeI sites to the right of this arrow are the same in both the amplified and unamplified DNA.

The same ladder of fragments is detected by the RI-Bam probe in both 3B and Rat-2 cells (see right hand panel above) implying that a partially cleaved NaeI site must exist between the Py insert and the RI-Bam fragment. The fact the Py probe also detects these fragments in 3B cells (see left hand panel above), implies that the nearby NaeI sites in the mouse DNA at the centre of the inverted duplication, are even more poorly cleaved than the sites in the rat DNA.

The NaeI sites that are question marked, represent possible sites that may be poorly cleaved in the rat DNA, thus accounting for the ladder of fragments of 100-300 kb detected by both the probes in the blots shown. Additional NaeI sites, not shown here, must also be present in the unamplified rat DNA to account for the larger partial fragments detected by the RI-Bam probe in both Rat-2 and 3B DNA.


λ markers
3B NaeI
3B (5μM 5-azaC) NaeI
3B NaeI (1mM spermidine)
Rat 2 NaeI
/kb
>1000
900
800
700
600
500
400
300
200
100
POLYOMA PROBE
RAT (RI-BAM) PROBE


AMPLIFIED DNA
(3B cells only)
NaeI?
100kb
NaeI
NaeI NaeI
NaeI
100kb
NaeI?
100kb
NaeI?
UNAMPLIFIED ALLEL(S)
(3B and Rat-2 cells)
NaeI
NaeI?
NaeI?
20kb
POLYOMA PROBE
RI-BAM RAT PROBE

the ability of a particular site to be cleaved by NaeI (Conrad and Topal,1989). In order to try and maximize the cleavage of the NaeI sites in the amplified DNA, the effects of both spermidine and 5-azaC (to minimise any NaeI site methylation) were tested on 3B genomic DNA, as shown in the PFG in Fig.3.6.3. Using either the RI-Bam probe or the Py probe it can be seen that neither spermidine (third track from left) nor 5-azaC (second from left) had any effect on the pattern of bands generated with NaeI. The RI-Bam probe hybridises to at least two NaeI fragments in Rat-2 DNA of about 150 kb and 300 kb, as well as to a faint ladder of larger fragments which indicate partial digestion of NaeI sites in this region. The same pattern of bands below 400-500 kb seen in Rat-2 DNA is found in 3B DNA using both the RI-Bam and Py probes. However, the Py probe also detects two additional faint bands of between 10-50 kb. This indicates that there must be another inefficiently cut NaeI site in the the rat DNA (in addition to the two mouse NaeI sites at the centre of the inverted duplication, see Fig. 3.6.2), which lies between the Py and RI-Bam probes. Hence, the majority of the amplified DNA remained unresolved on the PFG. It was concluded that NaeI would not be a useful enzyme for a chromosome jump as there were no efficiently cut NaeI sites to jump from to within the inverted duplication.

3.7 DISCUSSION.

Throughout the course of this long range mapping analysis, DNA from 3B cells that had been passaged for varying periods of time was used. The structure of the amplified DNA as well as the degree of amplification was found to be the same (as far as could be judged by Southern blot analysis) in all cases. This showed that the organisation of the amplified DNA in 3B cells was very stable, and could therefore be easily studied without the worry over possible heterogeneity occurring during the course of analysis.

Structure of the Amplified Units.

The discovery that large inverted duplications are often associated with gene amplification led to the proposal of a number of new models which would incorporate their formation into the amplification process as described in Chapter 1. These models made certain predictions concerning the structure and organisation of the amplification

units. The analysis described in this chapter of the size and homogeneity of the amplification units in the 3B cell line, as well as the overall organisation of the amplified DNA, attempted to test some of the predictions of these new models. It appears that gene amplification in mammalian cells involves extremely long stretches of DNA. This necessitates the use of long range mapping techniques to study of the structure of amplified DNA.

The amplification of a foreign (Polyoma) gene in 3B rat cells has certain advantages over other cell lines containing amplified cellular genes, in terms of analysis. The mouse and polyoma virus (Py) sequences at one end of all the amplification units in 3B provided probes which were specific for only the amplified DNA, while the rat RI-Bam probe isolated in Chapter 2 could detect the amplified copies of the region as well as the native unamplified copy. The use of the Py and RI-Bam probes in parallel was found to be extremely helpful in identifying the origins of the complex patterns of bands detected during the course of this PFGE analysis.

Long range DNA analysis with pulsed field gel electrophoresis relies on the use of enzymes which cut the mammalian genome infrequently. The sites of such enzymes are very susceptible to methylation, which can inhibit their cleavage and lead to multiple fragments as a result of partial digestion. The resulting complex patterns of bands were expected to create difficulties in the analysis of the amplified DNA. However, the use of 5-azacytidine (5-azaC) to produce hypomethylated genomic DNA enabled verification that the banding patterns observed were due to methylation of a number of sites within each unit of amplification rather than to heterogeneity between the units (Part 3.3). Indeed once this had been clarified, the partial digestion products of the enzymes SalI and XhoI were very useful for the construction of a long range restriction map of the inverted duplication (Fig. 3.3.3) and the prediction of its minimum size. The largest SalI band detected was 750-800 kb, implying that the minimum size of each palindromic arm is about 375-400 kb. The presence of either discrete bands (such as for ClaI, see Fig.3.2.1) or multiple bands due only to partial digestion (as in XhoI and SalI, see Fig.3.3.2) suggested that the units were homogeneous in structure, for at least 375-400 kb. Furthermore, the parallel analysis of DNA from the parental cell line, Rat-

2, did not detect any gross rearrangements in the region of rat DNA comprising the amplification unit. The PFGE analysis of 3B and Rat-2 genomic DNA enabled the construction of a map of the native, unamplified copy of the rat DNA (see Fig.3.3.3). Nearly all the restriction sites mapped in this region were consistent with those found in the inverted duplication. However, in the case of SmaI, one or more sites were detected which were only cleaved in Rat-2 DNA and the unamplified rat DNA in 3B but not in the amplified copies (see Part 3.5). This suggests that there may be differences in methylation between the amplified and unamplified copies of the DNA within 3B cells, as it seemed unlikely that only these sites should be absent from the inverted duplication, and none of the other sites detected in the same region (such as Sal^b, see Fig. 3.3.3). Furthermore, the (amplification specific) SmaI sites mapped in λ3B (see Part 3.6, Fig. 3.6.1) could also not be cut in 3B genomic DNA, again implying methylation of SmaI sites in the amplified DNA. Although 5-azaC treatment did not appear to have any effect on the cleavage of any of these sites, this could have been partly due to the fact that SmaI is an enzyme which cleaves DNA very inefficiently. All of the other rare cutting restriction sites found in the amplification unit appeared to be methylated to the same extents as their counterparts in the native copy of the region (eg. XhoI, SalI, ClaI in Parts 3.2 and 3.3, and NaeI, NarI in Part 3.6). In the case of EagI neither of the sites which generated the 450 kb native rat DNA fragment (see Fig.3.4.1) were cleaved in the amplified DNA. However, this was presumed to be due to the exclusion of the sites from the amplification unit (estimated to be less than 450 kb in size), rather than to their differential methylation.

Size and Organization of the Amplification Units.

The treatment of growing cells with 5-azaC, appeared to have an effect on some restriction sites that were inefficiently cut and which had led to the detection of ladders of partial products (such as with SalI and XhoI). However, hypomethylation (at the doses of 5-azaC used) did not reveal the presence of any new sites which had not previously been detected to some extent. Furthermore, 5-azaC had no effect on some

sites that were observed to be inefficiently cut such as NaeI, (see Fig.3.6.3) presumably because the problem with these sites was not methylation but the nature of surrounding sequences which prevented the enzyme from cleaving efficiently. In the case of NarI there appeared to be a number of sites in the rat sequences which were very poorly cut in both the inverted duplication in 3B genomic DNA as well as in Rat-2 genomic DNA (see Fig.3.4.1), and which were unaffected by 5-azaC treatment. In fact this was quite useful as there was an NarI site in the amplified (Py) sequence which was found to be cleaved (more efficiently than the rat sites) generating a discrete 900 kb fragment of the amplified DNA (see Fig.3.4.1). Thus, if NarI was cutting only at this site (ie at one end of each palindromic arm), the generation of a discrete 900 kb band implied that the centres of adjacent inverted duplications were 900 kb apart. In this case, each palindromic arm would be 450 kb long, (which is consistent with the size predictions of Part 3.3). Furthermore, this result indicates that the amplified DNA must be arranged as a tandem array of inverted repeats.

Both these predictions are supported by the results from the long range analysis using enzymes that did not cut within the amplified DNA (Parts 3.4 and 3.5). Such enzymes were in fact just as informative as those that did cut within the units. Using EagI it was possible to deduce the maximum size of each palindromic arm of the inverted duplication. A discrete 450kb fragment in the native copy of the region (ie present in 3B and Rat-2 DNA) was thought not to be present in the amplified copies (see Part 3.4 and Fig. 3.4.1). This strongly suggested that the inverted duplication did not contain these EagI sites. If this were the case, the size of each arm had to be less than 450 kb in length. This is consistent with the minimum size estimate of 375-400 kb derived from the SalI restriction map, as well as the prediction that the inversion centres are 900 kb apart, from the NarI map. The finding that no fragments of amplified DNA > 900 kb were ever resolved by PFGE with any of the enzymes used, also confirmed that the size of each inverted duplication was 900 kb, and furthermore that they were arranged as an array.

If the size of each palindromic arm is about 450 kb, the 20-40 fold amplification of this region would involve between 8-16 Mb of DNA. Using enzymes which did not cut

within the amplified DNA it was not possible to resolve any fragments containing amplified DNA below about 10 Mb (see Part 3.5). This suggests that all the amplified DNA is clustered together in at least one array of this size and that the inverted duplications are arranged in tandem (from the results using NarI on 3B DNA in Part 3.4), the centres being a fixed distance apart (see Fig. 3.4.4).

Strategies for Cloning the Amplified DNA in 3B cells.

These PFGE results would predict that there must be another amplified novel joint, at the other end of each palindromic arm. As the distance to the next joint was now known to be several hundred kilobases (350 kb from the part of the region already cloned), a suitable long range cloning strategy was needed. Chromosome jumping requires suitable sites to jump from and to as described in Part 3.6. It was unfortunate that the mouse region adjacent to the polyoma virus oncogene did not contain a cluster of unmethylated infrequently cutting sites (ie an HTF island). In most gene amplification systems studied, the genes amplified are "housekeeping" genes which have unmethylated HTF islands around their 5' regions. For example in DHFR gene amplification Shimada and Nienhuis (1986) have verified that all amplified copies have equally undermethylated 5' regions (and hence the genes are equally active). Sites such as these greatly facilitate the long range mapping and cloning of amplified DNA, as in the case of N-*myc* amplification (Amler and Schwab, 1989) where a cluster of unmethylated rare cutter sites at the 5' region of the gene was used in a detailed structural analysis of the amplified DNA. In the case of the 3B cell line, however, a viral (Py) "TATA"-containing gene was activated by a weak mouse enhancer. No HTF island was found to exist in either the Py or adjacent mouse sequences. The mouse sequences were found to contain two sites (SmaI and NaeI), but the SmaI site was methylated and could not be cut in genomic DNA (see Fig.3.5.2). The analysis described in Part 3.6 revealed that neither of the two possible jumping sites present in the cloned region (NarI and NaeI), could be used as they were both inefficiently cut. Forthermore, it can be seen from the map of NarI sites (see Fig. 3.5.3) that this enzyme

would not have provided a useful jump, as it cleaves (albeit inefficiently) only once per amplification unit. The inefficiency in cleavage of the NaeI site appeared to be due to its location rather than methylation, and so could not be overcome by 5-azacytidine treatment.

In the light of this information, alternative ways of cloning the amplified DNA needed to be found. A chromosome jump from a frequently cutting restriction enzyme site to one of the distally mapped rare cutter sites (such as XhoI or ClaI) might have been a possibility, but the manipulations required to do this would have been more difficult and time-consuming, and the amount of starting material necessary was far greater, thus making it a less efficient approach. Furthermore, the process of chromosome jumping meant that large regions of intervening DNA would not have been cloned and could not therefore be examined carefully for rearrangements in the amplified DNA; neither did it overcome the problems of cloning mammalian sequences in prokaryotic hosts (mentioned in Chapter 2). Instead the advent of Yeast Artificial Chromosomes (Burke et al, 1987) as vectors for cloning large stretches of genomic DNA provided a more attractive alternative.

In summary, the results presented in this chapter showed that the amplified DNA in the 3B cell line appears to be stably arranged as a single array of identical inverted duplications each of about 900 kb in size, spanning a region of more than 10 Mbp, (see Fig.3.5.3). This situation is in agreement with what has been found in all other mammalian gene amplifications so far described, where the unit size can be hundreds or even thousands of kilobases in length, and the copies are usually clustered together either in HSRs or DMs. The fact that the units of amplification in 3B appear to be homogeneous and that they are clustered in a regular array has certain implications for the possible mechanism of amplification, as was discussed in Chapter 1 and is discussed further in Chapter 6.

CHAPTER 4

4.1 INTRODUCTION:

The results of Chapter 2 illustrated a number of the problems inherent in the use of lambda 'phage vectors and a prokaryotic host to conduct a chromosome walk. However, the major drawback of this cloning system with respect to the analysis of the amplified DNA in 3B cells, was the small size of the steps taken (maximum 22 kb, but on average about 10 -15 kb) compared to the distance that would have to be cloned. As was discussed in Chapter 3, one possible way to overcome the size limitations of lambda or cosmid cloning would have been to conduct a chromosome jump. Apart from the lack of any convenient sites within the region already cloned from which to launch a jump, this method has been reported to be technically complex, and fails to isolate the DNA between the jumping sites. At around this time, a novel method for cloning exogenous DNA fragments several hundred kilobases long using yeast artificial chromosomes (YACs) had become available. This technique provided a number of advantages over the other molecular cloning methods considered, not only in the size of the fragment cloned (a minimum of 50 kb and no upper size limit, in theory), but also in the use of a eukaryotic host (yeast) rather than *E.coli* for the cloning of eukaryotic DNA.

The YAC system is based on the definition of a number of the functional units of yeast chromosomes ie centromeres, ARS's (autonomously replicating sequences, with properties expected of replication origins), and telomeres. Murray and Szostack (1983) showed that when these components were combined on an artificial chromosome of 50kb or longer, it exhibited mitotic and meiotic stability in yeast cells. In fact the larger the size of the chromosome the greater its stability, as is the case with natural yeast chromosomes (Murray et al, 1986). Based on this, Burke et al (1987) designed a YAC vector which incorporated all the necessary functions on a single plasmid that could be propagated in *E.coli*. This vector (shown in Fig. 4.1.1) supplies a cloning site within a gene (SUP4) whose interruption is easily detected phenotypically; a yeast centromere

(CEN4), an ARS, a selectable marker on each side of the centromere, and two Tetrahymena sequences which are capable of seeding telomere formation in yeast cells (TEL). Two chromosome arms are generated from the vector, as shown in Fig. 4.1.1, which are ligated on to large insert DNA molecules derived from the source DNA by partial EcoRI digestion (in the example shown) or complete digestion (with rare cutter enzymes such as NotI) and are then introduced into yeast cells.

Apart from the possibility of cloning very large DNA fragments, the other main advantage of the YAC system is the use of a yeast host. The replication mechanism of yeast should be much more compatible with the sequence organisation found in eukaryotic DNA than that of *E.coli*. In fact, the genomes of higher eukaryotes contain endogenous ARS's, unlike *E.coli* (Stinchcomb et al, 1980). In addition, yeast contains the usual eukaryotic spectrum of repetitive sequences, such as dispersed repetitive sequences, alternating purine-pyrimidine tracts, satellite simple sequences, long tandem arrays and perfect palindromes (see Burke et al 1987, for references). The existence of the last two types of sequence structures in yeast is particularly relevant to the use of YACs for cloning amplified DNA, such as that in 3B cells, which is probably in the form of a long tandem array of large palindromes. These sorts of repetitive sequences are not common in *E.coli* DNA, which is perhaps why they are found to be unstable or difficult to clone in this host (Wyman et al,1982), whereas in yeast they may be more likely to be tolerated. Indeed, an encouraging finding by Coulson et al (1988), was that in a YAC library of *C.elegans* DNA sequences had been cloned which were poorly represented in cosmid banks.

Despite the substantial biological advantages offered by YAC cloning, it has a number of technical drawbacks compared to cloning systems which use *E.coli* as a host. One of the major problems concerns the screening of YAC libraries. In YAC cloning, the primary transformants have to be embedded in agar which precludes direct colony screening. Also, they can only be put under selection for one of the two YAC markers (URA3) at this stage. The colonies therefore have to be transferred into liquid medium or on to plates for screening and in order to test for the presence of the selectable markers on both the YAC arms. Furthermore the selections for recombinants cannot be

FIGURE 4.1.1

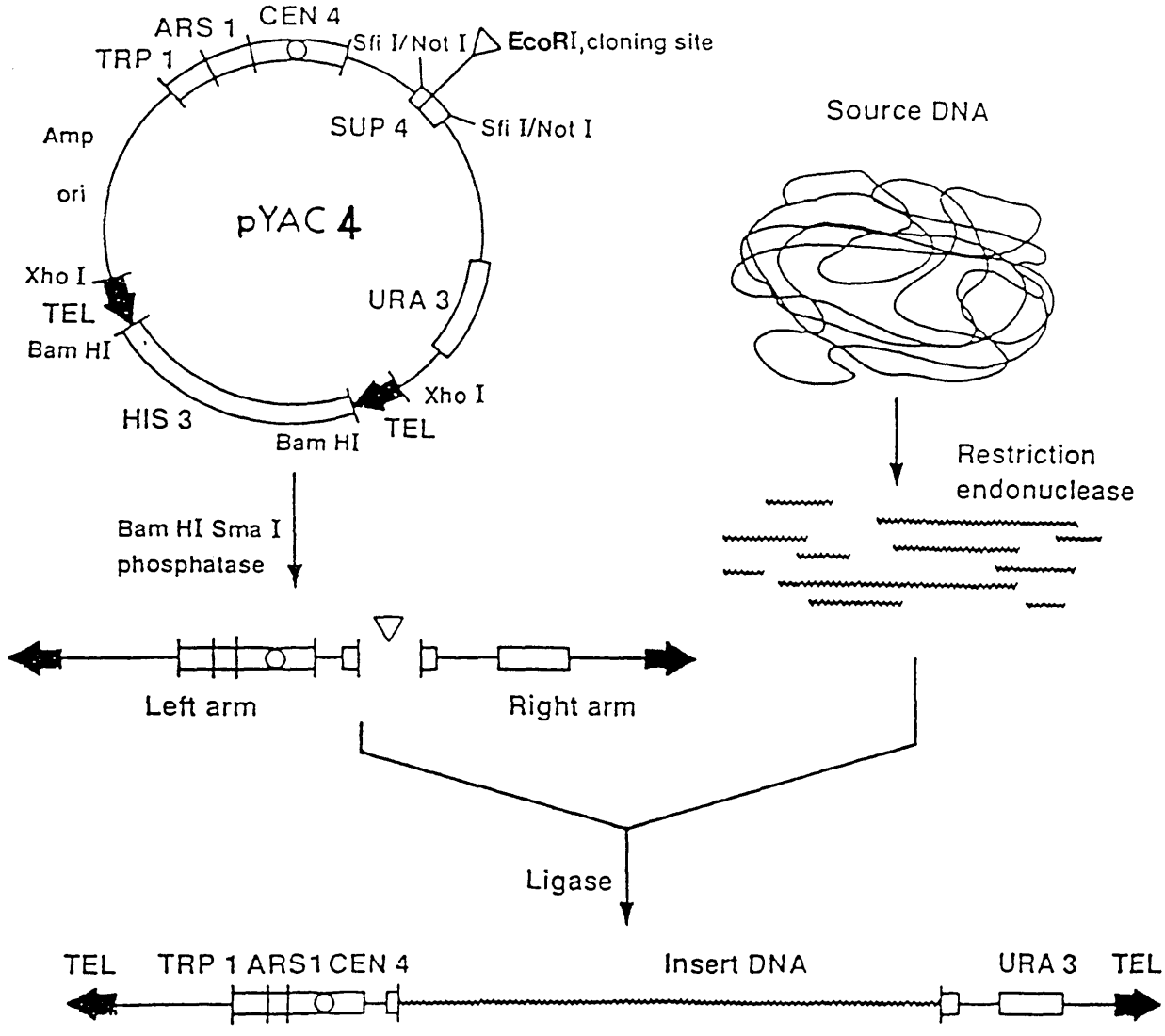
Strategy for the cloning of large fragments of genomic DNA in yeast artificial chromosomes.

A schematic representation of the procedure for cloning in yeast artificial chromosomes (YACs) is shown (adapted from Burke et al 1987). The functional units of the pYAC vector are described under Materials and Methods. The cloning site in the pYAC4 vector is EcoRI. As shown on the left, the vector is first linearized by digestion with BamHI, which exposes two telomere-seeding sequences (TEL) and releases a HIS 3 selectable marker. Two YAC arms are then generated by digestion with EcoRI. These are phosphatased to prevent subsequent vector religation. The left arm in the diagram, carries the yeast centromere and an ARS (both essential for autonomous replication), the TRP I selectable marker gene, and a TEL. The right arm carries only the URA 3 selectable marker gene, and a TEL.

On the right of the diagram, the preparation of the genomic DNA for the YAC inserts is shown. Very high molecular weight genomic DNA (prepared by a variety of possible means, see 4.1) is partially digested (in the case of EcoRI) with the enzyme to generate large DNA fragments.

The phosphatased vector arms are ligated to the restricted genomic DNA to generate a YAC (as shown at the bottom of the diagram). This must be at least 50 kb in length in order to be mitotically stable in yeast (Murray and Szostak, 1983).

The ligation products are transformed into yeast and primary transformants are selected for complementation of a ura3 marker in the host by the URA3 gene on the vector. Transformants can then be screened for complementation of a host trp1 marker which ensures that the YACs carry both vector arms, and for loss of expression of the ochre suppressor SUP4 , which indicates that the YACs are recombinant molecules (see Materials and Methods).


ARS 1
CEN 4
TRP 1
Sfi I/Not I
EcoRI, cloning site
SUP 4
Sfi I/Not I
Amp
ori
pYAC 4
Xho I
TEL
Bam HI
URA 3
HIS 3
Bam HI
Xho I
TEL
Source DNA
Restriction endonuclease
Bam HI Sma I phosphatase
Left arm
Right arm
Ligase
TEL
TRP 1 ARS1 CEN 4
Insert DNA
URA 3
TEL

used immediately at the primary transformation stage. In the case of canavanine resistance, this is because the growth conditions necessary for spheroplast regeneration rule out the presence of canavanine and the absence of arginine (see Materials and Methods) in the regeneration medium. In the case of the red/white colour phenotype, the redness of recombinant YAC-containing colonies only appears after a period of many days or even weeks. Screening by hybridisation requires that the yeast colonies are grown on filters. The YAC cloning system has the disadvantage that both the number of YAC molecules per cell (usually one) and the number of cells per yeast colony are much lower than in plasmid or cosmid cloning in *E.coli*. A further problem with this method of screening is that an extra step is required during the processing of the library filters. The yeast colonies on the filters have to be spheroplasted before their DNA can be released, and this process can be inefficient leading to underrepresentation of some colonies on the filters. In order to overcome some of these difficulties, a more rapid and efficient method for screening YAC libraries was developed (Heard et al, 1989), which was based on the use of the polymerase chain reaction (PCR) (Saiki et al,1988). This took advantage of the fact that some of the cloned region of the amplified DNA in 3B cells had been sequenced (Passananti et al, 1987). Primers based on known nucleotide sequence within the cloned region were used to screen pooled colonies by PCR. The method (see Appendix 1) thus avoided growth of the colonies on filters, or the need for hybridisation and provided a more sensitive and less time-consuming way of screening YAC libraries (Heard et al, 1989).

Apart from the technical issues which limited the practicality of YAC cloning, there were a number of unanswered questions about the system, as it had not yet been widely used. These concerned the stability of the clones, the extent to which the source DNA would be randomly sampled, and the biological activity of the cloned DNA. It is only after three years of widespread use of this system, that the answers to such questions are beginning to emerge.

The aim of the work described in this chapter was to construct a library of yeast artificial chromosomes from 3B cellular DNA, in order to clone large fragments

(>>50 kb) from the amplified inverted duplication. This would enable rapid progression along the length of the inverted duplication, and ultimately the cloning of another novel joint at the other end. One of the unanswered questions in the YAC system was whether large palindromes were stably propagated in yeast. Eukaryotic novel joints flanked by palindromic arms are notoriously difficult to clone in *E.coli* (eg. Looney and Hamlin, 1987), but the situation in yeast was not known. As one of the approaches used here involved cloning a specific DNA fragment (such as the 240 kb ClaI fragment identified in Chapter 3) which would span the complete centre of the inverted duplication, in a YAC, it was important to know whether such a structure would be stable and clonable in yeast. As a control the complete λ3B 'phage DNA was therefore cloned as a YAC in order to determine whether it could be stably propagated. Having shown this, construction and screening of YAC libraries of 3B DNA could be undertaken with more certainty that YAC cloning would be possible of the DNA around the inversion joint previously defined by Ford et al (1985) and Passananti et al (1987), as well as of the novel joint at the other end of the amplification unit which is also suspected to be in the form of an inverted duplication.

This chapter describes the development of YAC cloning from its most basic form (as outlined in Fig.4.1.1) into one adapted to overcome various technical difficulties as they were encountered, and as more information became available on ways to improve the system. The major problem during the course of this work turned out to be the inability to generate YACs with inserts of an adequately large average size. The causes for this only gradually came to light (and are still not completely understood) after a considerable amount of time and effort had been expended by many of the people involved in the construction of YAC libraries. Various different strategies to try and improve the size of the cloned DNA were used with the guidance of Z.Larin and T.Monaco from Hans Lehrach's laboratory (I.C.R.F, London). It became apparent that size fractionation was necessary for the long-range cloning potential of the YAC system to be realised.

RESULTS:

4.2 Construction of a YAC library of 3B EcoRI partial fragments, without DNA Size-Fractionation.

The YAC vectors available when this work was begun were the original pYAC vectors which were described by Burke et al (1987). The only possible cloning sites were therefore: EcoRI (pYAC4), NotI (pYAC5), SmaI (pYAC2) and SnaBI (pYAC3). The latter three were all unsuitable for the purposes of cloning the inverted duplication in 3B because no readily cleavable sites for these enzymes were detected within the region (see Chapter 3). In addition, SmaI and SnaBI both generate blunt ended fragments which are only inefficiently cloned. As this vector had already been successfully used to clone human DNA by Burke et al (1987), it was decided to construct a library of overlapping EcoRI fragments from partially cut 3B genomic DNA, using pYAC4.

The map of the vector is shown in Fig. 4.1.1. Throughout the work in this chapter, prior to the use of a YAC vector for cloning purposes, an aliquot was checked to see that its telomeres were intact by HindIII digestion (see Materials and Methods). The YAC arms were then prepared by digestion with BamHI and EcoRI, followed by two phosphatase treatments. After using some of this phosphatased DNA in a ligation reaction to check for the absence of any self-ligated products, the arms were ready for ligation to genomic DNA. These procedures are described in detail in Materials and Methods under "Yeast Artificial Chromosome Libraries".

The method used to generate large EcoRI partial fragments from 3B DNA was based on that described by Burke et al (1987), although their source DNA was prepared very differently to that used here. Their DNA preparations were in solution at all stages and size-fractionated by sucrose density gradient centrifugation prior to ligation. In this work, DNA was prepared from cells embedded in LMP agarose (as described in Materials and Methods). Cellular DNA generated by this method is much more straightforward to prepare and handle, as shearing is a major problem with HMW DNA

in solution. Furthermore, DNA in agarose was routinely being used in PFGE analysis as described in Chapter 3. PFGE analysis showed that the size of undigested DNA in blocks was at least several thousand kilobases in length. Starting DNA of this size is necessary for YAC library construction, where the aim is to clone DNA fragments which are still several hundred kilobases long following cleavage with restriction enzymes.

For the construction of this library, genomic DNA from 3B cells was prepared at a concentration of about 45 µg/80 µl agarose block (ie $5x10^6$ cells/block). To establish suitable conditions for EcoRI partial digestion, a series of reactions was set up using agarose blocks with various concentrations of the enzyme for a set time, (see Materials and Methods). This test was necessary prior to each large scale reaction as the size ranges of fragments obtained were quite variable, depending on the batch of agarose blocks and reaction solutions used. The EcoRI digestion conditions which only slightly lowered the size distribution (checked by PFGE) were chosen as optimal for the purposes of large DNA cloning, as suggested by D. Burke (personal communication). Having established the conditions (usually about 0.0001 units RI / 10 µg DNA) the large scale reaction was set up. Four or five blocks of DNA were individually digested at a time. After the reaction had been stopped with Proteinase K, the blocks were pooled together and treated in preparation for ligation to the vector arms (see Materials and Methods). Each ligation (three were made in total) usually involved 150-220 µg of partially-cut EcoRI insert DNA, and 100-150 µg of pYAC4 arms (BamHI-EcoRI cut and phosphatased).

The ligated DNA was used in this form (ie without further treatment such as agarase, or size-selection) for yeast transformations, by carefully removing and melting a 10-15 µl aliquot of it at 65°C for a few seconds prior to the addition of 100-150 µl of spheroplasts. The yeast spheroplast preparation and transformation procedures used are described under Materials and Methods. The (experimental) efficiency of each set of transformations was monitored using the YCp plasmid. At optimal conditions (about 500-1000 colonies/ng YCp), this YAC library was found to yield on average 200-300 colonies per µg source DNA (ie ligated DNA). This was in agreement with the

efficiencies reported by Burke et al (1987). In total, 900 μg of ligated DNA was generated, which in theory could provide up to $2x10^5$ colonies. However, the transformation efficiency varied considerabely from one transformation to another (for numerous technical reasons). Furthermore, only 3-4 μg of DNA could be used per transformation reaction, making the generation of sufficient numbers of colonies for screening quite a time-consuming process. It was hoped that the 20-40 fold amplification of the region to be cloned would mean that a correspondingly lower number of colonies would have to be screened than for a single copy sequence.

While the library was being plated out in steps, it was simultaneously checked for the proportion of recombinants present and the average size of the inserts. Random colonies were picked and grown up in selective liquid medium and minipreparations of their chromosomal DNA were then made in LMP agarose blocks (see Materials and Methods). The chromosomes of these yeast clones could then be analysed by PFGE. The blot shown in Fig. 4.2.1 is of one such PFG with 15 random YAC clones. The blot was hybridised with a YAC-specific probe and also with radiolabelled total rat DNA (which hybridises to both the rat DNA YAC inserts and the yeast chromosomes through cross-species sequence homology). It can be seen that nine out of the fifteen clones have a band that hybridises both with plasmid and rat probes, which represents a YAC. The different intensities of the YAC band between tracks was due to the varying quantities of DNA in each preparation. The average size of the YACs appeared to be about 50 kb. The YACs shown here are a representative sample of about 50 that were tested in this way. The library therefore appeared to contain only 60% recombinants, and the average size of the inserts was 50 kb (the minimum possible size for a stable YAC). However, as the plating efficiency of the ligated DNA was reasonably high, a library of about 6000 colonies had already been generated. It was decided to screen these colonies, estimating that if 60% of them were recombinants (ie 3600), and they contained inserts of 50kb, then assuming that this region is amplified 20-40 fold, the equivalent of 1-2 rat genomes would be screened.

The drawbacks in the screening of libraries of yeast colonies by filter hybridisation

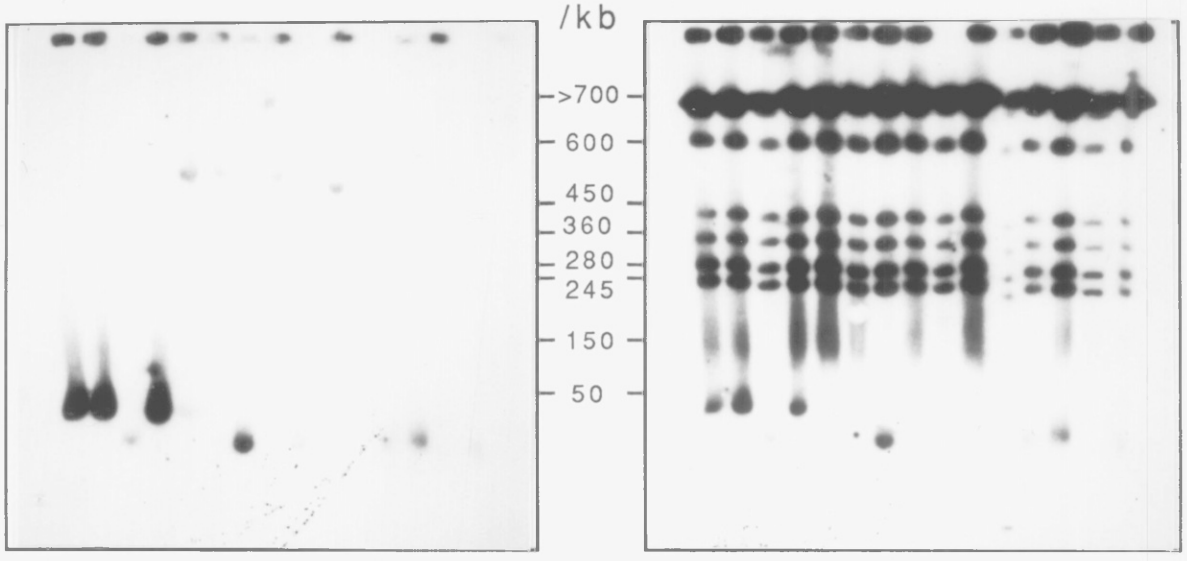
FIGURE 4.2.1

Size determination of random YACs from a library of unsize-selected, partially cleaved EcoRI fragments of 3B DNA.

The same southern blot is shown of the chromosomes of 15 yeast colonies from the YAC RI partial library, probed with YAC vector sequences on the left, and with total rat genomic DNA on the right. Cells from randomly picked YAC-containing clones were lysed in miniblocks of agarose, and their chromosomes fractionated on a pulsed field gel. Each track contains approximately 2 μg of yeast DNA. The pulsed field gel was run in a CHEF apparatus at 170 V for 24 hrs, in 0.5 x TBE using a 45 sec pulse interval. The limit of resolution under these conditions is 700 kb. The sizes indicated were deduced from both lambda concatemer markers and the known sizes of yeast (*S.cerevisiae*) chromosomes.

On the left, the YAC vector probe used for the specific detection of YACs consisted of plasmid pBR322 sequences which are present in part of the vector (see Materials and Methods). The complete YAC vector was not used as a probe because it contains yeast sequences (a centromere, an ARS and amino acid marker genes) which would hybridise to endogenous yeast chromosomes as well as YACs.

On the right, total rat genomic DNA was used as a probe to detect the presence of rat repetitive sequences in the YAC inserts. Cross-hybridisation with this probe has also visualised the yeast chromosomes.



YAC vector probe

Rat total genomic DNA probe

were described in Part 4.1. The new PCR-based method which had been developed to overcome some of these difficulties (Heard et al 1989, see Appendix 1), was therefore used to screen this library. This involved the pooling of 96 colony cultures from each of the microtitre dishes into which the library had been picked, followed by extraction of the DNA from each pool using a rapid miniprep method (so that 24 pools could be processed at a time). The DNA of each pool was then used for a single PCR with a pair of primers derived from the polyoma sequences near the centre of the inverted duplication. Only pools containing polyoma sequences (ie the amplified region in 3B cells) would therefore be expected to generate a band of the appropriate size. After testing a number of different oligomers as primers, one pair was chosen that efficiently and reproduceably generated a single PCR band of the expected size. The sequences of the oligomers used (PY2 and PY3) are shown under "PCR" in Materials and Methods. These primers produced a 2.2 kb PCR product in the presence of polyoma DNA. During screening, all the DNA generated from each miniprep (about 1 μg) was used in a single PCR, to ensure that sufficient DNA from each of the 96 clones would be present in the reaction. It was subsequently found that large amounts of miniprep DNA were inhibitory, and that using one tenth of this amount gave a much stronger PCR band. A positive band was not detected by ethidium bromide staining after running one tenth of the PCR from each pool of yeast colony DNAs on an agarose minigel. Therefore, all the PCRs of the library were rechecked by running them on large agarose gels, using half of the total PCR per slot. These gels were blotted and hybridised with a polyoma probe. Thus, if any of the pools had generated a band too faint too be seen by ethidium bromide staining, it would be detected by radioactive hybridisation. One faint band was seen by hybridisation. The appropriate dish was replica-plated onto filters and the colonies grown up and screened by hybridisation. A single colony gave a signal on duplicate filters with the Py probe. However the signal, although clear, was again faint, and it was impossible to isolate a recombinant YAC from the corresponding colony (after growing it under URA and TRYP selection, and testing for canavanine resistance).

It was therefore concluded that this 3B EcoRI partial YAC library contained no

clones for this part of the inverted duplication. One possibility was that any YAC clones containing the polyoma gene (used as a probe) were also likely to contain the nearby centre of the inverted duplication. These fragments are only amplified 10-20 fold, not 20-40 fold as the fragments from one side of the inverted duplication are. Furthermore, novel joints flanked by palindromic sequences are difficult to clone in *E.coli* as described in Part 4.1. If such structures are also unstable in yeast this could explain why this region might have been underrepresented in a YAC library. This possibility seemed much less likely following the work described in the next section (Part 4.3). However, as this library had contained a relatively low number of recombinants and the inserts had been of the minimum possible size, it was deduced that the equivalent of 1-2 rat genomes in YAC inserts that had been estimated (assuming that all sequences are equally clonable), was insufficient despite the 20-40 fold enrichment expected for the region sought.

The low proportion of recombinants may have been due to a fraction of uncut vector remaining after YAC arm preparation, and/or to insufficiently phosphatased vector fragments, which reannealed during the ligation (although this was unlikely, as double phosphatase treatment was used, and the fragments were carefully checked for religation afterwards). The small size of the inserts was thought to be mainly due to the fact that the average size of the EcoRI partially-cut starting DNA which was used in the ligation, was less than 100 kb. Partial restriction digestions of DNA in agarose blocks are very difficult to control or monitor, owing to the extra time factor required for the enzyme solution to be absorbed by the agarose, and slight variations in the quantity and quality of DNA in different blocks . A more controllable method for the partial digestion of genomic DNA, and more importantly, some form of size-selection were obviously necessary for generation of large insert DNA when using EcoRI. Such modifications were made during the construction of a second EcoRI YAC library (see Part 4.6).

At this stage, however, it seemed that one way to overcome the problem of insert size was to make use of a newly available YAC vector, pYAC-RC (Marchuck and Collins, 1988), to construct a library from totally digested ClaI or XhoI fragments, and thus

clone a specific fragment within the inverted duplication.

4.3 Stability of a large inverted duplication present as a YAC in yeast cells.

The pYAC-RC vector developed by Marchuk and Collins (1988) was derived from the pYAC3 vector of Burke et al (1987), and contains a polylinker inserted at the SnaBI cloning site. The polylinker consists of a number of unique restriction sites for enzymes that cut mammalian DNA infrequently (see Fig. 4.3.1). YAC libraries of totally cut genomic fragments can therefore be constructed, as the average size of fragments generated by such enzymes is usually in the order of 100 kb or more. The ClaI site in this new vector could therefore be used to try and clone the amplified 240 kb ClaI fragment in 3B DNA (see Fig. 3.3.3). However, this would mean trying to isolate a YAC which contained a large palindrome. Palindromic sequences are known to be fairly abundant in yeast nuclear DNA (Klein and Welch, 1980) with about 250 of them per haploid genome being estimated, but these inverted repeats appear to be at most only a few kilobases long. It was not clear that an inverted repeat of the size found in 3B DNA would necessarily be stable. To try and obtain some indication of this, it was decided to clone the complete 'phage λ3B DNA in a YAC and thus test the stability in yeast cells of an inverted duplication of at least 18 kb in size.

The complete λ3B 'phage DNA was used as an insert, as the minimum size for a stable YAC is about 50 kb and the λ 'phage arms would therefore provide the extra DNA. In addition, the lambda 'phage long arm contains a SacII site which was convenient for use with the pYAC-RC vector. The SacII cloning site in the polylinker of pYAC-RC was therefore used in conjunction with the single site found in λ 'phage DNA to clone λ3B in a YAC (the λ3B insert contained no SacII sites as shown in Chapter 3, Part 3.4). The protocol used is described in Fig. 4.3.2. One quarter of the ligation DNA (ie 0.75 μg) was used in a single yeast transformation, which generated between 10^3-10^4 colonies. Over 100 of these were randomly picked and grown on selective medium (-URA, -TRYP, + canavanine) to test for the proportion of stable recombinants. All grew, indicating that 100% of them contained recombinant YACs.

FIGURE 4.3.1

The pYAC-RC vector.

A map of the pYAC-RC vector developed by Marchuk and Collins (1988) is shown. This vector was adapted from the original pYAC vector (Burke et al 1987) by the removal of the SfiI/NotI sites flanking the SUP 4 gene, and the insertion of a polylinker into the unique SnaBI site within the SUP 4 intron. This insertion does not disrupt the function of the SUP4 gene and thus still allows either the red/white colony colour screen, or canavanine selection for recombinant molecules to be used. The polylinker contains unique restriction sites for a number of rare cutter enzymes, as indicated. In addition to these, DNA fragments from BssHII, NaeI, NarI, NruI, SmaI and XhoI digestions can also be cloned owing to their compatibility with sites in the polylinker.

