Université de Sherbrooke

Integration of a polyomavirus recombinant containing highly repetitive sequences: Analysis of the junctions and genome rearrangements.

par

John C. Wallenburg

Département de Microbiologie

Thèse présentée à la Facultée de Médecine

en vue de l'obtention du grade de

Philosophiae Doctor (Ph.D.)

Novembre 1989

Permission has been granted to the National Library of Canada to microfilm this thesis and to lend or sell copies of the film.

L'autorisation a été accordée à la Bibliothèque nationale du Canada de microfilmer cette thèse et de prêter ou de vendre des exemplaires du film.

The author (copyright owner) has reserved other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without his/her written permission.

L'auteur (titulaire du droit d'auteur) se réserve les autres droits de publication; ni la thèse ni de longs extraits de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation écrite.

ISBN 0-315-61255-X

SUMMARY

Studies on the integration of exogenous DNA into the genomes of mammalian cells have established that three different recombination mechanisms are involved; site-specific, illegitimate and homologous recombination. Since no essential functions have been found to be necessarily associated with the exogenous molecule it is beleived that this process faithfully reflects the natural recombination mechanisms of the cell. The DNA tumour virus paradigm has led to the development of a linear "replacement" model for integration by illegitimate recombination - model which predicts an excision of host DNA similar to the length of DNA inserted.

RmI is a naturally occurring thermosensitive Py recombinant that contains an insertion (INS) of mouse cellular DNA. RmI should have the potential to integrate by any or all three of site-specific, illegitimate, or homologous recombination mechanisms. Firstly, RmI contains viral Py sequences known to integrate by illegitimate recombination. Secondly, RmI contains B2 and MT repetitive sequences both of which have 100,000 homologous copies dispersed throughout the rat genome any one of which could serve as a target for homologous integration. Finally the junctions between Ins and the Py sequences of RmI display features that suggest that RmI may be capable of site-specific recombination.

My previous work on the integration of RmI into the genomes of rat cells showed that at the temperature non-permissive for replication, integration was non random with respect to the sequences of RmI. One region of RmI was underrepresented, whereas two regions, including the region of Ins containing the repetitive elements B2 and MT, were significantly overrepresented. Furthermore, the lengths of the integrated genomes tended importantly toward the unit length of RmI contrary to what was observed at the permissive temperature.

In an effort to further define the integration mechanism and in particular to determine the nature of the viral-cellular junctions at the molecular level as well as to characterize the host site before and after integration, the cellular DNAs flanking 6 junctions from 3 transformed cell lines were cloned (4 in this study, 2 were previously cloned) and the sequences across 4 of these junctions were determined (one previously determined, 3 were determined here). The cloned DNAs were used to map the rearrangements of the cellular DNA caused by the integration of RmI. Furthermore, the fate of the rearranged intervening host sequences in one of the clones was determined.

The results show that the overrepresentation of the repetitive sequences was due to their acting as hotspots for illegitimate recombination. Even though the host

genome contained over 10^5 potential targets for homologous recombination we found no evidence for homologous integration, and conclude that it is not an efficient procedure in the case of RmI. Furthermore, contrary to previous reports which suggested that exogenous DNA integrated preferentially into repetitive sequences of the host, we found that the DNA flanking all but one junction was unique sequence DNA.

Even though RmI was transfected in the presence of carrier DNA, the results presented here show that it integrated directly into the rat cellular DNA. This shows that transgenomes are not necessary intermediates for the integration of (circular) DNA even when transfected in the presence of carrier.

The minimum lengths of the rearrangements caused to the host DNA by RmI's integration into two of the clones was determined to be 12 kbp for an insertion of 7 kbp and 55 kbp for an insertion of 6 kbp. It was further established that the intervening host DNA in the latter clone had been deleted. We show (for the first time by hybridization) that in the case of the third clone the DNAs flanking the insertion on either side were linked before integration. The rearrangement caused by RmI's integration is complex but can be partially explained by an inversion event of approximately 20 kbp for an insertion of 7 kbp.

The length and nature of the rearrangements are incompatible with a linear insertion-replacement model, and are more easily explained by an integration mechanism in which the incoming exogenous DNA recombines with a looped structure of chromosomal DNA. Such a model, which takes into account recent findings on chromatin structure, is presented.

For Marie, Marika and Catherine, with much love.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES iv

ABBREVIATIONS v

LIST OF APPENDICES vi

INTRODUCTION 1

Site-specific recombination. 1

Homologous recombination. 3

Illegitimate recombination. 8

Enzymes involved. 14

Previous studies. 16

MATERIALS AND METHODS 22

Cell lines. 22

DNA extraction. 23

Genomic DNA: 23

Plasmid DNA: 24

Restriction enzyme analyses. 24

Restriction enzyme digestion: 24

Electrophoresis: 25

Transfer to membranes: 25

Probes: 25

Hybridization and autoradiography: 27

Cloning of the RmI-Host DNA junctions. 28

Fragment enrichment: 28

Vector preparation: 29

Ligation and in vitro packaging: 30

Subcloning into plasmids. 31
Plasmid preparation: 31
Ligation: 32
Transformation and screening: 32
Other plasmids: 33
Sequencing. 33
12L 33
12R 34
14L 34
14R 34
Sequence comparisons 35
FIGE 35

RESULTS 37
Cloning of the integrated RmI sequences 37
Choice of the integrated RmI molecules. 37
Cloning strategy. 39
Clone W94.10. 41
Clone W98.14. 42
Clone W98.12. 44
Subcloning into plasmid vectors. 45
Initial analysis. 45
Localization of repetitive DNA 47
11.6R 51
14R 51
14L 54
12R 57
Detailed maps of the junctions. 60

12R 60

12L 60

14R 61

14L 61

Preparation of unique sequence probes. 62

Source of the sequences flanking the junctions. 64

Sequences. 67

Rearrangements of the host sequences. 73

W98.12 74

W98.14 77

11.6 85

Fate of the rearranged DNA in 11.6. 87

The special case of 11.6R. 87

DISCUSSION 93

ACKNOWLEDGEMENTS 109

BIBLIOGRAPHY 111

APPENDICES

FIGURES AND TABLES

Figure 1. Model for the integration of linear DNAs.....14

Figure 2. The physical structure of RmI. 17

Figure 3. Comigration of the cloned junctions with the corresponding cellular fragments. 46

Figure 4. The physical maps of the cloned RmI-cellular junctions. 48

Figure 5. Repetitive sequences in p11.6R. 52

Figure 6. Repetitive sequences in p14R and p14L. 55

Figure 7. Repetitive sequences in p12R. 58

Figure 8. RmI integrated directly into host DNA. 65

Figure 9. The sequences across and near the junctions..69

Figure 10. p2.6 contains a rat LINEs (L1Rn) sequence. ..71

Figure 11. Maps of the cellular DNAs rearranged by the integration of RmI. 75

Figure 12. Restriction enzyme analysis of the rearrangement in W98.14. 79

Figure 13. Potential model for the rearrangement in W98.14. 81

Figure 14. Duplication in W98.14 and deletion in 11.6. .84

Figure 15. Rearrangement in 11.6. 86

Figure 16. The special case of 11.6R. 89

Figure 17. Model for the integration of exogenous DNA. 105

Table 1. Lengths of the EcoRI junction fragments. 40

ABBREVIATIONS

M	molar, (moles/litre)
mM	millimolar (10^{-3} M)
TE	10 mM Tris, 1 mM EDTA, pH 7.9
DNA	deoxyribonucleic acid
g	gram
ug	microgram (10^{-6} gram)
ng	nanogram (10^{-9} gram)
ml	millilitre
ul	microlitre (10^{-3} ml)
l	litre
bp	base pair (of DNA)
Kbp	kilobase pair (1000 bp)
ts	temperature sensitive
ccc	covalently closed circular
FIGE	Field Inversion Gel Electrophoresis
Py	polyomavirus
SV40	Simian virus 40
HSV	Herpes simplex virus
TK	thymidine kinase
phenol	a solution of distilled phenol saturated with an aqueous solution of 100mM Tris, 1 mM EDTA.
chloroform	chloroform:isoamyl alcohol (24:1)
SSC	20X SSC is 3 M NaCl, 0.3 M Sodium Acetate
SDS	sodium dodecyl sulfate

APPENDICES

Appendix A: Wallenburg, J.C., A. Nepveu, and P. Chartrand (1984) Random and nonrandom integration of a polyomavirus DNA molecule containing highly repetitive cellular sequences. J. Virol. 50: 678-683

Appendix B: Wallenburg, J.C., A. Nepveu, and P. Chartrand (1987) Integration of a vector containing rodent repetitive elements in the rat genome. Nucleic Acids Res. 15: 7849-7855

Appendix C: Roy, G., J.C. Wallenburg, and P. Chartrand (1988) Inexpensive and simple set-up for FIGE. Nucleic Acids Res. 16: 768

Appendix D: Computer program written in IBM Basic for controlling FIGE gels

INTRODUCTION

Since the advent of DNA mediated gene transfer, (reviewed in Scangos and Ruddle, 1981) the mechanisms for inserting the transferred genes into the genomes of the recipient cells have come under study. These studies reveal that there are three different mechanisms active in the integration of exogenous DNA into the genomes of mammalian cells -- site-specific, illegitimate, and homologous recombination.