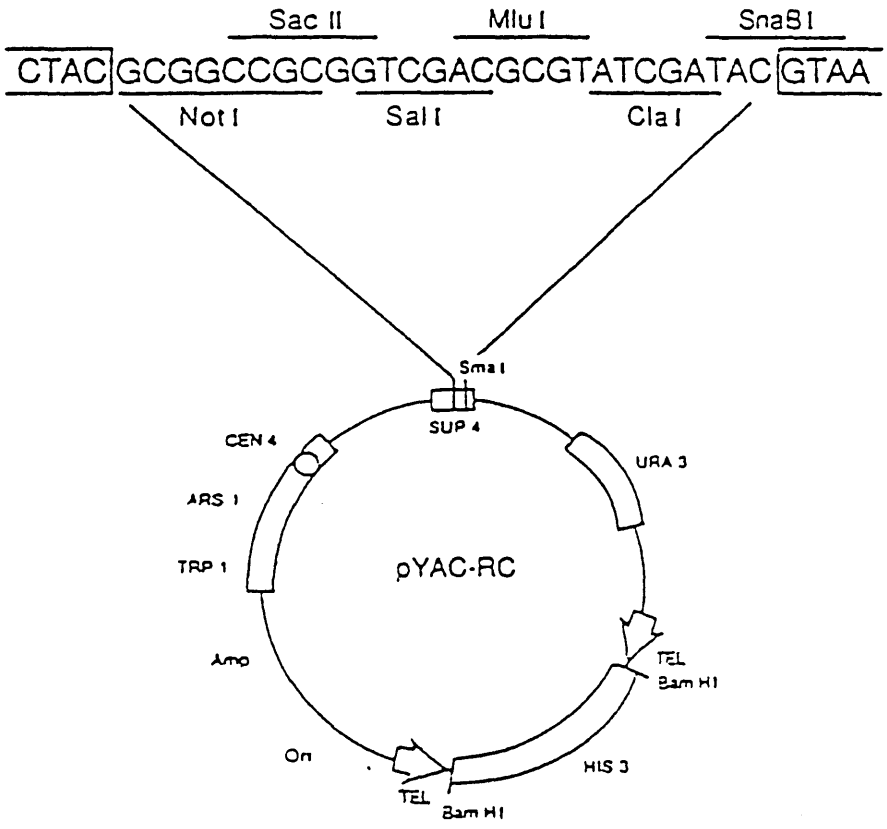

Sac II
MluI
SnaBI
CTAC GCGGCCGCGGTCGACGCGTATCGATAC GTAA
NotI
SalI
ClaI
SmaI
SUP 4
CEN 4
ARS 1
TRP 1
Amp
Ori
URA 3
TEL
Bam HI
HIS 3
TEL
Bam HI
pYAC-RC

FIGURE 4.3.2

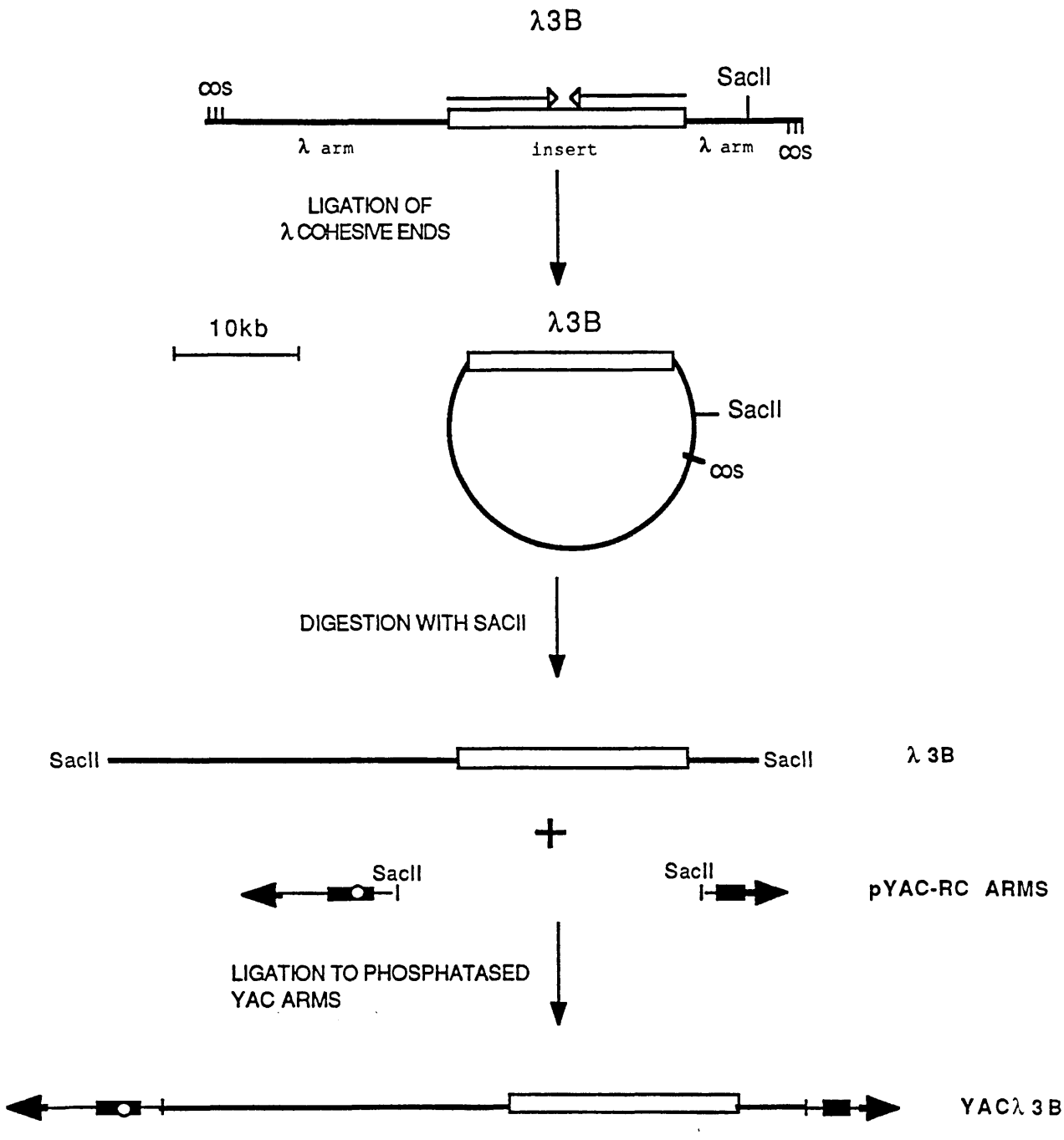
Construction of a YAC containing the complete λ3B 'phage DNA.

The diagram outlines the procedure used to insert the complete 'phage DNA of λ3B in to YAC vector arms . At the top, a simplified map of the λ3B clone is shown (see also Fig.1.1). The 18 kb insert, which is in the form of an inverted duplication, is indicated by the two arrows above. The size of the complete clone (insert + lambda arms) is about 46 kb.

The first step involves ligation of the cohesive ends (cos ends) of the lambda arms to generate circular DNA molecules. (A proportion of intermolecular ligation events will occur in addition to the intramolecular events, which may result in linear dimers as well as monomers after the next step).

The next step involves cleavage of the unique SacII site that exists in the lambda 'phage DNA. The λ3B insert contains no SacII sites (see Fig. 3.6.1), thus a single linear molecule will be generated, with SacII overhangs. This is then ligated to phosphatased YAC vector arms. The polylinker in the pYAC-RC vector contains a SacII site (see Fig. 4.3.1) which was used for cloning in this case. The two YAC arms are approximately 6 kb and 4 kb in length. Thus, a ligation involving a single λ3B insert (46 kb) will result in a YAC of 56 kb in length. A YAC of this size (> 50 kb) should be mitotically stable in yeast. A proportion of YACs containing multimers of λ3B as inserts might also be obtained through intermolecular ligation of the unphosphatased SacII ends.

Transformation of these ligation products in to yeast spheroplasts enables the investigation of the stability of large palindromes in yeast nuclei.


λ3B
cos
SacII
λ arm
insert
λ arm
cos
LIGATION OF
λ COHESIVE ENDS
10kb
λ3B
SacII
cos
DIGESTION WITH SACII
SacII
SacII
λ 3B
+
SacII
SacII
pYAC-RC ARMS
LIGATION TO PHOSPHATASED
YAC ARMS
YACλ 3B

When filter lifts were taken from this plate and hybridised with a polyoma probe, it was verified that all the colonies contained the λ3B insert. Nine of them were grown up in selective liquid culture to check the size of the YACs as well as their stability by PFGE analysis. All of the isolates grew readily in liquid culture and yeast agarose block minipreps ('yeast miniblocks') were made from them after 2 days growth, as described under Materials and Methods. The blot in Fig. 4.3.3 is of the PFG on which the yeast chromosomes from each isolate were resolved, hybridised with a polyoma virus (Py) probe. A strongly hybridising YAC band of 50-60 kb can be seen in the majority of the tracks (the size expected for a single λ3B insert), although in two or three of the tracks an additional YAC band of approximately 120 kb is seen (the size expected for a dimeric λ3B insert). This indicated that a small proportion of the double-sized insert was present in some cultures. This PFGE analysis suggested that at least one pair of palindromic arms spanning 18 kb appeared to be stable and could be easily propagated in yeast. However, although the size of the insert indicated that no gross deletions had occurred and all the isolates contained Py sequences, the integrity of the structure of the inverted duplication had to be checked. This was done for 10 isolates by digestion of the DNA from each with the enzymes KpnI as well as KpnI and BclI together, as the bands generated with these enzymes would reveal whether the structure was intact. The blot shown in Fig. 4.3.4 is of the gel on which each pair of digested YACλ3B products was run, hybridised with a Py probe. The map indicates the expected fragments generated if the structure of the λ3B clone had been maintained in the YAC. The pattern of bands was as expected in every YACλ3B isolate (see legend to Fig.4.3.4), indicating that the complete structure of λ3B was maintained intact in all isolates. As an inverted duplication of at least 18 kb could be stably propagated in yeast cells, this suggested that the cloning of a larger fragment spanning the inverted duplication (such as the 240 kb ClaI fragment) might be feasible.

4.4 Attempted cloning of the 240 kb ClaI fragment spanning the centre of the inverted duplication.

The polylinker of the pYAC-RC vector (see Fig.4.3.1) contained appropriate sites for

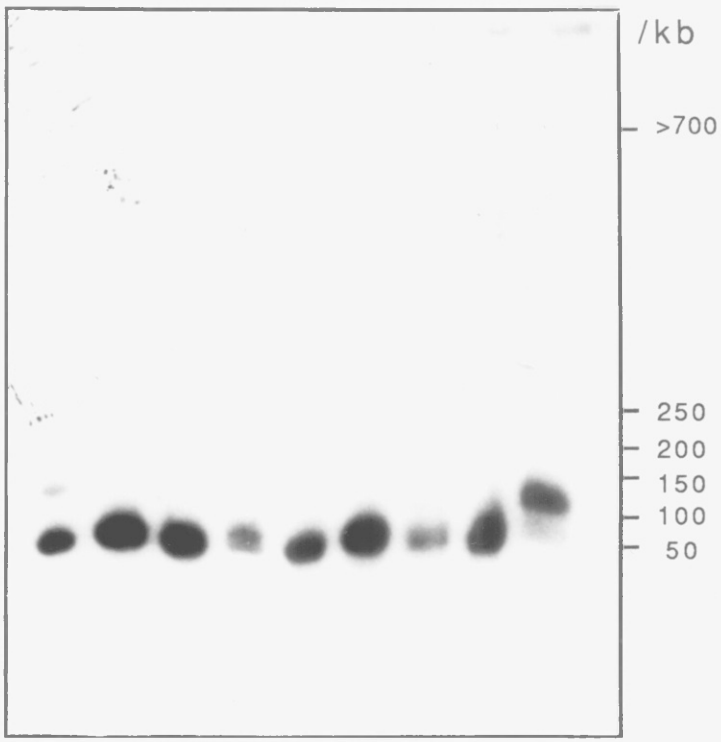
FIGURE 4.3.3

Analysis by PFGE of the size of nine independent YACλ3B isolates.

Above, a southern blot of the chromosomes from nine YACλ3B-containing yeast colonies is presented. Cells from each independent isolate were lysed in agarose miniblocks, and their DNA was resolved on a pulsed field gel. Each track contains about 2 μg of yeast DNA, and the pulsed field gel was run in a CHEF apparatus at 170 V for 20 hrs in 0.5 x TBE running buffer, using a pulse interval of 50 secs. The limit of resolution is 700 kb under these conditions. The sizes indicated were determined from both lambda concatemer markers and the known sizes of the yeast chromosomes.

Polyoma virus DNA was used as a probe in order to detect the presence and size of the λ3B insert in each YAC. At the bottom, a simplified map of the λ3B phage clone shows the positions of the polyoma DNA sequences (Py). The inverted duplicated structure of the insert is indicated by the two arrows.

YAC λ 3B



Py probe

λ 3 B



FIGURE 4.3.4

Investigation by southern blotting of the structural integrity of the λ3B DNA insert in independent YACλ3B isolates.

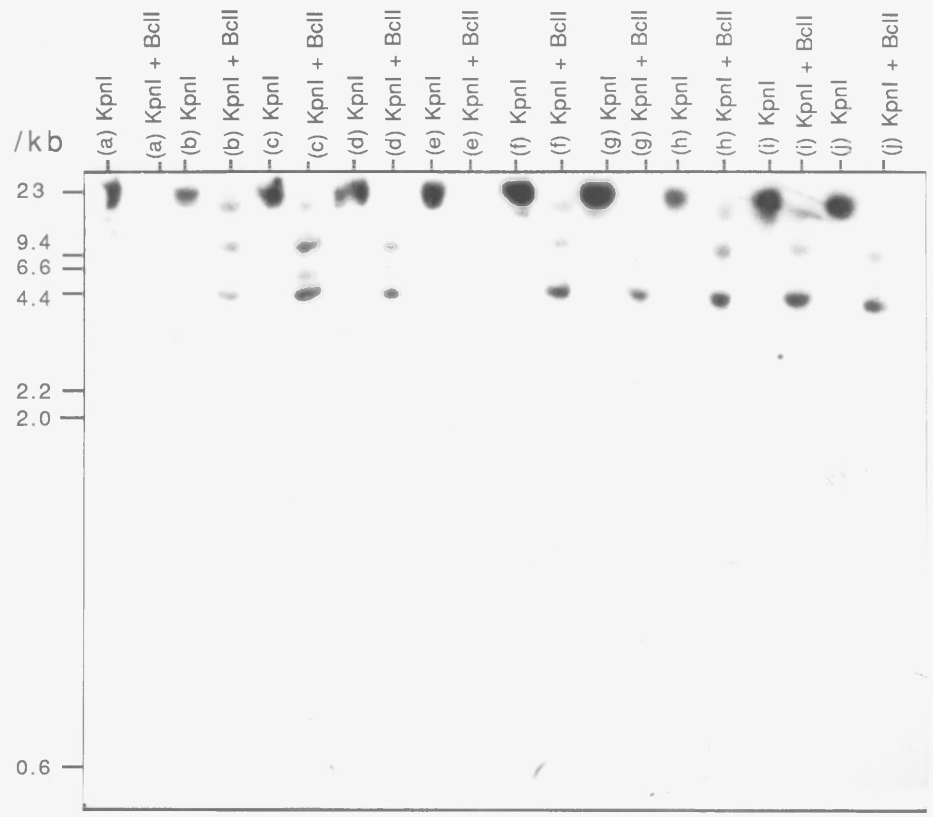
Above a southern blot is presented of restricted yeast genomic DNA from 10 independent isolates containing YACλ3B. The blot was probed with polyoma virus (Py) DNA. For each of the isolates [(a) to (j)], 2 μg of miniprep DNA was digested either with KpnI alone, or with KpnI and BClI as indicated above the lanes. The sizes indicated on the left hand side were deduced from lambda HindIII fragments used as markers.

A simplified map of the λ3B 'phage DNA is presented below. The sites for the two enzymes used above, KpnI and BClI, are indicated (as defined by Ford et al, 1985). The palindromic structure of the λ3B insert is shown by two arrows. The sizes of relevant restriction fragments which should be detected by the Py probe if the structure of the inverted duplication is unrearranged in the YAC, are indicated above the map (KpnI fragments) and below it (KpnI + BClI fragments).

Digestion with KpnI in each case, although incomplete, produces a 17.5kb band of double intensity detected by the Py probe, which spans the centre of the inverted duplication. In addition, a single copy band of ≈ 2.5kb is generated from the KpnI site in the λ long arm. The λ short arm and attached YAC arm remain unresolved (> 23 kb).

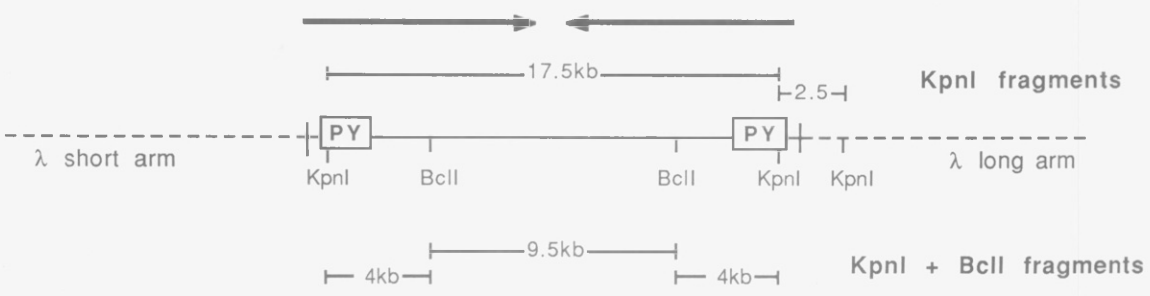
Double digestion with KpnI and BclI generates a 4kb BclI-KpnI band of double intensity, a 13.5kb band due to incomplete BclI digestion (ie 9.5kb + 4kb), a 6.5kb band (in some cases) due to incomplete KpnI digestion (from the KpnI site in the lambda long arm rather than the site in Py), and as before the single copy 2.5kb KpnI band.

The restriction fragment sizes generated by incomplete KpnI and BClI digestion of each of the 10 YACλ3B isolates shown in the blot above, thus correspond to the correct structure of the inverted duplication in λ3B 'phage DNA.



Py probe

MAP OF λ3B



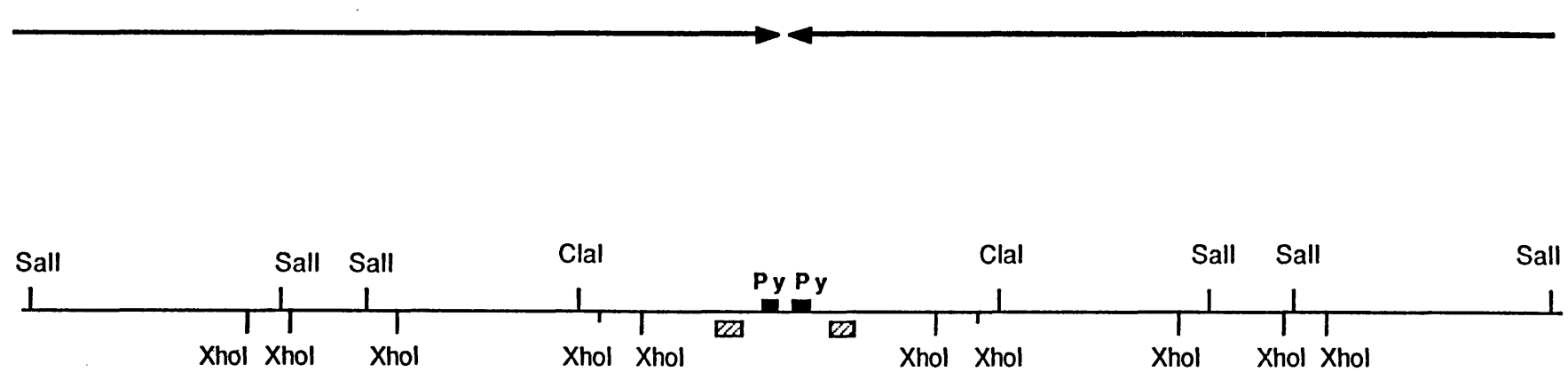
the cloning of either the 240 kb ClaI fragment or the 180 kb XhoI fragment, which were known to span the centre of the inverted duplication (see Fig.4.4.1), from the PFGE analysis described in Chapter 3. It was hoped that a directional cloning approach would avoid some of the size limitation problems encountered when trying to clone overlapping EcoRI fragments in YACs. Directional cloning means that a chromosome step of the desired size will be taken, unlike the cloning of partial restriction products where the size of a step can be unpredictably small. The majority of rat DNA fragments generated by ClaI are100-300kb in size (as detected on a PFG by ethidium bromide staining), so that a YAC library of genomic DNA fragments derived from total digestion by such an enzyme should contain YACs with inserts of a much larger average size than the EcoRI partial library made in Part 4.2. The enzyme ClaI generates a discrete fragment from within the inverted duplication in 3B genomic DNA (see Fig. 3.2.1). Thus these ClaI sites are cut efficiently enough for cloning purposes and, furthermore, the isolation of this fragment would provide a substantial chromosome step (of 120 kb) along the inverted duplication.

The amplification of the inverted duplication in 3B cells would provide a 10-20 fold enrichment for this fragment. The most straightforward cloning strategy was therefore to construct a YAC library of unfractionated 3B ClaI fragments. Genomic DNA from 3B cells in agarose blocks (at a concentration of 15-20 µg/block) was digested with ClaI and checked for complete digestion as well as no nuclease degradation by PFGE. The pYAC-RC vector was cut with BamHI and ClaI (checking carefully that there was minimum uncut vector present) and then phosphatased and tested for its inability to be ligated. The ligation reaction consisted of approximately 50 µg of 3B DNA and 50 µg of vector DNA in a final volume of 500 µl (ie DNA concentration 0.2 µg/µl). The ligated DNA was transformed into yeast spheroplasts as described for the EcoRI YAC library in Part 4.2. The transformation efficiency for this DNA was quite high, yielding about 1000 colonies per transformation (ie 300 colonies / µg of ligated DNA). Random colonies were transferred to selective plates containing canavanine (which selects for YAC recombinants) and over 90% were found to grow. Thus there would be over 90%

FIGURE 4.4.1

Long range restriction map of ClaI, XhoI and SalI sites within the amplified inverted duplication in 3B cells.

LONG RANGE MAP OF INVERTED DUPLICATION IN 3B CELLS.



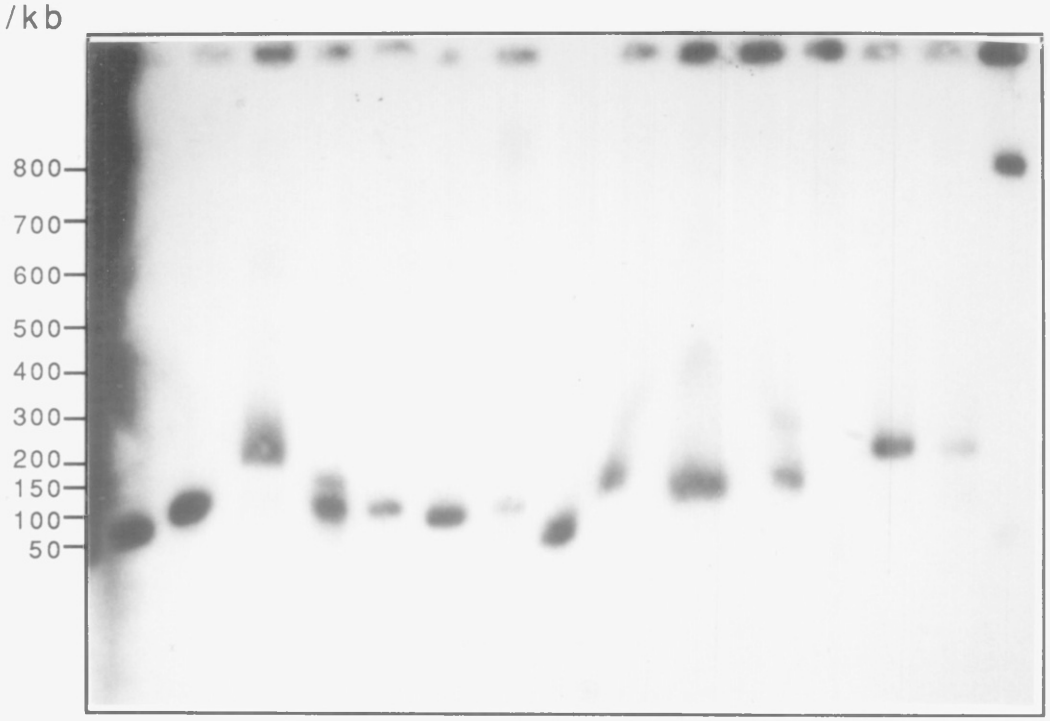
recombinants in the library, and this was confirmed by PFGE of more than 30 randomly picked colony minipreps. Fifteen of these yeast minipreps can be seen in Fig. 4.4.2. The blot of the PFG was hybridised with plasmid vector probe and in each track a single YAC can be seen. However, the average size of these YACs was 100 kb, although it had been expected that they would range from 100 kb to 300 kb, as the majority of fragments of rat DNA digested to completion by ClaI appear to be in this size range by PFG analysis. Although the transformation efficiency of the ligated DNA and the porportion of recombinants in it were both very good, it would not have been worthwhile transforming the rest of it into yeast and screening for the 240 kb ClaI fragment, as the average insert size was well below this. The 10-20 fold enrichment of this amplified region had therefore been counter-balanced by the lower than predicted average size of the inserts. As will be discussed later, it was concluded that preferential cloning of YAC inserts of 50 - 100 kb occurred when unfractionated DNA was used.

It was therefore decided to try and size-select the ClaI DNA to be cloned. The extra preparative steps for this would involve more handling of the DNA and thus might increase the risk of DNA shearing during the construction of the library. Therefore great care had to be taken in the planning of the size fractionation method to minimize the number of steps. The approach chosen was to isolate ClaI-digested DNA in the correct size range (around 240 kb) from a PFG, before its ligation to the vector. This form of size fractionation is described in Materials and Methods. Eight agarose blocks of DNA (15-20 μg/block) were digested with ClaI, and after stopping the reaction they were loaded into a preparative 1% LMP agarose gel (0.25 x TBE), which was run for 16 hours at 120 V, 11^{0}C, using a pulse interval of 50 seconds. A gel slice containing one track of ClaI cut genomic DNA and the λ size marker track was cut away and stained with ethidium bromide so that the size range of DNA in different regions of the gel could be identified. Five slices of agarose across the width of the gel, containing ClaI fragments in the size ranges: 50-150 kb (fraction 1), 150-250 kb (fraction 2), 250-400 kb (fraction 3), 400-600 kb (fraction 4) and > 600 kb (fraction 5), were isolated. All the steps were conducted at 4°C using sterile instruments throughout to try and minimise the shearing of the DNA. The gel slices were equilibrated to an appropriate

FIGURE 4.4.2

Size determination of random YACs from an unsize-selected library of ClaI-cut 3B DNA.

The YACs in 15 randomly picked colonies from a library of unfractionated ClaI fragments can be seen in this southern blot of a pulsed field gel. The blot was probed with YAC vector DNA, which selectively hybridises to artificial and not endogenous yeast chromosomes. Each track contains 2 μg of yeast (agarose miniblock) DNA. The pulsed field gel was run at 170 V for 18hrs, in 0.5 x TBE running buffer, with a pulse interval of 55 secs. The limit of resolution is 800 kb under these conditions. The sizes indicated were deduced from both lambda concatemer size markers and the known sizes of the yeast chromosomes.



YAC vector probe

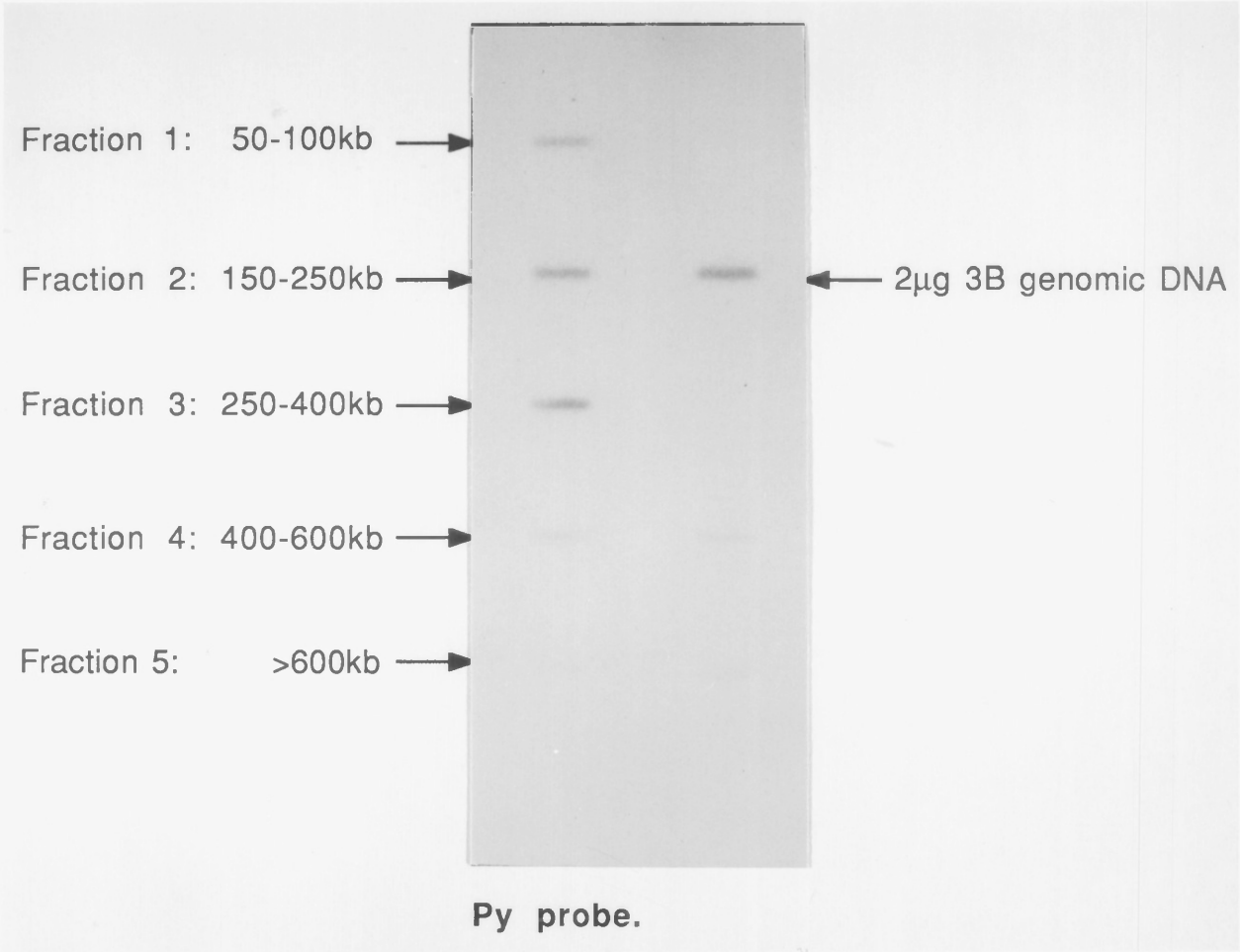
salt concentration prior to melting of the agarose at 68°C (to prevent melting of the double stranded DNA). The molten agarose slices were then transferred to 37°C and treated with agarase overnight. In order to determine which fraction contained the maximum amount of the desired 240 kb ClaI fragment, aliquots of each were analysed using the "Slot Blot" technique (see Materials and Methods). It was estimated that the DNA concentration in each fraction was about 15-20μg/ml. Thus, 0.2 ml (about 3.5 μg) of each was loaded per slot, in addition to 2 μg of 3B (uncut) DNA as a standard and was hybridised with a Py probe (Fig. 4.4.3). The most intensely hybridising fractions were 2 (150 -250 kb) and 3 (250 - 350 kb), and the intensity of the signals for these fractions compared to the 3B standard verified that the amount of DNA was approximately what had been estimated. Fractions 2 and 3 were thus used in ligations with the ClaI cut YAC vector arms. Each ligation consisted of 100 μl (ie approximately 1.75 μg) of fractionated DNA and 3μg of vector arms in a final solution of 200 μl. Initially, 15μl of the ligation reaction was used per transformation as before, but as the concentration of the DNA (about 24 ng/μl) was 10 times lower than it had been in the total ClaI YAC ligation the number of colonies produced per transformation was correspondingly lower. However, increasing the volume (> 20 μl) of ligated DNA in a transformation resulted in a drastic decrease in the transformation efficiency. This was thought to be due to the osmotic shock experienced by the spheroplasts in a diluted buffer solution (see "Yeast Transformation" in Materials and Methods). To overcome this, spheroplasts in 2M Sorbitol solution were used at the DNA transformation stage, enabling up to 100 μl of ligated DNA to be mixed with 100-150 μl of spheroplasts without any substantial decrease in transformation efficiency. The efficiency was in fact about 700-800 colonies per 100 μl ligation mixture, ie 300 colonies per μg DNA, which was as high as had been previously achieved. The reduced quantity of ligated DNA available (compared to the non-size selected libraries) was not expected to be a problem at this efficiency of transformation, because of the enrichment provided by the size fractionation as well as the amplification of this region. Random colonies were picked from the YAC transformants of each of the fractions and agarose block

FIGURE 4.4.3

Slot blot analysis of size-fractionated ClaI-cut genomic DNA from 3B cells.

A slot blot of various size fractions of ClaI-cut 3B cellular DNA, probed with polyoma virus DNA is presented. The slots on the left each contain about 3.5 μg of DNA from one of four fractions (as indicated) containing the different ClaI fragment size ranges. The slot on the right contains 2 μg of undigested 3B genomic DNA as a standard.

The Py probe hybridises to the polyoma sequences at the centre of the 240 kb ClaI fragment which spans part of the inverted duplication (see Fig. 4.4.1) in 3B cells. This fragment is present mainly in fractions 2 and 3, as expected.


Fraction 1: 50-100kb
Fraction 2: 150-250kb
Fraction 3: 250-400kb
Fraction 4: 400-600kb
Fraction 5: >600kb
2µg 3B genomic DNA

Py probe.

minipreps of their chromosomes were run on a PFG and hybridised with a plasmid probe to reveal the YACs (see Fig. 4.4.4). It can be seen that there appear to be 100% recombinants in each ligation. However, although fraction 2 supposedly contained DNA in the range of 150 - 250 kb, the average size of YAC insert was only 50 - 100 kb; and in the case of fraction 3, the size range that had been selected was 250 - 400 kb, yet the resulting average size of the YAC inserts was only 150 - 200 kb. Although these were the two fractions that had contained most of the 240kb ClaI fragment (see Fig.4.4.3), it seemed unlikely that a fragment of this size would be present in the YAC transformants produced by either of the ligations containing these fractions. The possible reasons for the lower than expected size are discussed in Part 4.7. However, it had become apparent that the upper size limit of YACs capable of being transformed when prepared by such methods (involving agarose at every stage) was only in the order of 200 kb (see above; Z.Larin and T.Monaco, personal communication). The library of size-fractionated ClaI fragments was therefore not screened.

4.5 Attempted Cloning of the 180 kb XhoI Fragment Spanning the Centre of the Inverted Duplication.

The apparent difficulty in obtaining YACs over approximately 200 kb in length using the available methods eliminated the possibility of cloning the discrete 240 kb ClaI fragment. It was decided that the XhoI fragment of 180 kb which spans the inverted duplication (see Fig. 4.4.1)might be a more reasonable choice for cloning as it was within the size range possible for YAC cloning.

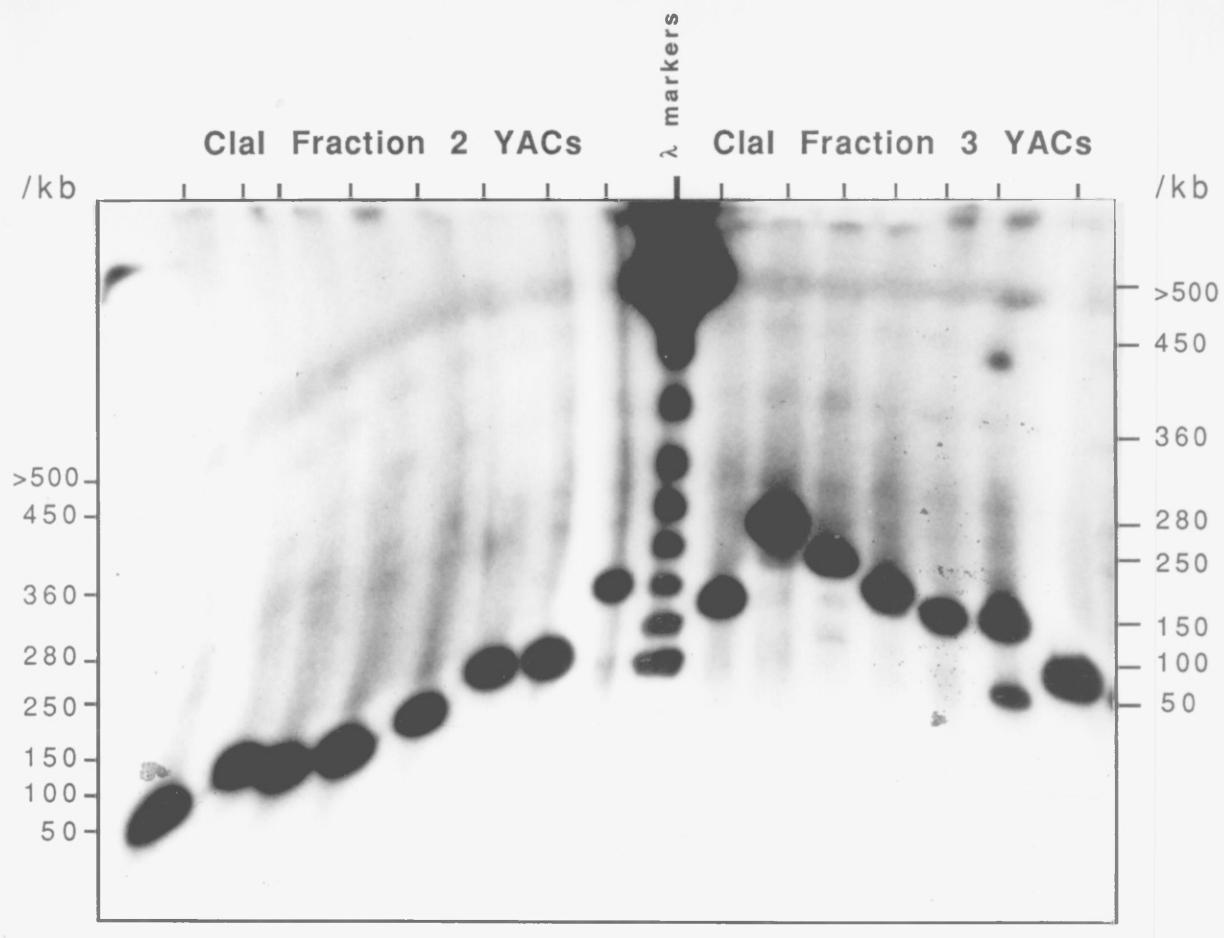
However, a number of additional factors had to be considered during the construction of a YAC library of XhoI fragments from 3B DNA. As the 180 kb fragment was only the smallest in a series of XhoI partially digested fragments (as described in Chapter 3), construction of a YAC library of complete XhoI fragments had the advantage that it would contain further XhoI fragments of a clonable size (about 50 -150 kb) for a continued chromosome walk along the length of the inverted duplication, after the initial 180kb fragment had been isolated. Also, construction of a library of unfractionated DNA involved much less handling and possible degradation, as mentioned before. On

FIGURE 4.4.4

Size determination of random YACs from a library of size-selected ClaI-cut 3B DNA fragments.

The YACs of randomly picked colonies from libraries of ClaI fragments made with size-fractionated genomic DNA can be seen on the southern blot presented. The blot was probed with YAC vector DNA which selectively hybridises to artificial rather than natural yeast chromosomes. The eight tracks on the left each contain 2 μg of (agarose miniblock) DNA from colonies in the YAC library of "fraction 2" ClaI fragments (see Fig. 4.4.3), and the seven tracks on the right contain DNA from colonies in the "fraction 3" ClaI YAC library. The lambda concatemer size markers are shown in the middle, and the sizes along each side of the blot were deduced from the sizes of yeast chromosomes. Note that, as this pulsed field gel was run in an OFAGE apparatus, the tracks are curved and the chromosome sizes are thus not parallel across the width of the blot. The gel was run at 350 V for 20 hrs, in 0.5 x TBE running buffer using a pulse interval of 30 secs.

Although "fraction 2" ClaI fragments were isolated from DNA in a size range of 150-250 kb, and "fraction 3" from DNA in a size range of 250-400 kb, from this blot it can be seen that YACs made from fraction 2 are only 50-150 kb in length, and YACs from fraction 3 are only 100-250 kb in length.



YAC vector probe

the other hand, however, the enrichment through amplification of this region in 3B cells would be lower than in the case of the ClaI library attempted, as the proportion of 180 kb XhoI fragments would be reduced through partial cleavage. To compensate for this, a size fractionated XhoI library was also attempted in order to enrich for the 180 kb XhoI fragment. A final consideration in the construction of an XhoI YAC library was that the vector did not contain an XhoI cloning site in the polylinker. The SalI cloning site was therefore used as this provides YAC arms with compatible ends for XhoI ligation. However, pYAC-RC and other YAC vectors (see Fig, 4.1.1) contain two endogenous XhoI sites, one at the border of each of the TEL sequences (ie the telomeres). Thus digestion of a vector with XhoI would cut away the telomeres from the arms. It was therefore crucial to ensure that the XhoI enzyme used to digest the genomic DNA was thoroughly inactivated before combining it with the vector arms in a ligation reaction.

Agarose blocks of 3B DNA at a concentration of 15-20 µg/block were individually digested with XhoI. A single block was checked for complete digestion and no DNA degradation by PFGE prior to using the others for library construction. The blocks were each washed once in T.E for 30 mins, and then each was transferred into 100 µl of T.E containing 1mg/ml Proteinase K and 50 mM EDTA and incubated for at least 30 minutes at 37°C. This would inactivate the XhoI enzyme. The blocks were then washed and treated with PMSF twice to inactivate the Proteinase K.

The pYAC-RC BamHI-SalI phosphatased vector arms were prepared and tested as described for previous YAC arm preparations. For the YAC ligation involving complete XhoI digestion products, the reaction was set up using identical vector and insert quantites of DNA as described for the complete ClaI library. The ligated DNA was either transformed directly into yeast spheroplasts or treated overnight with agarase (see "Preparation of Ligations for Yeast Transformation" in Materials and Methods), as this had been found to increase the efficiency of transformation. The transformation efficiency of this YAC ligation was found to be very high, producing between 300 and 400 colonies per µg ligated DNA. Using the canavanine resistance test, almost 100% of transformants were found to contain recombinant YACs. This was consistent with the

finding that transformation of spheroplasts using the self-ligated vector arms alone (phosphatased and then self ligated) gave a very low transformation efficiency (20 colonies/μg DNA), indicating very little background uncut or unphosphatased vector present in the ligation.

In the case of the size selected XhoI YAC ligation, fractionation of the DNA fragments was carried out by essentially the same procedure as that described for ClaI. Only the fractions in which the majority of the polyoma-containing fragments were seen by slot blot analysis were used in ligation reactions (see Fig.4.4.3).

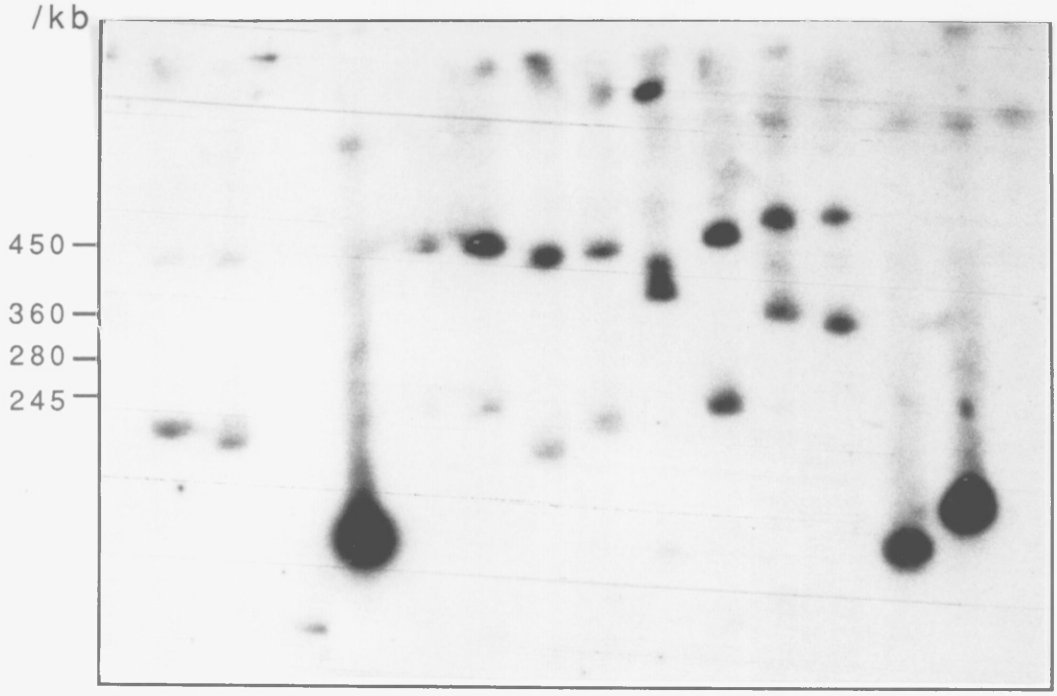
The resulting transformants from both these ligations were tested by PFGE on minipreps of randomly picked colonies. Fig.4.5.1 shows a blot of one of many such PFGs. Only a small minority of the YACs (hybridised with a plasmid probe) give the expected intense single band seen in all the previous YAC libraries tested. The rest of the tracks each contain a number of weak bands, the lowest of which might represent a true YAC, in the size range 50-150 kb, with at least one higher molecular weight band, usually about 300-400 kb in size. Fig. 4.5.2 is a blot showing further YAC XhoI minipreps, but in this case it was hybridised first with a plasmid probe and then with radio-labelled total rat DNA to verify that the the lower bands in each track contained rat DNA, and to enable detection of the yeast chromosomes. Again the majority of the YACs seem to be represented by multiple faint signals rather than one intense signal. From the comparison using the two probes it was possible to see that the upper bands detected by the plasmid probe corresponded to natural yeast chromosomes, the lower band probably being the autonomous YAC. In such cases the YAC appeared to be unstable, and integration of part of it (including the selectable markers) into a yeast chromosome in a proportion of the cells presumabely allowed its propagation. Thus the majority of these YACs were integrating into yeast chromosomes leaving only a minority of cells that contained an autonomous artificial chromosome. This phenomenon will be discussed in Part 4.7. As there was no obvious cause for this instability, use of the pYAC-RC vector had to be abandoned. Obviously, the XhoI YAC ligations which had been prepared could not be used, as only truly autonomous

FIGURE 4.5.1

Analysis by PFGE of random YACs from a library of unsize-selected XhoI-cut 3B DNA fragments.