Site-specific recombination.

Integration by site specific recombination involves specific DNA sequences or structures on the "donor" or "recipient" molecule or both. In mammalian cells this type of integration seems limited to RNA-mediated transfers of information (reviewed in Weiner _et al_, 1986). The avian and rodent retroviruses which serve as prototypes for the viral superfamily of retroposons, are identified by the presence of LTRs (which have common structural features). They transpose by reverse transcription of an RNA genome or transcript. Insertion is precise to the nucleotide level with respect to viral or transposon DNA, but host sites are random or semi-random (Furano _et al_, 1986; Voliva _et al_, 1984). Short direct repeats of the host sequences flank the element at the site of integration. The lengths of these repeats are specific to the element indicating the

involvement of viral- or element-encoded enzymes. Not all transposing viruses or elements are capable of coding for their own enzymes however, indicating that defective members of the family can "borrow" enzymes supplied _in trans_ by non-defective members.

The members of the nonviral superfamily of retroposons are partial or complete DNA copies of processed cellular RNAs. These include the long interspersed repetitive sequences (LINEs e.g.:L1Md, L1Rn) and the short interspersed repetitive sequences (SINEs) such as the Alu and Alu-type elements (e.g. rodent B1) and the B2, MT and ID repeats in rodents. (Note that the previously identified _KpnI_, Bam 500, and RI families are segments of LINEs). Insertion generally generates a duplication of target sequences but these duplications are not characteristic of the element, nor are there any consistent structural similarities. This leads to the supposition that these elements (with the possible exception of LINEs) transpose passively- i.e. that they do not code for the enzymes involved.

SINEs are RNA polymerase III products present on the order of 10^5 copies per genome. SINEs have well-defined 5′- and 3′-ends and create target duplications from 7-21 bp long. Significantly they all contain highly conserved "internal" RNA polymerase III promoters suggesting that each transposition generates another potentially active

copy. The degree of homology within the Alu and Alu-like (e.g. B1) families is ~85%, (Weiner et al, 1986; Bandyopadhyay et al, 1984) whereas that of the tRNA-derived families (such as B2 and ID in rat) is ~70%. (The latter families may be subject to a different set of constraints-see Weiner et al, 1986, for a discussion)

Homologous recombination.

Integration in eucaryotic cells by homologous recombination was first reported in yeast (Hinnen, Hicks and Fink, 1978). Plasmids carrying yeast genes integrate into the endogenous copy by homologous recombination in transformed cells (Orr-Weaver et al, 1981). The resulting structure has the plasmid sequences integrated at the gene locus flanked by recombinant/chimaeric copies of the gene in direct repeat.

Subsequent studies have shown that similar events occur in *Aspergillus nidulans* (Miller et al, 1985), *Dictyostelium discoideum* (DeLozanne and Spudich, 1987; Witke et al, 1987) and also in mammalian cells in culture although at much reduced frequencies (see below).

Over the last few years a number of assays have been developed to study recombination in mammalian cells. Most of these assays involved the reconstruction of some selectable marker by either intra- or intermolecular

homologous recombination between overlapping gene segments. These studies have demonstrated efficiencies of (i) intermolecular homologous recombination between cointroduced extrachromosomal molecules ranging from .11% to 50% (Ayares et al, 1986; Bandyopadhyay et al, 1984; Dorsett et al, 1983; Folger et al, 1982; Folger et al, 1984; Lin et al, 1984; Pomerantz et al, 1983; Rubnitz and Subramani, 1984; Small and Scangos, 1983; Smithies et al, 1984; Subramani and Berg, 1983), (ii) intramolecular homologous recombination within extrachromosomal molecules from .004% to 13.4% (Kirchhausen et al, 1985; Rubnitz and Subramani, 1984; Rubnitz and Subramani, 1985; Rubnitz and Subramani, 1986; Subramani and Berg, 1983), and (iii) intramolecular homologous recombination between sequences previously introduced and integrated into the chromosome at a frequency of 10^{-4} to 10^{-8} per generation (Lin et al,1984; Liskay and Stachelek, 1983; Liskay et al,1984; Rubnitz and Subramani, 1986; Smith and Berg, 1984; Subramani and Rubnitz, 1985). Efficiency of homologous recombination was found to be dependent on the length of the overlapping homology (Bandyopadhyay et al, 1984; Kirchhausen et al, 1985; Lin et al, 1984; Rubnitz and Subramani, 1984; Shapira et al, 1983; Subramani and Berg, 1983). A double-strand break within or just flanking the homology increased the rate of recombination from 10 to 100-fold (Lin et al, 1984; Kirchhausen et al, 1985; Rubnitz and Subramani, 1985; Smithies et al, 1984).

The studies just described proved that mammalian cells in culture are highly efficient in homologous recombination, that all the enzymatic machinery necessary is supplied by the host and that the method used to introduce the DNA (infection, calcium phosphate precipitation, DEAE transfection, microinjection, and protoplast fusion) is not a limiting factor. They furthermore showed that homologous recombination could proceed by reciprocal recombination, (Liskay et al, 1984; Small and Scangos, 1983; Smith and Berg, 1984, Smithies et al, 1984, Lin et al, 1985, and Subramani and Rubnitz, 1985) non-reciprocal recombination (Brouillette and Chartrand, 1987, Folger et al, 1984, and Smithies et al, 1984) and gene conversion (Folger et al, 1984, Smith and Berg, 1984, Smithies et al, 1984, Lin et al, 1985, and Subramani and Rubnitz, 1985). Finally there have been a number of reports in which endogenous chromosomal sequences have been "rescued" (presumably via gene conversion) by defective extrachromosomal vectors leading to the correction of the defective functions (Bandyopadhyay et al, 1984, Jasin et al, 1985; Schwartzberg et al, 1985; Shaul et al, 1985). Thus mammalian cells in culture have been shown to be highly proficient in homologous recombination and the chromosomal sequences shown to be available targets.

Notwithstanding the above, early attempts to detect homologous integration without selection were unsuccessful (Graessmann et al, 1979; Steele et al, 1984). Botchan and

coworkers (1979) used SV40 to retransform a revertant cell line already containing an integrated copy of SV40 but were unable to detect homologous integration. In another attempt Steele *et al* (1984) transfected cultured mouse cells with a plasmid in which the HSV-tk gene was surrounded by rDNA sequences. They hypothesized that the 200 copy-number of the endogenous rDNA sequences could compensate for the 200 times greater size of the mammalian genome when compared to yeast. No homologous integration was detected.

More recently a number of groups have used re-construction of a selectable marker to detect homologous integration (Lin *et al*, 1984, Smith and Berg, 1984a, Smithies *et al*, 1984, Thomas *et al*, 1986). The principle of these assays was to first establish cell lines in which a defective selectable marker had been stably integrated into the chromosome. Subsequent transfection of these cell lines with a complementary defective marker allowed selection of clones that were positive for the reconstructed selected marker. Southern analysis of the DNAs of these clones demonstrated that they contained a "diagnostic" recombinant fragment which corresponded to a "repaired" copy of the marker gene (Lin *et al*, 1984, Smithies *et al*, 1984, Thomas *et al*, 1986) and that in at least some of the isolates the gene had been repaired by homologous integration (Lin *et al*, 1984, Thomas *et al*, 1986). Importantly, not all of the established cell lines

were productive in these assays. Lin et al (1984) detected gene targeting in only one of ten transfected cell lines whereas Smithies et al (1984) detected it in two of four transfected cell lines. Thomas et al (1986) did not observe this phenomenon in similar experiments when the DNA was introduced by microinjection. They furthermore claim that gene targeting is not sensitive to target size, based on the observation that targeting frequency was equal in lines containing only 1 or 4-5 targets.

Gene targeting of an endogenous host gene without selection for the modification was successfully demonstrated for the first time in 1985 by Smithies et al. Using an elaborate scheme involving the "rescue" of plasmid sequences they succeeded in demonstrating the planned modification of the human B-globin locus by homologous integration in 10^{-3} successfully transfected cells whether or not the locus was expressed in those cells (Ibid and O. Smithies, personal communication).

In 1987 two groups (Doetschman et al, 1987; Thomas and Capecchi, 1987) demonstrated homologous integration into a natural endogenous gene with selection. Using mouse embryonic stem (ES) cells both groups targeted the HPRT gene. Thomas and Capecchi (1987) mutated the host gene by interrupting the coding sequences by targeted integration of a neo gene, whereas Doetschman et al (1987) corrected a naturally occurring deletion by targeted integration of a

human/mouse construct. Using suitable media to select for the desired phenotype both groups were able to detect targeted integration by homologous recombination at frequencies of ~10^{-6} (10^{-3} successfully transfected cells).

Finally, Mansour et al (1988) demonstrated homologous integration into an endogenous gene without selection for that gene. Using what they called positive-negative selection (PNS) they succeeded in selecting for the integration of the transfected DNA but against those integration events which were non-homologous (illegitimate). Using this system they were able to enrich 2,000 fold for homologous integration (targeting) events. The frequency of targeting (which was not itself modified) was nonetheless very low, ranging from 2.5×10^{-4} to 2.5×10^{-5} successfully transfected cells.

Illegitimate recombination.

Recombination is said to be illegitimate when the interacting sequences neither share significant homology nor show any specific structure or sequence. Since virtually any exogenous DNA can be integrated into the genomes of mammalian cells in culture, be it prokaryotic or eukaryotic, the functions involved are surely cellular. Although the technique used to introduce the DNA into the cells influences the efficiency of stable transformation, it is generally accepted that any mode of DNA introduction

can lead to integration by illegitimate recombination. A number of studies on the fate of the DNA following transfection with the calcium phosphate precipitation technique (Ruddle and Scangos, 1981; Scangos et al, 1981; Loyter et al, 1982; and Perucho et al, 1980) followed the progress of the transfected markers to stable integration. They demonstrated that only a small proportion of the transfected DNA was ever found in the nucleus, and that the selected marker was initially found to be associated with carrier DNA in a high molecular weight molecule termed transgenome (Scangos and Ruddle, 1981) or pekelosome (Perucho et al, 1980). The association between carrier and marker DNA is potentially formed by end to end ligation (Scangos and Ruddle, 1981) or illegitimate recombination (Brenner et al, 1984). Within the transgenomes the selected markers are frequently organized in head to tail tandems, formed by homologous recombination (Brenner et al 1984, Pomerantz et al, 1983). The transgenomes are unstably maintained within the cells or eventually integrated into the chromosome resulting in a stable transformant (Scangos et al, 1981). The analysis of the DNA flanking the markers in the stable transformants showed that it was carrier DNA (Perucho et al 1980).

The integration of the papovaviruses SV40 and Py has served as a paradigm for integration by illegitimate recombination. Initial studies using restriction enzyme mapping showed that the integration sites of SV40 (Botchan

et al 1976, Ketner and Kelly 1976,) and Py (Birg et al 1979) were not specific with respect to the viral nor the cellular DNA. The integrated genomes were shown to be associated with a number of different chromosomes (Croce 1977; Folger et al, 1982; Croce et al, 1973; Hwang and Kucherlapati 1980; Kucherlapati et al 1978). Finally, the sequences across a number of viral-cellular junctions (Botchan et al, 1980; Hayday et al, 1982; Ruley et al, 1982; Ruley and Fried, 1983; Stringer, 1981; Stringer, 1982) have not revealed the consistent involvement of any specific sequences or structural features. Some of these analyses detected limited amounts of "patchy" homology between the involved molecules in the vicinity of the crossing-over (Botchan et al, 1980; Ruley et al, 1982; Stringer, 1981; Stringer, 1982) and it has been proposed that this "patchy" homology may play a role in the stabilization of the recombination intermediate (Marvo et al 1983). The statistical significance of this "patchy" homology, however, has been questioned (Smith et al, 1985; McLachlan and Boswell, 1985; Savageau et al, 1983).

The role that the state (i.e. linear, ccc or replicating) of the exogenous DNA might play in the integration process has yet to be clarified. Certain studies have shown linear molecules to transform more efficiently than circular molecules following microinjection (Folger et al, 1984) whereas others have found no difference when using the calcium phosphate

precipitation technique (Gusew et al, 1987; Steele et al, 1984) including one extensive study (Graham et al, 1980). Folger et al (1982) compared the efficiency of transformation as well as the organization of the integrated DNA following microinjection of plasmid DNA directly into the nucleus of mouse L cells. They observed that at low copy numbers (<50 molecules/ nucleus) linear molecules transformed better than ccc molecules but that when >50 molecules per nucleus were injected the efficiencies were equivalent. Furthermore, both linear and ccc molecules formed head to tail tandems by homologous recombination but linearized plasmids integrated via their free ends. The suggestion that free ends are recombinogenic in homologous recombination has already been discussed (see above). In 1987 Gusew et al looked more closely at the involvement of free ends in illegitimate recombination and found that blocking the ends of linearized molecules with proteins (restriction enzymes in this case) dramatically reduced the efficiency of integration. These results support the hypothesis that linear molecules integrate via their free ends. They further observed that unprotected molecules were highly degraded and found a direct correlation between the extent of degradation and the ability to integrate, suggesting that exonucleolytic degradation of ends may generate a substrate favourable for integration.

My own studies (Wallenburg, 1984 and Wallenburg et al,

1984) and those of Gusew et al (1987) further suggest that circular molcules do not integrate via the same pathway as linear molecules and that replicating molecules probably behave like linear ones. Calcium phosphate precipitate mediated transformation by non-replicating circular molecules is just as efficient as that of linearized "unprotected" molecules (Graham et al, 1980; Gusew et al, 1987; Steele et al, 1984; Wallenburg, 1984; and Wallenburg et al, 1984). However in direct contrast to the results on the integration of linear molecules where high efficiency is associated with greater degradation, the integrated sequences of non-replicating circular substrates have suffered little loss due to degradation, and tend toward unit length (Gusew et al, 1987; Wallenburg, 1984; and Wallenburg et al, 1984). This strongly suggests that circular molecules do not pass via a linear intermediate before integration. In support of this hypothesis Konopka (1988) recently compiled and examined 496 non-homologous (illegitimate) recombination regions. Ninety-eight of the parental sequences compiled were the result of recombination events that occurred during transfection of linear DNAs (TLD). Alternating repeats of purines and pyrimidines are significantly overrepresented in the other 398 illegitimate recombination regions but not in TLD. He suggests that this difference reflects a difference in the integration pathways and concludes that more than one pathway for illegitimate recombination exists in somatic cells.

My previous studies, (Wallenburg, 1984 and Wallenburg et al, 1984) and those of Gusew et al (1987), examined the integration of a recombinant Py mutant (which is thermosensitive for replication) at the permissive and non-permissive temperatures. The differences seen at the different temperatures are interpreted as being the result of the capacity of the mutant to initiate replication at the lower, permissive temperature. From these studies it was determined that circular molecules that can replicate integrate (transform) as efficiently as linear and non-replicating circular molecules. The pattern of the integrated structures is similar to that of linear molecules in that there is significant degradation of the integrating molecule in both cases (Wallenburg, 1984; and Wallenburg et al, 1984). However, whereas sequence loss from linear molecules is centered around the site of linearization, the sequence loss from circular replicating molecules is random and distributed equally throughout the genome (Wallenburg, 1984; and Wallenburg et al, 1984). These results are consistent with the idea that circular replicating molecules integrate via a mechanism similar to that of linear ones, and suggest that the replicating molecule is linearized at random sites.