A southern blot of the chromosomes of 15 randomly picked YAC-containing colonies, from a library of unfractionated XhoI fragments, has been probed with YAC vector DNA. This probe should selectively hybridise to artificial chromosomes, as it consists of pBR322 sequences which are only present in the YAC vector and not in natural yeast chromosomes.

Each track contains 2 μg of yeast (agarose miniblock) DNA. The pulsed field gel was run in a CHEF apparatus at 170 V for 22 hrs, in 0.5 x TBE, using a pulse interval of 45 secs. The limit of resolution is 600 kb under these conditions. The sizes indicate the positions of the four smallest natural yeast chromosomes.



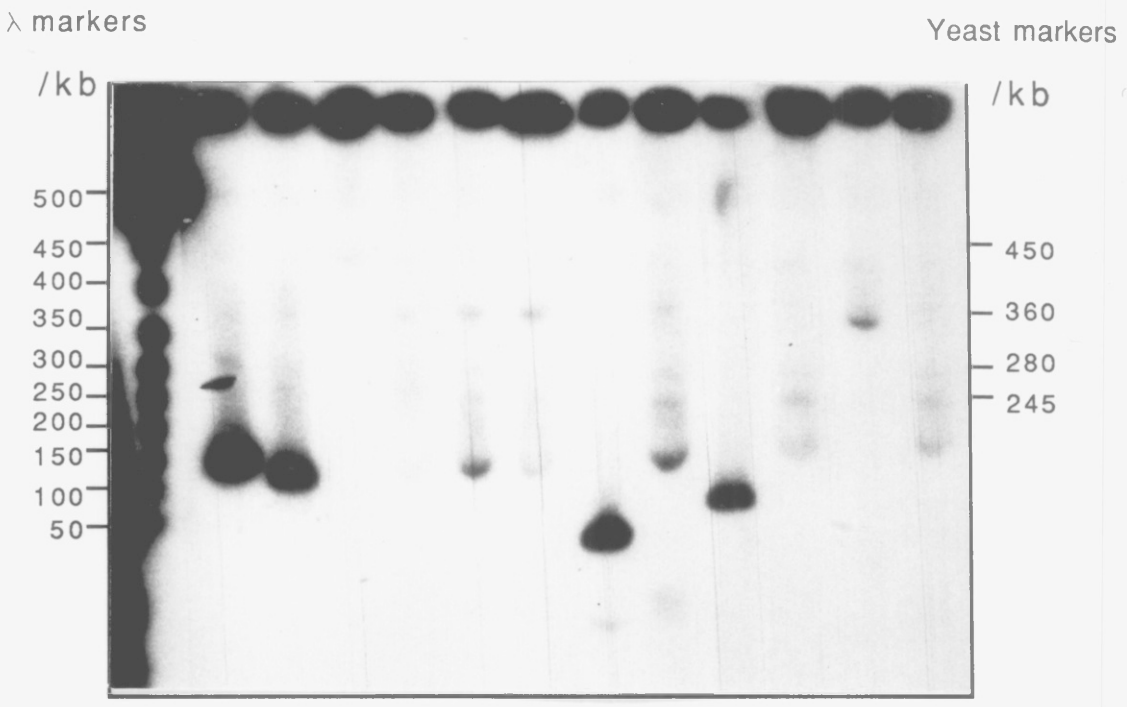
YAC vector probe

FIGURE 4.5.2

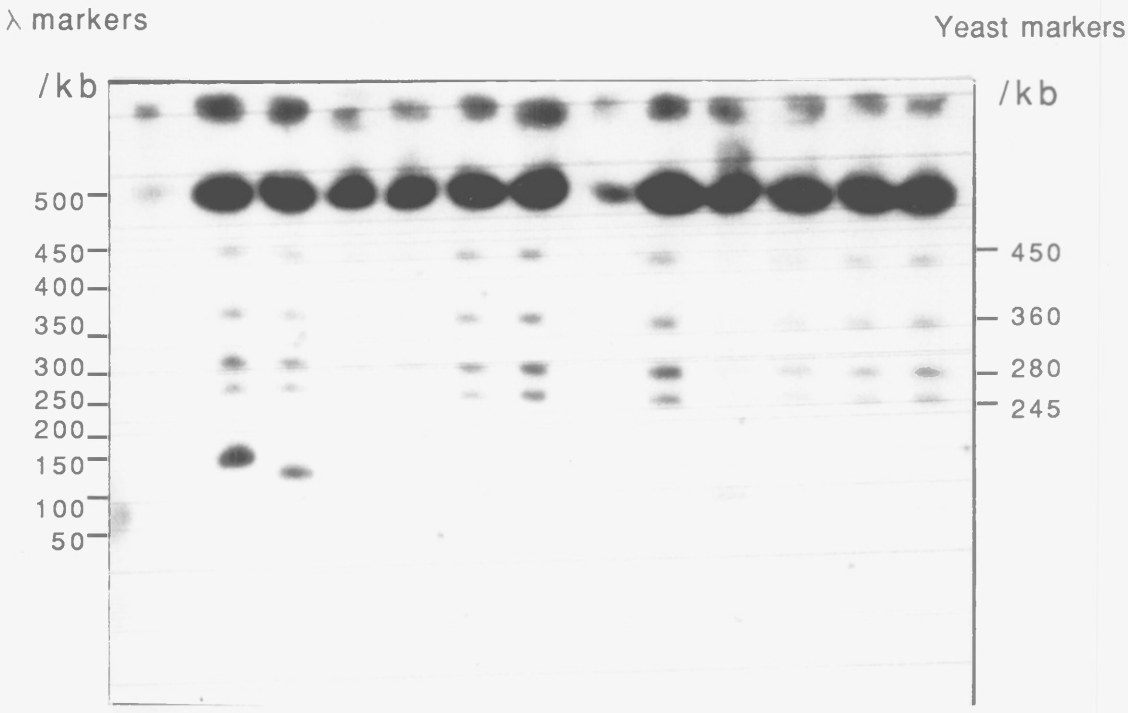
Further PFGE analysis of YACs from an unsize-selected library of XhoI-cut 3B DNA fragments.

A southern blot of the chromosomes from 12 randomly picked colonies from the YAC library of XhoI fragments, has been hybridised with a YAC vector probe (shown above) and with a total genomic rat DNA probe (shown below). The tracks each contain 2 μg of yeast DNA, and this pulsed field gel was run under identical conditions to those described for the previous figure (Fig.4.5.1). The sizes shown were determined from lambda concatemers (shown on the left) and natural yeast chromosomes (on the right).

The YAC vector probe (pBR322 sequences) should selectively hybridise to artificial chromosomes. The total rat DNA probe hybridises to the rat sequences in the YAC inserts, and it also cross hybridises to the repetitive sequences in the yeast chromosomes. The YAC vector probe hybridises to a number of faint bands in addition to the lowest (presumed) YAC band in some tracks in the upper blot. By comparing the same tracks in the lower blot, the faint bands in the upper blot can be seen to correspond to the positions of natural yeast chromosomes in every case.



YAC vector probe



Total rat genomic DNA probe

artificial chromosomes would yeild useful cloned inserts.

4.6 Construction of a size-selected YAC library with EcoRI overlapping fragments of 3B DNA.

Shortly after these attempts to clone the ClaI or XhoI fragments of the amplified DNA using the pYAC-RC vector, some modifications of the YAC cloning protocol were developed and tested by Z. Larin and T. Monaco in Hans Lehrach's laboratory. These modifications enabled them to generate YAC libraries of EcoRI fragments with a much larger average size than before (over 200 kb). Furthermore, the pYAC4 vector used for the construction of EcoRI libraries, was known to be reliable for the production of stable YACs, unlike the pYAC-RC vector (see Part 4.5).

One of the modifications developed was in the technique used to generate EcoRI partial fragments. The new method (see Materials and methods) ensured much greater control of the reaction and the consistent production of fragments within the same size range in independant EcoRI digestions. It involved the use of a large excess of EcoRI methylase simultaneously with the EcoRI enzyme. This limits the extent of the DNA restriction, as (random) methylation of EcoRI sites prevents EcoRI from cleaving.

The second modification was aimed at increasing the average size of the YACs in the resulting library and involved a size-selection step which was performed after the ligation reaction. In fact this step had two advantages over the size-selection of genomic DNA prior to ligation: it allowed elimination of actual YACs which were too small, so that only YACs in the desired size range would be transformed into yeast; in addition, size-fractionation eliminated any uncut or self-ligated vector, and this would theoretically decrease the proportion of non-recombinants in the resulting library. The combination of these two factors greatly reduced the number of YACs that would have to be screened for a particular sequence compared to the previous EcoRI YAC protocol used, (as described in Part 4.2). Theoretically only one or two thousand yeast colonies would have to be screened for a sequence that was amplified 20-40 fold in the rat genome. It was therefore decided to make one final attempt at cloning more of the amplified DNA from 3B in a YAC using the modified protocol.

A YAC library of overlapping fragments had the advantage that it could be used repeatedly to take chromosome steps, unlike a library made from a complete digestion of genomic DNA. For the new YAC library, large EcoRI fragments were generated by partial digestion with EcoRI in the presence of EcoRI methylase (see Materials and Methods). The ratio of the two enzymes used was such that the resulting fragments were mainly over 600 kb as judged by PFGE. Agarose blocks containing 3B DNA at a concentration of 10 μg/block were processed individually upto the size fractionation step. The EcoRI partial digestions were stopped and prepared for the ligation reaction, as was described for all previous YAC constructions.

The YAC4 BamHI-EcoRI vector arms were prepared as previously described (see Part 4.2). The arms were phosphatased twice and then checked by setting a small self-ligation reaction up. An aliquot of this (1 μg) was transformed into yeast as an assay for the proportion of non-recombinants there would be in the genomic ligations, although most of these should have been eliminated after the size fractionation step that was to be conducted. It was found that 1μg of vector DNA produced about 100 transformants, showing that some uncut or unphosphatased vector would be present in the genomic ligations. This reinforced the need for the fractionation step, as a typical transformation of ligated (unfractionated) YAC DNA contains 1-2 μg of vector.

Each ligation consisted of 10-15 μg of vector arms mixed with 10 μg of EcoRI-cut genomic DNA (1 block), and the reactions were carried out as before, except that 2.6 mM spermidine was included in the mixture as this had been found to prevent the shearing of large DNA molecules. After ligation instead of using the DNA directly (or after agarase treatment) for yeast transformations, the ligations were size fractionated on a PFG. The ligated DNA (still in the agarose) was loaded into a preparative gel made with 1% LMP agarose (Beckman). The DNA was sealed with 0.5% LMP agarose, and the PFG was run using the CHEF system in 0.25 x TBE buffer, at 170 V and 10°C with a pulse interval of 30 secs for 18-20 hrs. These conditions resolved DNA fragments below 600 kb, containing, it was hoped, all the smaller YACs and the non-recombinant YAC vector (as well as unligated genomic DNA). The compression zone

and slots which contained the size selected YACs were cut away from the rest of the gel as a single slice. Unlike the previous agarose size fractionations (in Part 4.4) this slice was then equilibrated with 1 M Sorbitol, 10 mM Tris-Cl, 2mM EDTA and 2.5 mM NaCl so that the YAC DNA would be appropriately buffered for transformation into yeast spheroplasts later. The agarose was then melted at 68°C and treated with agarase at 37°C overnight. The liquid DNA solution was then ready for transformation into yeast.

Test transformations using the unfractionated and the fractionated YAC ligations were carried out to determine whether the size selection step had indeed produced YACs of a greater size. Fig 4.6.1 is a blot of a PFG on which random colony minipreps from each of the two transformations have been run, hybridised with a plasmid probe that detects only YAC sequences. All the tracks contain a single intense YAC band, and the difference in the average YAC size between the two ligations can be seen. In the case of the unfractionated YAC ligation, the average size is about 50 kb, as was found in the first EcoRI YAC library made (see Part 4.2). On the other hand, the size fractionated YACs have a much larger average size of about 200 kb. Using the same method a similar increase in YAC size was found by members of Hans Lehrach's lab. However, these PFG conditions should have selected for DNA fragments greater than 600 kb in size, but YACs of this order of size were never seen using the size selected DNA. In order to determine what size range of YACs existed in the four fractionated YAC ligation DNA preparations that had been made (preparations RI YACs (a) and (b) were from one set of ligations, while RI YACs (c) and (d) were from another), 100 µl of each was run out on a PFG. A blot of this gel hybridised with plasmid probe can be seen in Fig. 4.6.2. Despite the fractionation of the DNA, there is still a large amount of vector present in the YAC DNA, as seen by the two intense bands of low molecular weight which represent unligated and/or non-recombinant YAC arms. The higher molecular weight YACs can be seen to range from over 600 kb down to as low as 100 kb in size, but there did not appear to be any gross DNA degradation.

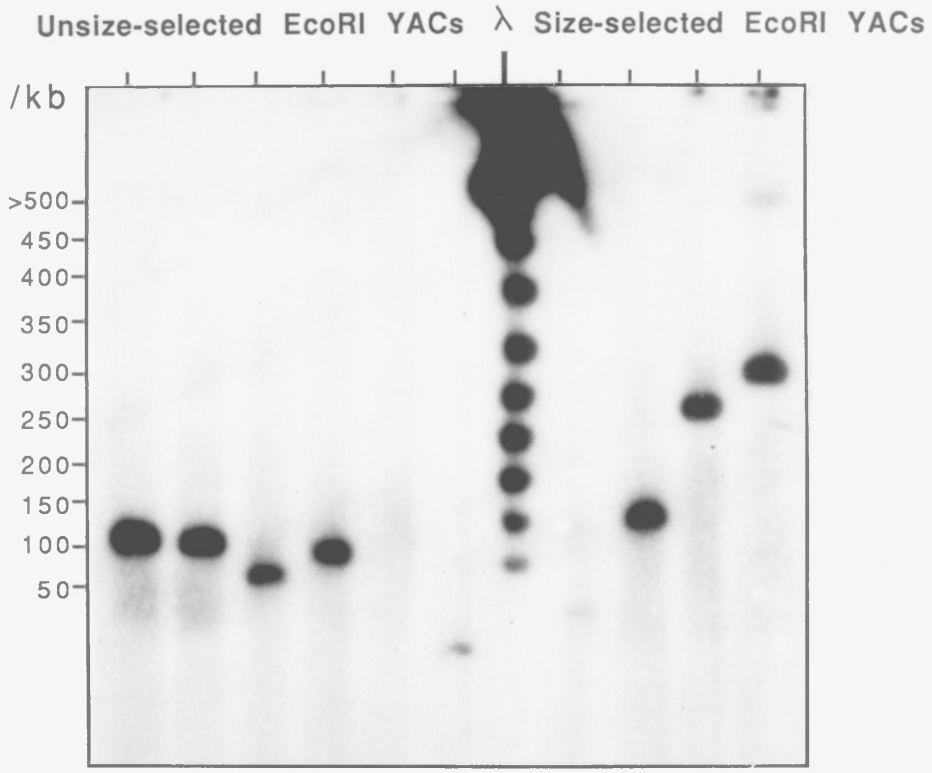
The fact that a single size fractionation for DNA over 600 kb in size had not eliminated all of the non-recombinant vector indicated that a large number of molecules

FIGURE 4.6.1

Size comparison of YACs in yeast transformants obtained with or without YAC size-selection prior to transformation.

The southern blot presented here has been hybridised with a YAC specific probe, and reveals the sizes of YACs (made from EcoRI-cut genomic DNA), which were either fractionated or not prior to transformation. Each track contains 2 μg of yeast (agarose miniblock) DNA from a randomly picked colony. The pulsed field gel was run in a CHEF apparatus at 170 V for 20 hrs, in 0.5 x TBE running buffer, using a pulse interval of 30 secs. The limit of resolution is 500 kb under these conditions.The lambda DNA concatemers, from which the sizes indicated on the left were deduced, can be seen in the lane marked λ.

As indicated above the lanes, the tracks on the left of the blot contain DNA from colonies that were transformed with YACs that had <u>not</u> been size selected. The tracks on the right of the blot contain DNA from colonies into which YACs that were size-selected (following ligation) had been transformed.



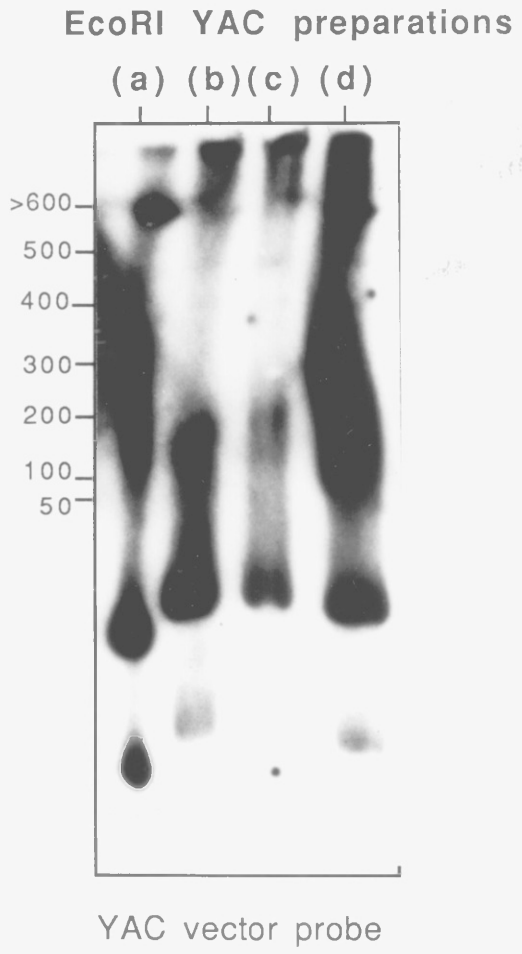
YAC vector probe

FIGURE 4.6.2

Southern analysis of the size-fractionated YAC ligations to be used in the construction of a genomic library of overlapping EcoRI fragments.

The southern blot of a pulsed field gel, shows the DNA from four independent YAC-ligation preparations, hybridised with a YAC vector probe. The four preparations [RIYAC (a) to (d)], consist of EcoRI genomic DNA fragments which were ligated to YAC vector armsand were then fractionated to select for YACs over 600 kb in length (see Part 4.6). Each track contains 100 µl of the size-fractionated preparation indicated above the lane, which corresponds to 1-2 µg of DNA. The pulsed field gel was run in a CHEF apparatus at 170 V for 18 hrs, in 0.5 x TBE running buffer with a pulse interval of 40 secs. The limit of resolution is 600 kb under these conditions. The sizes were deduced from lambda concatemers.

The YAC probe hybridises to the recombinant YACs containing genomic DNA inserts (signal seen > 50 kb in the blot), as well as to non-recombinant YAC vector which is either unligated or self-ligated (intense bands seen < 50 kb in blot). Note that a large proportion of hybridising DNA (from both recombinant YACs and non-recombinant vector) is found below 600 kb, despite the fact that these preparations had been size-fractionated.


EcoRI YAC preparations
(a) (b)(c) (d)
>600
500
400
300
200
100
50
YAC vector probe

of much less than 600 kb in size must have been trapped with the unresolved DNA. Moreover, the smaller YACs which were present in the DNA preparations (as described above), seemed to be preferentially transformed into yeast, so that the YAC size was again never over about 300 kb. The possible reasons for such size limitation are discussed in Part 4.7. Nevertheless, the average insert size was still greater than that obtained with unfractionated DNA, so that these YAC preparations were still useful.

As a high proportion of vector DNA still seemed to be present in the four DNA preparations, it was decided to check a number of yeast transformants from each for the proportion of non-recombinants by testing for canavanine resistance. RI YAC preparations (a) and (b) produced 60-80% canavanine resistant (recombinant YAC) colonies. This proportion actually varied from one transformation to another. However, RI YAC preparations (c) and (d) seemed to contain a very high proportion of non-recombinants. These preparations had been made in parallel using the same vector arm preparation as well as the same ligation conditions, whereas preparations (a) and (b) had been made separately. This presumably was the cause of the differences between the preparations.

RI YAC preparations (a) and (b) were used for the library as they contained the lowest proportion of non-recombinants and adequately large inserts. The efficiency of transformation of this DNA, however, was much lower than expected, ranging from 50-100 colonies per μg DNA (or 1-2 colonies per μl DNA solution, at a concentration of 10-20 ng/μl). The nonsize-fractionated YAC EcoRI ligation which had been set up simultaneously as a control to monitor the increase in YAC size (see Fig. 4.6.1), gave a much higher transformation efficiency of about 500 colonies/μg DNA. Thus, although the ligation reactions must have been successful, the efficiency of YAC molecules capable of being transformed after size fractionation was 5-10 times lower. It was also found that the volume of DNA used in a single transformation could not exceed 50 μl without the efficiency decreasing rapidly, even though this DNA was supposedly in an appropriate buffer to prevent osmotic shock of the spheroplasts. Varying degrees of yeast transformation inhibition had been noticed between the different types (or even

batches) used during the course of this work. Thus, trace impurities in the agarose might have been a contributing factor in decreasing both the size and efficiency of DNA taken up into yeast and this will be discussed further in Part 4.7.

The low transformation efficiency of this DNA combined with the limitation in the quantity it was possible to use per transformation, made its construction a very time consuming process. Of the 3552 colonies picked into microtitre wells, approximately 3000 grew well under -URA -TRYP double selection, implying that a proportion of them (15%) had only the URA3 YAC arm (presence of the URA3 marker being the only selection at the primary transformation stage) and not the TRP1 YAC arm. This was in fact tested by making a number of agarose block minipreps from random colonies which had been picked into either SD COM -URA medium (one arm) or SD COM -URA -TRYP medium (both arms). These were resolved on the PFG shown in Fig. 4.6.3. Again 15% would only grow in -URA medium, and the YACs from these colonies sometimes appeared unstable (on a PFG) giving an intense YAC band and a fainter higher band (see fourth lane from right) which could be explained by the lack of a telomere at one end of the YAC leading to recombination/deletion and reduction in YAC size.

In order to check that the average size of the YAC inserts from the complete library, agarose block minipreps were made from a number of random microtitre well cultures, and the yeast chromosomes resolved on a PFG. Fig. 4.6.4 shows a blot of this gel hybridised with a YAC-specific (pBR322 DNA) probe. As had been predicted from Fig. 4.6.1, the majority of the YACs are150-250 kb in size. Furthermore, about 75 % of the colonies (12/15 in this example, see Fig. legend 4.6.4) are true recombinants as predicted from the canavanine selection test above.

Thus of the 3000 colonies which had <u>both</u> YAC arms it was estimated that 70%-80% of these were recombinants (ie about 2300), and that these contained inserts of an average size of about 200 kb in size. Assuming that the rat genome contains approximately $3x10^6$ kb of DNA, then a library of this size should be equivalent to 2 -3 genomes, for a sequence which is amplified 20-40 fold. The library was therefore screened, by hybridisation (rather than PCR) so that the rat RI-Bam probe could be

FIGURE 4.6.3

Analysis of random RI YAC library transformants for the presence of both YAC vector arms.

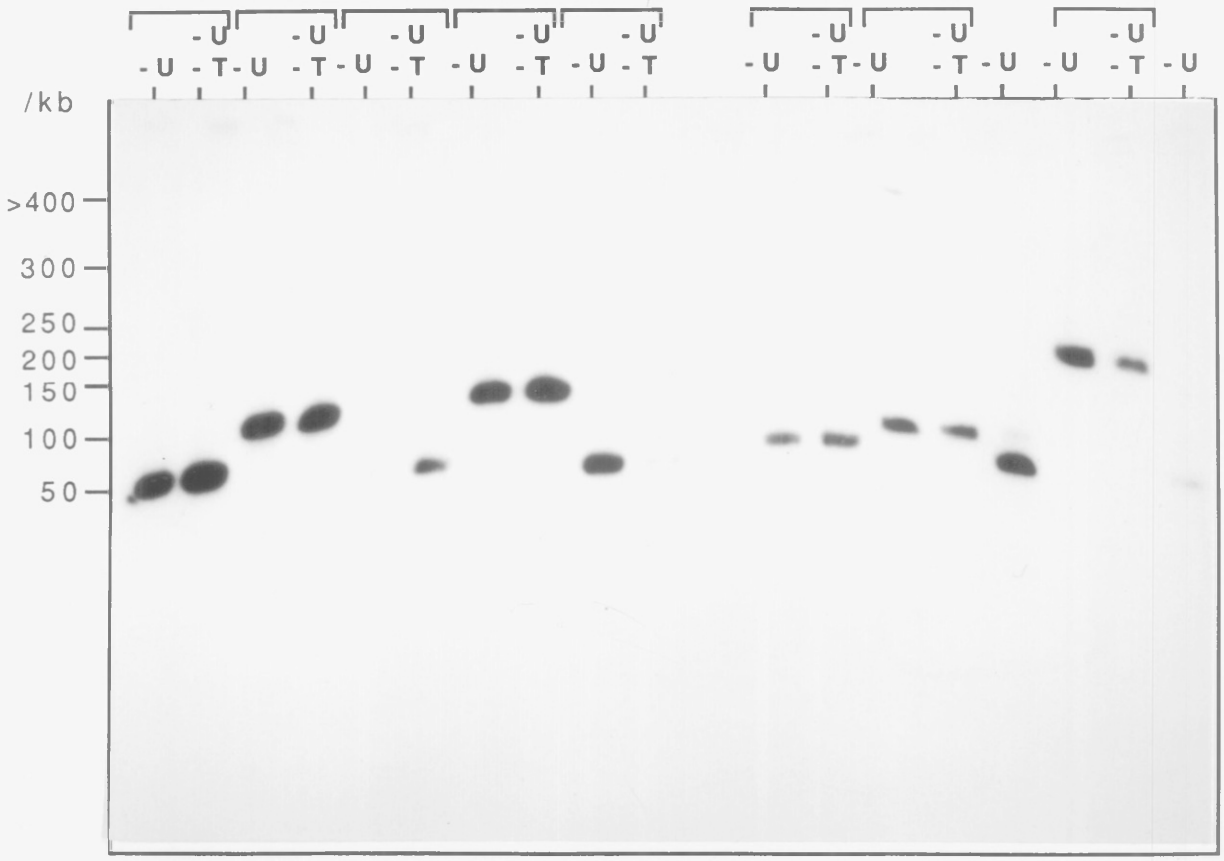
The YACs from colonies grown under single or double selection are seen on this southern blot probed with YAC vector DNA. Ten randomly picked colonies from the size-selected RI YAC library were grown in medium lacking uracil (-U) which selects for one YAC vector arm, and in medium lacking uracil and tryptophan (-U-T) which selects for both vector arms (see Fig.4.1.1). All ten grew in (-U) medium as expected, as this is the selection for primary transformants. Two out of the ten did not grow under double selection (-U-T), implying that they contained only one vector arm.

From the cultures in grown in either medium, about 2 μg of DNA (in an agarose miniblock) was run in each track. The DNA preparations from individual primary transformants grown in either -U or -U-T medium, were run in pairs as indicated above the lanes. The pulsed field gel was run in an OFAGE apparatus, at 250 V, for 20 hrs, in 0.5 x TBE running buffer with a pulse interval of 30 secs. The limit of resolution was 400 kb under these conditions. Lambda concatemers were used as size markers.

Note, that for transformants which grew in both -U and -U-T media, a single YAC of the same size is seen in both tracks, whereas for one of the transformants that only grew in -U medium (fourth track from right), the YAC appears to be slightly unstable as a minor proportion of it exists in a higher molecular weight form. As this YAC does not appear to have the vector arm on which the yeast centromere and ARS (as well as the telomere) are located, it might be expected to exhibit instability.

- U = growth in SD COM -URA

- U
- T = growth in SD COM -URA -TRYP



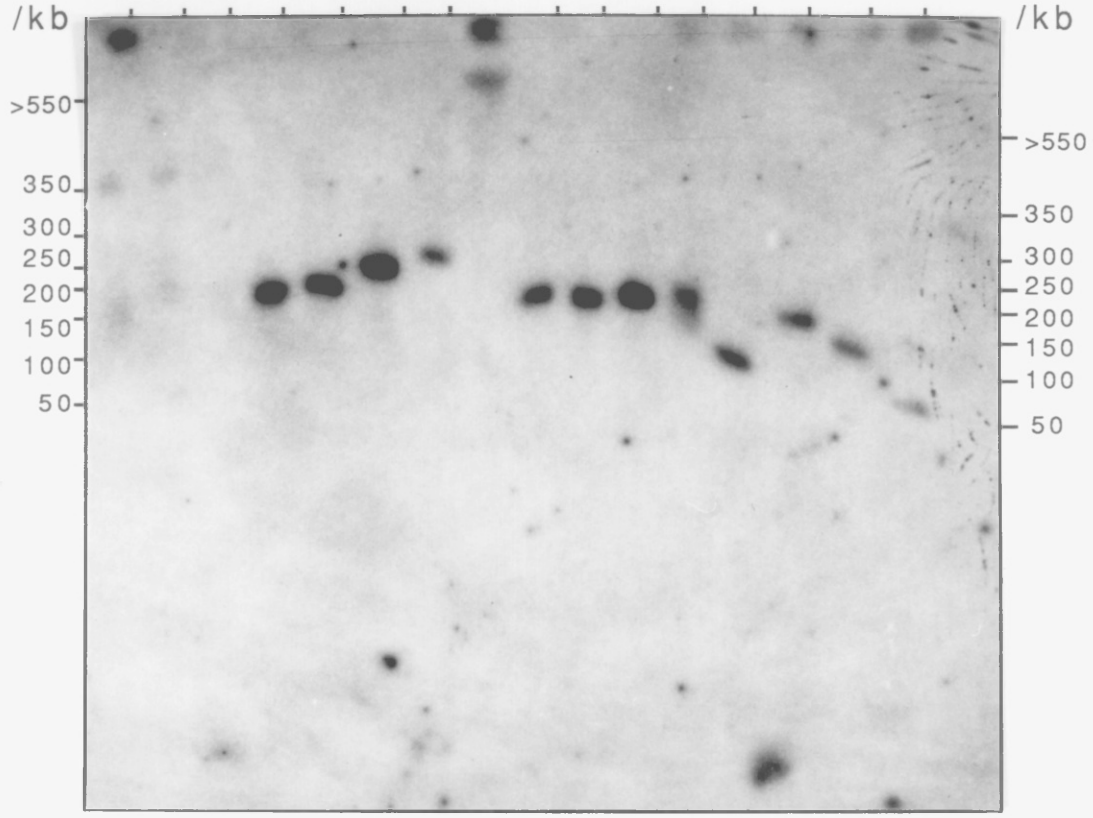
YAC vector probe

FIGURE 4.6.4

Size determination of random YACs from the size-selected EcoRI YAC library of 3B DNA.

The southern blot shown is of a pulsed field gel on which the chromosomes of 15 colonies randomly picked from the size-selected RI YAC library have been resolved. The blot has been hybridised with a YAC vector probe. Each track contains approximately 2 μg of yeast (agarose miniblock) DNA. The pulsed field gel was run in an OFAGE apparatus at 300 V for 20 hrs, in 0.5 x TBE running buffer, using a pulse interval of 35 secs. The limit of resolution is 550 kb under these conditions. The sizes indicated were deduced from lambda concatemers. Note that this gel was run in an OFAGE apparatus, so that the tracks are curved, and the sizes are not parallel across its width.

In the majority of the tracks, the YAC-specific probe detects a single YAC, usually of 150-250 kb in length. In the first three tracks on the left however, the characteristic hybridisation pattern of non-recombinant YACs can be seen, with the vector sequences integrated into two natural yeast chromosomes (the same two in each case).



YAC vector probe

used. This probe would isolate YACs which extended furthest in to the uncloned region. The nucleotide sequence of the RI-Bam probe had not been determined, thus the library could not be screened by PCR, as in Part 4.2. Duplicate filters were screened, on each of which the yeast colonies from six, 96-well microtitre dishes had been grown. The YAC library filters were in fact simultaneously hybridised with both the rat RI-Bam probe and the polyoma virus DNA probe. The probes used were always checked for their specific activity, and autoradiographs were left on the filters for 3-7 days. No positives were detected using both the probes. Owing to lack of time, no more of the RI-YAC ligations could be transformed into yeast and screened.

4.7 DISCUSSION.

The advent of YAC technology seemed to provide an ideal method for the cloning of large stretches of eukaryotic DNA, such as those found in mammalian gene amplifications. However, the YAC system was relatively untested when the work in this chapter was begun. Certain difficulties which were intrinsic to the use of yeast for cloning were encountered. However, these were enhanced by the particular DNA preparative methods used for this work. In addition, there were unexpected complications with one of the two vectors used. These problems are discussed below.

Practical Issues in the Construction and Screening of YAC Libraries.

The use of yeast as a host for cloning has a number of unique problems associated with it. The introduction of exogenous DNA into yeast is technically more complex than it is for *E.coli*, as it requires the preparation of spheroplasts immediately prior to the transformation, if the efficiency is to be maximal. Unfortunately, there is a limit to the amount of DNA and the number of spheroplasts that can be used in a single transformation. In the case of YACs, it was found that each transformation could yield a few hundred transformants. Furthermore, yeast spheroplasts are quite fragile so that the efficiency of the transformation process was very sensitive to the nature of the DNA solution added. Throughout this work, large genomic DNA was prepared and

fractionated in LMP agarose. However, the presence of agarose during the transformation step seemed to reduce the transformation efficiency. This could be partly overcome by the use of agarase and an appropriately buffered solution to prevent osmotic shock, as were used for the construction of the size selcted EcoRI YAC library in Part 4.6. It was apparent, however, that certain types or batches of agarose still had an inhibitory effect, even after this type of treatment. For example, one batch of "BRL ultrapure LMP agarose" used during the ClaI fragment size fractionation (see Part 4.4) produced a very poor transformation efficiency. When tested with control YCp plasmid, the presence of this particular agarose reduced the transformation efficiency by about 100 fold. Eventually, Beckman LMP agarose was found to be the most reliable in grade. When this inhibitory effect was encountered, even greater numbers of transformations had to be set up in order to generate enough transformants for a library. This was especially true if the YAC DNA concentration was fairly low, for example after size fractionation.

The use of agarose for the preparation of genomic DNA was found to have further disadvantages. This method was originally used because it was a straightforward method for preparing high molecular weight DNA with minimum mechanical shearing. Also, DNA embedded in agarose could be stored for months or even years without any signs of degradation. However, the use of agarose seemed to impose an upper limit on the size of YACs possible. This was probably for a number of different reasons. In the case of the EcoRI libraries (Parts 4.2 and 4.6), the extent of the partial digestion by EcoRI was difficult to control when the DNA was in this form, leading to the generation of smaller fragments than expected. Also, the use of agarose gels as a means of size-fractionating DNA, was found to be misleading and inefficient, compared to the use of sucrose density gradients. When DNA of a particular size range was isolated after gel electrophoresis, it was found to contain a high proportion of smaller DNA molecules which had migrated aberrantly probably as a consequence of being trapped with the larger molecules in the agarose matrix. This could explain why a proportion of the YACs would be of a smaller size than had been selected, but not why the

overwhelming majority should be as was found with both the size-selected ClaI and EcoRI YAC transformants. There seemed to be preferential transformation of YACs below 200 kb in size in both cases. Either the agarose imposed this limit by inhibiting the spheroplast transformation step, or else trace impurities in the agarose meant that very large molecules (> 300 kb) were being sheared. Even a very slight amount of degradation, not detectable on an agrose gel (see Fig.4.6.2), would have been likely to shear the larger YACs. This might also explain why the transformation efficiency was very low in the case of the size-fractionated EcoRI YAC library (Part 4.6). The fractionation of very large DNA molecules (over 600 kb in size) which had already been inserted into YAC arms (unlike in the case of the ClaI size fractionation), probably meant there were very few intact YACs left at the transformation stage.

The problems inherent in the yeast transformation process itself combined with the limited size of the YACs obtainable, made the generation of sufficient numbers of clones for screening a very time-consuming process. The steps required for the isolation and screening of the clones themselves, then provided additional technical difficulties. Yeast spheroplasts have to be regenerated in agar which precludes direct screening of primary transformants. Furthermore, only the selection for one YAC arm can be imposed at this stage (URA). Thus, for screening and for double selection (URA,TRYP) to be applied, the colonies had to be picked into microtitre wells. This was crucial as a proportion of the YACs were found to lack the TRYP YAC arm (see Fig. 4.6.3) which probably rendered them unstable and would have led to complications later in their analysis. The selection for recombinants is also problematic, and can only be applied on a random selection of colonies. (The red/white colony colour change only appears after days or weeks on the plates, and canavanine resistance involves growth on a selective medium (see Materials and Methods) which is unsuitable for either the regeneration of spheroplasts or the long-term storage of picked colonies).

The screening of yeast colonies on filters was inefficient for the reasons already mentioned in Part 4.1 of this chapter. However, although the PCR screening method was more rapid and sensitive, it required a knowledge of the sequence being used as a probe. This could be inconvenient during a long chromosome walk. Furthermore, the

primers used have to be carefully tested prior to screening, to ensure that the PCR efficiency was optimal for use on yeast miniprep DNA.

Specific problems encountered with each of the YAC libraries constructed.

Chromosome walking in YACs was attempted using the same two types of approaches described for lambda vectors in Chapter 2. These involved either the isolation of overlapping DNA fragments, which was a useful long-term approach; or the isolation of specific restriction fragments, where the size of the step taken could be large.

The first attempt at making a library of overlapping EcoRI fragments from 3B genomic DNA, did not involve a size fractionation step. This meant that the average size of the partial EcoRI fragments was only about 50 kb, and that a high proportion of non-recombinant vector was present at the transformation stage. This library was therefore probably not adequate in size, and the one putative positive detected may have been an artefact. Screening by PCR was used for this library, which is probably far more sensitive than hybridisation. Thus, a highly unstable YAC whose insert was only maintained in a small minority of yeast cells when grown under selection, could have produced the faint PCR signal detected (part 4.2).

It was thought that cloning a specific restriction fragment detected by the screening probe (eg the 240 kb ClaI fragment), would avoid the problems encountered with EcoRI partial digestions. As the average insert size of a library of completely digested genomic ClaI fragments would have been only about 100 kb, an enrichment step was required for the isolation of a 240 kb fragment. It was during the attempted purification of the 240 kb ClaI fragment the apparent 200 kb size limitation of YACs was observed. Numerous attempts using PFGE DNA fractionation consistently produced YACs of a smaller size than expected. As discussed above this was probably due to the use of agarose causing shearing of the DNA fragments and thereby reducing the number of large fragments with clonable ClaI overhangs in any fraction isolated. On the other hand, it could have been due to the presence of agarose inhibiting the uptake of larger YACs during the transformation. In fact it was at this point that the effect on

transformation efficiency of the different types of agarose used was noted.

As it was not technically possible to isolate YACs over 200 kb in size, the only alternative using this approach was to clone the 180 kb XhoI fragment. However, this was prevented by the discovery that the cloning of XhoI fragments into the SalI site of the pYAC-RC vector produced highly unstable YACs for unknown reasons. Instead of being propagated autonomously, most of these YACs integrated into natural yeast chromosomes (see Figs. 4.5.1 and 4.5.2). The possibility that a high proportion of non-recombinant YACs was the cause for this was eliminated. Moreover it did not appear to be due to inefficient removal of the HIS marker between the telomeres either. It was unclear why the vector should show a tendency to produce unstable YACs when using one of the cloning sites (SalI, in Part 4.5) in this vector but not another (ClaI, in Part 4.4). One possibility might be that the use of the SalI cloning site in the SUP4 gene had some kind of destabilising effect on one of the chromosomal functional units of the vector (the ARS, CEN or TEL sequences), as these are essential for the stable propagation of autonomous artificial chromosomes. On the other hand, it could simply have been a consequence of incomplete XhoI enzyme inactivation after the digestion of 3B genomic DNA. Any residual XhoI activity would have cleaved away a proportion of the vector's telomeric ends when the YAC arms were mixed with the genomic DNA for ligation. This should have been unlikely because Proteinase K treatment as well as heating at 68°C for over 10 minutes (to melt the agarose block prior to mixing with the vector arms for ligation) should be sufficient to inactivate XhoI (as recommended by manufacturers). It turned out that other laboratories using this vector have noticed a similar phenomenon, ie a high rate of instability in the YACs produced using different cloning sites in this vector.

In the final attempt to clone more of the amplified DNA from 3B cells, the YAC4 vector was used, as it had by then been shown to be free of the problems in stability found with pYAC-RC. Several modifications were made to try and improve the size of the EcoRI YAC inserts from those that had been previously obtained. The size fractionation did not produce as large an increase in size as had been hoped, and furthermore, it led to a reduction in the transformation efficiency. Presumably this was

due to the use of agarose, as described in the first part of this discussion. Nevertheless, an increase from an average YAC size of 50 kb to one of 200 kb was achieved, so that considerably fewer clones should have had to be screened in order to take a chromosome step of up to 200 kb. The consistency in insert size and proportion of recombinants were checked throughout this library. However, an average YAC size of 200 kb also meant that the degree of amplification of the region being sought was lower than 20-40 fold, compared to the library of 50 kb inserts. This was due to the increased likelihood that the RI-Bam probe, which is only 28 kb away from the centre of the inverted duplication, would detect some EcoRI fragments which actually spanned the centre. These EcoRI fragments would only be amplified to half the extent (10-20 fold) as fragments which lie on either side of the centre of the duplication. This library might therefore have contained the equivalent of only about 1.5 rat genomes, rather than the 2 to 3 originally estimated (Part 4.6). Despite this, the number of colonies that were actually screened should have led to the isolation of at least one positive clone. The absence of any positive clones in this library and in the first (unsize-fractionated) EcoRI YAC library, suggested that this region of the inverted duplication might be difficult to clone in yeast. It seemed unlikely that this was due to the palindromic structure of the region, as it was shown that at least 18 kb of the inverted duplication (around its centre) could be stably propagated in yeast (see Part 4.3). Nevertheless, much larger palindromes might not exhibit the same stability. On the other hand, there may be mammalian DNA sequences in this region, which are not tolerated in yeast and are thus difficult to clone. The systematic cloning of large stretches of mammalian DNA in a yeast host is still a relatively new procedure. Although it would be expected that there are fewer "unclonable" mammalian sequences in yeast compared to *E.coli*, this has only begun to be tested recently. The much greater stretches of DNA per clone when using YACs also mean that if there are certain sequences that yeast cannot tolerate, there is an increased likelihood that a larger DNA fragment would contain one.

Unfortunately, it was not possible to pick and screen more of this library in case the region was only slightly underrepresented, owing to a lack of time.

Summary.

In theory it should be possible to create YACs many hundreds of kilobases long, and in fact this has been achieved in some laboratories but using different methods of DNA preparation to those described here. It had not been suspected that the use of agarose would hinder the process of YAC cloning when this work began. However, this problem combined with the inherent difficulties of the system, meant that the construction of a useful library took much longer than intended. Having eventually succeeded in making a library of what was thought to be an adequate size, it was unfortunately found that this region of DNA in 3B cells was probably difficult to clone in yeast. This came as a surprise, as the use of a eukaryotic host rather than *E.coli* had been expected to overcome most of the problems with "unclonable" sequences. The same region had proved to be quite difficult to clone in *E.coli*, as described in Chapter 2, and it could be that a similar intolerance for this region is exhibited by yeast.

This work has illustrated some of the unexpected difficulties in YAC cloning. After its widespread use it is becoming apparent that this cloning system has a number of specific drawbacks that need to be overcome. However, on-going technical improvements are making it a more efficient, and less time-consuming process. It is now clear that size fractionation of the DNA to be cloned, both before and after the ligation step (ie two size fractionations rather than one), as well as careful optimalization of all the materials used during the yeast transformation step, lead to a substantial increase in the average YAC size (4-500 kb at the time of writing). Further examples of its use in different systems should provide a deeper understanding of its limitations and potentials.

CHAPTER 5

5.1 INTRODUCTION.

The molecular analysis of the amplified DNA in the 3B cell line described in Chapter 3 indicated that all the amplification units were clustered at one (or a few) chromosomal location(s). In order to confirm these results, and also to examine the chromosomal consequences of the events which gave rise to this DNA amplification, a cytogenetic analysis was carried out. As in the molecular analysis, the fact that the amplification units in 3B rat cells consist of foreign DNA as well as coamplified rat sequences, was again exploited. The different probes derived from the amplification units would enable the location(s) of the amplified sequences to be distinguished from those of the native (unamplified) copies of the rat DNA. For this cytogenetic analysis, the technique of fluorescence *in situ* hybridisation was used which allows the detection of single copy sequences on metaphase chromosomes with high efficiency (Pinkel et al, 1986). This work was carried out with the guidance of Dr. Sarah Williams in the Cytogenetics Laboratory, I.C.R.F. The use of this highly sensitive technique would ensure that the location of every copy of the amplified DNA could be detected in 3B cells. Moreover, this method would enable localization of the unamplified, native rat alleles of the region in both 3B and Rat-2 cells.