In order to explain the variable subgenomic lengths of the insertions, initial models for the integration of papovavirus genomes proposed the integration of linear (rather than circular) molecules via double crossings-over (Figure 1).

```
             A  B  C  D  E  F  G  H
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
                X           X
               _______________


             A  B              G  H
~~~~~~~~~~~~~~~~_____________~~~~~~~~~~~~~~~~~~~~
```

Figure 1.

Model for the integration of linear DNAs into the host chromosome. Wavy line (~~~~) represents chromosomal DNA. Thin line (——) represents the linear exogenous DNA. Xs indicate the sites of crossing-over. (Redrawn from Chia and Rigby, 1981)

This would lead to a deletion of cellular sequences of similar length to the length of the sequences inserted (Topp *et al* in Tooze 1981, Chia and Rigby, 1981). Analyses of the genomic DNA surrounding integrated molecules have in fact frequently revealed rearrangements in the genomic sequences (Botchan *et al*, 1980; Hayday *et al*, 1982; Neer *et al*, 1983; Stringer, 1982). However, except for one study (Neer *et al*, 1983), the nature of these rearrangements (deletion, translocation etc.) has not been determined. Based on the similarity of the restriction enzyme patterns of the DNA flanking the Py insertion and unrearranged DNA in the host cell line, Neer *et al* (1983) determined that a fragment of ~ 3.1 Kbp of host DNA was deleted at the time of integration. In this case the length of the deletion was shorter than the length of the inserted DNA (5.9 Kbp).

Enzymes involved.

Little is known about the functions mediating the homologous and illegitimate recombination events. Presumably the cell is capable of encoding all of the required enzymes since no essential functions have been found to be necessarily associated with donor molecules.

It has been proposed that cellular topoisomerases may play an important role in illegitimate recombination (Been and Champoux 1981; Bullock et al, 1984; Bullock et al, 1985; Champoux et al, 1984; Halligan et al 1982; Konopka, 1988) and in fact, in a recent study in which he compiled and examined 496 previously published illegitimate recombination regions, Konopka (1988) found topoisomerase I trinucleotide recognition sequences at or near the cross-over sites in over 92% of them. Furthermore, both topoisomerases I and II have been found intimately associated with SV40 and Py chromatin (Krauss et al 1984, Waldeck et al 1984).

The "fixation" of the transformed state, (presumably by integration into the genome (Scangos et al, 1981)) requires that the target cells go through at least one round of replication (Topp et al in Tooze, 1981). Presumably, the replicative form of DNA, with its nicks and single-strand gaps, is a more accessible substrate for integration. However, as it seems unlikely that eukaryotic

cells would have developed a set of enzymes intended to integrate exogenous DNA, the enzymes involved in the illegitimate integration process are probably normally involved in other cell functions, such as replication, and so integration would be limited to those periods where the required enzymes are available. Topoisomerase II has in fact been shown to be a specific and sensitive marker for cell proliferation being rapidly lost as cells progress from mitosis into G_1 and upon the cessation of mitotic activity of some cells (Heck and Earnshaw, 1986; Heck, Hittelman and Earnshaw, 1988). In support of a role for topoisomerase II in illegitimate integration Bae et al (1988) have shown that the eukaryotic enzyme can catalyze illegitimate recombination in vitro and suggest a subunit exchange model as a potential mechanism.

On the basis of experiments using Py tsa mutants in which the activity of Py large T antigen was shown to increase the number of integration sites and number of head to tail tandems, it has been suggested that large T is at least indirectly involved in the integration process of Py (Dailey et al, 1984; Della Valle et al, 1981). The authors interpreted its effect as being the result of its role in replication.

Previous studies.

In order to examine the different recombination

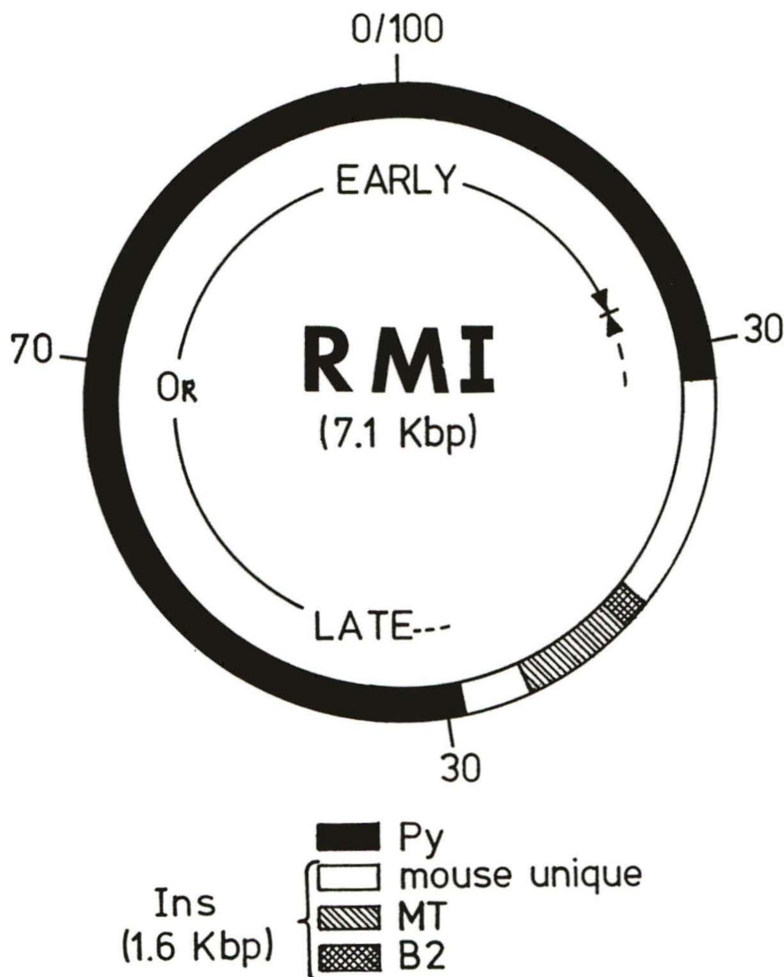


Figure 2.

The physical structure of RmI as previously determined (Bourgaux et al, 1982). The numbers refer to standard polyoma map units and OR identifies the polyoma origin of replication. The early and late coding regions have been indicated with the late region interrupted by Ins. Ins is an insertion of mouse cellular DNA derived from the cellular DNA flanking the Py insertion in the host cell line, (Cyp clone C12a1; Bourgaux et al, 1982). B2 and MT refer to the B2 (Sylla et al, 1984) and MT (Bastien and Bourgaux, 1987) families of rodent repetitive DNA. Throughout this thesis base positions are numbered sequentially from 5′ to 3′ on the early gene coding strand. Base #1 is the first base of the Py genome (A2 strain) as described in "DNA Tumour Viruses", (1981, J. Tooze, editor). Thus EcoRI cuts after base 1560 (this also corresponds to map unit 0/100); INS covers bases 3274-4901; and the B2 and MT sequences are from 4127-4245 and 4282-4655 respectively.

mechanisms involved in integration I had previously established a number of cell lines resulting from the transformation of FR3T3 rat cells with a modified Py vector, RmI (Figure 2).

RmI is a naturally occurring Py recombinant produced upon the induction of C12a1, a mouse cell line carrying integrated sequences from tsP155, a Py mutant (Sylla et al 1980). Mutant tsP155 (and therefore RmI) is thermo-sensitive for replication but is not affected in its transforming ability (Eckhart, 1974). RmI is a circular molecule composed of 1.03 copies of the tsP155 genome (including an uninterrupted early region and origin of replication) with 1,628 bp of mouse genomic DNA inserted into the coding region of the late viral genes. This insertion (Ins) is flanked by a direct repeat of Py DNA of 182 bp, and contains at its boundaries an imperfect 7 bp inverted repeat of cellular DNA (Bourgaux et al, 1982). Ins begins with the nucleotides TG and ends with AC. These nucleotides are the first and last of retrovirus genomes as well as repetitive elements of Drosophila and yeast capable of site-specific recombination (Jelinek and Schmid, 1982). Ins contains two partial mouse SINEs elements, 120 base pairs of B2 (Sylla et al, 1984) and 374 base pairs of MT (Bastien and Bourgaux, 1987), both of which are highly dispersed and repeated on the order of 10^5 copies per genome.

The RmI molecule should have the potential to integrate by any or all three of illegitimate, homologous, or site-specific recombination mechanisms. Firstly, the viral Py sequences are known to integrate by illegitimate recombination. Secondly, the B2 and MT repetitive sequences both have 100,000 homologous copies dispersed throughout the rat genome any one of which could serve as a target for homologous integration. Finally the junctions between Ins and the Py sequences of RmI (RmI Py/Ins junctions) display features (flanking direct and inverted repeats) which are in common with eukaryotic transposable elements (such as copia elements), retroviruses and bacterial transposon cointegrates and suggest that RmI may be capable of site-specific recombination. Indeed RmI excision is most probably the result of site-specific recombination (Bourgaux-Ramoisy et al, 1986; Huberdeau et al, 1985; Sylla et al, 1984). Furthermore, Piché and Bourgaux (1987a and b) have recently presented results that strongly suggest that tsP155 can excise from RmI via a site-specific mechanism in which Py large T antigen plays a direct role.

My previous work on the integration of RmI into the genomes of rat cells (Appendix A) showed that by shifting the temperature down to 33°C for 20 hrs. immediately after transfection (W3 series of clones) then the pattern of integration produced is random with respect to both the length of the integrated sequences and the distribution of

the crossings-over within the RmI sequences. Because of the large number of individual clones examined, these results gave statistical support to the hypothesis that the integration of papovaviruses is mediated by a random process. However, when the transfections were maintained at the nonpermissive temperature 39°C (W9 clones) then integration was non random with respect to the sequences of RmI. One region of RmI (between Py m.u. 40 and 70) was underrepresented, whereas two regions, from m.u. 30 to 40 and the region of Ins containing the repetitive elements B2 and MT, were significantly overrepresented. Furthermore, the lengths of the integrated genomes tended importantly toward the unit length of RmI contrary to what was observed at 33°C. In other words, whereas the integrated molecules in the W3 clones had suffered substantial degradation (up to 80% of the original material), the integrated molecules in the W9 clones had, in general, suffered little sequence loss (< 20%). This raised the possibility that the observed differences were the result of different integration mechanisms. The previously described model for the integration of papovaviruses in which a linear molecule integrates via two crossings-over (Figure 1) could explain the insertions in the W3 clones. Unit-length insertions are more readilly explained by the integration of a circular molecule by a single cross-over. In the case of the W9 clones it would nonetheless be necessary to postulate that integration was followed by rearrangements since the integrated sequences have two distinct endpoints.

These were the first studies on the integration of a vector containing 1) a very large number of homologous targets in the genome and 2) SINEs sequences. The only studies to date addressing the question of target size in homologous integration are those of Botchan *et al*, Steele *et al* (1984) and Thomas *et al* (1986) which were discussed previously. Studies using target sizes several orders of magnitude greater than unique copy genes have never been done. Furthermore, SINEs are not only present in a large number of copies (10^5) dispersed throughout the genome but Rogers (1985) and others have suggested that special mechanisms (such as gene conversion) may be involved in maintaining sequence homogeneity. SINEs also contain internal RNA polymerase III promoters which may be active in the dispersal of these sequences throughout the genome. Gene transcription has recently been invoked to explain the recombination stimulating activity of HOT 1 in yeast (Voelkel-Meiman *et al*, 1987) and has been suggested to stimulate genetic exchange in a number of systems (Klar *et al*, 1981; Blackwell *et al*, 1986). How might these activities affect the recombination potential of the vector, RmI?

Due to the particular qualities of RmI and the non-random pattern of integration produced by transfection at 39°C it was of interest to further define the integration mechanism and in particular to determine the nature of the viral-cellular junctions at the molecular level as well as to characterize the host site before and after integration.

MATERIALS AND METHODS

Cell lines.

Cell line LTA (Graham *et al* 1980) was a generous gift from S. Bacchetti. Fisher rat 3T3 (FR3T3) cells (Seif and Cuzin, 1977), and clones W94.10, W98.12, W98.14 and the procedures used to isolate them have been described elsewhere (Wallenburg, 1984; Wallenburg *et al*, 1984). The isolation of clone 11.6 by A. Nepveu is described in his thesis (Nepveu, 1984). Briefly, FR3T3 cells (at $39^{o}C$) were transfected with RmI using the calcium phosphate technique (Graham *et al* 1980) and either mouse LTA DNA (W94.10, W98.12, and W98.14) or salmon sperm DNA (11.6) as carrier. The cells were split 48 hours post-transfection and reseeded at 5×10^{4} cells per 60 mm dish. One to two weeks after transfection isolated foci of transformed cells were picked. These clones were purified by at least one more round of colony isolation before analysis. Culture temperature at all times was maintained at $39^{o}C$.

Cells were grown in Dulbecco's modified minimum essential medium (DMEM, supplied by Flow, Gibco) supplemented with L-glutamine (Sigma, 5,8 x 10^{-6} gm/ml), 10% fetal bovine serum (Flow and Bockneck), and gentamycin (U.S.B., 55 ug/ml) at $39^{o}C$ in an atmosphere containing 5% CO_2.

DNA extraction.

Genomic DNA:

Cells were grown to semi-confluence (about 80%) in 100mm dishes and the DNA extracted by one of two methods; 1. using proteinase K digestion followed by phenol/TE, phenol/chloroform and chloroform extractions as previously described (Wallenburg, 1984) or 2. using guanidine HCl (G-HCl) as described by Jeanpierre (1987) with minor modifications. Briefly, monolayers of cells were washed with phosphate buffered saline (PBS) and then lysed in a solution containing 6 M G-HCl (Fisher Biotech), 0.5 M ammonium acetate (NH_4OAc), 1% sarkosyl, and .2 ug/ml proteinase K (made up fresh from stock solutions of 7.5 M G-HCl, 5 M NH_4OAc, 10% sarkosyl and proteinase K). The lysate was collected with rubber policemen and incubated at 65°C for 1 to 2 hours. Two to 2.5 volumes of ethanol (95%) were layered on top and the DNA precipitated by gentle inversion at room temperature or by spooling onto glass rods. The precipitated DNA was collected and washed sequentially in 70%, 80%, 90%, and 100% ethanol to remove the water and salt, <u>briefly</u> air dried and dissolved in TE to a final concentration of 0.5 to 1.0 mg DNA/ml by incubation at 65°C overnight with gentle tube inversion. DNA prepared in this way has an average molecular weight of 100-200 Kbp and is readily digested.

Salmon sperm DNA was purchased from Sigma and dissolved into TE at 5 mg/ml.

RmI was prepared as previously described (Bourgaux _et al_, 1982).

Plasmid DNA:

Mini-preparations were prepared using the alkaline lysis technique as described in Maniatis _et_ _al_ (1982). Large preparations of plasmid DNA were also prepared using the alkaline lysis technique as described in Maniatis _et_ _al_ (1982) but were further purified by cesium chloride density gradient centrifugation (Maniatis _et_ _al_, 1982).

Restriction enzyme analyses.

Restriction enzyme digestion:

Plasmid and cellular DNAs were digested with restriction enzymes (Amersham, BRL, Boehringer Mannheim, New England Biolabs, and Pharmacia) following the recommendations of the manufacturer and using a 5 time excess of enzyme. For Southern analysis, 5 ug of cellular DNA in solution were digested in a volume small enough (20-30 ul) to be loaded directly onto the gel following addition of a 10X stop/loading buffer (20% Ficoll, 250 mM EDTA pH 8.0, 1.0% sarkosyl, 0.4% bromophenol blue, 0.4%

xylene cyanol). Samples were heated to 56°C for 10 minutes before loading.

Electrophoresis:

Unique field electrophoresis of plasmid and cellular DNAs was done in agarose gels ranging from 0.3 to 1.5% agarose (BRL, Pharmacia, FMC) in 1X TAE (40 mM Tris acetate, 2 mM EDTA, pH 8.4) buffer at field strengths of 1 v/cm for 16 to 20 hours or of about 5 v/cm for 1-4 hours. Very small DNA fragments (less than 1000 bp in length) were separated on high concentration (3-4%) Nuseive agarose (FMC) gels.

Transfer to membranes:

Following electrophoresis, plasmid or genomic DNAs were transferred to nylon membranes (Hybond N, Amersham or Gene Screen Plus, Dupont/NEN) by the methods of Southern (1975) or Reed and Mann (1985). Plasmid DNAs were blotted bidirectionally onto membranes using the technique of Southern (1975) except that the gel was "sandwiched" between two membranes and no buffer reservoir used.

Probes:

In order to prepare fragments for use as probes, plasmid DNA was digested using the appropriate restriction