In this way a number of important questions concerning the gene amplification in 3B cells could be investigated. Firstly, hybridisation of an amplification-specific probe (λ3B, which contains mouse and polyoma virus sequences, see Fig. 1.5) on 3B karyotypic spreads, would reveal the chromosomal manifestation of the amplified DNA, thus testing the predictions of Chapter 3.

Secondly, using a probe that detects both the amplified and unamplified rat sequences (λBam), the native alleles of the amplified region could be identified. By comparing the locations of these single copy signals in 3B and Rat-2 cells, the original chromosomal region into which the transfected mouse and Polyoma DNA had integrated, and which was subsequently ampliifed, could be identified. It would be particularly interesting to

determine whether or not the amplified rat DNA resides at the same chromosomal location as the native locus from which it arose. If the amplifed DNA was found to reside at a new chromosomal location, it might be possible to address the question of whether the original region of DNA involved in the amplification, has been preserved or deleted. As outlined in Chapter 1, deletion or preservation of the original locus is one prediction by which the current models for gene amplification can be distinguished. This information would therefore be extremely valuable for an understanding of the amplification process which occurred in the 3B cell line.

Thirdly, a comparison of the G-banded karyotypes of 3B and Rat-2 cells would reveal the chromosomal differences (gains, losses and rearrangements) which may have occurred as a result of the transfection/integration and amplification events. Any karyotypic differences observed would be useful indicators of the recombination processes involved in the generation of the amplified DNA in the 3B cell line. It was expected that an analysis of the chromosomes of Rat-2 and 3B cells would be less complex than the analysis of some cell lines which contain amplified genes, for a variety of reasons. The Rat-2 cell line from which 3B was derived, is essentially euploid (Topp et al, 1981) which would probably simplify the precise localization of the region involved in the amplification in 3B cells. Also, the amplified DNA in 3B cells appears to consist of a relatively low number of simple, homogeneous units, which are completely stable both in structure and size (from the southern blotting analysis described in Chapter 3). This indicates that the amplified DNA was not generated by multiple rearrangement events, and also that its structure is not evolving. Thus, complicated karyotypic changes in 3B were not expected to be seen when compared to Rat-2. Finally, many studies on gene amplification involve the drug selection of cultured cells, and as described in Chapter 1, such drugs (eg. methotrexate, PALA, coformycin) are thought to increase the frequency of gene amplification and induce other chromosomal aberrations. The amplification in 3B arose in the absence of chemical stress produced by such drug treatment. Furthermore, the amplification of the transforming oncogene was concommitent with the cellular change from an immortalized stste (with strong growth control) in Rat-2 cells, to a transformed state

(loss of growth control) in 3B cells. This situation probably reflects more closely the mechanisms of gene amplification (of proto-oncogenes) that take place during neoplastic progression *in vivo*.

RESULTS.

5.2 *In situ* hybridisation of 3B cell chromosome spreads with the amplification-specific λ3B probe.

The hybridisation of the λ3B probe onto 3B spreads was expected to reveal whether the amplified DNA was extrachromosomal or intrachromosomal, and whether it was clustered in one location or dispersed at many locations. This analysis would test the predictions made in Chapter 3, that the amplified DNA exists as an array (or arrays) of over 10Mb in length. Metaphase chromosome spreads were prepared (see Materials and Methods) from 3B cells which had been isolated soon after transformation (approximately 25 cell divisions), as well as from cells which had been passaged for longer periods of time. Any changes that may have occurred in the copy number or location of the amplified DNA during cell culture would therefore be detected.

The technique of fluorescence *in situ* hybridisation was used, which involves the use of biotinylated DNA probes for hybridisation onto chromosome spreads and fluoroscein-labelled avidin for signal detection (see Materials and Methods). Computer enhancement of fluorescent signals was **never** used throughout the fluorescence hybridisation analysis described in this chapter.

The λ3B clone (Ford et al, 1985) was chosen for use as a probe as it contains an insert which is adequately large (>10kb) to enable fluorescent signal detection using this method. The insert is 18kb in length and spans the centre of the inverted duplication (see Fig. 1.5). It consists almost entirely of polyoma virus (Py) and mouse sequences, with the rat DNA at the two ends of the insert being less than 250 bp in length (Passananti et al, 1987). This probe will therefore detect only the amplified DNA in 3B rat cells. However, the 12 kb of mouse DNA in λ3B, contains a number of repetitive sequences which cross-react with rat DNA (Passananti et al, 1987). The

probe was thus pre-annealed with competitor DNA (sonicated, total genomic rat DNA) prior to its hybridisation onto rat chromosomes. Chromosome spread hybridisation with biotinylated probes, detection with fluoresceinated avidin and signal amplification (when necessary) using biotinylated anti-avidin and further fluoroscein-avidin layers, were carried out as described in the Materials and Methods. The fluorescent signals and the chromosomal DNA (counter-stained with propidium iodide), were examined under the microscope and photographed, also as described in Materials and Methods.

Six metaphase spreads of 3B chromosomes hybridised with the λ3B probe are shown in Figures 5.2.1 and 5.2.2. In all spreads looked at, the signal representing the amplified DNA was always chromosomally located. No signs of hybridisation to extrachromosomal elements, such as double minutes or other acentric fragments can be seen. The spreads shown in Fig. 5.2.1 are derived from early-passage 3B cells, and those in Fig. 5.2.2 from 3B cells grown for several passages. Thus, in each case, the amplifed DNA resides on a single, acrocentric chromosome of the same length (relative to the other chromosomes in the spread). Furthermore the position of the signal along the chromosome (approximately two thirds of its length, away from the centromere) seems to be the same in all the spreads. The predictions made in Chapter 3 that the amplified DNA is clustered in an array, and that it spans at least 10 Mb are confirmed with the data shown here. The proportion of the chromosome covered by fluorescent signal indicates that the amplified region probably does not span more than 10-11 Mb of DNA. This was estimated be measuring the length of chromosome covered by signal, and expressing it as a percentage of the haploid autosome complement (as reported by the Committee for a Standardized Karyotype of *Rattus norvegicus*, 1973) and thus of the haploid DNA content of rat cells (approximately 3×10^9 bp). The estimate of 10-11 Mb for the size of the amplified region is consistent with an approximately 20-25 fold amplification of a unit which is about 450 kb long (as described in Chapter 3).

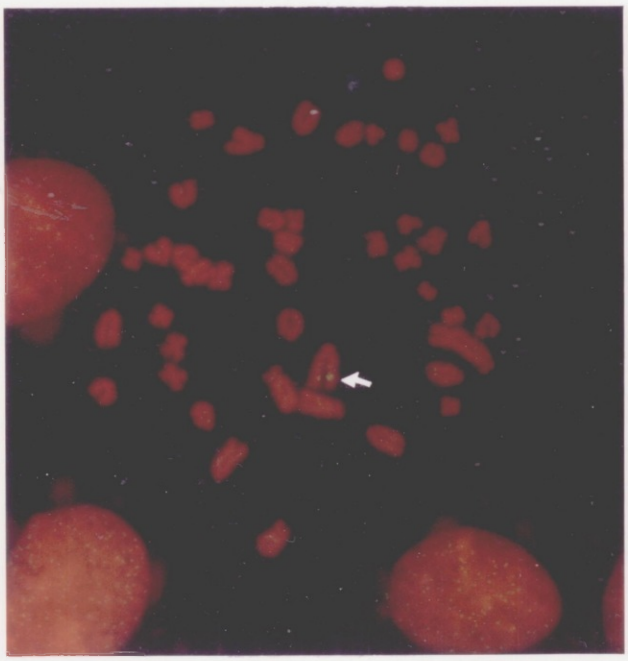
The amplified DNA appears to be stably located at a single chromosomal position. However, to verify that this location was truly constant a more rigorous identification of the chromosomes in 3B cells, than just their relative size, needed to be made. The G-

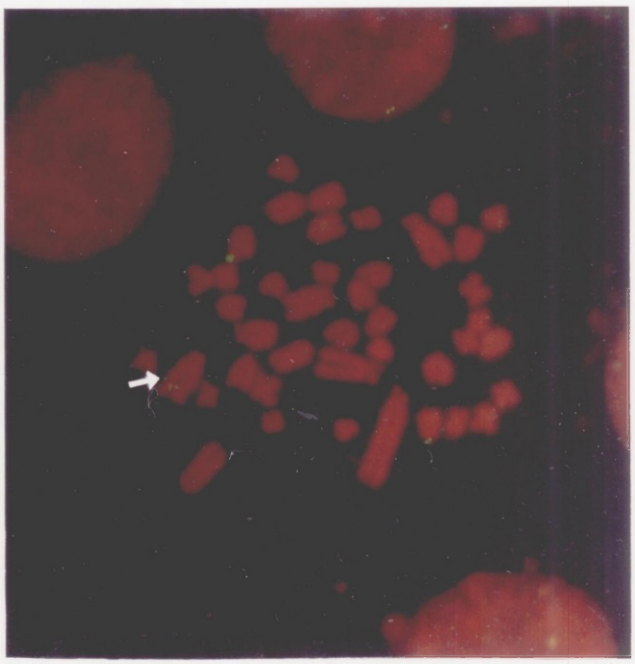
FIGURE 5.2.1

Fluorescence *in situ* hybridisation of metaphase spreads from early-passage 3B cells with the λ3B probe.

Three representative metaphase spreads from early-passage 3B cells hybridised with the λ3B probe, are presented. The metaphase spreads were prepared and hybridised with biotinylated probes as described in Materials and Methods. Sites of hybridisation, indicated by the arrows, were revealed by binding fluoresceinated avidin and are seen as yellow fluorescein signals located on both chromatids at a particular site. As the 18 kb λ3B probe consists almost entirely of mouse and polyoma virus sequences which lie at the centre of the inverted duplication in 3B cells (Fig. 1.5). This probe only detects the amplified DNA in 3B cells. The λ3B clone contains repetitive mouse sequences which can cross-react with rat repetitive sequences, thus the probe was preannealed with sonicated rat genomic DNA.

Note the presence of signal, representing the amplified DNA, at the same position on, a single chromosome of the same size, in all three spreads.





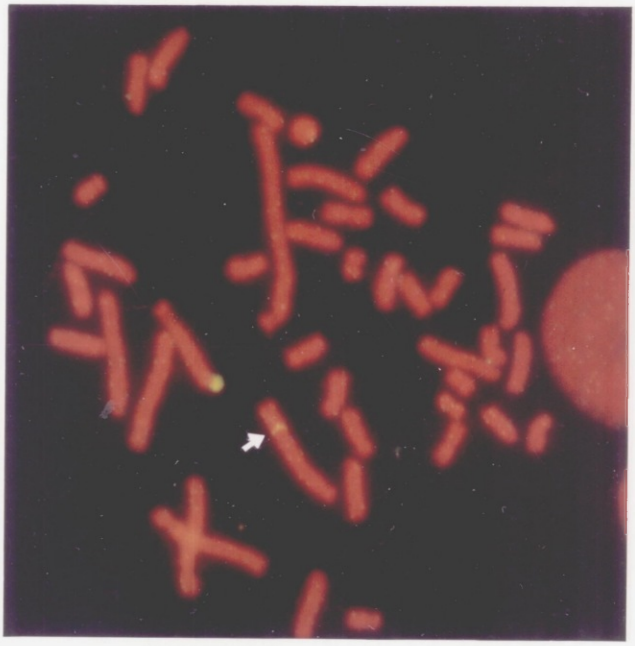
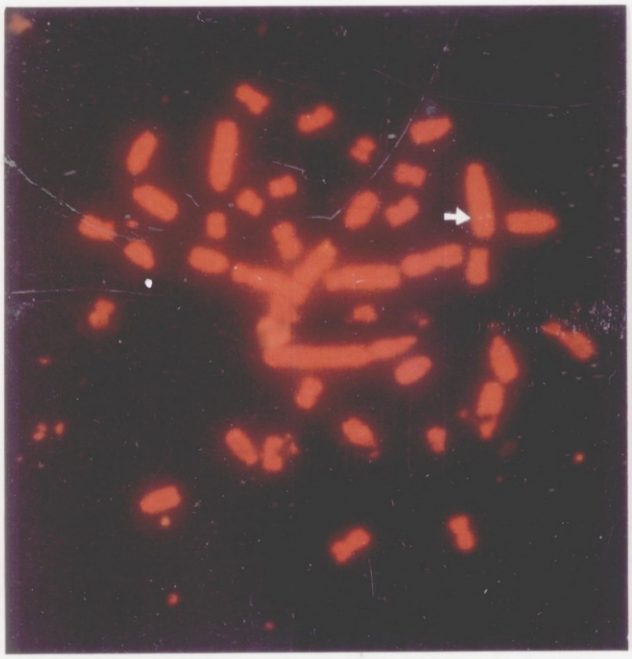


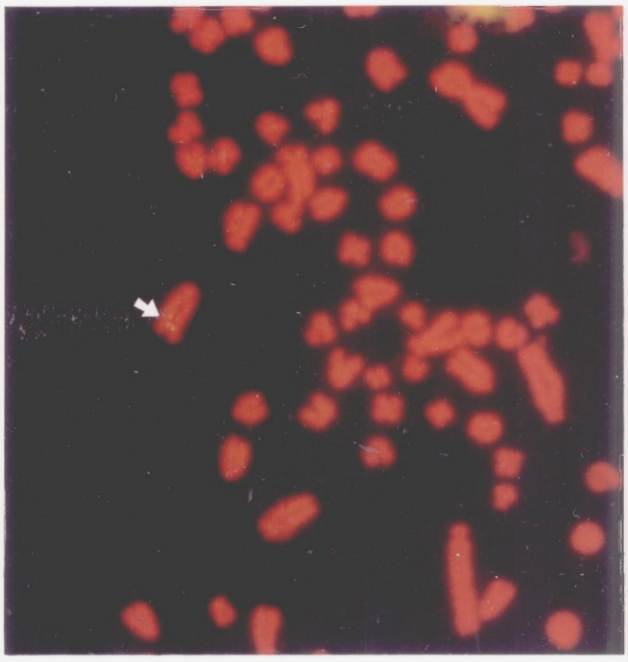
FIGURE 5.2.2

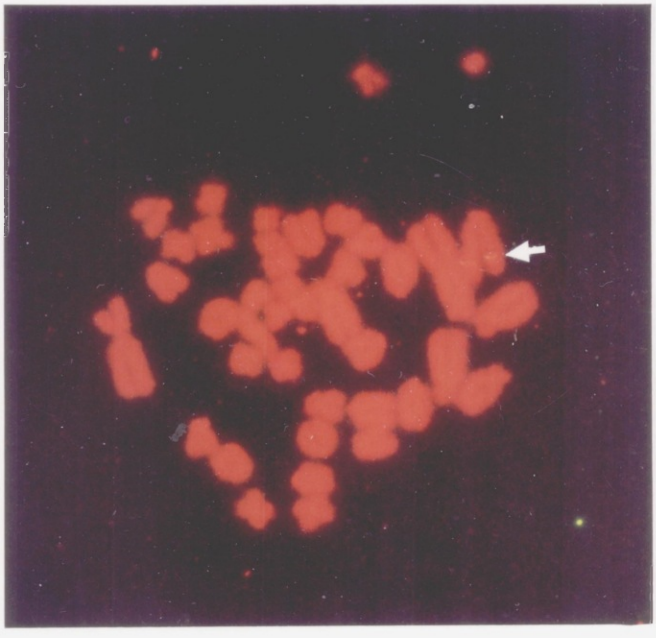
Fluorescence *in situ* hybridisation of metaphase spreads from late-passage 3B cells with the λ3B probe.

Three representative metaphase spreads from late-passage 3B cells hybridised with the λ3B probe, are presented. The chromosome spreads and biotinylated probe were prepared and hybridised exactly as described in Fig. 5.2.1. The locations of fluorescent hybridisation signal are indicated by arrows.

The pattern of hybridisation in late-passage 3B cells appears to be identical to that found in early-passage cells (see Fig.5.2.1).







banding of 3B metaphase spreads was therefore carried out and is described in Part 5.4. Furthermore, the data presented here provided no information about the chromosomal origin of the amplified DNA and hence the mechanism by which it arose. The results of the next section address these questions.

5.3 *In Situ* Hybridisation of 3B and Rat-2 Chromosome Spreads with the λBam Rat Probe.

A probe which was derived from the amplification unit but which consists of rat sequences, would identify the unamplified native allele(s) as well as the amplified DNA when hybridized to 3B chromosome spreads. When hybridized to Rat-2 chromosome spreads, this probe would identify all the native alleles of the region amplified in 3B cells. Hence, by comparing the chromosomal sites detected by this probe in 3B and Rat-2 cells, the actual region into which the transfected mouse and polyoma virus DNA integrated and which was subsequently amplified could be identified. It would be seen whether or not the amplified DNA resides on the same chromosome as that from which it arose, and also what changes have occurred in the native rat loci during the production of the 3B cell line.

The λBam rat clone was isolated as part of the chromosome walk along the length of the amplification unit, described in Chapter 2 (see Fig. 2.2.1). It contains a 15 kb insert which consists solely of rat sequences. As the *in situ* hybridisation technique used here is sensitive enough to detect single copy sequences, all copies of amplified and unamplified rat DNA should be detected by this probe. The complete lambda clone was biotinylated but as the insert contains some repetitive rat sequences, it was pre-annealed with rat competitor DNA as described for λ3B. Chromosome spreads from 3B and Rat-2 cells were prepared and hybridised with the biotinylated λBam probe.

In Fig. 5.3.1, two spreads of 3B cells hybridised with the λBam probe are shown, which are representative of all those analysed. As with the λ3B probe, spreads from early and late-passaged cells showed an identical pattern of hybridisation. The λBam probe hybridises to three different chromosomes in 3B. The chromosome on which

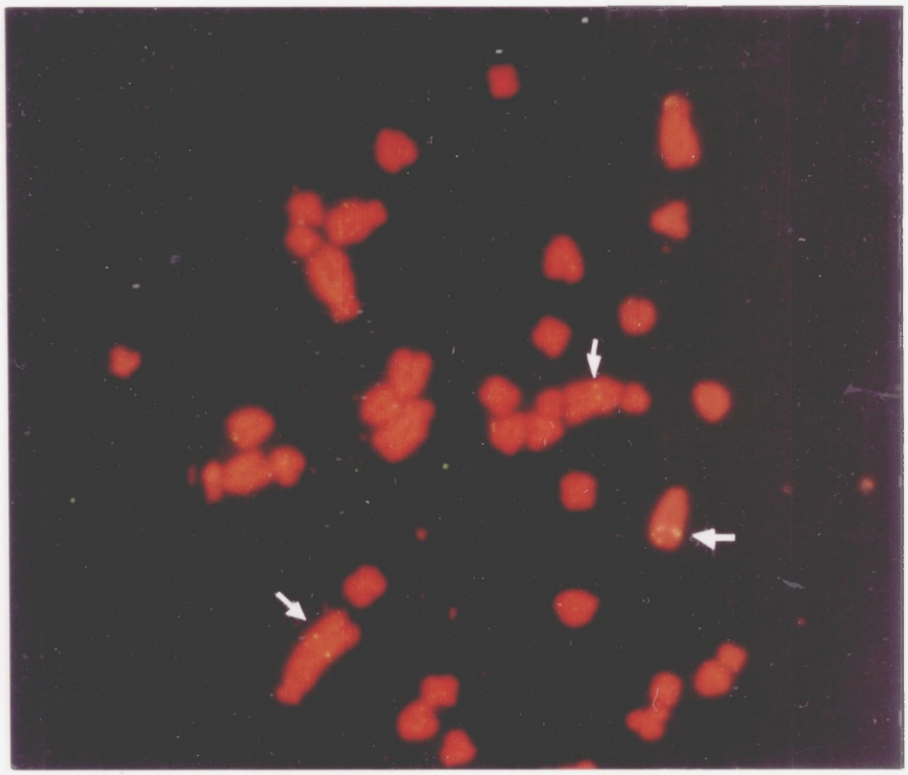
FIGURE 5.3.1

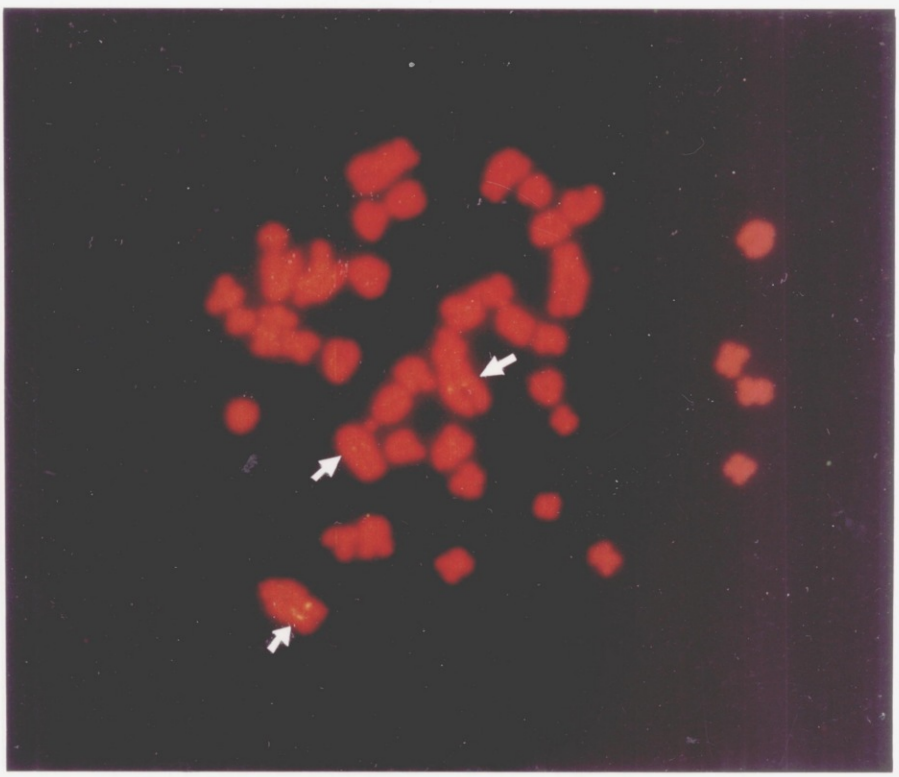
Fluorescence *in situ* hybridisation of metaphase spreads from 3B cells with the λBam rat probe.

Two representative metaphase spreads from 3B cells hybridised with the λBam rat DNA probe, are presented. The chromosome spreads and biotinylated probe were prepared and hybridised as described in Fig. 5.2.1. The isolation of the 15 kb λBam probe is described in Chapter 2. As this probe consists entirely of rat DNA sequences it will detect not only the amplified inverted duplicated sequences, but also the unamplified native alleles of this region of the rat genome. The λBam insert contains repetitive rat sequences, so the probe was preannealed with sonicated rat genomic DNA.

In each spread the arrows indicate the sites of fluorescein signal present on both chromatids. The amplified DNA is detected as a cluster of signals at the same chromosomal location as that found with the λ3B probe (Figs. 5.2.1 and 5.2.2). Single-copy signal, representing the native unamplified alleles, is detected as a single fluorescent spot on each chromatid. This is seen on <u>two</u> different chromosomes, one of which is much longer than the other and exists in only one copy per diploid cell.

Note that the λBam probe detects 15 kb of sequence on each side of the amplified inverted joint. Thus it detects twice as many copies of the amplified DNA as the 18 kb λ3B probe, which spans the centre of the inverted duplication. For this reason the intensity of the amplified signal detected by λBam probe tends to be stronger than that detected by the λ3B probe (in Figs. 5.2.1 and 5.2.2).





most signal is detected (ie the amplified region), resembles that detected previously by λ3B (Part 5.2), both in its relative size and in the location of the signal along its length. In addition, single copy signal is detected on two chromosomes which are very different in size to each other. One of these appears to be an abnormal, long marker chromosome (M1), of which there was only a single copy in all spreads examined. This long marker chromosome can also be seen in the spreads shown in Fig. 5.2.1, but no consistant hybridisation was ever detected on it using the amplification-specific λ3B probe. The other chromosome with single copy signal is of a similar size to that on which the amplified DNA is detected, but the position of the signal on it is very different, being much more proximally located.

In Fig. 5.3.2, three spreads of Rat-2 cells hybridised with the λBam probe are shown. In all the spreads, single copy signal is detected on three chromosomes. One of these is the M1 chromosome, also seen in the 3B spreads of the previous figure. The other two chromosomes appear to be identical both in size and in the position of signal along them, and furthermore the same chromosome (with single copy signal) is also present in 3B cells, but only in one copy per diploid cell (see previous figure).

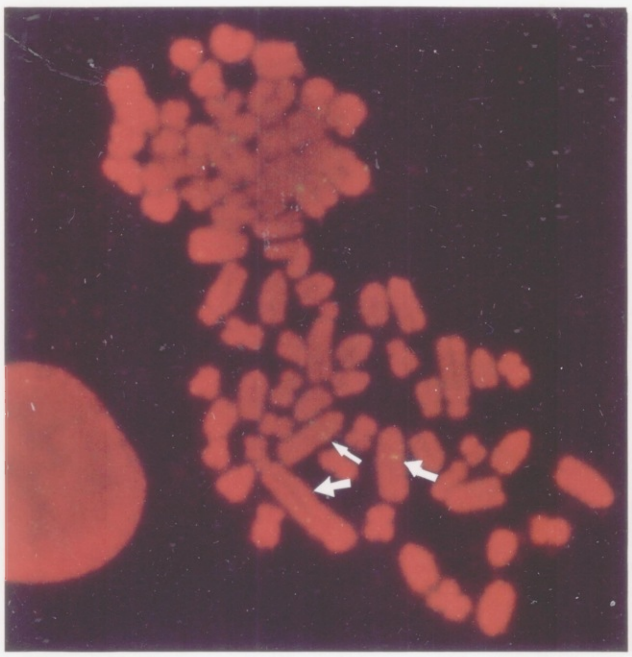
Then 3B and Rat-2 spreads shown hybridised with the λBam probe are representative of all those examined. The comparison between the two cell lines described here revealed that the single copy signal detected on one of the Rat-2 chromosomes was absent in 3B cells. Assuming that the 3B clone arose from a Rat-2 cell containing two copies of this chromosome, there are a number of possibilities. Either the amplified DNA in 3B cells is located on the same chromosome but at a different position; or else the amplified DNA is located on a completely different chromosome, and the Rat-2 chromosome was lost or rearranged. Both of these explanations take into account the fact that the original region of rat DNA, into which the transfected DNA integrated and which was then amplified, is apparently no longer present in 3B cells. The region has either been deleted or the chromosome on which it lay has been lost or rearranged. In order to investigate these possibilities further, the chromosomes involved would have to be characterised more carefully.

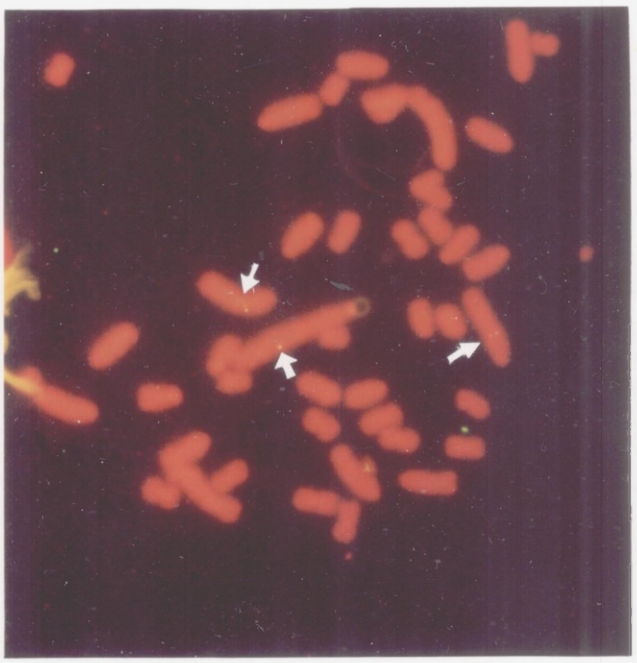
FIGURE 5.3.2

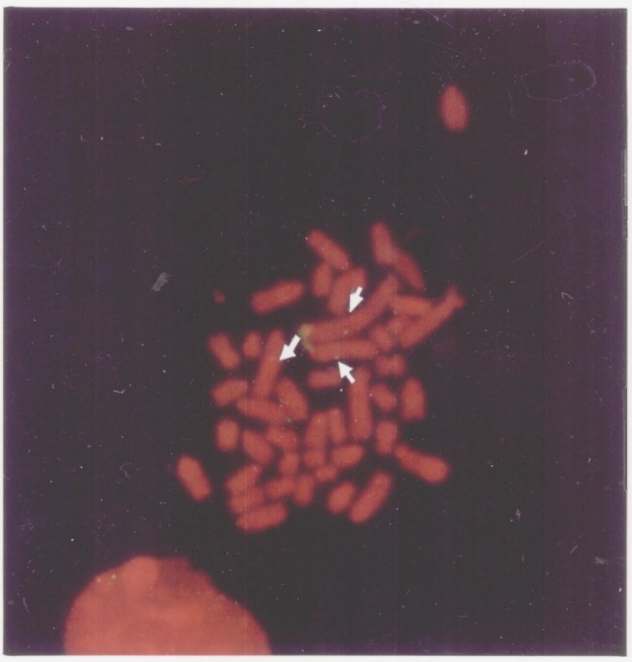
Fluorescence *in situ* hybridisation of metaphase spreads from Rat-2 cells with the λBam rat probe.

Three representative metaphase spreads from Rat-2 cells hybridised with the λBam probe, are shown. This rat DNA probe detects, in Rat-2 cells, the original location of the sequences involved in the amplification in 3B cells, as well as the other native alleles of this region. The probe was preannealed with sonicated rat genomic DNA, as in Fig. 5.3.1.

Single-copy fluorescent signal is detected on three chromosomes in Rat-2 cells (as indicated by the arrows). One of these is the abnormally long marker chromosome also noted in 3B cells (Fig. 5.3.1). The other two chromosomes appear to be identical, and the same as the other chromosome in 3B cells that was found to contain a native allele (Fig. 5.3.1).







5.4 Characterisation of 3B and Rat-2 Chromosomes by G-banding Followed by *In Situ* Hybridisation with the λBam probe.

The relationship between the two apparently identical chromosomes and the marker chromosome (M1) on which single copy signal was detected in Rat-2 cells could not be deduced from the analysis in Part 5.3. Neither could the origin of the chromosome on which the amplified DNA was located be determined. Thus, for an understanding of the possible chromosomal changes (such as rearrangement, deletion or chromosome loss) that occurred during the evolution of the 3B cell line, these chromosomes had to be characterised and distinguished more rigorously.

For this purpose, the technique of G-banding (using Wright's stain, as described in Materials and Methods), which produces characteristic patterns of bands on mammalian chromosomes, was applied to 3B and Rat-2 chromosome spreads. Following photography of the banded chromosomes, the same spreads were fluorescence hybridised with the λBam probe, and rephotographed. In this way, the banding patterns of the different chromosomes on which signal was detected could be compared, and also the position of the hybridisation signal on a chromosome could be more precisely determined by reference to its band location.

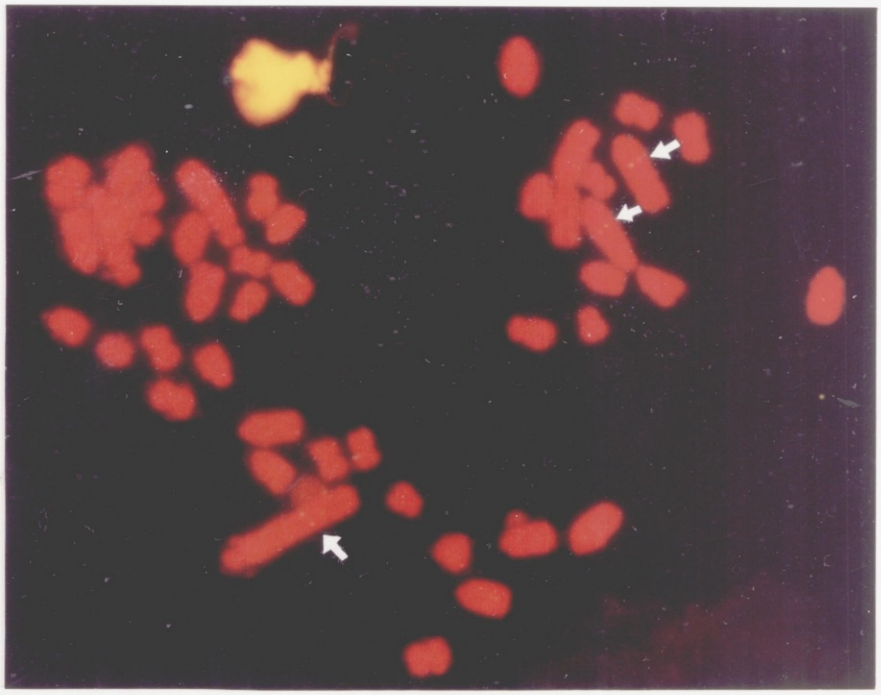
In Fig.5.4.1 a Rat-2 metaphase spread which has been G-banded and then fluorescence hybridised with the λBam probe is shown. The fluorescent signal is located on two chromosomes of similar size and on the marker chromosome M1, as seen previously in Fig. 5.3.2. From their banding patterns it can be that the two chromosomes of similar size are in fact homologues. It was deduced by reference to the standard banded karyotype of the normal rat (Committee for a Standardized Karyotype of *Rattus norvegicus*, 1973 and Fig. 5.4.3 from Levan, 1974) that these were homologues of rat chromosome 2. The Rat-1 cell line (and hence the Rat-2 cell line, see below) was derived from embryo fibroblasts of the Fischer rat (Botchan et al, 1976), which is a strain of the Norway rat. The position of the signal was located in band 2q23 (using the banding nomenclature proposed by Levan, 1974, see Fig. 5.4.4). From its banding pattern it appears that the distal half of the marker chromosome M1

FIGURE 5.4.1

G-Banding of Rat-2 metaphase spreads followed by fluorescence *in situ* hybridisation with the λBam rat probe.

The Rat-2 metaphase spread shown, was Giemsa banded and photographed, after which it was hybridised to the biotinylated λBam probe by fluorescence hybridisation and rephotographed (see Materials and Methods). In this way, the chromosomes to which the λBam probe hybridises in Rat-2 cells can be identified and compared to each other.

In the photograph on the right, the fluorescein signals indicated by the white arrows, are located on three chromosomes as seen in Fig. 5.3.2. The equivalent positions are indicated by black arrows on the G-banded chromosomes seen in the photograph of the same spread on the left. These positions represent the sites of the native alleles of the region found amplified in 3B cells. From the G-banding it can be seen that two of the chromosomes with signal are homologues with identical banding patterns. The distal part of the abnormal marker chromosome M1 also appears to be homologous to these two chromosomes. The signal lies in the same G-band position on all three chromosomes.





(from the striking non-staining gap down to the telomere), is also homologous to chromosome 2, stretching from the proximal 2q1 region to the centromere. The signal is located in an identical position as well, at 2q23. Chromosome M1 appears to be the product of a translocation between a proximally broken chromosome 2 and most of rat chromosome 3 including its centromere, as determined from the standard rat karyotypes (Fig.5.4.3). The majority of Rat-2 cells are thus trisomic for most of chromosome 2.

In Fig.5.4.2, a 3B metaphase spread is shown that was treated in a similar fashion and is a representative example of other spreads examined. Fluorescent signal is again detected on three chromosomes. From the banding patterns of these chromosomes, it can be seen that single copy signal, representing unamplified alleles, is located on one normal homologue of chromosome 2 and on the marker chromosome M1. However, the amplified copies of the region are located on a chromosome with a very different banding pattern to chromosome 2. In all 3B spreads examined in this way only one copy of this chromosome was ever present, and it bears no resemblance to any chromosome seen in Rat-2 spreads. Furthermore, the second normal homologue of chromosome 2, seen in Rat-2 cells, is absent in 3B cells, as judged from the G-banded chromosomes of all 3B spreads examined.

Partial karyotypes of representative 3B and Rat-2 metaphase spreads showing the chromosomes on which signal was detected can be seen in Fig. 5.4.5. This illustrates that the formation of the 3B cell line was probably accompanied by loss or rearrangement of one copy of chromosome 2, and the gain of a new marker chromosome (M2), which bears no obvious resemblance to chromosome 2, or indeed any other endogenous rat chromosomes (see Part 5.5). Any resemblance (or lack of) between the normal and rearranged rat chromosomes mentioned so far can be seen in the schematic representation of these chromosomes shown in Fig. 5.4.6.

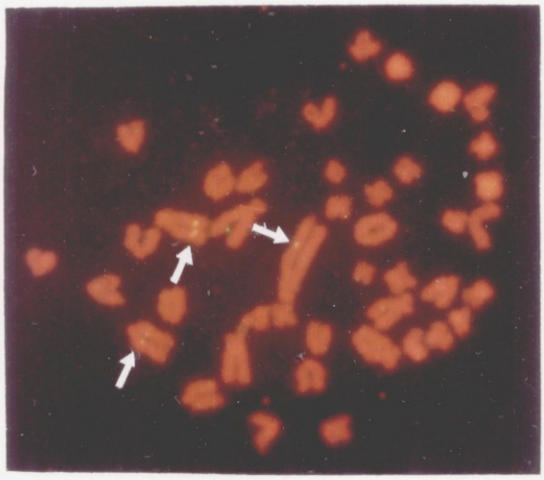
The above analysis revealed a number of unusual points concerning the region of the rat genome which is amplified in 3B cells. Firstly, rather than containing two alleles of this region on homologous chromosomes as would be expected, Rat-2 cells were found to contain an extra allele on a translocated marker chromosome (M1). The M1 chromosome was also present in 3B cells and hence probably did not play a part in the

FIGURE 5.4.2

G-Banding of 3B metaphase spreads followed by fluorescence *in situ* hybridisation with the λBam rat probe.

The 3B metaphase spread shown, was G-banded and then fluorescence hybridised with the λBam probe as described in Fig. 5.4.1. The chromosome which carries he amplified copies of the region detected by the λBam probe, as well as the chromosomes carrying the unamplified native copies, can be identified and compared to each other from their G-banding patterns.

In the photograph on the right, the fluorescein signals indicated by the white arrows, are located on three chromosomes as seen in the 3B metaphase spreads in Fig. 5.3.1. These positions are also indicated by black arrows on the G-banded chromosomes seen in the photograph of the same spread on the left. The G-banding reveals that single-copy signal is located at the same position on <u>one</u> of the two similarly-sized homologues in Rat-2 cells and on the long marker chromosome (M1). However, the pattern of G-bands on the M2 chromosome carrying the amplified DNA (the cluster of fluorescein signals indicated in the photograph on the right), appears to be very different to the other two chromosomes.



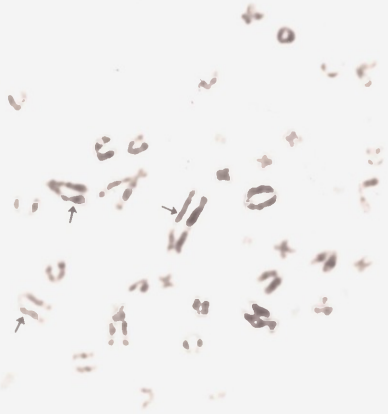


FIGURE 5.4.3

G-banded karyotypes of the Norway rat, *Rattus norvegicus*.

Diagrams of G-banding patterns of rat chromosomes are shown, with each chromosome represented at four different degrees of chromosome condensation (modified from Levan, 1974). Note that the Rat-2 cell line was derived from the Fischer rat (Topp, 1981), which is a strain of *Rattus norvegicus*.

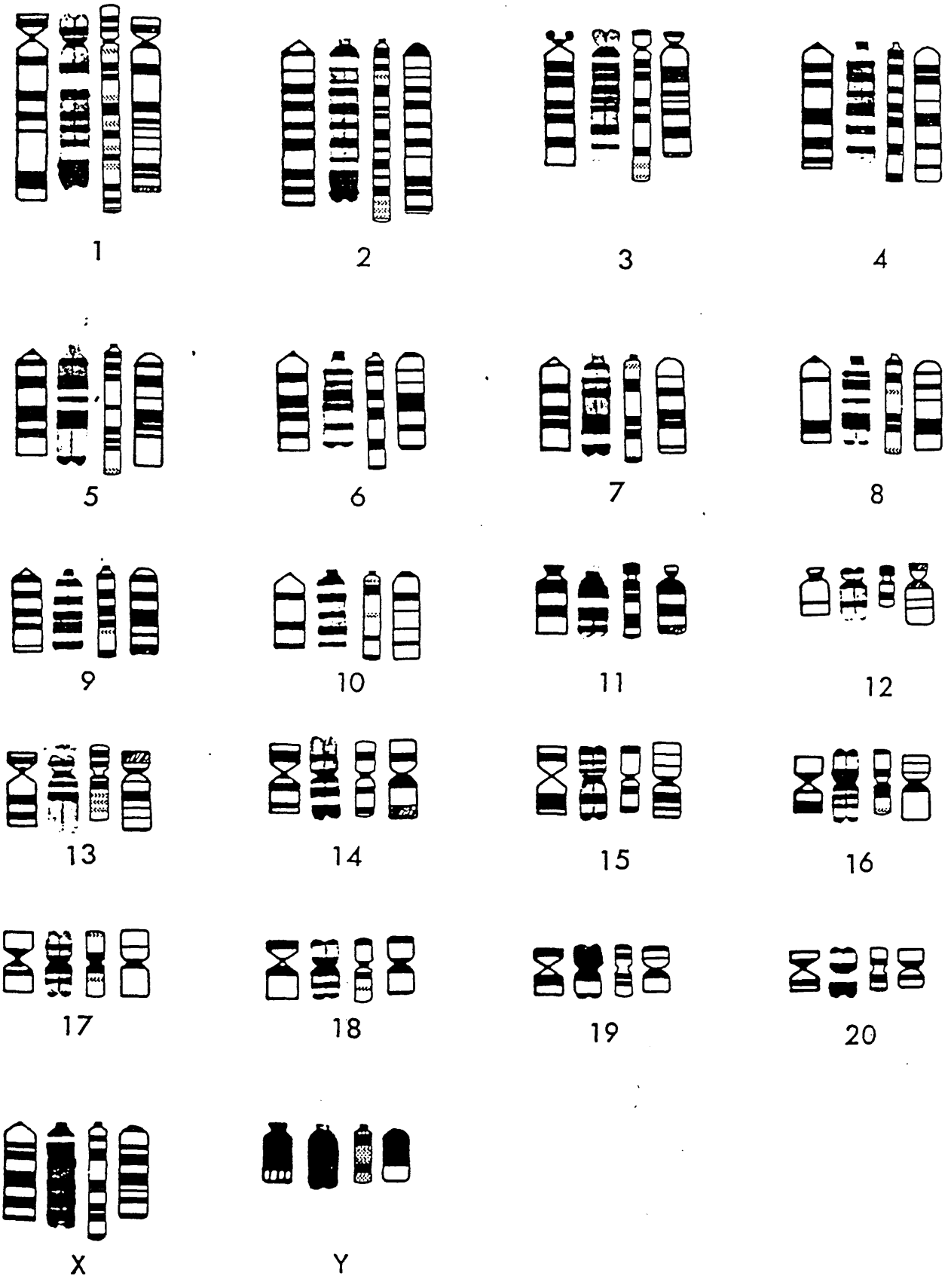

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
X
Y

FIGURE 5.4.4

G-band nomenclature of rat chromosome 2 (adapted from Levan, 1974).