enzyme (RE) cleaned by one extraction each with phenol/chloroform and chloroform:isoamyl alcohol(24:1), precipitated with 2 volumes ethanol, dried, redissolved in TE and electrophoresed on agarose gels. The desired fragment was cut from the gel and extracted by one of a number of techniques; 1. electroelution into a 7.5 M ammonium acetate "cushion" using an IBI electroeluter and following the directions of the manufacturer, 2. migration onto DEAE-cellulose paper (DE 81 paper, Whatman) and reelution (Dretzen _et al_, 1981), 3. adsorption onto glass powder (Vogelstein and Gillespie, 1979; and Gene Clean, from Bio 101, used as directed), or 4. by passage on NACS columns (BRL) as directed by the manufacturer. When the fragment was intended for use in the random hexamer primed synthesis protocol it was occasionally left in the low melting point (LMP) agarose in which it was migrated as suggested by Feinberg and Vogelstein (1984), although probes made from DNA fragments in LMP agarose gave highly variable results.

Probes were prepared using the nick translation procedure of Rigby _et al_ (1977) or using random hexamer primed synthesis (Feinberg and Vogelstein, 1983) kits (Amersham, and Pharmacia).

In order to screen for repetitive DNA total FR3T3 DNA was first sheared by multiple passages through a 27g needle, then labeled by nick translation.

Hybridization and autoradiography:

All hybridizations were performed at 68°C. Gene Screen Plus membranes were prehybridized, hybridized and washed following the recommendations of the manufacturer. Based on the procedure described in Singh and Jones (1984) (with minor modifications) Hybond membranes were prehybridized in 4X SET (.4M NaCl, .12M Tris pH7.4, 8mM EDTA), .1% sodium pyrophosphate, .5% SDS, and 50 ug/ml heparin, and hybridized in the same solution using probe concentrations. Blots of genomic DNA were hybridized in the same solution except that dextran sulphate (Pharmacia) was added to 10% and the heparin concentration increased to 500 ug/ml. Washes were 2 times 15 minutes in 1 X SSC, .1% SDS at 68°C, 2 times 20 minutes in 0.3 X SSC, 0.1% SDS at 68°C, and 2 times 5 minutes in 2 X SSC at room temperature.

Membranes used to detect repetitive DNA were hybridized using standard conditions, however post-hybridization washes were limited initially to 3 times 30 minutes in 2 X SSC, 0.1% SDS at 65°C. The membranes were autoradiographed wet, then washed 3 times 30 minutes in 1 X SSC, 0.1% SDS at 65°C, reexposed, and finally washed 3 times 30 minutes in 0.3 X SSC, 0.1% SDS at 65°C and autoradiographed.

Membranes were autoradiographed between two intensifying screens (Dupont Lightning Plus or Kyokko Super-HS)

using pre-flashed XAR-5 (Kodak) or XRP (Kodak, Fuji) autoradiography films at $-70^{o}C$.

Relative copy numbers of DNA fragments from Southern analyses were determined by densitometry scanning of autoradiograms using an LKB Ultroscan XL scanning densitometer and accompanying Gelscan XL software.

Cloning of the RmI-Host DNA junctions.

Fragment enrichment:

100 ug of cellular DNA were digested with EcoRI and loaded onto 2 preformed sucrose (10-40% w/v) gradients and centrifuged (Maniatis *et al*, 1982). Fractions of 0.5ml were collected, and 15-20 ul aliquots loaded onto agarose gels, electrophoresed and analysed by Southern blotting to determine the distribution of the junction fragments. The fractions enriched for junction fragments were extensively dialysed against TE and the DNA recovered by ethanol precipitation, or alternatively, the enriched fractions were diluted with 2 volumes of water, precipitated with 2 volumes of ethanol, resuspended into 0.1 volumes of 30% sodium acetate, ethanol precipitated again and washed in 70% ethanol. The recovered DNA pellets were dried and resuspended into a minimum volume of TE.

Vector preparation:

λgtWES.λB (Leder *et al* 1977) was grown on *E.coli* strain LA 101 in NZCYM medium and the phage collected and DNA prepared as described in Maniatis *et al* (1982). *EcoRI* arms linked by their cohesive ends, and purified on preformed sucrose gradients as in Maniatis *et al* (1982) were used to generate partial libraries with junction-fragment enriched W98.14/*EcoRI* and W94.10/*EcoRI* DNAs.

The λgtWES. λ B *EcoRI* arms used to ligate junction-fragment enriched W98.12/*EcoRI* DNA had not been purified away from the *EcoRI* insertion but rather the insertion had been further digested by *SstI* which cuts the insertion twice more. Following the *SstI* digestion the arms were cleaned by phenol extraction, ethanol precipitated and used directly in the ligation. In order to be able to test their cloning efficiency, pBR322 linearized at its unique *EcoRI* site (pBR322/RI) was cloned into these gtWES. B *EcoRI* arms. The arms alone gave a titer of 3.5 x 10^7, the pBR322 ligation gave a titer of 3.0 x 10^6/ml. When the pBR322/RI library was hybridized with a pBR322 probe, 82% of the plaques hybridized. This shows that although the background was high when no DNA with two *EcoRI* ends was present this type of DNA competed effectively for the arms and gave good percentages of recombinant phages.

Ligation and *in vitro* packaging:

EcoRI digested cellular DNAs enriched for the junction fragments were ligated into λgtWES.λB *EcoRI* arms overnight at 12°C in a volume of 5 ul at a final concentration of 180-230 ng/ul. The vector to insert ratio was varied (4:1, 2:1, 1:1, 1:2) where possible.

Sonic extracts (SE) for packaging were prepared from bacterial strain *E.coli* 154; freeze-thaw lysates (LFT) were prepared from *E.coli* strain 131. Packaging extracts were prepared according to protocol II in Maniatis *et al* (1982) except that SEs and LFTs were not mixed before freezing.

The *in vitro* packaging reaction was performed as in Maniatis *et al* (1982) with minor modifications. The reactions contained:

(ul)		
5	DNA	(~1 ug total, ligated vector plus insert)
14	buffer A	(20 mM Tris pH8.0, 20mM $MgCl_2$, 1mM EDTA, 10 mM 2-mercaptoethanol)
4	M1 buffer	(prepared freshly; 6 mM Tris pH7.5, 18 mM $MgCl_2$, 30 mM spermidine HCl, 60 mM putrescine HCl, 15 mM ATP, 0.2% 2-mercaptoethanol)
12	SE	
30	LFT	(or 24 ul, the quantity of LFT required was determined empirically for each batch)

The reactions were incubated at room temperature for 60 minutes, a second volume of LFT was added and the reactions incubated another 60 minutes at room temperature. One ml of SM (20 mM Tris pH 7.5, 0.1 M NaCl, 1 mM $MgSO_4$, 0.01% gelatin) was added per tube and the number of phages recovered was titrated on E.coli strain LA101 on NZCYM (Maniatis et al, 1982) agar plates.

The partial libraries thus prepared were screened for the presence of junction containing recombinants using the technique of Benton and Davis (1977) and a Py probe. Positive clones were picked and purified to homogeneity as determined by hybridization to the Py probe. The DNAs of these phage preparations were extracted, digested with EcoRI and tested for the presence of Py containing fragments of the appropriate length by hybridization.

Subcloning into plasmids.

Plasmid preparation:

13 ug of pBR322 DNA was digested with EcoRI, and the 3′ ends dephosphorylated according to Maniatis et al (1982). DNA prepared in this way, then religated, transformed bacteria 3 orders of magnitude less efficiently than circular pBR322. The same 3′-dephosphorylated pBR322/EcoRI (pBR322/RI-OH) preparation was used for plasmid subcloning of all the junction fragments.

Ligation:

Junction containing fragments from purified phages were subcloned onto pBR322/RI-OH by "shotgun" ligation. Total phage DNA was digested with EcoRI and ligated overnight to pBR322/RI-OH at $4^{o}C$ in the appropriate buffer at a final DNA concentration of 40-90 ng/ul and a ratio of phage:plasmid DNA of 1:1.

Transformation and screening:

Ligated DNA was used to transform *E.coli* strain DH-1 by the technique of Hanahan (1982). Transformed bacteria were plated onto LB agar plates containing ampicillin (Boehringer Mannheim, 35 mg/ml). Junction-fragment containing colonies were identified using a Py probe and the technique of Grunstein and Hogness (1975) except that duplicate agar plates were not prepared. Rather, the original colony bearing agar plate was blotted directly onto Hybond-N membranes (Amersham), then incubated overnight at room temperature to allow the transferred colonies to regrow before being transferred to $4^{o}C$ for storage. Positive colonies were picked, grown in liquid media (LB, Maniatis *et al*, 1982) and rapidly screened using plasmid minipreparations and restriction enzyme analysis.

Other plasmids:

The right junction of 11.6 was subcloned into pBR322/RI from the phage vector using the same general protocol as described above (N. Gusew, unpublished data).

Plasmid p2.6 contains the 2.6 Kbp EcoRI cellular DNA fragment (IS) that is the original arrangement of the cellular sequences at the left junction of the RmI insertion in clone 11.6, before integration (Figure 4), cloned into pBR325 linearized at its unique EcoRI site (A. Nepveu, 1984). Plasmid p6.4 contains the 6.4 Kbp of celluar DNA which immediately flanks IS to the right (A. Nepveu, 1984; and Figure 4).

Sequencing.

Fragments were purified for sequencing by RE digestion, phenol extraction, ethanol precipitation, agarose gel electrophoresis, and extraction from the gel by migration onto DEAE-cellulose paper or passage on NACS (BRL) columns as described previously in the probe section. End labelled DNA was sequenced using the chemical cleavage method of Maxam and Gilbert (1977 and 1980).

12L

p12L was digested with EcoRI, labelled with E.coli DNA

polymerase I (Klenow fragment) and α-^{32}PdATP (Amersham) for 5 minutes at room temperature, then digested further with HincII. The 950 bp fragment was purified from agarose gels and sequenced.

12R

p12R was digested with PvuII, and the 330 bp fragment was extracted from agarose gels. It was further digested with HinfI, labelled with E.coli DNA polymerase I (Klenow fragment) and α-^{32}PdATP (Amersham) for 5 minutes at room temperature, and the 100 and 200 bp fragments separated on Nuseive (FMC) agarose gels, extracted and sequenced.

14L

p14L was digested with BglII, labelled with E.coli DNA polymerase I (Klenow fragment) and α-^{32}PdGTP (Amersham) for 5 minutes at room temperature, recut with KpnI, and the 750 bp labelled fragment extracted from agarose gels and sequenced.

14R

p14R was digested with BglII, labelled with E.coli DNA polymerase I (Klenow fragment) and α-^{32}PdGTP (Amersham) for 5 minutes at room temperature. It was then further digested with PvuII and the 80 bp fragment isolated from a

4% Nuseive (FMC) agarose gel, then sequenced. Several attempts to determine the sequence in the other direction (across the junction) were unsuccessful.

Sequence comparisons

The determined sequences were compared to published sequences of the B1 (Krayev et al, 1980), B2 (Krayev et al, 1982), MT (Heinlein et al, 1986), and L1Rn (D'Ambrosio et al, 1986) families of repetitive DNA, using the DOTMATRIX (stringency=60%, 70%, or 80% over 10 bases) and HOMOLOGY (stringency=65% over 20 bases) programs of Stephens (1985).

FIGE

Field Inversion Gel Electrophoresis (FIGE) was performed under two sets of conditions. Initial analyses were performed using the technique described in Carle, Frank and Olson (1986). Restricted DNA was migrated through 1% agarose (NA, Pharmacia) gels, prepared and run in 0.5X TBE buffer (45mM Tris, 45mM boric acid, 12.5 mM EDTA) in an IBI HRH electrophoresis bath with cooling plate. Buffer temperature was maintained at approximately 15°C by connecting the cooling plate to a constant temperature recirculating water bath set at 10°C. The voltage gradient throughout the runs was maintained constant (from 4-6 volts/cm) and the current switched using a homemade switch box (Roy et al, 1988; Appendix C). The

ramping of the pulse times and the period number were done with an IBM PCjr microcomputer interfaced with the switchbox. The program (Appendix D) was written in IBM Basic.

For improved resolution FIGE gels were run as described in Lalande et al (1987) with minor modifications. Briefly the DNA was run into the gel until the bromophenol blue had penetrated to about 1 cm , after which the forward and reverse fields were alternated using an Intervalometer (Sound Scientific, Seattle). Field strengths were generally +2 v/cm and -1 v/cm. Pulse times had a ratio of +2.5 to -1, and varied depending on the lengths of the fragments to be separated. Gels (0.8% agarose) were run at room temperature with buffer recirculation. The buffer used was 1X TBE (90 mM Tris, 90 mM boric acid, 25 mM EDTA, pH 8.3); the temperature was maintained at about 25°C by cooling the recirculating buffer in ice when necessary. Total migration time ranged from 36 to 72 hours.

Plugs of lambda phage DNA concatamers (Waterbury and Lane, 1987) and phage T4 DNA (Amersham) in solution were used as molecular weight markers.

RESULTS

Cloning of the integrated RmI sequences with the flanking cellular DNA from the W9 clones.

Choice of the integrated RmI molecules.

Factors which were considered when choosing the transformed cell lines from which I should clone the integrated RmI molecules were: 1. Since the nonrandom pattern of integration had only been seen in the clones produced at 39°C any attempt to identify the specific mechanisms involved in the nonrandom integration had to be restricted to the W9 clones. 2. The analysis of the W9 clones had not only demonstrated a nonrandom distribution of crossover points but also of insertion lengths; the insertions in the W9 clones were preferentially long, approaching the unit length of RmI. The implication was that since RmI could not replicate at 39°C the substrate in the integration reaction was an intact circular molecule and that only little or no sequences were lost from both RmI and the host at the time of integration. I therefore wanted to clone insertions approaching unit length.
3. Since I had shown that integration of RmI occurred preferentially in the region containing both the B2 and MT sequences of RmI, I wanted to examine in more detail clones in which at least one of the junctions was in this region.
4. I also wanted to clone at least one junction in which

the integration occurred at or near the RmI Py/Ins junctions in order to be able to examine the possibility of a site-specific event.

Bearing in mind the above factors, I chose to clone the junctions of clones W98.12, W98.14 and W94.10. Clone W98.12 and W98.14 were of particular interest since both junctions had originally been mapped to overlapping regions containing the repetitive sequences (see Figure 2, Appendix A). This raised the intriguing possibility that RmI had integrated as a circular molecule via a single homologous event into cellular repetitive sequences. Clone W94.10 has one junction mapped at an RmI Py/Ins junction. (Figure 2, Appendix A).

I also included in my study clone 11.6 which had been produced by Alain Nepveu using the same general protocol as the W9 clones (A. Nepveu, 1984). The left and right junctions of this clone had been mapped to the regions containing the repetitive sequences and an RmI Py/Ins junction respectively. Both the left (A. Nepveu, 1984) and right (N.Gusew, unpublished results) junctions of clone 11.6 as well as the unoccupied cellular site flanking the left junction (A. Nepveu, 1984) had already been cloned. Furthermore the left junction and the crossover site of the unoccupied site had been sequenced (A. Nepveu, 1984). This had established that the crossover had occurred in the B2 sequences of RmI but not in homologous sequences of

the host. I wanted to further characterize the host sequences at the integration site.

Cloning strategy.

In order to clone the integrated RmI with the flanking cellular sequences I had to find a restriction enzyme which cut once in RmI, and far enough beyond the RmI cellular junction on each side of the insertion to allow an extensive analysis of the cellular sequences present. Although I could have used a restriction enzyme which does not cut RmI, these were relatively rare and would not allow a precise measurement of the flanking cellular sequences. Furthermore the restriction enzyme had to allow for convenient cloning in one of the phage vectors available.

An initial screening of the restriction maps that I had already constructed permitted me to choose amongst a restricted number of enzymes. I subsequently digested the cellular DNA with the prospective enzymes and verified the length of the junctions produced by Southern analysis (Southern, 1975). The information in Table 1 shows that digestion with EcoRI produces in all cases a junction containing enough Py sequences to be easily detected with a Py probe, and enough cellular sequences to permit extensive analysis as well as isolation of a unique sequence probe.