RAT CHROMOSOME 2

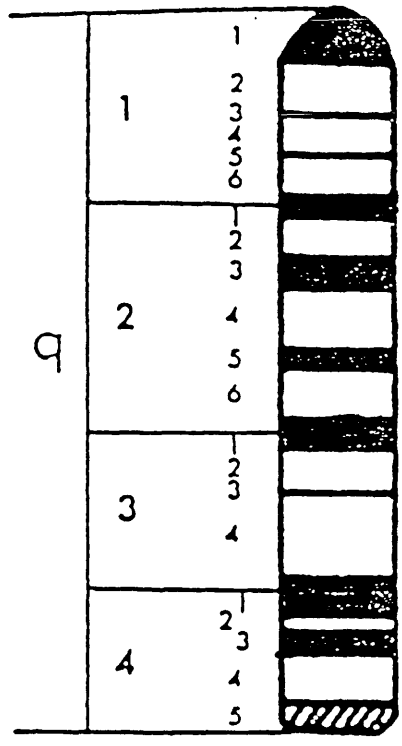


FIGURE 5.4.5

Partial karyotypes of Rat-2 and 3B cells.

The chromosomes to which the λBam probe hybridises, from two Rat-2 and two 3B metaphase spreads, are shown. The upper Rat-2 and 3B partial karyotypes are presented together, as are the lower ones, to facilitate comparison of the G-banding patterns as the degree of chromosome condensation is similar for each pair. The arrows indicate the positions of fluorescein signal detected on each chromosome.

Note that the marker chromosome M2 on which the amplified DNA is located at the position indicated, has a very distinctive G-banding pattern that bears no obvious resemblance to the chromosomes carrying the unamplified alleles (ie chromosomes 2 and t(2;3) / M1

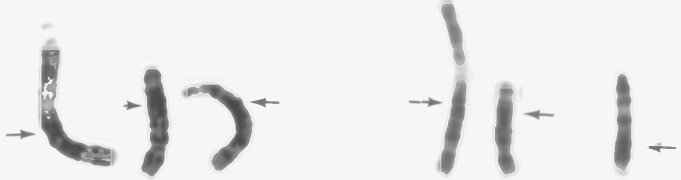
RAT 2

3B



RAT 2

3B



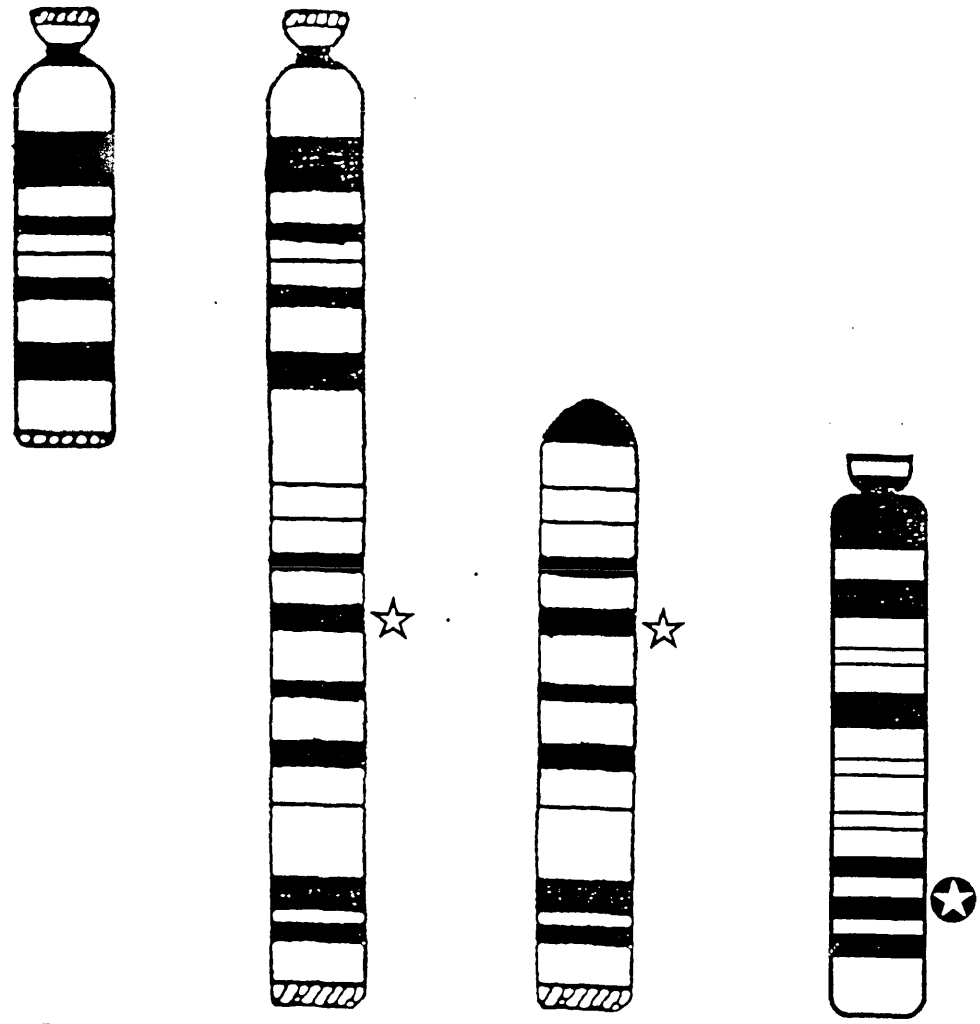
Chromosome: t (2;3) (M l) 2 2 (2;3) 2 M2

FIGURE 5.4.6

Diagramatic representation of G-banding patterns of rat chromosomes 2 and 3, and marker chromosomes M1 / t(2;3) and M2.

The diagrams of rat chromosomes 2 and 3 presented are adapted from Levan (1974). The aberrant M 1 marker chromosome is found in Rat-2 and 3B cells (see Fig. 5.4.5) but not in euploid Rat-1 cells (see Part 5.6). From its G-banding it appears to have been formed by a translocation between the proximal q1 region of chromosome 2 (shown on its right) and the telomeric region of chromosome 3 (shown on its left). The asterisks indicate the positions of native unamplified alleles of the region that hybridises with the λBam probe in Rat-2 and 3B cells.

The marker chromosome M2 is only present in 3B cells. The position of the amplified DNA detected by both the λBam and λ3B probes is indicated by the bold asterisk. The M2 chromosome, although similar in size, has a very different G-banding pattern to rat chromosome 2 (shown on its left).


3
t2;3
(M1)
2
M2

evolution of the amplified DNA. However, it did seem striking that a part of chromosome 2 close to the region which was amplified in 3B cells, had previously been involved in another type of chromosomal rearrangement (a translocation). The exact origins of chromosome M1 are investigated in the next section. It was also noted that the point of translocation on M1 is a non-staining gap, and furthermore, in a number of spreads chromatid breaks were suspected to have occurred in the same region. These observations were investigated further, as described in Part 5.6.

The appearance of the amplified DNA on a new, rearranged chromosome (M2) was coincident with the loss or disruption of a rat chromosome 2 which had carried a native allele of this region, the other two native alleles remaining intact. However, it was not possible from this analysis to determine whether the chromosome 2 present in Rat-2 but not 3B cells, had been rearranged or completely lost. It was also not clear from which endogenous rat chromosomes (including chromosome 2) the new marker M2 carrying the amplied DNA was derived. The rearrangement, or gain or loss of other rat chromosomes in 3B cells compared to Rat-2 cells, would provide useful information about the origins of chromosome M2, and therefore about the chromosomal events which took place when 3B was created.

5.5 Karyotypic Analysis of the 3B and Rat-2 Cell lines.

In order to identify the chromosomal changes that occurred during the evolution of the 3B cell line from Rat-2 cells, G-banded metaphase spreads from both cell lines were karyotyped. The modal chromosome number was also determined for each cell line, as this would indicate any consistent chromosome loss or gain in 3B cells. The aim was that this karyotypic analysis would enable identification of the endogenous chromosome(s) from which the marker chromosome M2 was derived. The exact origins of the translocated marker chromosome (M1), present in both cell lines, could also be investigated.

Metaphase spreads from 3B and Rat-2 cells were G-banded and photographed as before. Some of these spreads were subsequently rephotographed after fluorescence hybridisation with the λBam probe. Furthermore, about forty chromosome spreads

from each cell line were counted. The chromosomes from some of these spreads were ordered for karyotypic identification. The standard karyotype of the Norway rat (see Fig.5.4.3) was used as a reference for chromosome ordering, in addition to G-banded karyotypes of *Rattus norvegicus*, kindly provided by Professor P.H. Gallimore.

Rat-2 is a Tk^- cell line which was isolated after 5'-bromo-deoxyuridine (BrDU) treatment of Rat-1 cells (Topp, 1981), a treatment which is known to cause chromosome breaks and rearrangements. Rat-1 is a fibroblastoid 3T3-like cell line (Botchan et al, 1976) with a modal chromosome number of 2n=42, which is the normal chromosome complement of *Rattus norvegicus*. The Rat-2 cell line, like Rat-1, is density-dependent and contact-inhibited, producing flat monolayers *in vitro*. It is thus reported to be phenotypically "normal" by all criteria tested (Topp, 1981). As mentioned in the Part 5.1, this cell line is near-diploid, its modal chromosome number being 2n=44.

An analysis of Rat-2 chromosomes revealed that the modal chromosome number was 44, which is consistent with the results of Topp (1981). The chromosome number in 3B cells was found to be more variable however. The majority of the spreads examined contained between 40 and 43 chromosomes. A small proportion (< 5%) of those examined contained less than 40 chromosomes, and two were polyploid (4n=83 see Fig. 5.5.5 and 3n=61). None of the (near) diploid 3B spreads examined contained more than 44 chromosomes. This implies that the karyotypic differences in 3B compared to Rat-2 cells are due to chromosome losses. The greater variation in chromosome number compared to Rat-2 was expected in 3B, as it is a transformed cell line and such cells have a much greater tendency to become aneuploid and evolve karyotypic abnormalities than "normal" cells or immortalised cell lines.

A number of ordered karyotypes of Rat-2 chromosomes are shown in Figs. 5.5.1, 5.5.2 and 5.5.3. In all cases, the spread from which each was derived is also shown. From these karyotypes it can be seen that Rat-2 is aneuploid (as expected from its modal chromosome number of 2n=44). An extra two copies of chromosome 7 are usually present, and these are responsible for the increased chromosome number of

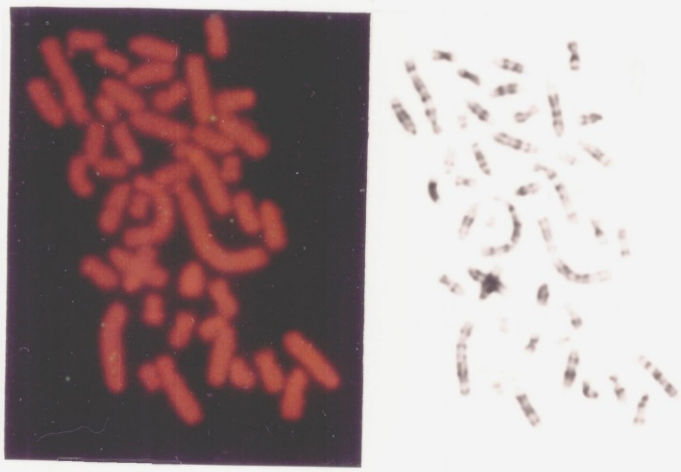


FIGURE 5.5.1

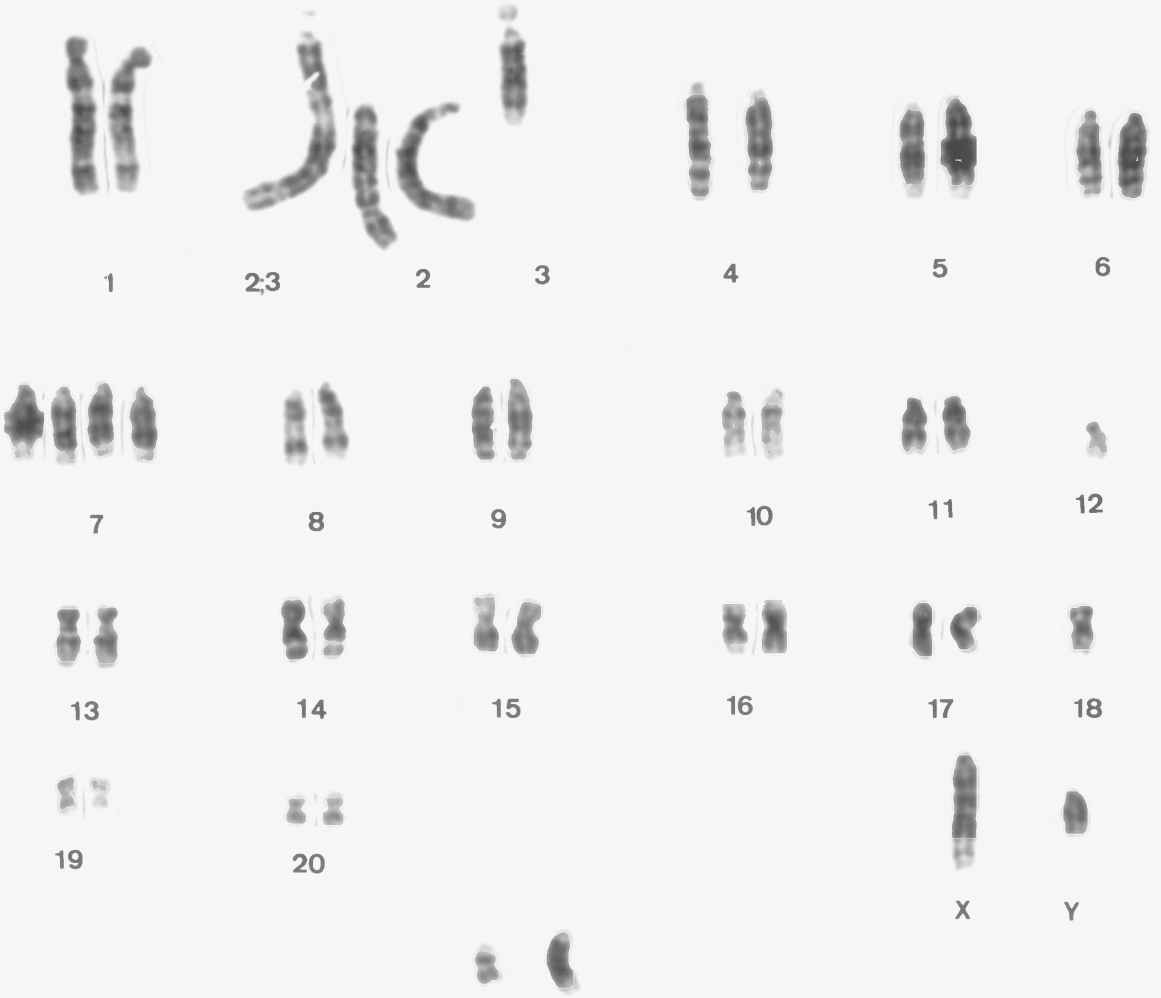
G-banded karyotype of a Rat-2 cell.

On the facing page two photographs are presented of the metaphase spread from which the karyotype overleaf was derived. The same spread was G-banded (left photograph) and stained with propidium iodide following fluorescence *in situ* hybridisation with the λBam probe (right photograph). Note that the efficiency of hybridisation was poor in this case so that signal is not clearly visible on both chromatids of the two chromosomes 2 and chromosome t(2;3).

Overleaf, the G-banded rat chromosomes from this spread have been identified and ordered according to the standard karyotype of the Norway rat (Committee for a Standardized Karyotype of *Rattus norvegicus*, 1973). This was carried out by reference to Fig. 5.4.3 as well as photographs of G-banded rat karyotypes kindly provided by Professor P.H. Gallimore.

The diploid chromosome complement of this Rat-2 spread is 44. Note the presence of two normal chromosome 2 homologues, the single M1 marker chromosome t(2;3), and a single normal chromosome 3. Furthermore, note the four copies of chromosome 7. Two unidentified chromosomes are shown at the bottom, and are probably rearrangements involving small portions of chromosomes, such as the telomeric region of chromosome 4 which appears to be absent in one of the two homologues.

RAT 2



RAT 2

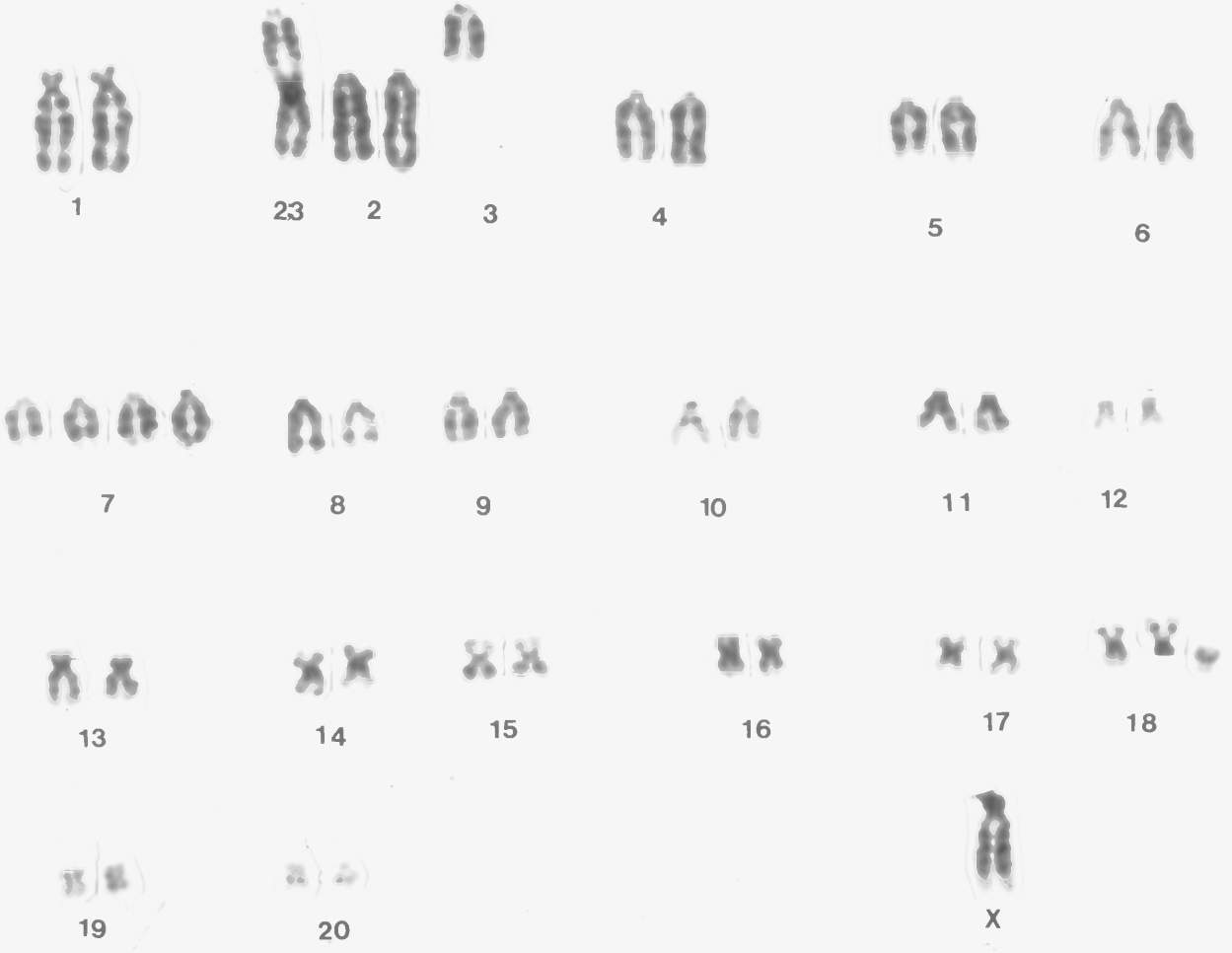


FIGURE 5.5.2

G-banded karyotype of a Rat-2 cell.

On the facing page the G-banded rat chromosomes from a Rat-2 metaphase spread have been identified and ordered as described for Fig. 5.5.1. As in Fig. 5.5.1, two normal chromosome 2 homologues, a single M1 marker chromosome t(2;3), and a single normal chromosome 3, as well as four copies of chromosome 7 are present. However, the Y chromosome appears to be absent which accounts for the chromosome complement of 43 rather 44.

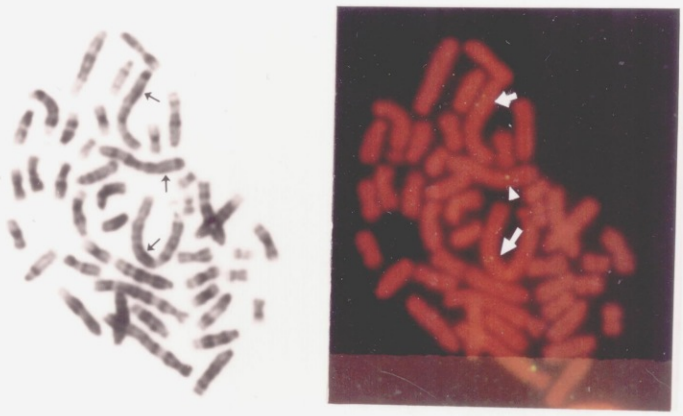


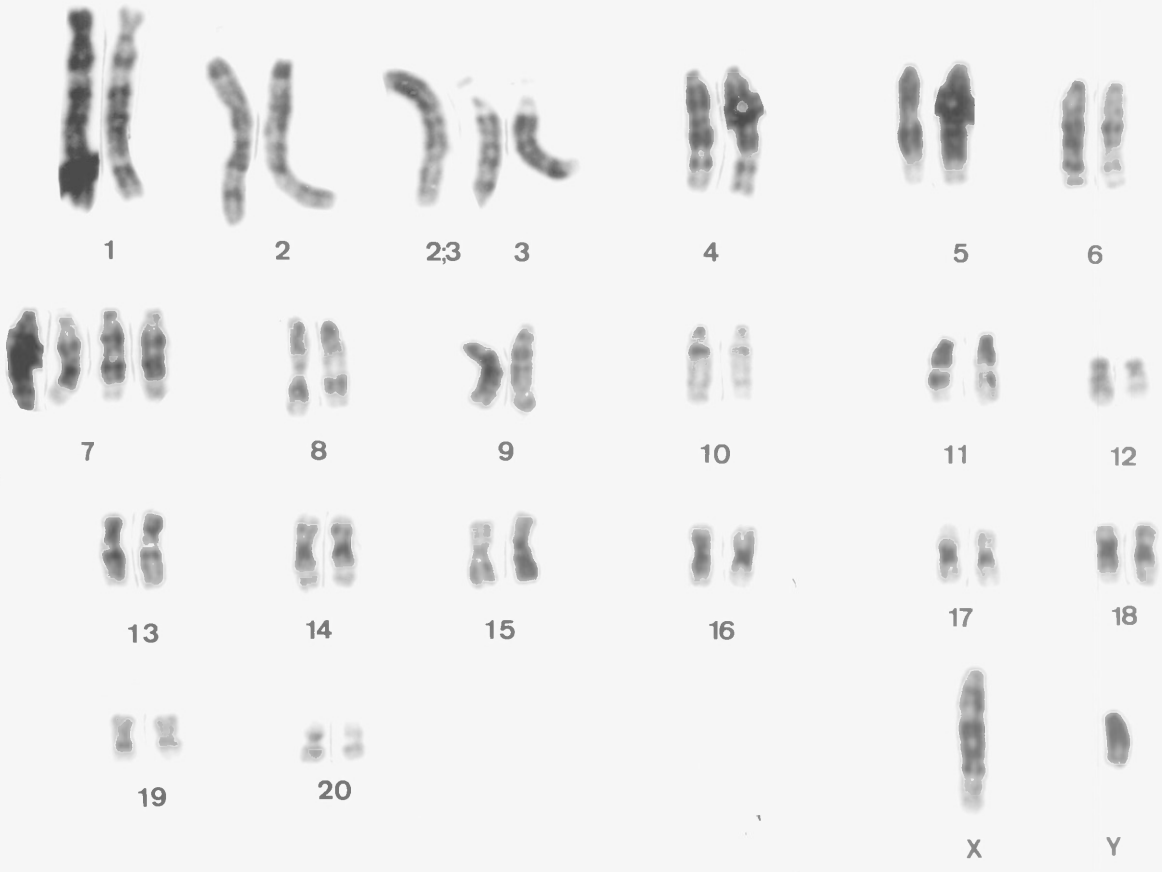
FIGURE 5.5.3

G-banded karyotype of a Rat-2 cell.

On the facing page, two photographs are shown of the metaphase spread from which the karyotype overleaf was derived. The same spread was G-banded (left photograph) and stained with propidium iodide following fluorescence hybridisation with the λBam probe (right photograph). The efficiency of hybridisation was poor, thus fluorescein signal is not visible at the expected positions.

Overleaf, the G-banded rat chromosomes from this spread have been identified and ordered as described for Fig. 5.5.1. As in Fig.5.5.1, two normal chromosome 2 homologues, a single M1 marker chromosome t(2;3), and a single normal chromosome 3 are present. Note that in the karyotype, chromosome t(2;3) has been cut into two halves (see facing page for its true appearance) to facilitate its comparison with chromosomes 2 and 3. Four copies of chromosome 7 are again present, and in this case the Y chromosome can be seen. The chromosome complement is thus 44.

RAT 2



2n=44 in Rat-2 over the normal 2n=42. The two copies of rat chromosome 2 in addition to the translocated marker chromosome M1 are all present, as was shown in the previous section. However, here it can be seen that the proximal region of M1 (from the centromere to the translocation point) corresponds to a complete rat chromosome 3. The complete banding pattern of the normal homologue of this chromosome is present on M1 (see Fig. 5.4.6 also). As there is only one (untranslocated) copy of chromosome 3 per diploid cell, this implies that the presence of M1 is essential for Rat-2 cells to have a correct chromosome complement. Chromosome M1 appears to be the product of a fusion between the telomere of a complete chromosome 3 and most of a (proximally) broken chromosome 2. Its correct nomenclature is therefore t(2;3) and some unusual features of this chromosome are analysed further in Part 5.6.

The Rat-2 karyotype appears to be that of a normal diploid male (XY) apart from the presence of the translocated chromosome M1 and the aneuploidy due to the two extra copies of chromosome 7. The X chromosome does appear to have a telomeric light staining chromosome segment not present on the X chromosome of the standard rat karyotype. However, according to the Committee for a Standardized Karyotype of *Rattus norvegicus* (1973), the rat X chromosome displays a high degree of polymorphism.

Similarly ordered karyotypes of 3B cell chromosomes are shown in Figs. 5.5.4 - 5.5.7. The corresponding chromosome spread from which each was derived is also shown. There is usually only one extra chromosome 7 in 3B cells, unlike Rat-2 cells where there are two. However, in 3B cells an extra copy of what appears to be chromosome 6 is often seen, but this may in fact represent a rearranged extra copy of chromosome 7, as the banding of these two chromosomes is quite similar (refer to standard karyotype in Fig. 5.4.3). As noted in Part 5.4, one of the two normal homologues of chromosome 2 present in Rat-2 cells is absent here, and a new marker chromosome M2 has appeared. In Fig. 5.5.6 the fluorescence hybridised chromosome M2 is shown aligned with its G-banded counterpart. The exact location of the amplified DNA can be seen in this case, as shown diagrammatically in Fig. 5.4.6. The Y



FIGURE 5.5.4

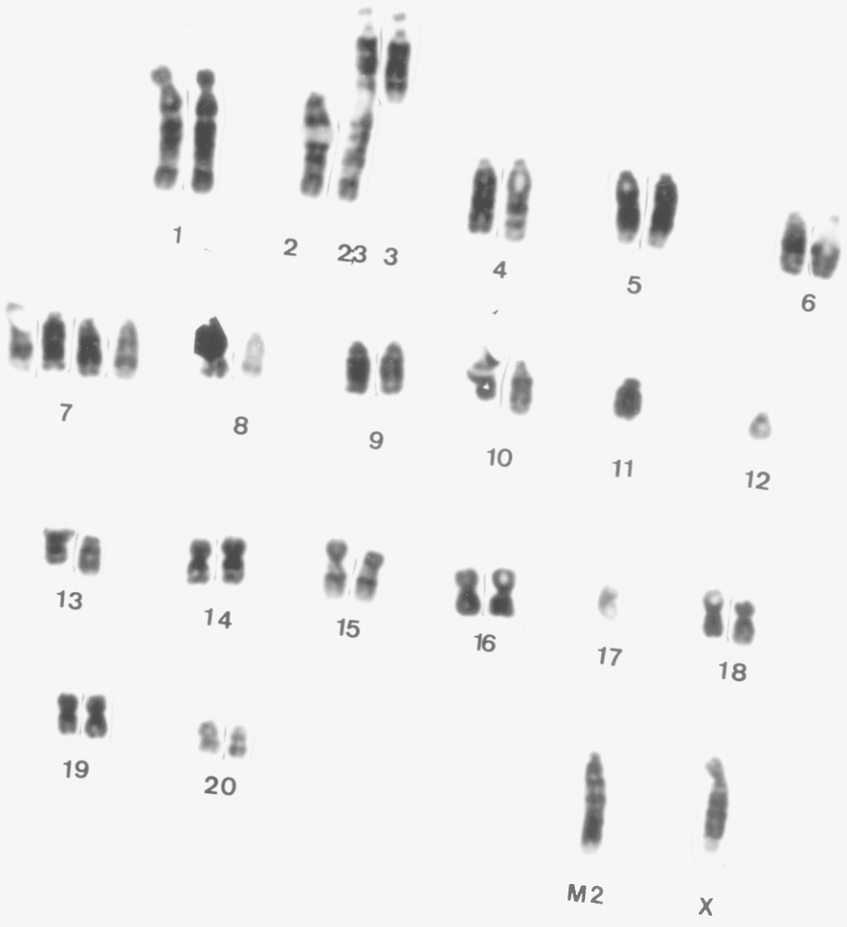
G-banded karyotype of a 3B cell.

On the facing page is shown, a photograph of the G-banded metaphase spread from which the ordered chromosome karyotype overleaf was derived.

Overleaf, the G-banded rat chromosomes from this spread have been identified and ordered as described for the Rat-2 spreads in Figs. 5.5.1-5.5.3. In this 3B cell the chromosome complement is 40. Only one normal chromosome 2 homologue is present, unlike Rat-2 cells, but the M1 marker chromosome t(2;3) and a single normal chromosome 3 are present as in Rat-2 cells. Four copies of chromosome 7 are also present as in Rat-2.

Note the appearance of the rearranged marker chromosome M2 on which the amplified DNA is located, and the concomitant disappearance of one homologue of chromosome 11. The proximal half of chromosome M2 appears to be identical to the remaining copy of chromosome 11. Finally, note that the Y chromosome is absent in this spread, and that the X chromosome appears to be centromerically fused to another chromosome.

3 B





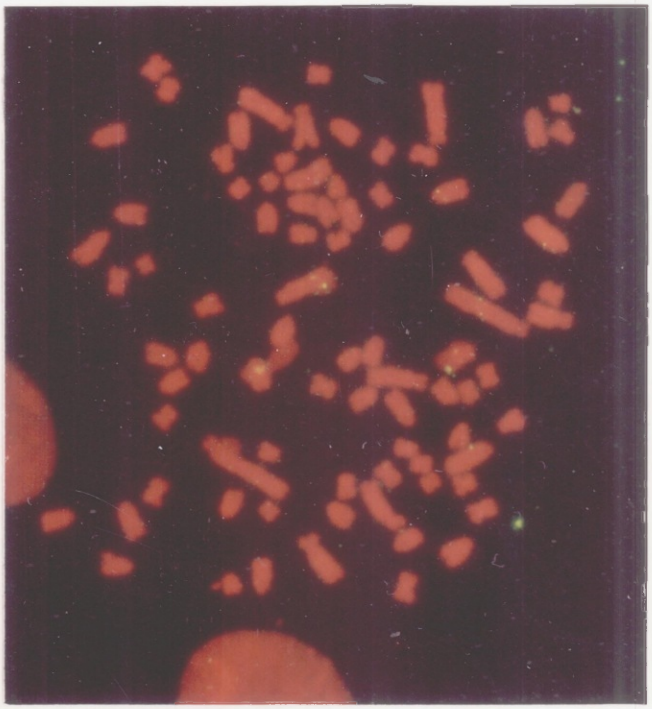


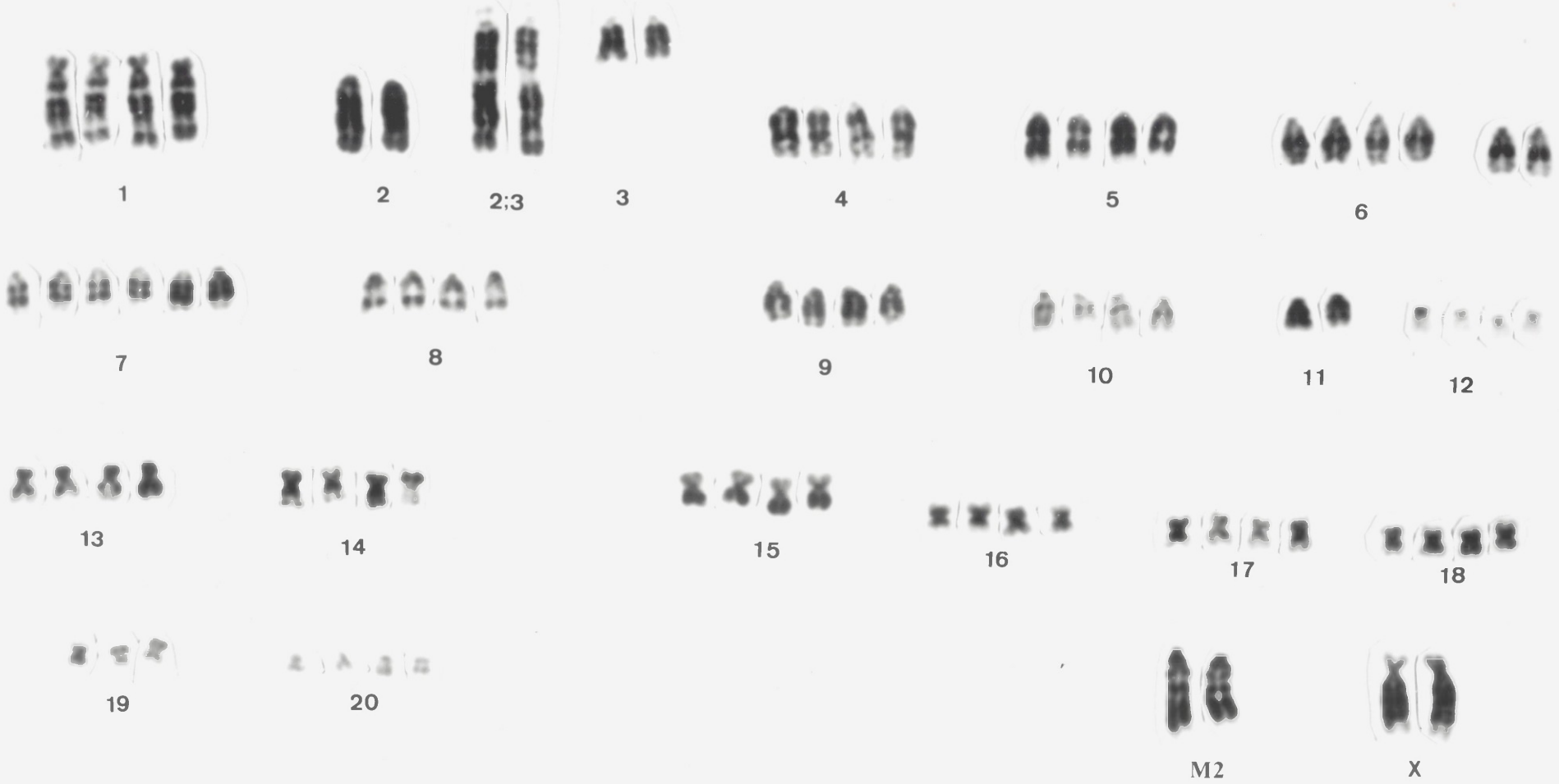
FIGURE 5.5.5

G-banded karyotype of a 3B cell.

On the facing page, two photographs are shown of the metaphase spread from which the karyotype overleaf was derived. The same spread was G-banded (left photograph), and stained with propidium iodide following fluorescence hybridisation with the λBam probe. The positions of hybridisaton signals on the expected chromosomes are indicated by arrows, although fluorescein signal cannot be seen on both chromatids in each case, as the efficiency of hybridisation was poor.

Overleaf, the G-banded rat chromosomes from this spread have been identified and ordered as described for the Rat-2 spreads in Figs. 5.5.1-5.5.3. This 3B cell is approximately tetraploid, with a chromosome complement of 83. As in Fig. 5.5.4, the presence of two M2 marker chromosomes is accompanied by the absence of two normal copies of chromosome 11 (in a tetraploid cell). The relative numbers of chromosomes 2, 3 and t(2;3) are also the same in this spread as in the others shown in Figs. 5.5.4, 5.5.6 and 5.5.7. Also note that the Y chromosome is again absent, and the X chromosome appears to be centromerically fused to another chromosome (perhaps different to that in Fig. 5.5.4). There is also heterogeneity in numbers for some of the smaller chromosomes, such as 19.

3 B



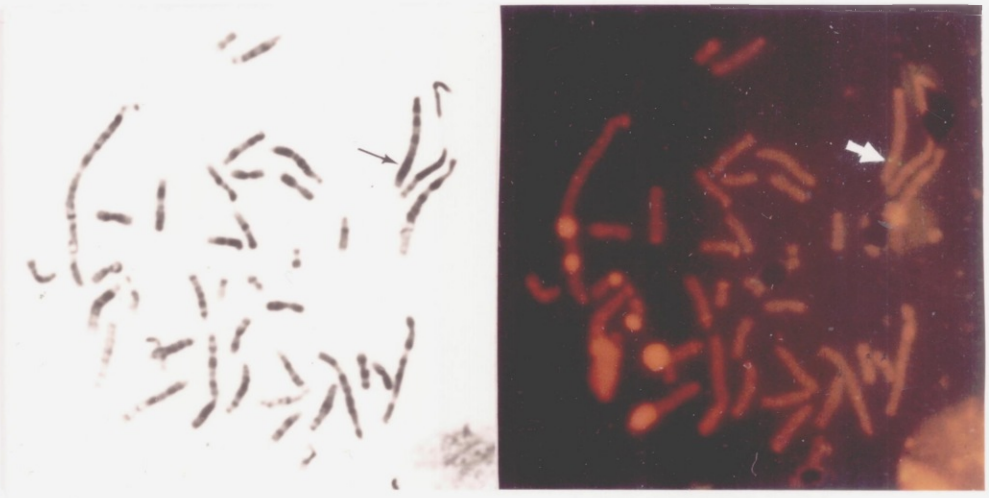


FIGURE 5.5.6

G-banded karyotype of a 3B cell.

On the facing page, two photographs are shown of the metaphase spread from which the karyotype overleaf was derived. The spread was G-banded (on the left), and stained with propidium iodide following hybridisation with the λBam probe (on the right). The hybridisation efficiency was poor so that fluorescein signal is not clearly visible at all the expected locations. However, the signal on chromosome M2 representing the amplified DNA is visible. This chromosome is therefore shown below together with its G-banded counterpart, as the G-banding is well-resolved and enables fairly precise localisation of the signal on the chromosome, as shown schematically in Fig. 5.4.6.

Overleaf, the G-banded rat chromosomes from this spread have been identified and ordered as described for previous karyotypes (see Fig. 5.5.1). The chromosome complement is 40 in this cell. Note (as in Figs. 5.5.4 and 5.5.5) the presence of a single chromosome 2, the M1 marker chromosome t(2;3) and a single chromosome 3, as well as the rearranged marker chromosome M2, but the absence of one homologue of chromosome 11. In this spread only three copies of chromosome 7 are present. The Y chromosome is present, but translocated with one chromosome 4 homologue. The X chromosome is centromerically fused with another chromosome (as in Figs. 5.5.4 and 5.5.5). Note also that this cell is monosomic for chromosomes 12 and 19.

3 B

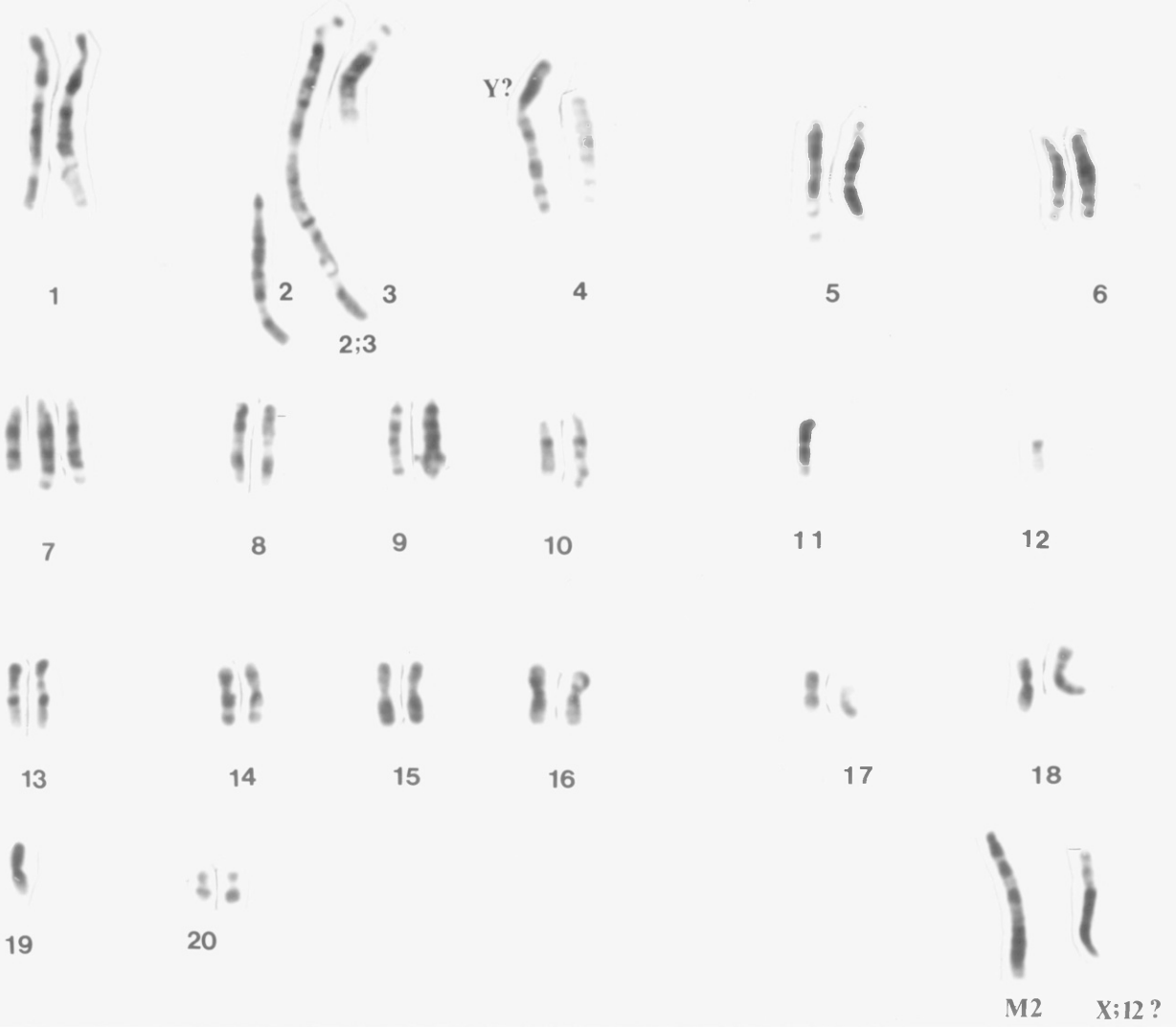






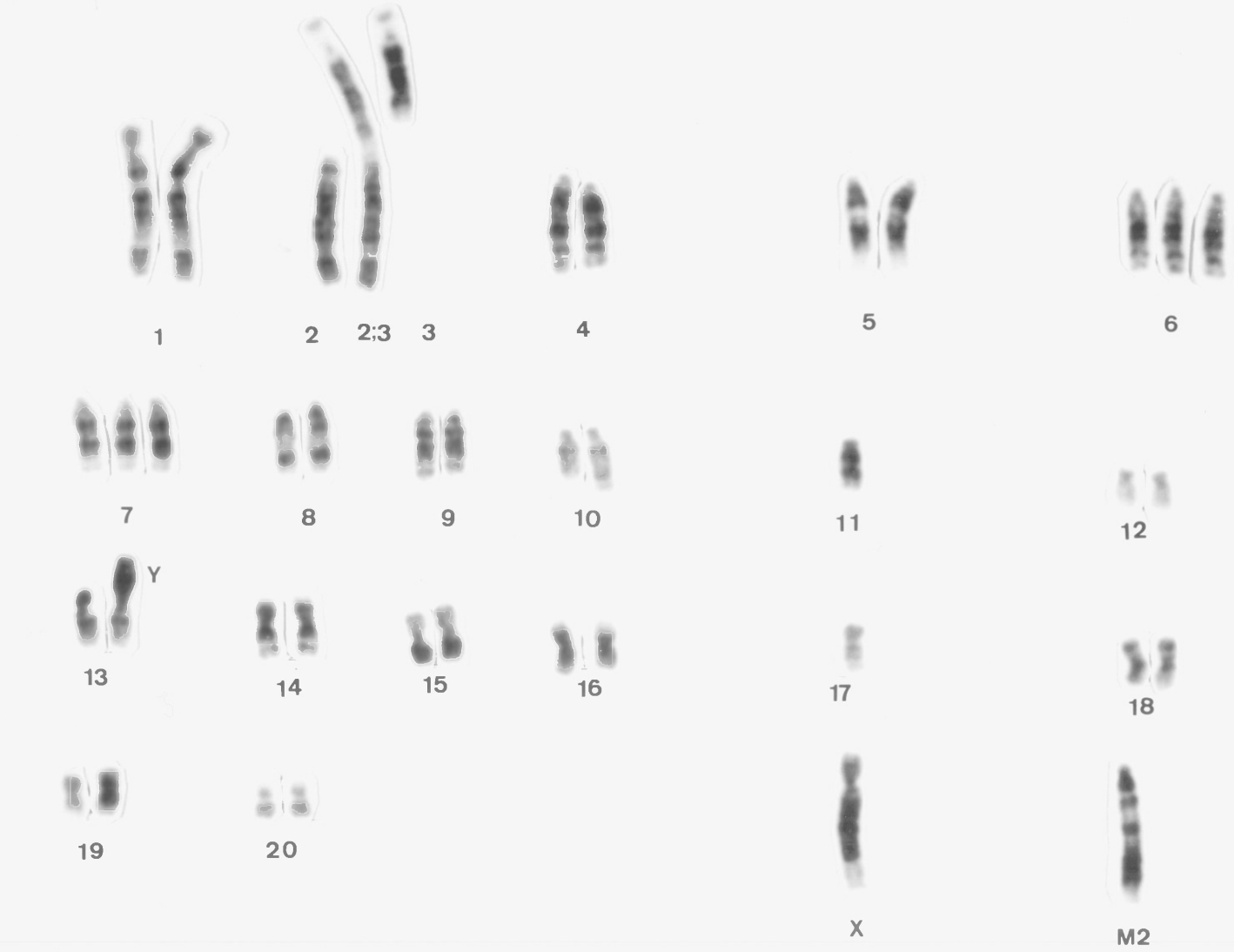
FIGURE 5.5.7

G-banded karyotype of a 3B cell.

On the facing page, two photographs are shown of the metaphase spread from which the karyotype overleaf was derived. The spread was G-banded (on the left), and stained with propidium iodide following hybridisation with the λBam probe (on the right). The hybridisation efficiency was poor and the fluorescein signal is not clearly visible at all the expected locations.

Overleaf, the G-banded rat chromosomes from this spread have been identified and ordered as described for previous karyotypes (see Fig. 5.5.1). The chromosome complement is 39 in this cell. Note (as in the 3B karyotypes in Figs. 5.5.4 and 5.5.5) the presence of a single chromosome 2, the M1 marker chromosome t(2;3) and a single chromosome 3, as well as the rearranged marker chromosome M2, but the absence of one homologue of chromosome 11. As in the previous 3B karyotype (Fig. 5.5.7), only three copies of chromosome 7 are present, and there also appear to be three copies of chromosome 6, one of which might be a deleted or otherwise rearranged chromosome 7. The Y chromosome is present, but translocated with part of chromosome 13 in this case. The X chromosome is centromerically fused with another, different, chromosome (as in Figs. 5.5.4 - 5.5.6). Note that this cell is monosomic for chromosome 17.

3B



chromosome in 3B cells was usually either absent (see Figs. 5.5.4 and 5.5.5) or else fused with the centromeres of a variety of other chromosomes (see Figs. 5.5.6 and 5.5.7), indicating that its centromeric functions were, for some reason, less stable than they had been in Rat-2 cells. The X chromosome, which is acrocentric in Rat-2 cells was found to be biarmed in 3B cells. This appears to be the result of a Robertsonian fusion with another, smaller chromosome or chromosome segment of unclear origin.

A further difference between the Rat-2 and 3B karyotypes is the consistent absence of one copy of chromosome 11 in 3B cells. In all the diploid karyotypes examined where the chromosome banding was adequately clear for chromosome 11 to be distinguished, only one copy of it could be found. Closer examination of the M2 chromosome (on which the amplified DNA is located) reveals that its proximal region displays a G-banding pattern reminiscent of chromosome 11. Chromosome 11 is the only chromosome in the rat karyotype with very intense areas on both sides of the centromere and over the middle of the long arm (see Fig. 5.4.3 and Levan,1974). The proximal region of M2 also has these distinctive features. However, the distal region of M2 (below the putative chromosome 11 portion) in which the amplification resides (see Fig. 5.4.6), has a banding pattern which does not obviously resemble that on any other rat chromosome or part of a chromosome. The region around the ampliifed DNA consists of three very intensely stained bands. This chromosome segment could conceivably be derived from the q2 region of chromosome 2 (see Fig. 5.4.4) which contains the native allele of the amplified DNA at 2q23. The four bands in the 2q2 region are described as being of medium intensity by Levan (1974). However, this probably does not correspond to the chromosome segment which contains the amplified DNA on M2, as the bands of the q2 region in the copy of chromosome 2 are never as darkly stained or as well resolved as those in the M2 chromosome in the same 3B spread (see Figs.5.4.5 and 5.5.4-5.5.7). In addition, the presence of a pale band and another very dark band proximal to this region in M2, is a feature which is not observed in the 2q2 region. However, a segment that has been transferred to a new chromosomal location may well exhibit bands of different intensity, and indeed other

complex rearrangements may also have occurred thus altering the banding pattern especially around the region of translocation.

The consistent absence of a chromosome 2 and a chromosome 11, and the presence of the new M2 marker chromosome in 3B cells when compared with Rat-2 cells (2n=44), would lead to a modal chromosome number of 43. However, in 3B cells, the fusion of the X chromosome with another small chromosome as well as the frequent absence or fusion of the Y chromosome reduce this number further to 42 or, more often, 41. Furthermore, in 3B cells one or two other chromosomes (usually smaller ones, such as chromosomes 12, 17 or 19 see Figs. 5.5.4-7) are sometimes absent or rearranged so as not to be discernable. This accounts for the variation in chromosome number between 39 and 43 found in 3B cells, when compared to the more constant number of 44 in Rat-2 cells.

From this analysis it was deduced that the new M2 chromosome which carries the amplified DNA was almost certainly the product of a fusion between the telomere of chromosome 11 and another chromosome fragment. The amplification of a region of DNA present on chromosome 2 appeared to be correlated to the disappearance of this chromosome in 3B cells. However, from its banding pattern, the fragment on which the amplified DNA on the new chromosome M2 was located appeared not to be necessarily derived from the disrupted chromosome 2. None of the other chromosomal differences or rearrangements consistantly found in 3B cells, when compared to Rat-2 cells, could easily be seen to have arisen as a consequence of the disruption of chromosome 2 and/or the formation of chromosome M2.

5.6 Investigation of the Translocated Marker Chromosome M1 / t(2;3).

The marker chromosome (M1) present in Rat-2 and 3B cells appeared to have arisen by a translocation between the (2q2 - 2qter) region of chromosome 2 and the (3cen - 3qter) region of chromosome 3 (see Fig. 5.4.6). Its correct nomenclature is therefore: t(2;3)(3cen - 3qter: : 2q21 - 2qter). The breakpoint on chromosome 2 (within 2q1) which produced this translocation, is thus close to the 2q23 region (of a different copy of chromosome 2) which was used in the DNA amplification in 3B cells. The fact that

the same part of chromosome 2 which was involved in the integration of transfected DNA, the formation of a large inverted duplication and its amplification, had also been involved in a translocation event at another time, suggested that the region was highly recombinogenic. A further observation which reinforced this idea, was that *in situ* hybridisation with the λBam rat probe occasionally revealed the presence of two adjacent copies of the unamplified native allele in both Rat-2 and 3B cells (for one example see top left hand spread in Fig. 5.3.2). This suggested that some kind of duplication of this chromosomal region was occurring. For these reasons it was decided to investigate chromosomes 2 and M1 [or t(2;3)] further, in order to gain some insight into the recombinogenic nature of this region.

The Rat-2 cell line was derived from Rat-1 cells via BUdR treatment (see Part 5.3), and slection for Tk^- phenotype. The agent BUdR is known to alter normal DNA replication and cause chromosome breaks and rearrangements (Hsu and Somers, 1961), chromosome spreads from Rat-1 cells were therefore examined for the presence or absence of chromosome M1 [t(2;3)], to determine whether the translocation event occurred as a consequence of the BUdR treatment. Metaphase spreads from Rat-1 cells were prepared and fluorescence hybridised with the λBam probe in the same way as described previously for 3B and Rat-2 cells. The results of hybridisation of Rat-1 chromosome spreads with probe are shown in Fig.5.6.1. The probe hybridised to two chromosomes of the same size in all the Rat-1 spreads examined. From their size these chromosomes appear to be the two homologues of chromosome 2. The chromosome M1 [t(2;3)], which is readily distinguished by its abnormal length as well as hybridisation to the λBam probe, was not detected in any of the Rat-1 spreads looked at (>30). Furthermore, two copies of a normal chromosome 3 always appeared to be present in Rat-1 cells. The t(2;3) translocation may therefore have occurred during the generation of the Rat-2 cell line, with the break in chromosome 2 perhaps being caused by the BUdR treatment of Rat-1 cells. The fact that the break occurred in this part of chromosome 2 suggests that the region is recombination-prone, as mentioned before.

In the two Rat-1 spreads shown in the two upper spreads in Fig. 5.6.1, it may be

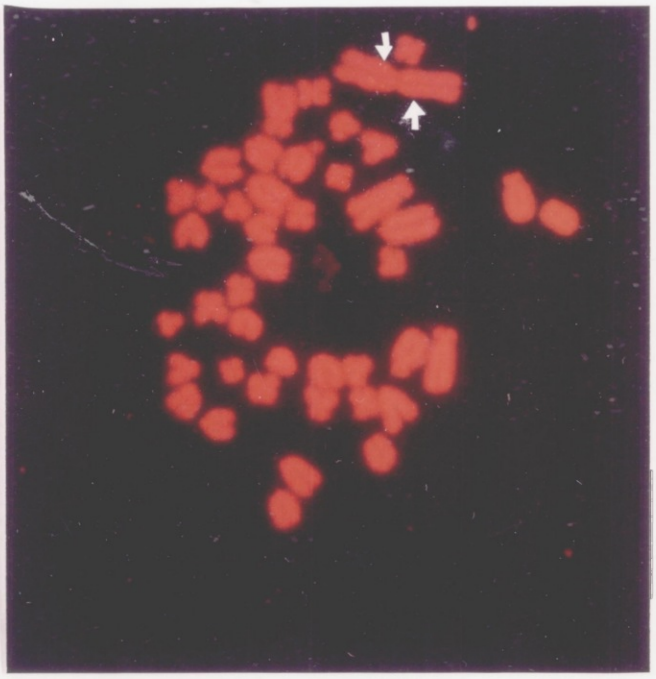
FIGURE 5.6.1

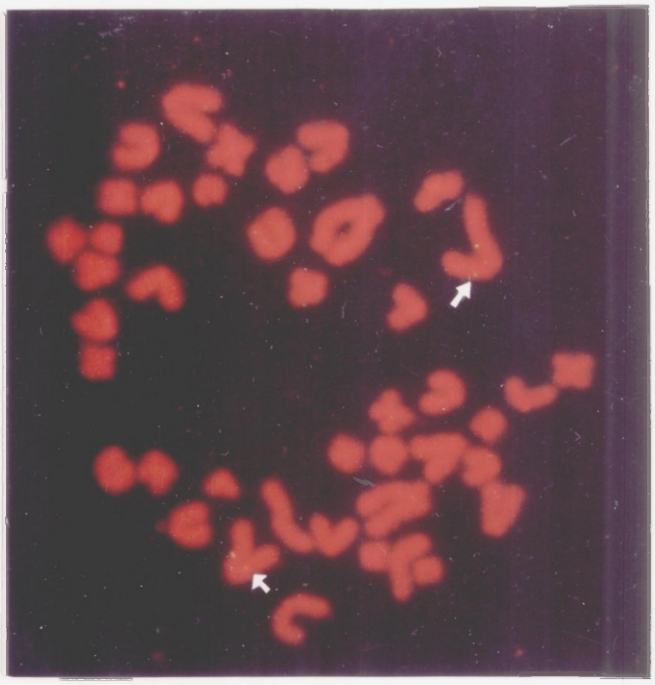
Fluorescence *in situ* hybridisation of metaphase spreads from Rat-1 and Rat-2 cells with the λBam rat probe.

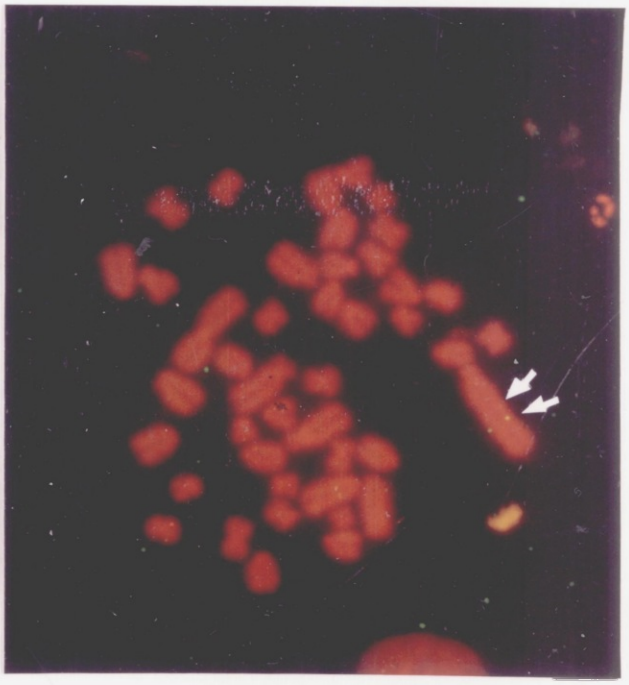
The two upper photographs are metaphase spreads of Rat-1 cells, whilst the lower photograph is a Rat-2 cell. All three spreads were fluorescence hybridised with the λBam probe. The chromosome spreads and biotinylated probe were prepared and hybridised exactly as described in Fig. 5.2.1. The locations of fluorescent hybridisation signal are indicated by arrows.

In Rat-1 cells, signal is only detected on two chromosomes of a similar size (upper photographs). In Rat-2 cells, signal is additionally seen on the long translocated marker chromosome M1 / t(2;3) (see lower photograph and previous figures). This distinctively long chromosome was absent in all Rat-1 spreads examined.

In each of the upper Rat-1 spreads, two fluorescein signals (rather than one) are just visible on a single chromatid of one of the chromosomes showing hybridisation with the probe. This is even more clearly seen in the lower Rat-2 spread, where double fluorescein signals are seen on both the chromatids of the M1 / t(2;3) chromosome, as indicated by the two arrows.







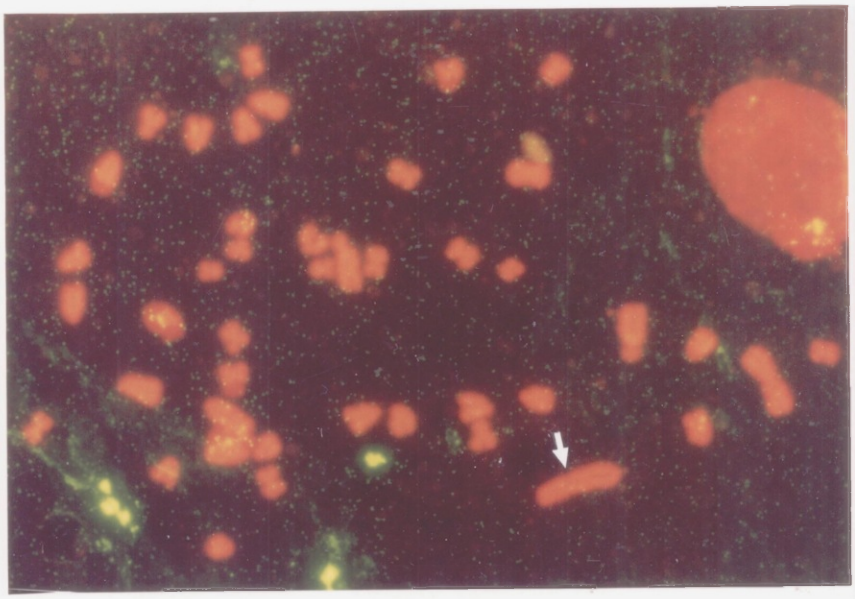
seen that two signals are detected on a single chromatid of one of the copies of chromosome 2 in both cases, implying a duplication of the region has occurred here. This was not seen in all spreads and rarely on both the homologues, either because the level of detection was not adequate or because the duplication event only took place occasionally. This phenomenon was also observed in Rat-2 and 3B cells, where copies of both chromosome 2 and chromosome t(2;3) were occasionally found to have double signals on one or both chromatids (see lower spread in Fig.5.6.1). Furthermore, in some cases where the same spread had been G-banded and then fluorescence hybridised, the presence of two signals on a chromatid corresponded to the presence of an extra G-band (or bands) if the resolution of the banding pattern was good enough. Thus, it seemed that a large duplication of this chromosomal region (the signals are several hundred thousand kilobase pairs apart) sometimes occurred. It was not certain whether the same phenomenon occurred in the chromosome carrying the amplified DNA (M2), as the signals due to the multiple amplified copies of the region obscured the detection of single duplication events.

G-banding of chromosome t(2;3) always revealed the presence of a non-staining gap at the putative translocation site. Actual chromosome breaks had also occasionally been detected in the same region, in spreads which had been G-banded or propidium-iodide stained (after fluorescence hybridisation), as shown in Fig. 5.6.2. Such chromosome breaks suggested the presence of a fragile site in this region. Fragile sites are characterized cytologically as specific regions which exhibit chromosome gaps or breaks (see Sutherland and Hecht, 1985, for a review). Breakage of one or both chromatids of a chromosome usually occurs spontaneously at fragile sites at a much higher frequency than random chromosome damage. Fragile site expression can also be induced by a number of chemical agents such as BUdR, FUdR or aphidicolin (Sutherland and Hecht, 1985), all of which are known to interfere with DNA synthesis. In the case of the putative fragile site on chromosome t(2;3), the frequency of spontaneous chromatid breaks was high enough for scoring without the need for chemical induction. As a chromosome break cannot easily be distinguished from a lightly stained band on a Giemsa-banded chromosome, solid Giemsa staining of Rat-2

FIGURE 5.6.2

Detection of a fragile site on marker chromosome M1 / t(2;3).

The 3B metaphase spread shown, was G-banded and then fluorescence hybridised with the λBam probe. However, fluorescein signal is not visible as the hybridisation efficiency was poor. The arrow indicates the position of a chromatid break on chromosome t(2;3), which is visible in both the G-banded spread as well as the propidium iodide stained spread. From the G-banded spread it can be seen that the chromatid break (the fragile site) is adjacent to, but not at the non-staining gap which represents the translocation point between chromosomes 2 and 3 in the formation of the M1 chromosome. The fragile site (arrowed) in fact seems to be located in the same band as the position of the signal detected by the λBam probe on this chromosome (see Figs. 5.4.5 and 5.4.6)





chromosome spreads was carried out. This technique (described in Materials and Methods) reveals chromatid breaks very clearly. Over one hundred Rat-2 spreads were scored for the presence of a chromatid break at the location of the putative fragile site on chromosome t(2;3). A frequency of 3% was found, and two representative positively scored breaks are shown in Fig. 5.6.3. It was not certain whether the chromatid breaks seen were located at the translocation site (2q1), or at the site of the region involved in the amplification (2q23). However, careful examination and measurement shows that the breaks are located more distally on chromosome t(2;3) than the lightly staining translocation breakpoint, which is still discernable with solid staining (Fig. 5.6.3). The non-staining gap is consistantly seen at a point that is 0.4 x the length of the chromosome away from the centromere, whilst the fluorescent signal is seen at a position between 0.6 and 0.7 x the length of the chromosome. All the chromatid breaks examined had occurred at a position which was about 0.65 x along the length of the chromosome. In Fig.5.6.2, close inspection of the G-banded spread also shows that the non-staining gap representing the translocation point is intact, whilst the chromatid break is located near or within the 2q23 band. These results imply that a fragile site is located at or very close to the region detected by the λBam probe, and not at the translocation breakpoint as had been expected. Spontaneous fragile site expression was not detected in this chromosomal region in the two unrearranged homologues of chromosome 2 in any of the spreads examined. However, the fragility of the region on chromosome t(2;3) may be enhanced, both by its close proximity to a translocation junction and the abnormal length of this chromosome.

5.7 DISCUSSION.

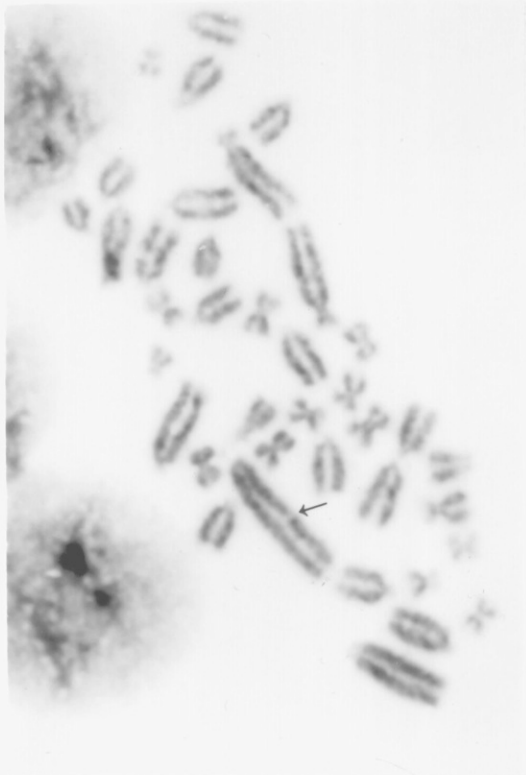
The cytogenetic analysis of the amplified DNA in 3B cells described in this chapter, confirmed many of the predictions of the molecular analysis in Chapter 3. Fluorescence *in situ* hybridisation of 3B metaphase spreads with the amplification-specific λ3B probe revealed that all the amplified DNA was located at a single position on a chromosome. The chromosomal location of the amplified DNA was found to be

FIGURE 5.6.3

Detection of the fragile site on chromosome t (2;3) using Giemsa solid staining.

Two Rat-2 metaphase spreads which have been stained with Giemsa are shown. Solid staining (see Materials and Methods) reveals the presence of chromosome breaks, as opposed to non-staining gaps which are seen after Giemsa banding and can be mistaken for breaks. Over one hundred such spreads were examined, and these are two of the three in which a chromatid break, indicated by the arrow, was detected on chromosome t(2;3).

The appearance of a chromatid break in 2-3 % of cells in a single position on a chromosome, is the rate expected for spontaneous expression of a fragile site. Note that the position of this fragile site is more distal than the expected position of the non-staining gap / translocation junction on this chromosome (as described in Part 5.6).





constant in early and late passage cells, and no acentric fragments such as DMs were ever observed in any of the cells examined. Furthermore, there was no evidence of frequent aberrant chromosome structures such as dicentrics and ring chromosomes which are characteristic of unstable, chromosomally-located amplified DNA, which is still evolving karyotypically (as described in Chapter 1). The observations in 3B cells are thus consistent with the stability in both the structure and copy number of the amplification units found in Chapter 3. The size of the amplified region estimated cytologically was 10-11 Mb, confirming the PFGE analysis described in Part 3.5 which predicted that the amplified array was at least 10 Mb in size. An array of 10-11 Mb is consistent with a twenty-fold amplification of units 450 kb in size. The relatively small size of the array in 3B cells, compared to other cell lines which contain more highly amplified genes and/or much larger amplification units, accounts for the absence of an ECR or HSR at its chromosomal location.

In Part 5.3, fluorescence *in situ* hybridisation analysis using the λBam rat DNA probe revealed the chromosomal locations of the unamplified alleles of the rat sequences amplified in 3B cells. The parallel G-banding and fluorescence *in situ* hybridisation of both 3B and Rat-2 cells (Part 5.4), identified rat chromosome 2 (G-band q23) as the original location of the amplified sequences. This chromosome band was therefore the integration site of the transfected mouse and polyoma virus DNA.

Hybridisation with the λBam probe also revealed that in Rat-2 cells there are three unamplified alleles of the region, not two as might have been expected. The extra allele in Rat-2 cells was present in a translocated marker chromosome t(2;3) which contained most of chromosome 2 (from q1 to qter) and all of chromosome 3. In addition two normal homologues of chromosome 2 were also present in Rat-2 cells.

In 3B cells <u>two</u> unamplified alles were detected by λBam, in addition to the amplified DNA. One of the two normal homologues of chromosome 2 was missing in 3B cells, implying that this chromosome may have been the original site of integration and of the subsequent amplification, and had been disrupted by one of these events. The fact that Rat-2 cells are almost trisomic for chromosome 2 [through the presence of chromosome

t(2;3)], may have meant that the integration and amplification events that led to the disruption and loss of one homologue in the 3B clone, were not deleterious to the cell as they did not reduce it to monosomy for chromosome 2. Furthermore, the absence of any recognizable fragments of the disrupted chromosome 2 in 3B cells, might be explained by the lack of any selection for their maintainence for the same reason as above.

The amplified DNA in 3B cells was found to be located on a rearranged marker chromosome M2. The proximal part of this chromosome bore a close resemblance to rat chromosome 11. The karyotypic analysis described in Part 5.5 confirmed this, as one homologue of chromosome 11 was consistently found to be missing in all 3B cells, unlike in Rat-2 cells where two were always seen. However, the part of chromosome M2 in which the amplified DNA resided was not derived from chromosome 11. This portion of the chromosome could not be identified definitively. It bore a slight resemblance to the q2 region of chromosome 2, but the relative intensities of the bands were quite different in these two chromosomal regions, when examined in the same metaphase spread. Apart from the disappearance of one normal copy of chromosome 2 and of chromosome 11, and the appearance of the marker chromosome M2, the comparison between 3B and Rat-2 karyotypes revealed no other consistent gross chromosomal differences that could be linked with the amplification event in 3B cells.

Implications for the possible mechanism of DNA amplification that took place in the 3B cell line.

The order of events which were proposed to have led to the genesis of the transformed 3B rat cell line (Passananti et al, 1987) were the integration of transfected mouse and polyoma virus sequences into a rat chromosome (chromosome 2 from this analysis), followed by amplification of the integrated DNA and flanking rat sequences. Disruption of chromosome 2 could either have been been caused by the integration of the transfected DNA, or the amplification event. The absence of the intact chromosome prevents the question of whether the (presumed) original location of the integrated DNA was preserved or deleted following the amplification event from being answered. As

mentioned in the introduction (5.1), this would distinguish between conservative models of replication, such as rereplication (Schimke et al,1986) or strand switching at nicked palindromes (Hyrien et al, 1988) which predict preservation of the original locus; or non-conservative models, including that proposed by Passananti et al (1987) based on the study of 3B cells, as well as others proposed by Carroll et al (1988) and Wahl (1989) as described in Chapter 1, which involve recombination within a replication bubble resulting in the excision (ie deletion) of the chromosomal region to be amplified, in the form of replicating circular molecules.

The amplified DNA is located on a rearranged chromosome, which consists of a translocation between chromosome 11 and another fragment of unknown chromosomal origins. As this unidentified fragment actually contains the amplified DNA, it cannot be determined at this stage whether the amplification proceeded via an extrachromosomal intermediate, as predicted by the extrachromosomal double rolling circle and deletion / episome models (see Chapter 1); or whether it occurred *in loco*, as predicted by the chromosomal spiral model (see Chapter 1). Chromosome microdissection of the region around the amplified DNA would enable identification of the origins of these sequences. If the sequences immediately flanking the amplified array on chromosome M2 are not derived from chromosome 2, this would strongly suggest that the amplification involved transposition of the amplified DNA, via an extrachromosomal intermediate. If, on the other hand, the amplified DNA is surrounded by chromosome 2 sequences, one possible explanation would be that amplification occurred intrachromosomally on chromosome 2, and that this somehow disrupted the chromosome leading to translocation of the fragment carrying the amplified DNA. However, there is also the possibility that extrachromosomally amplified DNA could become reassociated with chromosome 2 sequences. The disruption of chromosome 2 (for example by excision of an episome, see Chapter 1) would generate chromosome fragments with broken ends, and such broken ends may provide preffered sites for recombination with the extrachromosomal DNA, as is the case in yeast (Haber and Thorburn, 1984). Stabilisation of such acentric and / or telomere-lacking fragments

could then occur by fusion to other chromosomes, ie by translocation. The relocation of an amplified gene on to a translocated chromosome, at a site close to or at the chromosomal region from which it arose, has been proposed by Biedler et al (1988). In the case of the amplified DNA in 3B cells, a translocation to chromosome 11 has occurred, whether the amplification proceeded intrachromosomally or extrachromosomally. However, as the amplified DNA does not lie at (or close to) the translocation junction between the unidentified chromosome fragment and chromosome 11, this implies that more complex rearrangements must have preceeded its location on chromosome M2, possibly via breakage-bridge-fusion as described in Chapter 1; or, that the extrachromosomally amplified DNA integrated into a pre-existing M2 chromosome.

The recombinogenic nature of rat chromosome 2.

The q23 region of rat chromosome 2, which was involved not only in the integration of transfected DNA and in its subsequent amplification, was found to exhibit a number of other features (Part 5.6) suggesting that it is a region particularly prone to recombination.

The region adjacent to q23 was found to have been involved in a translocation which led to the formation of the M1 chromosome t(2;3), during the generation of the Rat-2 cell line. Duplication of a normal chromosome 2 by non-disjunction must also have occurred, as Rat-2 cells contain two copies of chromosome 2 in addition to chromosome t(2;3).

The chromosome 2q23 region was also found to be at or close to the location of a fragile site (Part 5.6). This fragile site was only observed on chromosome t(2;3), but not at its expected position on normal chromosome 2 homologues. A possible explanation for this is that the altered chromatin structure surrounding a translocation junction might enhance the fragility of a site, allowing it to be spontaneously expressed at a higher frequency than it would be at its normal chromosomal location. It has been suggested that fragile sites tend to be found in regions prone to recombination events such as DNA integration, translocation and mitotic cross-over (Laird et al, 1987).

Consistent with this hypothesis was the finding (Part 5.6) that the q23 region of rat chromosome 2 was frequently involved in spontaneous, large duplication events possibly by unequal sister chromatid exchange (USCE, or mitotic cross-over). These duplications were observed to occur on all the chromosomes [ie both chromosomes 2 and t(2;3)] on which the (unamplified) q23 region was detected by the λBam probe. The distance between the double hybridisation signals suggested that the size of the duplication was many megabase pairs in length.

It was not determined, during the course of these studies, whether the q23 region of chromosome 2 contains constitutive heterochromatin (C-heterochromatin) or satellite DNA. This type of DNA is often found to be associated with USCE events, as well as fragile sites and chromosomal rearrangements (see Riva et al 1989, for references). Presumabely this is because such DNA is made up of clustered repetitive DNA sequences, which are very amenable to recombination events (Atkin et al, 1981). It would be of interest to determine whether the 2q23 region does in fact contain C-heterochromatin (by C-banding), as it exhibits all of the features described for this type of DNA.

Expression of the fragile site in the 2q23 chromosome region was only seen in the translocated t(2;3) chromosome, and this suggests that chromosomal environment may be an important factor in fragile site expression, as has previously been mentioned. Likewise, the presence of integrated foreign DNA (mouse and polyoma sequences) on a normal chromosome 2 homologue, might have had a similar effect on the chromatin architecture of this region. It may have enhanced the recombinogenicity of the region, leading to the illegitimate recombination events proposed to occur as a prelude to the amplification process (Passananti et al 1987, see Chapter 1). On the other hand, *in loco* amplification (Hyrien et al 1988, see Chapter 1) occurring in this region might also lead to structural alterations in the chromatin, thus enhancing its fragility and its tendency to break. This might then result in translocation to another chromosome.

It is interesting to note that rat chromosome 2 has been reported to be specifically involved in the development of certain tumours in the rat. Both Kurita et al (1969) and

Levan et al (1974) reported that in tumours (leukaemias and sarcomas respectively) induced in rats by treatment with 7,12-dimethylbenz-(a)-anthracene (DMBA), a high frequency of chromosome 2 trisomy and other rearrangements which resulted in the presence of additional chromosome 2 material were found. Levan et al (1974) suggested that it is the part of chromosome 2 between regions q2 andq3 (see Fig. 5.4.4) that is preferentially duplicated in the course of tumour development. This region is also found to be effectively trisomic in Rat-2 cells. Furthermore, sequences from within this region (or adjacent to it) are found amplified in 3B cells.

The agent DMBA is known to cause chromosome breaks in rat bone marrow cells when injected intravenously into rats. Sugiyama (1971) has reported that following such treatment, highly selective breakage of chromosome 2 at two points, 2q21 and 2q26, was found. The former of these two points is very close to the region involved in the recombination events observed in Rat-2 and 3B cells, especially to the translocation junction in chromosome t(2;3).

Thus it appears that the region of rat chromosome 2 including 2q23 is highly prone to breakage and rearrangement, and that this may contribute in some way to loss of growth control both *in vivo* in rat tumours, and *in vitro* in cell lines such as Rat-2 and 3B.

CHAPTER 6

CONCLUSIONS AND FUTURE PROSPECTS.

The 3B rat cell line has been shown to contain an amplified polyoma virus oncogene, whose overexpression is responsible for the transformed phenotype of the cells (Ford et al 1985, Passananti et al 1987). The finding that the amplified DNA in 3B cells consists of large inverted repeats, represented the first of many examples of this feature of mammalian amplified DNA, and was the basis for a new model of gene amplification. A number of different models for mechanisms of gene amplification in mammalian cells have been presented in Chapter 1. The stability and the uniform nature of the amplified DNA in the 3B rat cell line, have made it an attractive system for the investigation of the predictions of some of these models.

Summary of results.

Long range DNA analysis has revealed that the inverted duplications are homogeneous in structure as far as can be mapped, with no evidence for rearrangements found in the rat sequences within the amplified units. The size of the amplified unit (one side of the inverted duplication) is estimated to be 450 kb. The region of the rat genome from which the amplified units are derived seems to be quite highly methylated, and no clusters of unmethylated rare cutter enzyme sites (HTF islands) were detected. This may imply that the region is transcriptionally inert, and its amplification would therefore not lead to the overexpression of any deleterious genes present within it (see Section 1.13 of the General Introduction). This is consistent with the absence of any rearrangements or deletions in the rat sequences within the amplified units. There is also some evidence to suggest that the amplified DNA is in fact more highly methylated than its unamplified counterpart in 3B cells (see Chapter 3).

The amplified units are clustered in a single array which is approximately 10-11 Mb in size, as shown by both the molecular and cytogenetic studies. Furthermore, the array

appears to be stable in molecular structure and copy number over several passages in culture. This is confirmed cytogenetically by both the absence of any unstable extrachromosomal and acentric structures such as DMs, as well as the constant chromosomal location of the amplified cluster in all 3B cells examined.

The length of the array suggests that the number copies of the amplified units (assuming each unit is 450 kb in size) is approximately 20 to 25. The lack of success in cloning any sequences from the amplified DNA in YACs, might be partly explained by the fact that the copy number of the units was lower than expected. If the nature of the rat sequences or the presence of the inversion joint decreased the cloning efficiency of the region in a yeast host to any extent, then the region would be underrepresented in a YAC library. Thus, the YAC libraries screened may have been inadequate in size.

Finally, cytogenetic and karyotypic investigations of the 3B cell line and the Rat-2 cell line, from which 3B is derived, have demonstrated that the amplified sequences were derived from rat chromosome 2. The amplified DNA is located on a new rearranged chromosome, and one homologue of chromosome 2 is missing in 3B cells, suggesting that the amplification process may have led to chromosome disruption. The q23 region of chromosome 2, from which the amplified DNA is derived, is highly recombinogenic. It is adjacent to the site of a translocation in Rat-2 cells, it is also the location of a fragile site, and furthermore it appears to undergo frequent, spontaneous unequal sister chromatid exchange, resulting in very large duplications. It is not surprising therefore, that this region was the site of integration of the transfected DNA in 3B cells, and also of its subsequent amplification.

Possible Mechanisms of Amplification in the 3B Cell Line.

The presence of inverted rather than tandem repeats in the amplified array in 3B cells, and the apparent homogeneity of the units (as determined by long range mapping), suggest that disproportionate replication and unequal sister chromatid exchange (USCE) are not likely mechanisms of amplification in this case (see Chapter 1). Furthermore, the presence of all the amplified units at a single chromosomal location in 3B cells excludes excision and episome formation, followed by unequal segregation of

episomes and random chromosomal reintegration, as a feasible mechanism. This model would predict reintegration of episomes in more than one location. Even if some of the episomal amplification units became concatemerised and/or integrated at the same site by homologous recombination, it seems unlikely that they should all have become located at a single chromosomal site.

The models which are most suited to the molecular and cytogenetic observations made in the 3B cell line, are the extrachromosomal double rolling circle model (Passananti et al, 1987) and the chromosomal spiral model (Hyrien et al, 1988). Both of these models postulate that the creation of an inverted duplication is coupled with immediate gene amplification, and predict the formation of an array of inverted repeats, as was the case in 3B cells. However, the double rolling circle model involves excision (and chromosomal deletion) of the target sequences, and amplification via an extrachromosomal intermediate, thus predicting the transposition of the amplified array to a new chromosomal location. The chromosomal spiral model involves *in loco* amplification, predicting that the amplified DNA will remain at the same chromosomal location, unless a translocation of the chromosome fragment carrying the amplified array occurs thus moving it to a different chromosomal position.

The amplified DNA in 3B cells is located on a new and rearranged chromosome (M2). As the chromosome fragment in which it lies cannot be identified by G-banding, it is not possible to determine whether the relocation of the amplified DNA occurred via transposition or translocation. Furthermore, the loss of the intact chromosome 2, from which the amplified DNA is presumed to have arisen, prevents the investigation of whether the amplification event resulted in deletion of the chromosomal locus, as predicted by the double rolling circle model.

Future Prospects.

In order to verify that the amplification units in 3B cells are indeed homogeneous, that the other ends of the inverted duplication are precise and that another amplified inversion joint exists between them, it is essential that a complete amplification unit

should be cloned. The heterogeneity in structure of the amplified DNA found in many systems is due to the multiple amplification events and other rearrangements that may have occured, especially for cells under drug selection. In such systems, the cloning of overlapping DNA fragments of amplified DNA can be very misleading as sequences resulting from different amplification events may be mistakenly linked. The amplified DNA in 3B cells appears, from the results presented in this thesis, to be the result of a single event. This can only be confirmed by the cloning and analysis of a complete amplification unit. It has recently been reported that modifications made in YAC cloning, such as the incorporation of two size-fractionation steps (as discussed in Chapter 4), mean that YAC libraries with average inserts of 4-500 kb in length are obtainable. As the amplification unit in 3B is estimated to be 450 kb in length, it now seems feasible for the first time, that a complete amplification unit could be cloned in a YAC.

The chromosomal origins of the sequences which flank the amplified array on the M2 chromosome in 3B cells would provide important insights in to the mechanism of amplification. Chromosome microdissection (Scalenghe et al 1981, Pirotta et al 1983) of the M2 chromosome around the readily identifiable G-band containing the amplified DNA (see Figs. 5.4.5 and 5.4.6), would provide probes which could then be used for *in situ* fluorescence hybridisation on to rat cells (see Chapter 5), in order to identify the rat chromosome from which they originated. As discussed above, relocation of the amplified DNA could either have occurred by transposition or translocation. If the flanking sequences were found to be derived from a rat chromosome other than chromosome 2, this would strongly suggest that the amplification occurred by transposition, via an extrachromosomal intermediate as in the double rolling circle model. If, on the other hand, the sequences flanking the amplified array were found to be derived from chromosome 2, this could either mean that translocation of the *in loco* amplified DNA occurred, as in the chromosomal spiral model; or that extrachromosomally amplified DNA (double rolling circle model) recombined with a fragment of the disrupted chromosome 2 (as discussed in Chapter 5).

It might also be interesting to try and isolate 3B cells which have reverted to an

untransformed phenotype and regained cellular growth control, in order to examine the possible mechanisms of reversion. Such experiments are currently underway (M. Read and M. Fried). A number of possible mechanisms for reversion can be envisioned. One is the loss of the complete marker chromosome M2 on which the amplified DNA is located. However, complete loss may be lethal to the cell, as the M2 chromosome contains a complete copy of chromosome 11, of which there is only one other homologue in 3B cells. A more likely possibility might be that chromosome breakage, or rearrangement within chromosome M2 might result in the deletion of the amplified sequences containing the transforming polyoma virus oncogene. Another possibility might be a gene conversion event, enabling a mutation in one copy of the transforming oncogene to be transferred to all the other copies. This has been shown to occur in only one example (Roberts and Axel, 1982), where a cell line containing an amplified *aprt* gene, was found to revert to an APRT$^-$ phenotype by mutation and gene conversion of the amplified gene, without any loss of amplified DNA. However, as discussed in Chapter 1, in this case disproportionate replication was shown to be the mechanism of amplification, with the repeated use of an origin of replication derived from plasmid sequences. It can easily be imagined that gene conversion between the multiple DNA strands of an onionskin structure (see Fig. 1.1) might occur. However, in the case of 3B cells, the disproportionate replication model is unlikely to apply (for the reasons discussed above, and in Chapter 1), so that gene conversion may be an unlikely mechanism of reversion in this case.

Future investigations such as these, will provide important insights in to the mechanisms of amplification of oncogenes, the role of such amplification in mammalian cell transformation, and the genetic stability of chromosomally located amplified DNA *in vitro*.

CHAPTER 7

MATERIALS AND METHODS.

MATERIALS.

Agarose	FMC Bioproducts
Agarose LMP	Beckman (PFGE grade), BRL (Ultrapure)
Ampicillin	Penbritin; Beecham Research Laboratories
Amino-acids	Sigma
Bacto agar	Difco
Biotinylated anti-avidin (goat)	Vector Labs (BA 0300)
Boric acid	FSA
BSA	Sigma
BSA enzyme grade	BRL
CaCl2	BDH
CH3COOH	FSA
Citifluor	Citifluor Ltd
CsCl	Koch-Light
CHCl3	FSA
Chloramphenicol	Sigma
Dialysis Tubing	Spectrapor
DNA 123bp ladder	BRL
DTT	Sigma
EDTA	FSA
ETOH	James Burough
Genescreen plus	NEN Research Products
Glycerol	Fisons
Ficoll	Pharmacia
Films	Polaroid type 667 for photography of EtBr-stained agarose gels; Kodak XAR-5 and Fuji RX for autoradiography; Fujicolour HR400 for colour photography of fluorescence hybridized chromosome spreads.
Formamide	IBI (for fluorescence in situ hybridisation, no. 72024)
Genescreen plus	NEN Research Procucts
Glucose	BDH
HEPES	Sigma
Hexadeoxynucleotide primers pd(N)6	Pharmacia
H3PO4	BDH
Hybond N, Hybond N+	Amersham
Iso-amyl alcohol	Fisons

KCl	BDH
KH2PO4	BDH
KOAc	BDH
KOH	FSA
Lysozyme	Sigma
Maltose	BDH
Marvel non-fat dry milk	Marvel
MeOH	BDH
MgSO4.7H2O	BDH
MnCl2	BDH
Na2HPO4	BDH
NaCl	BDH
NaOH	FSA
Nitrocellulose	Schleicher & Schuell
Novozym 234	Novo Biolabs
Nucleotides, deoxynucleotides, and dideoxynucleotides	Pharmacia
PEG	Fisons, Sigma (yeast transformations)
PMSF	Sigma
Radiochemicals	Isotopically labelled chemicals were obtained from Amersham International or NEN. Their specific activities are referred to in the text.
Salmon Sperm DNA	Sigma
Sarkosyl	BDH
SDS	BDH
Sephadex G75	Pharmacia
Sephadex G50	Pharmacia
Sorbitol	Sigma
Sodium citrate	Sigma
Spermidine	Sigma
Sucrose	Fisons
Tetracycline	Sigma
Tris.HCl	Sigma
Triton X-100	Sigma
Tryptone	Difco
Tween 20	Pierce (no. 28320)
Urea	Fisons
Yeast extract	Difco
Yeast nitrogem base (w/o NH_2SO_4)	Difco
X-Ray intensifying screens	Du Pont Lightning Plus

EQUIPMENT

Laser Densitomerter	LKB Ultroscan XL
OFAGE and CHEF pulsed field gel electrophoresis apparatus	LKB
Photomicroscope	Zeiss (photomicroscope 3)

All other reagents and equipment were obtained from Sigma or ICRF Central Services.

ENZYMES

Calf intestinal alkaline phosphatase	Boehringer
DNA restriction enzymes	
(frequent cutters)	New England Biolabs, BRL, Boehringer and NBL
(rare cutters)	New England Biolabs
DNase I	Worthington Biochemical Corporation
DNA polymerase I	Boehringer
E.coli DNA ligase	New England Biolabs
Klenow DNA polymerase	BRL
Lysozyme	Sigma
Lyticase	Sigma
Novozym 234	NovoBiolabs
Proteinase K	Sigma
RNase A	Sigma
T4 DNA ligase	New England Biolabs
T4 DNA polymerase	P-L
T4 polynucleotide kinase	Pharmacia.

<u>Kits</u>

Agarose PFGE size marker plugs:	InCert™ FMC
Concatemerized Lambda DNA ladder	
S. cerevisiae chromosomes (YPH 80)	
S. pombe chromosomes (972)	
Bionick Kit	BRL
Gigapack gold	Strategene
Lambda Fix	Stratagene
Lambda Dash	Strategene
GeneAmp	Cetus

Vectors and Host Bacteria

Lambda Dash, BamHI ends (Stratagene)
Lambda Fix, XhoI ends (Stratagene)
pBR322 (Bolivar et al,1977)
Pxf3 (Hanahan 1983)
E.coli strain DH5
E.coli strain DH1

Cell Lines

Rat-1	Immortalised rat fibroblastoid cell line [aka F4208, Botchan et al, (1976)]
Rat-2	Immortalised, TK^- rat fibroblastoid cell\ line, derived from Rat-1 cells (Topp, 1981)
3B	Rat-2 cells transformed by transfection and amplification of the polyoma virus middle-T antigen (Ford et al, 1985)

Formulation of Frequently Used Solutions and Media

GS1	2M NaCl 0.5M NaOH
GS2	4M MaCl 0.5M TrisHCl adjusted to pH 7.5 with NaOH
Tris	24mM Tris-base
saline	0.8% w/v NaCl 0.38% w/v KCl 0.01% w/v Na2HPO4 0.1% w/v dextrose 0.001% w./ neutral red pH adjusted to 7.4 with HCl
PBS.A	1% w/v NaCl 0.025% w/v KCL 0.14% w/v Na2HPO4 0.025% w/v KH2PO4
PBS	10mM$CaCl_2$ 6mM $MgCl_2$ in PBS.A
Versene	0.54mM EDTA in PBS.A

TE	1mM EDTA 10mM Tris.Cl pH7.4
SM	10mM Tris.Cl pH7.4 10mM MgCl2 100mM NaCl 0.05% w/v gelatin
1xSSC	150mM NaCl 15mM sodium citrate pH to 7.5 with NaOH
1x TAE	40mM Tris Acetate 2mM EDTA
1x TBE	89mM Tris-base 89mM Boric acid 2mM EDTA
L Broth	1% w/v Bacto-tryptone 0.5% w/v Bacto-yeast extract 1% w/v NaCl Supplemented with 1.5%w/v Bacto-agar for solid media or, 0.6% agarose for an overlay for lambda

Yeast Media.

YPD	1% yeast extract 2% bactopeptone 2% dextrose Add 2% Bacto agar for plates
Synthetic media (SD)	2% dextrose 0.67% yeast nitrogen base w/o amino acids Add 2% Bacto agar for plates
Complete media (COM)	2% dextrose 0.67% yeast nitrogen base w/o amino acids Required nutrients at 20µg/ml Add 2% Bacto agar for plates

Regeneration plates	1M Sorbitol
	2% dextrose
	0.67% yeast nitrogen base w/o amino acids
	Required nutrients at 20μg/ml
	1.5% Bacto agar
TOP	1M Sorbitol
	2% dextrose
	0.67% yeast nitrogen base w/o amino acids
	Required nutrients at 20μg/ml
	2.5% Bacto agar

For preparation of sterile Regeneration and TOP media, the solution of dextrose and sorbitol was autoclaved seperately from the yeast nitrogen base and agar, the two being combined only prior to pouring the plates. Furthermore, sterile amino acids were added after the media had been autoclaved.

Nutrients, amino-acids: (Prepared as the following stock solutions and sterile filtered).
adenine hydrochloride 7.5 mg/ml
L-arginine hydrochloride 7.5 mg/ml
L-histidine monoHCl 7.5 mg/ml
L-isoleucine 7.5 mg/ml
L-leucine 7.5 mg/ml
L-lysine monohydrochloride 7.5 mg/ml
L-methionine 7.5 mg/ml
L-phenylalanine 7.5 mg/ml
L-tryptophan 2 mg/ml
L-tyrosine 2mg/ml (needs NaOH in order to dissolve)
uracil 2 mg/ml
L-valine 7.5 mg/ml

METHODS.

General Methods:

DNA Quantitation and Manipulation.

DNA was quantitated by determining the absorbance of dilutions of stock solutions at a wave length of 260 nm (1 OD_{260}=50 µg/ml DNA). Following enzyme treatment, plasmid and cellular DNA was normally extracted with phenol, followed by a 1:1 phenol:$CHCl_3$ mix and then $CHCl_3$ alone, and ethanol precipitated before further manipulation. In the absence of any salt, NaOAc pH5.4 was added to 0.3 M to enable DNA precipitation. For small quantities of DNA 1-2 µg of yeast tRNA was added to assist ethanol precipitation.

Restriction enzyme digestions were conducted as described by Maniatis et al (1982), although 5 units/µg of DNA were used. For cellular DNA the buffer solution and enzymes were mixed in gently to avoid shearing. Incubations for plasmid DNA were 2 hours at 37°C, and for cellular DNA overnight at 37°C.

Intermolecular ligations were performed in minimum volumes possible, generally 10 µl for a total DNA content of 100 ng. For ligations with plasmids, a 1:10 ratio of plasmid:insert DNA was used. If the plasmid was capable of self-ligation, the compatible ends were dephosphorylated (see below). Ligations involving λ vectors were set up at three different ratios of arms:insert (1:2, 1:1, 2:1 for the ratio of compatible arm ends : compatible insert ends) so that following packaging and titration the most suitable ratio was selected for construction of a library. Ligations were performed overnight at 14°C and proteins were not extracted before transfection into competent bacteria.

Dephosphorylation of DNA was performed using calf intestinal phosphatase as described in Maniatis et al (1982). Following incubation at 37°C the reaction was stopped by addition of STE (1/10 of reaction volume) and 0.5% SDS followed by heating at 65°C for 15 min, and then extracted twice with phenol and once with $CHCl_3$.

Ligation buffer: 50 mM Tris-Cl pH7.8

100 mM $MgCl_2$

20 mM DTT

50 µg/ml BSA

1 mM ATP

10xSTE: 100 mM Tris-Cl (pH 8.0)

1 M NaCl

10 mM EDTA

Agarose Gel Electrophoresis of DNA.

Large submarine agarose gels (150-200 ml) and minigels (50-70 ml) were made and run in 1xTBE. Agarose concentrations varied between 0.6-1.2% w/v depending on the DNA fragment size resolution required. Visualisation of the DNA was achieved by staining in ethidium bromide (at a concentration of 0.5 µg/ml both in the gel and in the running buffer) under long wave (360 nm) ultraviolet light. DNA samples were loaded after addition of 0.1 x volume of loading buffer. To resolve fragments of cloned plasmids or lambda DNA, 1 µg of DNA was digested with an appropriate restriction enzyme in 10-20 µl of reaction mixture. Loading buffer was added directly to the digested DNA and the gel run at 60 mA (both for large gels and minigels).

For resolution of genomic DNA from 0.5 to 30 kb in size, 5 µg of genomic DNA was digested in 100 µl of restriction enzyme buffer with 20 units of enzyme for 5 to 8 hours. After addition of loading buffer 60 µl of this was loaded into a large gel of 0.8% agarose, and run at 30 mA overnight.

10 x loading buffer 25% w/v Ficoll

50 mM EDTA pH8

0.1%w/v bromophenol blue

Isolation of Specific DNA Fragments from Gels.

DNA isolated in this way was used both in ligation reactions and for making radioactively labelled probes (see later). The excised gel slice containing the required band was inserted into a 0.5 cm diameter dialysis bag with 1 x TBE plus 100 µg/ml BSA. This was then submerged in a gel tank of 1 x TBE and the DNA electroeluted at 100 V for 30 mins. The field was reversed for 1 min prior to removal of the eluted DNA. The DNA was extracted sequentially with phenol, phenol/$CHCl_3$ and $CHCl_3$, ethanol precipitated, washed twice in 70% ethanol, dried and resuspended in TE.

Southern Blotting.

(Church and Gilbert 1984)

For blots of fragments upto 25-30 kb in size, 3-5 µg of genomic DNA was digested with the appropriate restriction enzymes and loaded onto a 0.8% agarose gel, run in 1 x TBE. Both gel and buffer contained EtBr so that the DNA could be visualised and photographed with UV light after running the gel. For larger fragments pulsed field gel electrophoresis was conducted as described below.

Depurination of the DNA in the gel was achieved by soaking in 0.25 M HCl for 20 mins on a shaking platform. The DNA was then denatured by soaking the gel in 0.5 M NaOH/1.5 M NaCl for 2x30 mins, again with shaking. Genescreen nylon membrane was floated on H_2O to moisten and then soaked in the denaturation solution above for 15 mins. DNA was transferred from the gel to the membrane by blotting using a reservoir of denaturation solution and a stack of paper towels.

Transfer was allowed to continue overnight, but in the case of PFG blots over 2 days, and then the filter was neutralized by submerging in 50 mM NaPi, pH 6.5 for 5 mins. After drying on the bench for a few hours the membrane was then exposed to 254 nm UV radiation for 2 mins from a distance of 50 cm (approximately 0.16 KJ/m^2).

Rapid Two-sided Southern Blotting

For the analysis of small plasmid or lambda fragments resolved on agarose minigels (where large quantities of unique fragments are present), following photography the DNA was denatured by soaking the gel in 0.5 M NaOH, 1.5 M NaCl twice for 20 mins each time. Two genescreen membranes were pre-wetted (as in Southern blotting method), and one membrane was placed on each side of the gel, which was then sandwiched between two stacks of paper towels, enabling transfer of the DNA in both directions, the transfer buffer being provided by the liquid trapped in the gel. Transfer was for 3-4 hours, following which the membranes were treated in the same way as in the previous Southern blotting protocol.

Hybridisation of Southern Blots.

Hybridisation of the filter was conducted in a sealed plastic bag with 20-30 ml of hybridisation solution and prehybridised for 20 mins at 65-68°C. The prehybridisation solution was removed and replaced with 10-15 ml of new hybridisation solution containing denatured probe. The probe was radioactively labelled by the random primer method (see below) and was used at a concentration of <10 ng/ml or $<1.5x10^6$ cerenkov counts/ml. It was denatured by addition of NaOH to 0.4 M for 5 mins.

Hybridisation was allowed to continue for 16-20 hours, and the filter was then washed in a solution of 40 mM NaPi buffer, 1% SDS. Washing was carried out for 3x5 mins at room temperature, and then for 30 mins at 68°C. The filter was then wrapped in Saran Wrap, and exposed at -70°C with intensifying screens for 1-7 days.

Filters were stripped of probe by washing for 30 mins at 80°C in 2 mM EDTA, 0.1% SDS, after which prehybridisation was not necessary.

1 M NaPi buffer: 142 g/l $Na_2HPO_4.2H_2O$, adjusted to pH 7.2 with conc.H_3PO_4

Hybridisation soln.: 0.5 M NaPi pH7.2

7% SDS

1 mM EDTA

Slot Blotting.

A Schleicher and Schuell slot blot apparatus was used. A piece of nylon membrane was soaked in 0.4 M NaOH for a few minutes. Two sheets of 3MM paper, previously wetted in 0.4 M NaOH, were placed on the top of the vacuum unit of the apparatus and the wet membrane was layed over this. The two parts of the manifold were clamped together. 200 µl samples of DNA, previously resuspended in NaOH to a final concentration of 0.4 M, were loaded into the slots. After they had adsorbed on to the filter (without vacuum) a further 200,µl of 0.4 M NaOH was added. Once this had also adsorbed, the vacuum was applied for approximately 10 mins. The apparatus was disassembled and the filter left to dry. The DNA was cross-linked on to it by baking or UV irradiation. Hybridisation was carried out as above.

Oligo-labelling of DNA fragments by Hexadeoxynucleotide Primers.

Up to 50 ng of DNA in a final volume of 13.5 µl of water was boiled for 5 mins, and then put on ice briefly. The labelling reaction was carried out at room temperature by addition of the following reagents in the order stated:

5 µl OLB buffer

1 µl of 10 mg/ml BSA (enzyme grade, BRL 5561)

13.5 µl DNA solution containing 25-50 ng denatured DNA.

2.5 µl ^{32}P a-dATP (10mci/ml, 3000Ci/mmol)

2.5 µl ^{32}P a-dCTP (10mci/ml, 3000Ci/mmol)

0.5 µl (2u) Klenow (large fragment) Polymerase I.

OLB buffer consists of a mixture of solutions A:B:C in the ratio 10:25:15

Solution A: 1 ml of 1.25 M Tris-HCl, 0.125 M $MgCl_2$

18 µl 2-mercaptoethanol

5 l of 0.1 M dGTP and 0.1 M dTTP

Solution B: 2 M HEPES (pH to 6.6 with 4 M NaOH)

Solution C: Hexadeoxynucleotides, suspended in TE at 90 OD units/ml

The reaction was incubated at room temp. for at least 2 hours. The labelled DNA was separated from unincorporated nucleotides by passing the reaction solution over a G75 sephadex column and collecting the peak of radioactivity as it came off the column. This method yields 25 ng DNA labelled to 10^8 dpm/mg DNA, or a total of 2.5×10^6 dpm.

DNA End-labelling

For the labelling of recessed 3' termini, purified DNA fragments (1 μg) were incubated in 30 μl of restriction enzyme buffer containing 20 μCi of the appropriate α -labelled dNTP (3000 Ci/mMol) and 1 unit of Klenow polymerase for 15 mins at room temperature. The volume was then increased to 250 μl, containing 0.3 M NaOAc pH 5.4, and the mixture extracted with phenol, then chloroform, and finally ethanol precipitated.

Oligonucleotides

The oligonucleotides used for polymerase chain reaction (PCR) were obtained from Dr. Iain Goldsmith of the ICRF. These oligonucleotides have been fully deprotected, have -OH groups at both the 3' and 5' ends and the stepwise efficiency was 98.6%. After dissolution in TE, the concentration of the oligonucleotide was calculated from the UV absorbance at 260 nm, assuming that a solution containing 33 mg/ml has an absorbance of 1.0.

Polymerase Chain Reaction

The polymerase chain reaction (PCR) was used to amplify fragments from the published sequence of the polyoma virus (Tooze, 1980). GeneAmp DNA Amplification Reagent Kit purchased from Perkin Elmer Cetus was used for most of the reagents. The reaction mixture contained <1 μg yeast miniprep DNA or rat (3B) cellular DNA, 200 mM each dNTP, 1.0 mM each primer, 2.5 units Taq polymerase,

50 mM KCl, 10 mM Tris-Cl (pH 8.3), 1.5 mM $MgCl_2$ 0.1% (w/v) gelatin, in 100 µl. This was placed in a small microfuge tube and covered with 100 µl of mineral oil to prevent evaporation. The PCR reaction was carried out overnight on a programmable Dri-Block purchased from Techne. 30 cycles of 2 mins at 94°C, 2 mins at 59°C, and 9.9 mins at 70°C were followed by an additional 9.9 mins at 70°C. The PCR reaction products can be stored at 4°C for several weeks.
The primers PY2 and PY3 used for the PCR screening of the YAC library in Chapter 4 (Part 4.2) had the following sequence:
PY2: TTCTTTCCCTTTGCAAAAATTCACACT
PY3: TCTGAGCAGAGCTGACAAAGAAAGGCT

Plasmid Vectors.

Plasmid vectors used were Pxf3 (Hanahan, 1983), pBR322 (Bolivar et al, 1977), pYAC4 (Burke et al, 1987), pYAC-RC: (Marchuk and Collins, 1988) and YCp50 (Hieter et al, 1985). Plasmids were propagated in *E. coli* strains DH5 (Pxf3, pBR322, pYAC-RC) or DH1 (pYAC4), which were grown in L broth supplemented with 50 µg/ml ampicillin. In the absence of plasmid these bacteria were grown in L-broth alone. For long term storage of bacteria, glycerol was added to a fresh overnight culture to 20% v/v and an aliquot stored at -20°C.

Plasmid Transformation of Competent Bacteria.

Frozen competent bacteria (prepared as in Hanahan, 1985) were thawed and up to 100 ng DNA was added. The bacteria were kept on ice for 30-45 mins then heat shocked at 42°C for 1-2 mins and cooled on ice. 1 ml of LB medium (containing no antibiotics) was added and the mixture incubated at 37°C for 45 mins. They were then plated out onto LB plates with suitable anntibiotic selection. The efficiency was usually > 10^6 cfu/µg DNA.

Small Scale Preparation of Plasmid DNA

From an overnight culture 1.5 ml were transferred to an Eppendorf tube and spun for 3 mins. The supernatant was aspirated off and the bacterial pellet resuspended in 100 µl of cold solution I containing 4 mg/ml lysosyme (added immediately prior to use). This was kept on ice for 5 mins, and 200 µl of solution II was then added and the tube inverted slowly until the contents turned clear. 150 µl of ice cold solution III was then added, mixing carefully again. The tube was then spun for 10 mins at 4°C. 400 µl of the supernatant was then transferred into a new Eppendorf tube, and extracted with an equal volume of a 1:1 mix of phenol: chloroform. After centrifugation for 10 mins at 4°C, 350 µl of the upper layer was transferred into a new tube and precipitated with 1 ml 95% ethanol. After centrifugation the pellet was washed in 70% ethanol, pelleted, dried and resuspended in TE with 0.1 mg/ml RNase A. The yield was usually about 30 µg of DNA.

Solution I	50 mM glucose
	10 mM EDTA
	25 mM Tris-HCl pH 8.0
Solution II	0.2 M NaOH
	1% SDS
Solution III	4 M Potassium acetate pH 4.8

Large Scale Plasmid Preparation.

A 20 ml overnight bacterial culture was used to innoculate 400 mls of LB media containing appropriate antibiotic in a 2 L flask and this was incubated, with shaking at 37°C overnight. The bacteria were pelleted by centrifugation at 10,000 g for 7 mins, washed in bugwash and repelleted. The cells were then resuspended in 10 mls of STET containing 0.4 mg/ml lysozyme added fresh, and immediately placed in a boiling water bath for 2 mins. The lysed material was then transferred to SW40 rotor tubes and centrifuged at 30 K for 30 mins at 4°C. The supernatant from this centrifugation was taken and 0.9 g CsCl added per ml and 0.4 ml of 10 mg/ml EtBr

added per 10 mls. This solution was transferred to 2 vti65 tubes, the tubes were sealed and centrifuged in a vti65 rotor at 64,000 rpm for 5 hours at 18°C.

After centrifugation the bands of plasmid DNA could easily be seen due to the ethidium bromide. The lower, supercoiled band was recovered using a 23 gauge needle and a 1 ml syringe. The EtBr was removed by repeated extraction with CsCl saturated isopropanol until no red colour could be seen. The salt concentration was reduced by the addition of twice the volume of water and the DNA precipitated by the addition of 2.5 times the volume of 95% ethanol. The DNA was pelleted by spinning at 10,000 g in a Sorvall rotor, washed with 70% ethanol, dried and then dissolved in TE buffer. A 400 ml culture yields about 2-3 mg of DNA.

Bugwash 0.15 M NaCl
10 mM Tris-Cl pH 8.0
1 mM EDTA

STET 50 mM Tris-Cl pH8.0
50 mM EDTA pH8.0
8% w/v sucrose
5% v/v Triton x-100

Mammalian Cell Culture.

All mammalian cell lines were grown in a 10% CO_2:90% air, humidified incubator, using Dulbecco-modified Eagles medium (DMEM) supplemented with 0.1% D-glutamine and 5% v/v FCS. Cells were grown on plastic tissue culture dishes (9 cm circular) or 24.5 x 24.5 cm bioassay trays (Nunc). Cells were removed from the dishes, following an initial versene wash, with a 3:1 mixture of versene : 0.1% w/v trypsin in Tris saline, and passaged at appropriate densities.

Preparation of High Molecular Weight DNA.

Eukaryotic cell DNA was prepared by a variation of the procedure of Blattner et al (1978). Cells were cultivated on large bioassay trays until almost confluent. After removal of the medium, cells were washed twice in PBS and drained thoroughly. Each tray of cells was then overlaid with 4 mls of lysis buffer and left at room temperature for 10 mins. The lysed mixture was then carefully scraped into a 50 ml Falcon tube, using a rubber policeman. The lysed cells were then extracted twice with phenol and once with chloroform. The DNA was subsequently dialysed against 2 l 1 M NaCl, TE for 12 hours at 4°C and then 4l TE for 48hrs at 4°C. The nucleic acid was then treated with 20 µg/ml RNase A at 37°C for 3 hrs, protein extracted as above and dialysed against 4L TE for 48 hrs at 4°C before storage at 4°C.

Lysis buffer 10 mM Tris-Cl pH 8
10 mM EDTA pH 8
5 mM NaCl
0.5% w/v SDS
0.1 mg/ml proteinase K

Size Fractionation of Restricted Total Cellular DNA.

In order to enrich total enzyme restricted cellular DNA for a particular fragment, the mixture of fragments was size fractionated on a sucrose gradient. Restriction enzyme digests of total cellular DNA (300 µg) were loaded onto 10-40% sucrose gradients and spun at 120,000 g (24,000 rpm)/10°C/20 hrs/SW27 rotor. Fractions of 500 µl were collected via a 19 gauge needle inserted through the base of the tube. In order to determine the average size of DNA fragments in each fraction, 50 µl aliquots of every third fraction were run on a 0.8% mini agarose gel using λ DNA digested with HindIII as a size marker. Having determined roughly which fraction should contain the fragment of interest, 200 µl aliquots of each fraction in the appropriate range were precipitated, run out on 0.8% agarose gels and blotted onto genescreen plus as described under Southern blotting. The fraction(s) containing the desired sequence

were then identified by hybridisation to radiolabelled DNA probes.

Sucrose solution: 5 mM EDTA
20 mM Tris-Cl pH 8.0
1 M NaCl
10% or 40% Sucrose

Size Fractionation of Partially Restricted Total Cellular DNA.

The conditions for the production of restricted DNA of the appropriate average size (>20 kb for cosmid or lambda libraries) were established using small scale test digestions where the amount of the enzyme was varied, keeping time and temperature constant. A reaction mixture of 10 µg DNA and restriction enzyme buffer to a final volume of 150 µl was mixed well and kept on ice. 30 µl was dispensed into an Eppendorf tube (tube 1) and 15 µl into 7 others (tubes 2-8). 4 units of the enzyme were added to tube 1 (ie 2 units/µg DNA) and mixed, after which 15 µl of this reaction mixture was transferred to tube 2 (ie 1 unit/µg DNA) and mixed. The 2-fold serial dilutions were continued upto tube 8. These tubes (apart from that containing no enzyme and the remainder of the DNA) were placed in a 37°C water bath for 30 mins, after which they were plunged into ice and EDTA was added to a final concentration of 20 mM to each. 5 µl of each reaction was analysed on a 0.4% mini agarose gel as described under "Agarose Gel Electrophoresis", using λ HindIII markers. After running the samples on the gel it is photographed and the degree of digestion that produces the maximum amount of DNA in the desired size range is established. Half of the amount of enzyme that produced the maximum fluorescence under UV was used in order to obtain the maximum number of molecules in this size range.

Having established the optimum conditions the reaction was carried out on 250-500 µg of the same high molecular weight eukaryotic DNA, using identical enzyme concentration, temperature and time as before. An aliquot of this DNA was checked after the digestion for the correct size distribution by running it on a 0.4% agarose

minigel. The digested DNA was carefully phenol extracted and then chloroform extracted, ethanol precipitated, washed and dried. It was finally resuspended in 300 µl TE and size fractionated on a sucrose density gradient as described above. After analysis of the gradient fractions by gel electrophoresis and EtBr fluorescence, those containing DNA in the desired(20 kb) size range were pooled.

Preparation of High Molecular Weight Genomic DNA in Agarose blocks.

Genomic DNA as prepared by the method described above contains fragments of approximately 100 kb in size due to shearing of the DNA and to the presence of trace amounts of DNase. To prepare and manipulate DNA of greater than 100 kb in size, the cells from which the DNA was to be obtained were first immobilized in low melting point agarose blocks and all steps were carried out using sterile instruments and buffers in order to avoid any DNase contamination.

DNA was prepared from tissue culture rat cells, which were grown to near confluency and then trypsinised to give a single cell suspension. This was then washed in PBS and the concentration of the cells measured with a Coulter counter, they were then resuspended in PBS at $1x10^6$ cells per 40 µl PBS at room temperature, $1x10^6$ mammalian cells is equivalent to 10 µg genomic DNA. A 1% solution of low melting agarose in PBS was melted and held at 42°C, an equal volume of this was mixed with an equal volume of the cell suspension at room temperature. This mixture was dispensed immediately into 80µl block formers and allowed to set on ice.

To prepare the DNA in the agarose blocks, the blocks were incubated in 50 mls of 1% Sarkosyl, 0.5 M EDTA pH 8, and 2 mg/ml Proteinase K for 48 hours at 55°C. The blocks were then rinsed three times in TE and placed in TE plus 0.04 mg/ml PMSF (prepared in propan-2-ol) for 30 mins at 55°C, this was repeated once. The blocks were stored in 0.5 M EDTA.

Prior to use, the blocks were rinsed three times for 30 mins in TE. To cut the DNA with restriction enzymes, a single block was placed in an Eppendorf tube with

restriction enzyme buffer recommended by the manufacturer to a final volume of 400 µl and equilibrated at 37°C for 30-40 mins. The buffer was then replaced with fresh buffer containing 20-30 units of enzyme, to a final volume of 200 µl. The tube was incubated at the recommended temperature for the enzyme overnight. The reaction was stopped by addition of 20 µl 0.5 M EDTA and the block was rinsed at least once in TE and kept on ice before loading into the gel.

For size markers between 50 kb and 1000 kb, ligated Lambda concatemers were used, for sizes from 240 kb upto approximately 1600 kb *S. cerevisiae* chromosomes, and for sizes between 3.5 and 5.7 Mb *S. pombe* chromosomes were used. In all cases these were obtained in agarose blocks from FMC.

5-Azacytidine (5-azaC) Treatment of Mammalian cells in Culture.

Rat cells from 2 almost confluent 90 mm dishes were plated into a single 22x22 cm dish (ie 1-2x10^6 cells), and after 24 hours growth 5-azaC was added at a final concentration of 0.5 µM, 2 µM, or 5 µM from a 10 mM stock solution. After a 24 hour exposure, the culture was washed, trypsinised and the cells embedded in agarose blocks as described under "Preparation of HMW Genomic DNA in Agarose Blocks".

Pulsed Field Gel Electrophoresis (PFGE).

To separate DNA fragments greater than 30 kb in size PFGE was carried out using an LKB Pulsaphor unit, either for orthogonal fixed angle gel electrophoresis (OFAGE) or, by using the hexagonal adaptor electrodes, for contour clamped hexagonal field gel electrophoresis (CHEF). A 1% agarose gel in 0.25x or 0.5x TBE was poured using the appropriate gel plate and former supplied by LKB. The agarose blocks were inserted into the gel slots and sealed in place with 0.75% low melting agarose. The gel was immersed in the tank containing circulating buffer of the same TBE concentration, which had been pre-cooling to the desired temperature (between 10-14°C) for 2 hours. OFAGEs were run at 350 V and CHEFs were run at various pulse intervals and temperatures depending on the size resolution required.

The gel was stained in running buffer containing EtBr for 15 mins and then destained for 10 mins in running buffer alone, and the DNA visualised with UV. The gel was then transferred to genescreen plus membrane by Southern blotting over a 36-48 hour period.

Bacteriophage Lambda Vectors.

To make the lambda library of size selected Bam HI fragments, the λ replacement vector EMBL 4 (Frischauf et al 1983) was used. The lambda DNA was restricted with Bam HI and SalI and then precipitated with 0.2M NaCl and an equal volume of isopropanol. In this way the small Bam HI-SalI fragments are not precipitated, so the 14 kb stuffer region cannot ligate to the arms. Having determined that the optimal arms:insert ratio for ligation as described under DNA Quantitation and Manipulation, 1 μg of insert DNA was ligated to 1 μg of lambda DNA and the mixture packaged with Gigapack Gold purchased from Stratagene. *E.coli* strains LE392 (on which both recombinant and non-recombinant EMBL4 gives lysis) and P2392 (a P2 lysogen of LE392 on which only recombinant EMBL4 gives lysis) were used for infection with small aliquots of the library which indicated that the library contained over 90% recombinants. A library of approximately 100,000 plaques was plated out on *E.coli* P2392.

The library of size selected partially cut Sau 3A fragments (average size >20kb) was constructed in lambda DASH cut with Bam HI, purchased from Stratagene. 1 μg of insert DNA was ligated to 2 μg of vector DNA and packaged in Gigapack Gold (Stratagene). Aliquots were plated out on P2393 and LE392 (as in previous EMBL4 Bam HI size selected library) showing that the library contained 100% recombinants out of a total of 120,000. One third of this library (40,000 plaques) was plated out on *E.coli* P2392 for screening.

The library of nonsize-selected partially cut Sau 3A fragments was cloned into lambda FIX cut with Xho I followed by fill in of the first two nucleotides of the Xho I site leaving 3-AG-5' overhangs (supplied by Stratagene). The Sau 3A partially cut

genomic DNA was partially filled in with Klenow Pol 1 giving 3'-CT-5' overhangs. 1 μg of insert DNA was ligated to 1 μg of lambda DNA aand packaged in Gigapack Gold (Stratagene). Aliquots were plated out on P2392 and LE392 showing that the library contained almost 100% recombinants out of a total of 2000,000. A fraction of the library was plated out on P2392, and 170,000 plaques were screened in total.

Bacteriophage Lambda Libraries: Plating out and Screening.

Lambda libraries were packaged using kits purchased from Stratagene using the protocols provided. To make competent bacteria, a single colony from a fresh plate of the appropriate strain (P2392) was innoculated into 25 ml LB containing 10 mM $MgSO_4$ and shaken overnight at 37°C. The culture was spun down and resuspended in 8 ml SM, and kept at 4°C (for up to 2 weeks). The required fraction of packaging mix was added to 1 ml of P2392 bacteria (in SM) in a plastic 15 ml Falcon tube and incubated at 37°C for 20 mins without shaking. The bacteria were mixed with 35 ml of molten top agarose at 42°C and quickly poured onto a bioassay tray containing 400 ml of solid L broth agar supplemented with 10 mM $MgSO_4$, pre-warmed to 37°C. Libraries were plated out on large bioassay trays to give about 40-50,000 plaques per plate. Between one quarter and one third of a packaging reaction was usually used per plate. Once set, the plates were inverted and incubated for 12 hours at 37°C. Plates were then transferred to 4°C for a minimum of 2 hrs to allow the top agarose to harden prior to the taking of nylon/nitrocellulose membrane lifts.

Membrane lifts of 'phage DNA were taken in duplicate to avoid false positives, and the filters and plates were marked with ink to allow later alignment. Membrane was placed on the plates for 30 secs for the first lift and for 60 secs for subsequent lifts. The filters were then processed by immediate immersion in a bath of GS1 for 30secs, GS2 for 2-5 mins and finally rinsed in 2xSSC before being air dried and baked under vacuum at 80°C for 2 hrs. In the case of nylon filters, the DNA was then cross-linked by exposure to 254 nm UV radiation for 2 mins from a distance of 50 cm.

Conditions for hybridisation and washing were the same as for Southern blots.

Positive plaques were picked and purified to homogeneity by replating out and screening before further analysis.

Small Scale Preparation of Bacteriophage λ DNA.

A 20 ml overnight culture of the appropriate bacteria (P2392) was spun down and resuspended in 8ml of SM. A single well isolated λ plaque was removed from an agar plate in a plug of agar and this was added directly to 5 ml of a 1:100 dilution of the above bacteria in LB supplemented with 10 mM $MgSO_4$. The culture was incubated overnight at 37°C and λ DNA was isolated from the resulting liquid lysate as described by Maniatis et al (1982). This produces about 5 μg of λ DNA.

Large Scale Preparation of λ DNA.

A 10 ml overnight culture of the appropriate bacteria was spun down and resuspended in 4 ml SM. These were then incubated with approximately $5x10^7$,phage for 20 mins at 37°C and then used to inoculate 400 ml of prewarmed LB supplemented with 20 mM $MgSO_4$ which was then shaken vigorously at 37°C overnight. Lysis of the bacteria is indicated by a granular appearance of bacterial debris in the culture rather than the normal silky appearance of *E.coli*. The purification of 'phage DNA from the liquid lysate by centrifugation through CsCl step gradients was as described by Maniatis et al (1982).

Removal of Repeated Sequences from Hybridization Probes by Competition.

Driver DNA was prepared by dissolving DNA from appropriate source at 2.5 mg/ml in TE buffer. It was then sonicated (six 1 min pulses separated by 15 sec. intervals). An aliquot of this was checked on a 1.2% agarose gel to see that the average size of the DNA fragments was 300-500 bp. The DNA was then extracted twice with phenol/chloroform and once with chloroform, after which NaOAc pH 5.4 was added to a final concentration of 0.3 M, and the DNA was ethanol precipitated at -70°C for 15min. Following centrifugation for 30 mins in a microfuge at 4°C, the

pellet was washed in 70% ethanol, dried and dissolved in TE at a final concentration of 20-25 mg/ml.

For pre-association of this DNA with a ^{32}P labelled probe the following reaction mixture was set up in a 1.5 ml Eppendorf tube:

^{32}P labelled probe (1-2x10 ^{6}cpm/ml hybridization mix, 20-50 ng probe DNA).

20 µl driver DNA (final concentration 2.5 mg/ml)

TE buffer to 175 µl.

The mix was boiled for 10 mins and 25 µl of1 M $NaPO_4$ pH6.8 (final concentration 250 mM) was added, prior to incubation at 68°C for 2-3hours. This was then either placed directly in the hybridization mix or stored at -20°C for later use (by quickly thawing and adding to hybridization mix).

Methods Used for Generating Yeast Artificial Chromosome Libraries.

Preparation and Restriction of High Molecular Weight Genomic DNA.

High molecular weight DNA was prepared in agarose blocks as previously described except that more cells per block were used for YAC cloning, ranging in DNA concentration per block from 20 to 40 µg (ie 2-5 x 10^6 cells per block). After lysis and Proteinase K treatment, followed by inactivation of Proteinase K with PMSF, the blocks are stored in 0.5 M EDTA pH 8.0 at 4°C.

Prior to restriction enzyme digestion the blocks, were equilibrated with TE by washing twice in 50 mls of TE at room temperature for 30 mins.

In the case of rare cutter enzyme digestions (Cla I and XhoI YAC library preparations) the reactions were carried out as described previously for PFGE. The blocks required for any cloning reaction were individually digested, with an extra one for analysis by PFGE. The reactions were incubated overnight and the enzyme inactivated the next day by addition of 50 mM EDTA pH 7.6 and 1 mg/ml Proteinase K, followed by incubation at 37°C for 30 mins. The digested DNA from one block was fractionated on a PFG at appropriate size resolution conditions to ensure that the

fragments were in the expected size range (ie no degradation, or incomplete digestion).

For the preparation of partially Eco RI cut DNA (for the YAC libraries of overlapping Eco RI fragments) two approaches were used to generate fragments in a suitable size range for YAC cloning. The first involved the use of an appropriate quantity of Eco RI enzyme after calibration of the agarose blocks to be used with different amounts of the enzyme. Calibration reactions were set up as follows:

1/4 of an agarose block (5-10μg DNA)	20 μl
10x Eco RI buffer	20 μl
BSA (5mg/ml)	20 μl
Eco RI (Biolabs)	0.001-10 units/10μg DNA
H_2O	to 200 μl final volume

Reactions containing different quantities of enzyme were set up on ice and then incubated for 1.5 hrs in a 37°C water bath. The reactions were stopped by placing on ice followed by addition of EDTA to 50 mM and Proteinase K to 1 mg/ml and incubation at 37°C for 30 mins. The various partially cut blocks of DNA were then fractionated on a PFG (with size resolution upto 600 kb). The conditions at which a decrease in the size of fragments was just seen, were selected for the large scale restriction digests. These consisted of the required number of 80 μl (whole) agarose blocks, individually treated in final reaction volumes of 500 μl, with the appropriate amount of Eco RI enzyme.

The second approach for generating Eco RI fragments of the appropriate size involved the use of Eco RI methylase (New England Biolabs) to compete for genomic Eco RI sites. The reaction was assembled as follows:

Agarose block (10-20 µg DNA)	80 µl
BSA (5 mg/ml)	50 µl
10x reaction buffer	50 µl
Spermidine (100 mM)	12.5 µl
Eco RI methylase (40 U/µl)	50 µl (2000 units)
Eco RI (0.4 U/µl)	2.5 µl (1 unit)
H_2O	254 µl

The solutions, including the enzymes, were first mixed by vortexing and then left on ice for 10 mins prior to addition of the agarose block. This was kept on ice for a further 1hour, to allow the block to become equilibrated with the reaction mixture and then placed at 37°C for 4 hours.

1x reaction buffer contains:

80 µM S-adenosyl methionine

2 mM MgCl2

100 mM NaCl

50 mM Tris-Cl pH 7.6

1 mM DTT

Under these conditions fragments ranged from 50 kb to > 1 Mbp as judged by PFGE.

Following digestion, the reaction was stopped as before (50 mM EDTA, 1 mg/ml proteinase K at 37°C for 30 mins). A control digest block was checked by PFGE for satisfactory digestion and no degradation of the DNA.

Preparation of Yeast Artificial Chromosome vector.

The pYAC vectors used contain a yeast centromere (CEN4), a yeast replicating sequence (ARS1), two cassettes of the telomeric repeat from Tetrahymena (TEL) connected by the HIS3 gene, two selectable markers (URA3 and TRP1) and the majority of pBR322 . The cloning site is in the intron of a SUP4 ochre gene. The

pYAC4 vector (Burke et al,1987) contains an Eco RI cloning site (see Fig. 4.1.1), whilst pYAC-RC (Marchuk and Collins, 1988) is a derivative of pYAC3 (SnaBI) (Burke et al,1987) and contains a rare cutter polylinker including sites for SacII, SalI and ClaI (see Fig. 4.3.1).

To prepare the vector for cloning, a number of colonies of the *E.coli* strain in which it has been cloned were picked from a freshly streaked LB agar plate supplemented with ampicillin. Plasmid vector DNA was prepared from each (as described under "Small Scale Plasmid Preparation"). As occasional isolates of these vectors have spontaneously deleted small amounts of the telomere sequences, the DNA from each culture was checked for this by digestion with HindIII. Four bands of 3.5 kb, 3.0 kb, 1.9 kb and a doublet of 1.4 kb were observed unless telomere DNA had been deleted, in which case the doublet was resolved and a fifth band would be seen. The culture from an isolate whose telomeres were intact was then used for large scale preparation (as described under "Large Scale Preparation of Plasmid DNA").

To prepare the YAC arms, the vector DNA was first of all digested with BamHI which removed the HIS3 gene, releasing the TEL sequences and generating a linear vector. The cloning site was then digested with the appropriate enzyme for 1-2 hours. After each digestion reaction an aliquot of the vector was checked for complete cutting of the sites by electrophoresis on an agarose gel. The vector arms were then phosphatased twice, the phosphatase inactivated and the DNA extracted with phenol and chloroform (all as described under "DNA Quantitation and Manipulation"), precipitated and resuspended in TE at a final concentration of 1-2 µg/µl. Total dephosphorylation of the YAC arms was checked by means of a self-ligation reaction (as described under "DNA Quantitation and Manipulationan) using 1 µg of the vector DNA, followed by agarose gel electrophoresis where visualisation of the DNA should reveal no trace of inter-ligation between the three fragments (the two arms and the HIS3 fragment). The vector DNA was ready for use in a ligation reaction in this form.

Ligation of Restricted Genomic DNA into the pYAC vectors.

The ligation reactions involved phosphatased vector DNA and unphosphatased genomic DNA. Thus, a fraction of ligation events between genomic fragments was expected, and in the case of the Eco RI partials especially, such partial scrambling would be difficult to recognize. To minimise this, as high a molar ratio of vector ends to genomic DNA ends was used as was available.

The agarose blocks of restricted genomic DNA were prepared for ligation by inactivation of the Proteinase K using PMSF (Phenyl-methyl sulphonyl fluoride). The blocks were washed twice in 15 ml TE containing 40 µg/ml PMSF at 37°C for 30 mins. The PMSF was prepared as described under "Size Fractionation of HMW DNA". Prior to setting up the ligation reaction the agarose blocks had to be melted. They were first equilibrated in TE containing 25 mM NaCl for 30 mins at room temperature. Each block was then placed in an Eppendorf tube, melted at 68°C for 5 mins and placed at 37°C to keep it molten. The ligation reaction was assembled as follows:

DNA	80 µl
10x Ligation Buffer	20 µl
H_2O	to a final volume of 200 µl
pYAC vector	as required
ATP (100 mM)	2 µl
T4 DNA Ligase(400 U/µl)	5 µl

The reaction was very carefully mixed by stirring two or three times with a pipette tip, and it was then incubated at 16°C overnight. If it was not to be used immediately EDTA was added to a concentration of 20 mM and the ligation kept at 4°C.

Size Fractionation of High Molecular Weight DNA for YAC Libraries.

Size fractionation of genomic DNA involved use of CHEF pulsed field gels of low melting point (LMP) agarose. The fractionation was performed at two different stages in the preparation of the digested genomic DNA for cloning.

Where a fragment of a particular size was to be cloned, the totally digested DNA (in blocks, after Proteinase K and PMSF treatment) was run out on a PFG in an agarose gel made with 1% LMP agarose (Beckman) under the appropriate conditions to enrich for the desired fragment size. After running the gel, thin agarose slices around the appropriate region of the gel (as determined by cutting away and EthBr staining the gel slice with the size marker track and after UV visualisation, realigning this with the remainder of the gel). The slices were soaked in TE containing 25 mM NaCl for 30mins, melted at 68°C for 5-10mins, and placed at 37°C. Agarase was added at 50 units/ml molten agarose, and they were left at 37°C overnight. A 100 µl aliquot of each fraction was then used in a slot blot (see above). This was hybridised with the appropriate probe, to determine which fraction contained the highest percentage of the fragments to be cloned. The appropriate fraction was then used in a ligation reaction with the appropriate YAC arms. The ligations were usually performed in a final volume of 1ml using the same vector arm: insert excess as in ligations of total unfractionated DNA, after calculating the approximate amount of genomic DNA in a particular fraction used. The resulting ligated DNA was used in this form for transformation of yeast spheroplasts. Upto 100 µl could be used per transformation if the spheroplasts had previously been resuspended in buffer containing 2 M sorbitol to prevent osmotic shock (this is described under "DNA Transformation into Yeast Spheroplasts").

The second size fractionation system used (based on a protocol provided by Z. Larin and T. Monaco, ICRF) involved size selection of Eco RI partial DNA already ligated into pYAC4 arms on a PFG. This was done to try and increase the average size insert of the YAC clones (non-size selected ligations gave an average size insert of only 50 kb), and to eliminate any non-recombinant vector. The ligated DNA (still in agarose solution, ie without agarase treatment) was run on a CHEF PFG made with 1% LMP agarose as before. The conditions chosen resolved fragments under 600kb, so that the region of the gel cut away for use contained all of the unresolved window of DNA (>600 kb) as well as the slots. This was usually a gel slice of 1-2 cm in width (and however many tracks used, about 8 normally, in length), which

was about 4-5 ml agarose in total. This gel slice was soaked in the appropriate buffer and treated in preparation for yeast transformation as described under "Preparation of Ligation for Yeast Transformation."

Preparation of Ligation for Yeast Transformation.

The presence of agarose lowers the efficiency of transformation into yeast spheroplasts, thus before transformation the agarose was removed using agarase. For the cases where the DNA had not been size-fractionated, the ligation reaction was melted and placed at 37°C, EDTA added (if necessary) to a concentration of 20 mM as well as 5 µl agarase (20 U/µl, purchased from Calbiochem, prepared in 50% glycerol and stored at -20°C). The agarasing reaction was carried out at 37°C overnight. The ligation was kept at 4°C in this form and was ready for use in yeast transformations.

If the ligated DNA was in an agarose gel slice after size fractionation on a PFG, the slice was soaked three or four times, 30 mins each, in the following buffer:

1 M Sorbitol
10 mM Tris-Cl pH7.6
2 mM EDTA pH8.0
25 mM NaCl

(This meant that the highly dilute DNA would be in an appropriate isotonic solution for transformation into yeast spheroplasts).

The agarose slice was melted at 68°C for 10-15 mins and then placed at 37°C. Agarase was added to a concentration of 50 U/ml, the mixture was carefully stirred and placed at 37°C overnight. The ligation solution was ready for direct transformation into yeast in this form, and was kept at 4°C.

Preparation of Yeast Spheroplasts.

(Essentially as in Burgers and Percival, 1987)

Yeast strain AB1380 (Burke et al, 1987): MATa, ade2-1, can1-100, lys2-1, trp1, ura3, his5, [psi$^+$].

In order to propagate the yeast for efficient spheroplasting and subsequent DNA transformation, 5 ml of YPD was innoculated with a single colony from a fresh agar plate (< 1 week old) and incubated as a standing culture overnight at 30°C. Next day, 100 µl of this standing culture was used to innoculate 100 ml YPD and the culture was incubated overnight at 30°C with shaking in a 500 ml conical flask. Such a dilution gave an OD_{600} (1:10) of approximately 0.2 (ie $3x10^7$ cells/ml) after 16 hours. The cells were harvested when the OD_{600} (1:10) was between 0.1 and 0.2 by centrifugation at 2 K for 5 mins. The cells were washed successively with 40mls of sterile water and 40 ml of 1 M sorbitol by resuspension followed by 5 min spins at 2 K. They were resuspended in 40 ml SCEM to which was added 2000 units of lyticase (lyticase was prepared fresh each time by resuspension of the appropriate weight, determined on a chemical balance, in 1 ml SCEM). The cells were incubated at 30°C with occasional inversions. Spheroplasting was followed by measuring the decrease in turbidity at OD_{600} of a 1:10 dilution of spheroplasts in water. Alternatively, spheroplasting was monitored visually using a microscope, by suspending a drop of the spheroplasts on a slide in 20% SDS under a cover slip and observing the degree of cell lysis or "ghosting" (cells with outer walls are not affected by SDS or water, whilst spheroplasts will lyse due to the osmotic shock, and are observed as virtually transparent "ghosts"). When spheroplasting had proceeded to about 90% (ideally within 15-30 mins) the spheroplasts were harvested by centrifugation at 800 rpm for 5 mins and gently resuspended in 40 ml of 1 M sorbitol, pelleted at 800 rpm for 5 mins and again gently resuspended, in 40 ml STC. At this stage the spheroplasts (1:10) were counted using a haemocytometer. 100 ml of cells grown to OD_{600} (1:10) produced $5\text{-}6x10^9$ spheroplasts. The spheroplasts were pelleted once more at 800 rpm for 5 mins and resuspended at a concentration of between 3.5 and $5x10^8$ per ml (depending on the concentration of the DNA to be used in the transformation) in STC. The spheroplasts were used as quickly as possible for maximum transformation efficiency, although they could be left on ice for a few hours.

DNA Transformation into Yeast Spheroplasts.

Generally for plating on a 90 mm petri dish, less than 300 ng of DNA for transformation is added to about $4x10^7$ spheroplasts in 100 µl. 2-5 µg of carrier DNA (sonicated, sterile salmon sperm DNA prepared at 1mg/ml in 1 M sorbitol) was also added which acted as a buffer against nucleases. In order to avoid an osmotic shock through addition of the DNA if it was not in a 1 M sorbitol solution, or if the volume of added DNA was to be more than 10 µl, the spheroplasts were resuspended in STC containing 2 M sorbitol at the final step of spheroplast preparation. This buffered them sufficiently for up to 100 µl of DNA to be added to 100 µl spheroplasts. Individual DNA transformations were carried out with a maximum of 150 µl spheroplasts and 100 µl DNA, as further increases in either caused a rapid decline in the transformation efficiency.

A number of control transformations were performed each time:

Carrier DNA only

YCp50 supercoil 10 ng

YCp50 supercoil 100 ng

YCp50 provided a reliable indicator of the transformation efficiency for each transformation experiment. The optimum transformation conditions achieved gave $5x10^5$ colonies/µg of YCp DNA. Such conditions of transformation were necessary for satisfactory levels of efficiency when using genomic YAC library DNA. However these efficiencies still varied quite widely depending on the particular ligation preparation used.

100 µl of spheroplasts were mixed with the DNA for the transformation and the carrier DNA (total 10-15 µl, or more when speroplasts are in 2 M sorbitol, in which case $CaCl_2$ must also be added to a final concentration of 10 mM) by gentle shaking in a 15 ml sterile centrifuge tube and allowed to stand at room temperature for 10 mins at 22°C. 1 ml PEG was added and the transformation incubated at room temperature for a further 10 mins. The spheroplasts were then pelleted at 900 rpm for 5 mins, and the PEG decanted carefully. The pellet was resuspended in 150 µl of

SOS and placed at 30°C for 30-40 mins. 8 ml top agar (-ura) previously melted and held at 48°C was added to the transformation, mixed by inversion and quickly poured over a 90 mm regeneration plate (-ura) that had previously been warmed to 37°C. The plates were incubated at 30°C for 3-5 days.

SCEM	1 M sorbitol
	0.1 M sodium citrate pH 5.8
	10 mM EDTA pH 7.5
(Added just before use)	30 mM 2-mercaptoethanol
STC	1 M sorbitol
	10 mM Tris-HCl pH 7.5
	10 mM $CaCl_2$

SCE and STC were autoclaved for use.

PEG	20% PEG 6000
	10 mM Tris-HCl pH 7.5
	10 mM $CaCl_2$

PEG was prepared fresh and filter sterilised.

SOS	1M sorbitol
	25% YPD
	6.5 mM $CaCl_2$
	10 µg/ml tryptophan
	10 µg/ml uracil

SOS filter sterilised for use.

Screening of YAC libraries.

Spheroplasts were plated out in an osmotic stabilising agar in order for them to regenerate an outer cell wall without lysing. Thus, the resulting colonies were embedded in the top agar and had to be individually picked first into microtitre dishes or directly onto plates of selectable medium before they could be processed for

screening.

Sterile toothpicks were used to pick individual colonies into wells of 96-well microtitre dishes, where each well contained 200 µl of SD com -ura -tryp medium. The dishes were incubated at 30°C for 2-3 days, after which the yeast cultures were screened, either by transferring them onto SD com -ura-tryp agar plates and then taking filter lifts after 3-4 days growth, which is a time consuming procedure with unreliable sensitivity, or by pooling aliquots of the 96 cultures of each dish and preparing DNA from them for screening by PCR, which is described in Appendix 1.

The conventional screening procedure involved the use of a 96-prong steel replicator, which was used to transfer droplets of the yeast cultures from the microtitre dishes onto duplicate filters of Hybond N+ (Amersham) resting on large (24 x 24 cm) plates containing 2% agar SD COM -ura-tryp medium. The microtitre dishes were stored at 4°C for the short term, or else were placed at -70°C after the medium in each well had been supplemented with glycerol to a final concentration of 30%. The filters were incubated at 30°C for 3-4 days. Following this the yeast colonies on the filters had to be spheroplasted (in order that the DNA could then be bound to the filter by alkaline lysis). This was done by transferring the filters onto 3 MM filter paper soaked in SCEM containing either 4 mg/ml Novozym 234 or 2 mg/ml Lyticase (Artherobacterium luteus, from ICN), in a container that was then sealed with nescofilm. The filters were incubated at 37°C (Novozym) or 30°C (Lyticase) overnight. Colonies were checked for complete spheroplasting by taking a fraction of a colony with a sterile toothpick, gently resuspending the yeast in 1% SDS, and observing this on a slide under a microscope for the presence of spheroplast "ghosts". If spheroplasting was incomplete it was repeated using fresh SCEM+enzyme soaked filters. It was crucial that sufficient DNA/colony finally remained on the filters for screening, as the quantity of YAC DNA present/colony is much less than in bacterial cloning.

Each filter was then processed for screening by placing it first onto Whatman 3MM paper soaked in denaturing solution:1.5 M NaCl, 0.5 M NaOH for 10 mins . Excess denaturant was removed from the filter by allowing it to rest on dry Whatman 3 MM

for 2 mins. It was then neutralised by floating it on a 1.5 cm deep neutralising solution of 1.5 M NaCl, 1 M Tris pH 7.4, for 2 mins. The filter was then incubated on 3MM soaked in 0.1x neutralising solution containing 200 µg/ml Proteinase K for 20-30 mins at 37°C. Cell debris was wiped off the filter using Kleenex tissues soaked in 0.1x neutralising solution and rinsed thoroughly. The filter was then floated on 2xSSC for 2 mins, air dried, baked in a vacuum oven at 80°C for 1 hour and UV cross-linked.

Prehybridisation and hybridisation of YAC library filters were carried out in the same solution and under the same conditions as described in "Hybridisation of Southern Blots". The probe concentrations used were approximately 10^6 counts per minute/ml. Each preparation of probe was tested for hybridisation efficiency with an accompanying control filter on which were bound different ng quantities (0.1, 1, 10, 100 ng) of the same fragment of DNA (as the probe). Signal on 0.1 ng was usually visible after 3-5 days indicating efficient hybridisation of probe to the library filters. YAC library films were also developed after 3-5 days, using the same conditions described under "Hybridisation of Southern Blots".

YAC Selection Systems and Colour Indicators.

The YAC vectors used contain URA3, TRP1, and HIS3 genes, HIS3 being discarded on preparation of the vector for cloning, leaving the potential of selection for ura and trp upon transformation into the ura^-trp^- host. However, the TRP1 gene in pYAC4, is under the control of a truncated promoter. and although transformed yeast would grow in tryptophan deficient medium they would do so slowly. Thus, initial selection (on regeneration plates) was always on medium deficient only in uracil. Subsequent growth of yeast transformants (in liquid and on solid media) was usually selected for both URA3 and TRP1 markers.

A number of forms of selection against non-recombinant plasmid exist in the pYAC vectors. The cloning site is within an intron of a supressor tRNA gene (SUP4 ochre). Disruption of this gene causes a number of selection systems:

a.) [psi+] factor - is a cytoplasmically inherited enhancer which enhances the use of suppressor tRNA molecules in translation. In AB1380 the effect of the psi+ phenotype is to reduce the transformation efficiency of a SUP4 ochre plasmid with respect to a plasmid carrying no such gene.

b.) Canavanine sensitivity: Strains containing the *can1-100* gene are resistant to canavanine, an arginine analogue and cell poison. However an efficient ochre suppressor can suppress the can1-100 mutation rendering the strain sensitive to canavanine. Non-recombinant YAC vectors carry an intact SUP4 genes, so cells grown under canavinine selection are killed, whereas cells carrying a recombinant YAC will be canavanine resistant. Growth medium for canavanine selection is:

-ura,

-arg,

3.3 ml/L 2% L-canavanine solution (canavanine solution was filter sterilised).

This selection system was used as a quick way of testing the proportion of recombinant YACs in a particular library by picking a number of colonies at random and growing them under the above selective conditions.

c.) ade2-1 colour selection (Hieter et al, 1985)

AB1380 has a suppressible mutation at the ade2 locus. If adenine is present in the medium in limiting quantities (4 mg/ml) such mutants will accumulate a red pigment. However, if the mutation is suppressed, the strain produces white colonies: hence, recombinants should be red and non-recombinants white. Such colour coding was used once the colonies were removed from top agar containing sorbitol and were transferred to SD COM -ura or SD COM -ura -trp plates.

However, accumulation of sufficient red pigment for scoring took over two weeks, thus this was a less reliable indicator than random canavanine testing for YAC recombinants.

Small Scale Preparation of Yeast Chromosomes from YACs (Yeast Miniblocks).

For PFG analysis of YAC size in a particular isolate or for small scale restriction digestions, chromosomal size DNA was prepared by picking a single yeast colony into 2.5 mls of SD COM -ura (or -ura-tryp) medium and leaving the standing culture for exactly 48 hours at room temperature or overnight at 30°C. The final concentration had to be approximately $5x10^7$ cells/ml. The culture was transferred into a 2.5 ml Eppendorf tube, and the cells pelleted in a microfuge at 6500rpm for 30 secs. The pellet was resuspended in 1 ml 1 M sorbitol, 50 mM EDTA and spun down again. The pellet was resuspended in SCE to a concentration of $5x10^7$ cells / 40 µl (ie about 2 µg DNA per block). Usually a 2.5 ml culture was sufficient for 2 blocks. The agarose blocks were made by adding an equal volume of 1.2% LMP agarose (1.2% LMP agarose in SCEM containing 4 mg/ml Novozym 234 that has been kept at 48°C) to the yeast cells, and dispensing 80 µl aliquots into precooled block formers. The blocks were left to set for 15-20 mins, and then pushed out into SCEM and incubated at 37°C for at least 1 hour with gentle shaking. The blocks were then transferred into 0.5M EDTA, 1% sarkosyl and 2 mg/ml proteinase K (about 1 block/ml) and incubated at 55°C for 48 hours. Blocks were loaded onto gels directly or stored at 4°C for later use.

Cytogenetic Methods.

Rat Chromosome Spread Peparation.

Rat cells which were almost confluent in 9 cm tissue culture dishes were treated with colcemid to a final concentration of 0.05 µg/ml 2-3 hours before harvesting. After this time the medium was removed and kept. The cells were rinsed once with versene, and then with a 1:1 trypsin:versene mixture. These rinses were also kept. A 1:1 trypsin:versene mix was then added and incubated at 37°C for a few minutes until the cells had detached. The medium from each step was then pooled and centrifuged at 1K for 5 mins. The supernatant was discarded into Chloros and the cells resuspended in 10 ml of 0.075 M KCl. This was incubated at 37°C for 10

mins, and then centrifuged at 1 Krpm for 5 mins. The supernatant was discarded into Chloros, and the pellet gradually resuspended in fresh 3:1 methanol:glacial acetic acid fixative. This was centrifuged at 1 Krpm for 5 mins and the supernatant removed (but not discarded into Chloros). Washing of the cells in fixative was repeated once more. The cells were finally resuspended in a small volume (<1 ml) of fixative. 2-3 drops were dropped onto a clean, wet, glass slide from a height of about 1 m. The spreads were immediately observed under phase contrast with a microscope. The slides were then stored dry at 4°C for about 1 month (minimum 1 week before use for banding and/or *in situ* hybridisation).

Giemsa Solid Staining of Rat Chromosomes.

Giemsa stain is a Romanowsky-type dye mixture. Azure B and eosin in the Giemsa bind to DNA. Solid staining was used for scoring chromosome breaks. The slide used was at least a few days old. It was placed in Wright's stain freshly diluted 1:3 in 50% Sorensons buffer for 5 mins. The slide was then rinsed in tap water and air dried. It was coverslipped with 50% Sorensons buffer immediately prior to examination as the buffer leached out the stain. A Zeiss photomicroscope 3 was used with a magnification of approximately 800x. The camera was set at 11DIN and a KODAK 2415 film was used to take photographs.

G-Banding of Rat Chromosomes.

This method allowed prebanding of rat chromosomes which were subsequently used for *in situ* hybridisation. The slide to be used was at least one week old. It was incubated in 2x SSC at 60°C for 7 mins, rinsed with water and then stained for 2.5 mins in Wright's stain freshly diluted 1:3 in 50% Sorensons buffer. It was then rinsed and allowed to air dry. It was coverslipped and photographed under the microscope exactly as described under "Solid Staining of Rat Chromosomes". The slides are destained by rinsing with fixative.

Sorenson's phosphate buffer:

0.025 M KH_2PO_4 (3.4 g/l) titrated to pH 6.8

with 50% NaOH.

Wright's stain (0.25% stock solution):

2.5 g stain dissolved in 1 l of anhydrous methanol

incubated at 37°C for 16 hrs.

Age solution for minimum of 7 days.

In situ Hybridisation on Rat Chromosomes using Biotinylated Probes.

(Based on the method of Sarah Williams, Human Cytogenetics Laboratory, ICRF).

Probe Labelling:

1 µg of lambda DNA was labelled with biotinylated dUTP by nick translation using the BRL Bionick Kit. The reaction was set up as follows:

1 µg lambda DNA

5 µl dNTP mix (includes bio-14-dUTP)

dH_2O to 45 µl

5 µl enzyme mix

The reaction was incubated at 15°C for 1 hour, and 5 µl Stop buffer was then added. The probe was purified through a Sephadex G50 column in a pasteur pipette, and eluted with TE. The probe was within the second 600 µl collected. To this, 50 µg *E. coli* tRNA and 50 µg salmon sperm DNA (sonicated to approximately 500 bp length) were added as carrier, and then the DNA was precipitated by addition of NaOAc to 0.3 M and 2 volumes of 95% ethanol. The precipitated DNA was pelleted, dried and resuspended in 50 µl TE or dH_2O.

Removal of Repeated Sequences from Probe:

Prior to hybridisation, probes containing repetitive sequences were preannealed with total rat DNA. 80 ng of the biotinylated DNA was mixed with 2.5 µg competitor DNA (rat DNA sonicated to approximately 500 bp length) with 2 volumes of 95% ethanol in an Eppendorf tube. This was vacuum dried and resuspended in 11 µl hybridisation mix:

10% dextran sulphate

2 x SSC

50% deionised formamide

1% Tween 20, pH 7.0

which was stored in 0.5 ml aliquots at -20°C or -40°C.

This mixture was then denatured at 70-100°C for 10 mins, chilled on ice, and preannealed at 37°C for 3 hours.

Hybridisation of biotinylated probe:

Each hybridisation event normally occupied half a slide. 80 ng of the biotinylated DNA was used, if appropriate, preannealed with competitor. Denaturing of the chromosomal DNA on the slides was done shortly before they were required, by immersion in 70% formamide, 2 x SSC pH 7.0 at 75°C for 3 mins. They were then dehydrated through an ethanol series of : cold 70%, 95%, absolute, for 3 mins each. The prepared probe in hybridisation mix was then spun down briefly, and placed on the slide and covered with a 22x22 mm coverslip. This was sealed with rubber solution and the slides placed in a moist chamber at 37°C for overnight or upto 4 days (usually around 65 hours).

Post Hybridisation washes:

The coverslip and rubber solution were removed from each slide with tweezers. The slides were then immersed in 50% formamide, 2x SSC pH 7.0 at 42°C for 5 mins, three times. They were then immersed in 2x SSC pH 7.0 at 42°C for 5 mins, three times.

Detection of signal:

Each slide was immersed in 4x SSC, 0,05% Tween 20 pH 7.0 for 3 mins. It was then preincubated for 10 mins with 100 µl of 4x SSC, 0,05% Tween 20 pH 7.0 plus 5% Marvel non-fat dry milk, (SSCTM), with the solids spun out before use, under a 22x50 mm coverslip.The slide was then washed for 3 mins in 4xSSC, 0.05% Tween 20 pH7.0. There then followed a 20 mins incubation with 100 µl of avidin-FITC (5 µg/ml in SSCTM, solids spun out before use) under a 22x50 mm coverslip.

This incubation was kept in the dark. The slide was then washed in 4xSSC, 0.05% Tween 20 pH 7.0 for 3 mins, three times. It was then incubated with 100 µl biotinylated anti-avidin 95 µg/ml in SSCTM, with the solids spun out before use, under a 22x50 mm coverslip, again, in the dark for 20 mins. It was then washed in 4x SSC, 0.05% Tween 20 pH 7.0 again for 3 mins three times. It was finally incubated with FITC-avidin as before for 20 mins, and then washed in 4x SSC, 0.05% Tween 20 pH 7.0 for 3 mins, once. Each slide was then washed in PBSA for 5 mins, twice. The slides were then dehydrated through an ethanol series of 70%, 95%, absolute ethanol, and finally mounted in antifade medium containing counterstain (Citifluor with 0.5µg/ml propidium iodide), under a 22x50 mm coverslip.

Viewing and Photography:

A Zeiss photomicroscope 3 was used. The fluorescent signal was detected by means of a fluorescence filter set with excitation filter BP450-490, beam splitter FT510 and barrier filter LP520 (Zeiss filter set 9). In this way it was possible to visualise the propidium iodide counterstain and FITC signal simultaneously. Photography was either onto slide (Scotch 3M 640T) or print (Fujicolour HR400) film. Automatic exposure with spot reading was approximately 2 mins on the microscope used.

REFERENCES.

Alitalo, K., and Schwab, M. (1987). Advances in Cancer Research **47**, 235-281.

Amler, L. C., and Schwab, M. (1989). Mol. Cell. Biol. **9**, 4903-4913.

Anand, R. (1986). Trends Genet. **2**, 278-283.

Ardeshir F, Giulotto E, Zieg J, Brison O, Liao WSL, Stark GR. (1983) Mol. Cell. Biol. 3, 2078-2088

Atkin, N. B., and Brito-Babapulle, V. (1981). Cancer Genet. Cytogenet. **3**, 261-272.

Ayusawa, D., Iwata, K., and Seno, T. (1983). Cancer Res. **43**, 814-818.

Beverley, S. M., Ellenberger, T. E., Iovannisci, D. M., Kaplar, G. M., Petrillo-Peixoto, M., and Sina, B. J. (1988). "Gene Amplification in Leishamania" in: P.T England and A.Sher (eds), The Biology of Parasitism, pp 431-448. New York: Alan R. Liss, Inc.

Biedler, J. L., Tien-ding, C., Scotto, K. W., Melera, P. W., and Spengler, B. A. (1988). Cancer Res. **48**, 3179-3187.

Bishop, J. M. (1987). Science **235**, 305-311.

Bird, A. P., Taggart, M., Frommer, M., Miller, O. J., and Macleod, D. (1985). Cell **40**, 91-99.

Blattner, F. R., Blechl, A., Denniston-Thompson, K., Faber, H., Richards, J. E., Slighton, J. L., Tucker, P., Smithies, O. (1978). Science **202**, 1279-1284.

Bolivar, F., Rodriguez, R. L., Betlach, M. C., and Boyer, H. W. (1977). Gene **2**, 75.

Borst, P., Van der Blick, A. M., Van der Velde-Koerts, T., and Hes, E. (1987). J. Cell. Biochem. **34**, 247-258.

Botchan, M., Topp, W., and Sambrook, J. (1976). Cell **9**, 269-287.

Brown, W. R. A. and Bird, A. P. (1986). Nature **322**, 477-481.

Brown, P. C., Tlsty, T. D., Schimke, R. T. (1983). Mol. Cell. Biol. **3**, 1097-1107.

Bullock, P., and Botchan, M. (1982). In "Gene Amplification", pp 215-224. Cold Spring Harbour Laboratory Press, New York.

Burgers, P. M. J., and Percival, K. J. (1987). Analytical Biochemistry **163**, 391-397.

Burke, D. T., Carle, G. F., and Olson, M. V. (1987). Science **236**, 806-812.

Carle, G. F., Frank, M., and Olson, M. V. (1986). Science **232**, 65-68.
Carle GF and Olson MV (1984) Nucl. Acids Res. 12 : 5647
Carman, M. D., Schornagel, J. H., Rivest, R. S., Srimatkandada, S., Portlock, C. S., Duffy, T., and Bertino, J. R. (1984). J. Clin. Oncol. **2**, 16-20.

Carroll, S. M., DeRose, M. L., Gaudray, P., Moore, C. M., Needham-Vandevater, D. D., and Whl, G. M. (1988). Mol. Cell. Biol. **8**, 1525-1533.

Carroll, S. M., Gaudray, P., DeRose, M. L., Emery,J. F., Meinkoth, J. L., Nakkim, E., Subler, M., VonHoff, D.D., and Wahl, G. M. (1987). Mol. Cell. Biol. **7**, 1740-1750.

Cherif, D., Lavialle, C., Modjtahedi, N., Le Coniat, M., Berger, R., and Brison, O. (1989). Chromosoma **97**, 327-333.

Chu, G., Vollrath, D., and Davis, R. W. (1986). Science **234**, 1582-1585.

Church, G. M., and Gilbert, W. (1984). Proc. Natl. Acad. Sci. USA **81**, 1991-1995.

Collins, F. S., and Weissman, S. M. (1984). Proc. Natl. Acad. Sci. USA **81**, 6812-6816

Committee for a Standardized Karyotype of *Rattus norvegicus*. (1973). Cytogenet. Cell Genet. **12**, 199-205.

Conrad, M., and Topal, M. D. (1989). Proc. Natl. Acad. Sci. USA **86**, 9707-9711.

Coulson, A., Waterston, R., Kiff, J., Sulston, J., and Kohara, Y. (1988). Nature **335**, 184-186.

Cowell, J. K. (1982). Ann. Rev. Genet. **16**, 21-59.

Debatisse, M., Hyrien, O., Petit-Koskas, E., and Robert de Saint Vincent, B., (1986). Mol. Cell. Biol. **6**, 1776-1781.

Debatisse, M., Saito, I., Buttin, G., and Stark, G. R. (1988). Mol. Cell. Biol. **8**, 17-24.

Endow, S. A., and Atwood, K. C. (1988). Trends Genet. **4**, 348-351.

Field, L. M., Devonshire, A. L., and Forde, B. G. (1988). Biochem. J. **251**, 309-312.

Foote, S. J., Thompson, J. K., Cowman, A. F., Kemp, D. J. (1989) Cell **57**, 921-930.

Ford, M., Davies, B., Griffiths, M., Wilson, J., and Fried, M. (1985). Proc. Natl. Acad. Sci. USA **82**, 3370-3374.

Ford, M. and Fried, M. (1986). Cell **45**, 425-430.

Fougere-Deschatrette, C., Schimke, R. T., Weil, D., and Weiss, M. C.(1984). J. Cell. Biol. **99**, 497.

Fried, M., Griffiths, M., Davies, B., Bjursell, G., LaMantia, G., and Lania, L. (1983). Proc. Natl. Acad. Sci. USA **80**, 2117-2121.

Frischauf, A. M., Lehrach, H., Poustka, A., and Murray, N. (1983). J. Mol. Biol. **170**, 827-842.

Futcher, A. B. (1986). J. Theor. Biol. **119**, 197-204.

Gerlach, J. H., Endicott, J. A., Juranka, P. F., Henderson, G., Sarangi, F., Deuchers, K. L., and Ling, V. (1986). Nature **324**, 485-489.

Giulotto, E., Saito, I., and Stark, G. R. (1986). EMBO Journal **5**, 2115-2121.

Glanville, N. (1985). Mol. Cell. Biol. **5**, 1456-1464.

Gudkov, A. V., Chernova, O. B., Kazarov, A. R., and Kopnin, B. P. (1987). Somatic Cell. Mol. Genet. **13**, 609-619.

Haber, J. E., and Thorburn, P. C. (1984). Genetics **106**, 207-226.

Hahn, P., Morgan, W. F., and Painter, R. B. (1987). Somatic Cell. Mol. Genet. **13**, 597-608.

Hamlin, J. L., Milbrandt, J. D., Heintz, N. H., and Azizkhan, J. C. (1984). Int. Rev. Cytol. **90**, 31-82.

Hanahan, D. (1983). J. Mol. Biol. **166**, 557-580.

Hanahan, D. (1985). Nature **315**, 115-122.

Heard, E., Davies, B., Feo, S., and Fried, M. (1989). Nucleic Acids Res. **17**, 5861.

Heartlein, M. W., and Latt, S. A. (1989). Nucleic Acids Res. **17**, 1697-1716.

Hieter, P., Mann, C., Snyder, M., and Davis, R. (1985). Cell **40**, 381-392.

Holden, J. J. A., Hough, M. R., Reimer, D. L. and White, B. N. (1987). Cancer Genet. Cytogenet. **29**, 139-149.

Holmes, D. S., and Quigley, M. (1981). Anal. Biochem. **114**, 193-197.

Horns, R. C., Dower, W. J., Schimke, R. T. (1984). J. Clin. Oncol. **2**, 2-7.

Hunt, J. D., Valentine, M., and Tereba, A. (1990). Mol. Cell. Biol. **10**, 823-829.

Hsu, T. C., and Somers, C. E. (1961). Proc. Natl. Acad. Sci. USA **47**, 396-403.

Hyrien, o., Debatisse, M., Buttin, G., and Robert de Saint Vincent, B. (1987). EMBO Journal **6**, 2401-2408.

Hyrien, O., Debatisse, M., Buttin, G., and Robert de Saint Vincent, B. (1988). EMBO Journal **7**, 407-417.

Jones, P. A., and Taylor, S. M. (1980). Cell **20**, 85-93.

Kafatos, F.C., Orr, W., and Delidakis, C. (1985). Trends Genet. **1**, 301-306.

Kaufman, R. J., Sharp, P., A., and Latt, S. A. (1983). Mol. Cell. Biol. **3**, 699-711.

Klein, H.L. and Welch, S K. (1982) Nucleic Acids Res. 8: 4657-4669

Koide, T., Ishiura, M., Hazumi, N., Shiroishi, T., Okada, Y., and Uchida, T. (1990). Genomics **6**, 80-88.

Kurita, Y., Sugiyama, T., and Nishizuka, Y. (1968). Cancer Res. **28**, 1738-1752.

Lacy, J., Summers, W. P., Watson, M., Glazer, P. M., and Summers, W. C. (1987). Proc. Natl. Acad. Sci. USA **84**, 5838-5842.

Laird, C., Jaffe, E., Karpen, G., Lamb, M., and Nelson. (1987). Trends Genet. **3**, 274-281.

Latt, S. A., Lalande, M., Donlon, T., Wyman, A., Rose, E., Shiloh, Y., Korf, B., Muller, U., Sakai, K., Kanda, N., Kang, J., Stroh, H., Harris, P., Bruns, G., Wharton, R., and Kaplan, L. (1986). DNA-based detection of chromosome deletion and amplification: diagnostic and mechanistic significance. Cold Spring Harbour Symp. Quant. Biol. **51**, 299-307.

Legouy, E., Fosser, N., Llomond, G., and Brison, O. (1990). Som. Cell Mol. Genet. **15**, 309-320.

Lemontt, J. F., Azzaria, M., and Gros, P. (1988). Cancer Res. **48**, 6348.

Levan, G. (1974). Hereditas **77**, 37-52.

Levan, G., Ahlstrom, U., and Mitelman, F. (1974). Hereditas **77**, 263-280.

Lin, C. C., Alitalo, K., Schwab, M., George, D., Varmus, H. E., and Bishop, J. M. (1985). Chromosoma **92**, 11-15.

Looney JE and Hamlin JL (1987) Mol. Cell Biol. 7, 569-577.

Looney, J. E., Ma, C., Leu, T. H., Flintoff, W. F., Troutman, W. B., and Hamlin, J.L. (1988). Mol. Cell. Biol. **8**, 2316-2327.

Ma, C., Looney, J., E., Leu, J-H., and Hamlin, J. L. (1988). Mol. Cell. Biol. **8**, 2316-2327.

Maeda, N., and Smithies, O. (1986). Ann. Rev. Genet. **20**, 81-108.

Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982). Molecular cloning. A Laboratory Manual, Cold Spring Harbour Laboratories. Cold Spring Harbour Press, New York.

Marchuk, D., and Collins, F. S. (1988). Nucleic Acids Res. **16**, 7743.

McClintock, B. (1951). Cold Spring Harbour Symp. Quant. Biol. **16**, 13-47.

Meinkoth, J., Killary, A. M., Fournier, R. E. K., and Wahl, G. M. (1987). Mol. Cell. Biol. **7**, 1415-1424.

Meuth, M. in "Mobile DNA". (1989). Berg, D., E., and Howe, M. (eds), American Soc. for Microbiology, Washington DC.

Moore, G. , Hedge, P. J., Rider, S. H., Xu, W., Hing, S., Palmer, R., Sheer, D., and Solomon, E. (1989). Genomics **4**, 152-161.

Mouches, C., Pasteur, N., Berge, J. B., Hyrien, O., Raymond, M., De Saint Vincent, B. R., De Silvestric, M., and Georghiou, G. P. (1986). Science **233**, 788.

Mouches, C., Pauplin, Y., Agarwal, M., Lemieux, L., Herzog, M., Abadon, M., Beyssat-Arnaouty, V., Hyrien, O., Robert de Saint Vincent, B., Georghiou, G. P., and Pasteur, N. (1990). Proc. Natl. Acad. Sci. USA **87**, 2574-2578.

Murray, A. W., Schultes, N. P., and Szostak, J. W. (1986). Cell **45**, 529-536.

Murray, A. M., and Szostak, J. W. (1983). **305**, 189-193

Nalbantoglu, J., and Meuth, M. (1986). Nucleic Acids Res. **14**, 8361-8371.

Osheim, Y. N., and Miller, O. L. Jr. (1983). Cell **33**, 543-553.

Ottagio, L., Agnese, C., Bonatti, S., and Cavolina, P. (1988). Mut. Res. **199**, 111.

Pallavicini, M. G., DeTeresa, P. S., Rosette, C., Gray, J. W., and Wurm, F. M. (1990). Mol. Cell. Biol. **10**, 401-404.

Passananti, C., Davies, B., Ford, M., and Fried, M. (1987). EMBO Journal **6**, 1697-1703.

Pinkel, D., Straume, T., and Gray, J. W. (1986). Proc. Natl. Acad. Sci. USA **83**, 2934-2938.

Pirotta, V., Jaekle, H., Edstrom, J. E. (1983). In Setlow, J. K. and Hollaender, A. (eds), Genetic Engineering: Principles and Methods. Plenum Press, Vol. **5**, pp 1-17.

Popescu, N. C., and DiPaolo, J. A. (1989). Cancer Genet. Cytogenet. **42**, 157-171.

Poutska, A., and Lehrach, H. (1986). Trends Genet. **2**, 174-179

Prody, C., A., Dreyfus, P., Zamir, R., Zakut, H., and Soreq, H. (1989). Proc. Natl. Acad. Sci. USA **86**, 690-694.

Razin, A., and Riggs, A., D. (1980). Science **210**, 604-610.

Riva, P., De Giuli Morghen, C., and Latizza, L. (1989). Somat. Cell Mol. Genet. **15**, 377-385.

Roberts, J. M. and Axel, R. (1982). Cell **29**, 109-119.

Roberts, J. M., Buck, L. B., and Axel, R. (1983). Cell **33**, 53-63.

Robins, D. M., Ripley, S., Henderson, A. S., and Axel, R. (1980). Cell **23**, 29-39.

Rochaix, J. D., Bird, A., and Bakken, A. (1974). J. Mol. Biol. **87**, 473.

Ruiz, J. D., Choi, K., Von Hoff, D. D., Roninson, I. B., and Wahl, G. M. (1989). Mol. Cell. Biol. **9**, 109-115.

Ruiz, J. C., and Wahl, G. M. (1988). Mol. Cell. Biol. **8**, 4302-4313.

Ruiz, J. C., and Wahl, G. M. (1990). Mol. Cell. Biol. **10**, 3056-3066.

Sager, R., Gadi, I. K., Stephens L., and Grabowy, C. T. (1985). Proc. Natl. Acad. Sci. USA **82**, 7015-7019.

Saiki, R. K., Gelfand, D. T., Mullis, K. B. and Ehrlich, H. A. (1988). Science **239**, 487-491.

Saito, I., Groves, R., Giulotto, E., Rolfe, M., and Stark, G. R. (1989). Mol. Cell. Biol. **9**, 2445-2452.

Saito, I., and Stark, G. R. (1986). Proc. Natl. Acad. Sci. USA **83**, 8664-8668.

Scalenghe, F., Turco, E., Edstrom, J. E., Pirotta, V., and Melli, M. (1981). Chromosoma **82**, 205-216.

Schwartz, D. C., and Cantor, C. H. (1984). Cell **37**, 67-75.

Schimke, R., T. (1984). Cell **37**, 705-713.

Schimke RT, Sherwood SW, Hill AB, and Johnston RN, (1986) Proc. Natl. Acad. Sci. 83 2157-2161.

Shen, D-W., Fojo, A., Chin, J. E., Roninson, I., B., Richert, N., Pastas, I., and Gottesman, M. M. (1986). Science **232**, 643-645.

Shimada, T., and Nienhuis, A. W. (1986). J. Biol. Chem. **260**, 2468.

Spradling, A. C. (1981). Cell **27**, 193-201.

Stallings, R. L., Munk, C., Longmire, J. L., Carle, E., Hildebrand, C. E., and Crawford, B. D. (1984). Mol. Cell. Biol. **4**, 2932-2936.

Stark, G. R., Debatisse, M., Giulotto, E., and Wahl, G. M. (1989). Cell **57**, 901-908.

Stark, G. R., Debatisse, M., Wahl, G.M. and Glover, D. M. (1990). DNA amplification in eukaryotes. In "Gene Rearrangement", Hames, B. D., and Glover, D. M. (eds), Oxford University Press, New York.

Stark, G. R. and Wahl, G. M. (1984). Annu. Rev. Biochem. **53**, 447-491.

Stinchcomb, D. T., Struhl, K., and Davis, R. W. (1979). Nature **282**, 39-43

Sugiyama, T. (1971). J. Nat. Cancer Inst. **47**, 1267-1275.

Sutherland, G., and Hecht, F. (1985). Fragile Sites on Human Chromosomes (Oxford Monographs on Medical Genetics No.13), Oxford University Press.

Taya, Y., Terada, M., and Sugimura, T. (1987). Advances in Viral Oncol. **7**, 141-153.

Tlsty, T., Brown, P. C., Johnston, R., and Schimke, R. T. in Schmke, R., T.(1982). In "Gene Amplification", pp 231-238. Cold Spring Harbour Laboratory Press, New York.

Tlsty, T., Brown, P. C., and Schimke, R., T. (1985). Mol. Cell. Biol. **4**, 1050-1058.

Tokino, T., Fukushige, S., Nakamura, T., Nagaya, T., Murotsu, T., Shiga, K., Aoki, N., and Matsubara, K. (1987). J. Virol. **61**, 3848-3854.

Tooze, J. (ed). (1981). Molecular Biology of Tumor Viruses, 2nd edition, revised: DNA Tumour Viruses. Cold Spring Harbour Laboratory Press, New York.

Topp, W. C. (1981). Virology **113**, 408-411.

Trask, B. J., and Hamlin, J. L. (1989). Genes and Dev. **3**, 1913-1925.

Trent, J. M., Buick, R. M., Olson, S., Horns, R. C., and Schimke, R. T. (1984). J. Clin. Oncol. **2**, 8-15.

Vogelstein, B., Pardoll, D., M., and Coffee, D. S. (1980). Cell **22**, 79-85.

Volkert, F. C. and Broach, J. R. (1986). Cell **46**, 541-550.

Von Hoff, D. D., Needham-VanDevanter, D. R., Yucel, J., Windle, B. E., and Wahl, G. M. (1988). Proc. Natl. Acad. Sci. USA **85**, 4804-4808.

Wahl, G. M. (1989) Cancer Res. **49**, 1333-1340.

Wahl, G. M., Padgett, R. A., and Stark, G. R. (1979), J. Biol. Chem. **254**, 867-989.

Wahl, G. M., Robert de Saint Vincent, B., and DeRose, M. L. (1984). Nature **307**, 516-520.

Wyman, A. R., Wolfe, L. B., and Botstein, D. (1985). Proc. Natl. Acad. Sci. USA **82**, 2880-2884.

Yeung, C. Y., Frayne, L. G., Al-Ubaidi, M. R., Hook, A. G., Wright, A. G., and Kellems, R. E. (1983). J. Biol. Chem. **258**, 15179-15185.

APPENDIX 1

Volume 17 Number 14 1989 Nucleic Acids Research

An improved method for the screening of YAC libraries

Edith Heard, Brendan Davies, Salvatore Feo and Mike Fried

Eukaryotic Gene Organisation and Expression Laboratory, Imperial Cancer Research Fund, PO Box 123, Lincoln's Inn Fields, London WC2A 3PX, UK
Submitted June 12, 1989

The cloning and mapping of large regions of genomic DNA has been greatly facilitated by Yeast Artificial Chromosome (YAC) vectors[1]. However, the yeast system causes certain problems in the screening of YAC libraries by colony hybridization: (1)Primary yeast transformants must be embedded in agar, which precludes direct colony screening. (2)The number of YACs per cell and the number of cells per yeast colony are much lower than in cloning in bacteria. (3)The spheroplasting of the colonies on the filters prior to lysis can be inefficient leading to under-representation of some colonies. Here we describe a rapid and sensitive method for screening YAC libraries which involves screening of DNA from pools of colonies using the Polymerase Chain Reaction (PCR)[2].

Primary yeast transformants are picked into 96 well microtitre dishes for library storage (0.2ml selective medium/well). DNA from the colonies grown up in each dish can be prepared using a rapid miniprep method, after pooling together 0.1 ml of saturated culture ($5x10^5$cells/well) from each of the 96 wells using a 12-channel micropipette. The pools are spun down, washed once in SCE buffer (1M sorbitol, 0.1M sodium citrate, 10mM EDTA), resuspended in 0.32 ml SCEM (SCE plus 30mM 2-mercaptoethanol) and spheroplasted by incubation at 30°C for about 30' with Lyticase (2.85units/ml). The spheroplasts are gently spun down (30" in microfuge), resuspended in 0.4 ml of TE containing 40mM EDTA, 75mM Tris (pH7.4),0.4% SDS, and lysed by incubation at 68°C for 30'. 0.1ml of cold 4M K acetate is added and the minipreps kept on ice for one hour. Precipitated protein is spun down (15' at 4°C), the supernatant transferred to a fresh tube and the DNA is ethanol precipitated. The pellet is resuspended in 0.3 ml TE, treated with RNAase and then the DNA is reprecipitated, washed in 80% ethanol, and resuspended in 10λ of TE (final conc. 0.2-0.5mg DNA/ml). Thus, DNA for screening can be prepared from 24 pools (representing 2304 colonies) simultaneously in just a few hours. The DNA from the pools is stored for subsequent screenings.

For screening, 1µl of DNA from each of the pools is used in a single PCR using a pair of oligomers derived from a known target sequence as primers. If one of the 96 colonies represented in each PCR contains the sequence between the primers, a band of appropriate size will be amplified. The figure shows a number of pools screened for a YAC containing a region of Polyoma virus DNA intergrated into rat DNA, using 27 bp primers derived from Polyoma sequence that were about 1 kb apart. Only the pool containing this sequence amplifies the 1 kb fragment, giving a band on an ethidium bromide stained gel. In the event of inefficient PCRs (bad primers, poor growth of positive colony in well), a blot of this gel may allow detection of the band by hybridization. The positive colony is identified by PCRs on row and column minipreps of the appropriate microtitre dish, and this can be verified by PCR on the single (untreated) colony.

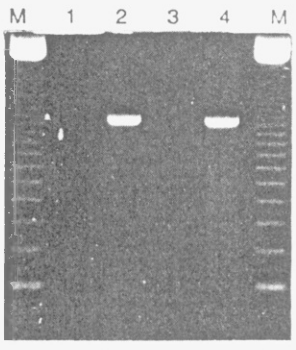


Fig.Legend :
1% agarose gel of PCR products from a selection of screened pools. Lane 2: Pool containing positive colony. Lanes 1and 3: Two random negative pools. Lane 4: PCR product from 2µg genomic rat DNA containing the target sequence.
M: 123bp ladder markers.

References.
1. Burke *et al* (87), Science 236 806-812.
2. Saiki *et al* (88), Science 239 487-491.

© IRL Press