TABLE 1

Lengths of the EcoRI junction fragments.

Clone	junction	Fragment sizes: (in Kbp) total	RmI	host
W98.12	L	5.0	4.2	0.8
	R	6.0	2.7	3.3
W98.14	L	13.0	4.5	8.5
	R	5.5	2.45	3.05
W94.10	L	7.6	3.35	4.25
	R	4.1	3.3	0.8

Total cellular DNAs from the indicated clones were digested with EcoRI, electrophoresed through agarose gels beside EcoRI-digested RmI and Py standards, transferred to membranes and hybridized with a Py probe. The lengths of the Py-bearing fragments (fragment size: total) were estmated by comparison to the standards. The left and right junctions could be distinguished by the intensity of the hybridization signals (left junctions contain approximately twice as much Py sequence as do right junctions). The lengths of the RmI sequences within the junction fragments were calculated from the maps which had been determined previously, (Figure 2, Appendix A).

The phage vector λgtWES.λB was chosen to clone the junctions because it could accommodate EcoRI fragments from 2.13 to 15.13 Kbp in length which was inclusive of all the junctions desired (Table 1).

The cellular DNA was digested with EcoRI and enriched for the junction fragments. The enriched cellular DNA fragments were ligated to λgtWES.λB EcoRI arms and the phages thus prepared were packaged *in vitro* (Materials and Methods).

Clone W94.10.

A fraction which contained both the 8 and 4 Kbp *EcoRI* junction fragments was picked from the sucrose gradient and used for the ligation to λgtWES.λB *EcoRI* arms, then packaged *in vitro.*

The 1.5 X 10^7 recombinant phages produced were amplified on LA101 in liquid culture to a final titer of 10^{11} phages. The phage DNA was extracted from the amplified banks and analyzed by Southern hybridization. Unfortunately, only the band at 8 Kbp was detected (data not shown). We were mostly interested in the 4 Kbp right junction since it contained the RmI Py-INS junction. Screening of four partial libraries failed to detect the desired fragment.

Nevertheless to test our procedures the amplified bank was screened by the method of Benton and Davis (1977). A total of 10^6 phages were screened on 20 large (529 cm^2) petri dishes, and two positive signals (Lgt94.10a and Lgt94.10b) were identified. These were plaque purified to Lgt94.10a1 and Lgt94.10b3. The DNAs of these phages preparations were extracted and digested with *EcoRI*, then hybridized with a Py probe. Isolate a1 gave a strong signal at approximately 8 Kbp while b3 gave only a weak signal at the same fragment length (data not shown). These phages have not been analyzed further.

Clone W98.14.

Sucrose gradient fractions enriched for the 5.3 (R) and 13.5 (L) kbp EcoRI junction fragments were ligated to λgtWES.λB EcoRI arms then packaged *in vitro*. The titer of the R-ligation was 10^5, that of the L-ligation was 6.5 x 10^5. After subtraction of the background (see Materials and Methods) the titers of recombinant phages for the R- and L-ligations were zero and 5.5 x 10^5 respectively. To determine experimentally if these libraries did contain recombinants, small semiconfluent petri dishes were blotted (Benton and Davis, 1977) and hybridized to an RmI probe. Since RmI contains two repetitive sequences each present in 10^5 copies dispersed throughout the rat genome, we can expect to find one copy for about every 15-30 Kbp of genomic DNA. Thus when cloning fragments of about 5 and 13 Kbp we can expect to find approximately 15-30% and >50% of the clones hybridizing to RmI respectively. Of the R-ligation (approximately 5.3 Kbp average length) 18% of the clones hybridized to RmI, and of the L-ligation (approximately 13.5 Kbp average length) 13% hybridized. Thus, although initial indications were that the R-ligation had not produced recombinant clones and that the L-ligation had, further examination showed that proportionally the R-ligation had been the more efficient.

The L-recombinants were amplified on 11 large petri dishes and recovered into SM. The final titer was 5 x 10^{10}

phages. The R-recombinants were amplified on two large petri dishes and recovered into SM. However, rather than discarding the petri dishes used to amplify the R-recombinant bank I chose to blot them to screen for Py-containing recombinants by the method of Benton and Davis (1977). Following the transfer to two large membranes each petri was blotted a third time with a small round filter. The large filters were then hybridized against a Py probe and the small filters against an RmI probe. Since I had already shown that 18% of the plaques would hybridize with RmI the small filters served as positive controls, and indeed I was able to detect hybridization to the RmI probe, although the signals were weak. From the large filters I detected one strong and one weak signal. The strong signal was purified by two rounds of plaque purification, and baptized Lgt98.14Ra. Analysis of the DNA of this clone revealed a Py-DNA containing fragment which comigrated with the 5.3 Kbp right junction of W98.14, and which hybridized to both a whole Py probe and a probe specific for the right side (data not shown).

A total of 10^6 L-recombinant phages were screened and two positive signals were detected. One of these was lost during the first step of plaque purification, the other required four rounds of plaque purification before the isolate (Lgt98.14L) was pure enough to subclone. The plaques of this recombinant were miniscule, indicating that it grew only very poorly. After only two rounds of plaque

purification the DNA from the phage culture was analyzed and found to contain a Py-containing fragment that comigrated with the 13.5 Kbp left junction of W98.14. This fragment hybridized with a probe made with Py in its entirety but not to a probe which was specific for the right side (data not shown).

Clone W98.12.

One fraction from the sucrose gradient was enriched for both the 5 and 6 Kbp EcoRI junction fragments of W98.12. The DNA from this fraction was ligated to λgtWES. λB EcoRI arms. The λgtWES.λB EcoRI arms used in this ligation had not been purified away from the EcoRI insertion but rather the insertion had been further digested by SstI which cuts the insertion twice more. Following the SstI digestion the arms were cleaned and used directly in the ligation. Following *in vitro* packaging and infection, 5.4×10^5 recombinant phages were recovered, 16-19% of which hybridized to an RmI probe.

These phages were screened for Py containing clones by the method of Benton and Davis (1977). Positive signals were detected and plaque purified. The DNA was extracted from these clones and analyzed by Southern blotting following EcoRI digestion. Clone #4 had one Py positive fragment which comigrated with the 5Kbp left junction; and clone #5 had one Py-containing fragment which comigrated

with the 6 Kbp right junction. Clones #4 and #5 were chosen for subcloning to plasmid vectors.

Subcloning into plasmid vectors.

The insertions in Lgt98.12 #4 and #5 and Lgt98.14Ra and Lgt98.14L6.2 were subcloned by "shotgun" ligation into pBR322/EcoRI treated with alkaline phosphatase. Py containing clones were detected by the technique of Grunstein and Hogness (1975), the plasmids were screened for appropriate size with cracking gels (which were blotted and hybridized to Py probes), and the insert lengths confirmed by digesting DNA from plasmid minipreparations (Materials and Methods) with EcoRI (see below). The plasmids were baptized p12L and p12R, and p14R and p14L respectively.

Initial analysis.

In order to verify that the fragments that had been cloned corresponded to the insertions as they occurred in clone W98.12 and W98.14, the plasmids were first digested with EcoRI and the fragment lengths compared to the junction lengths of the cellular DNA (also digested with EcoRI) by Southern analysis (Figure 3). As further confirmation the plasmids were mapped by restriction enzyme analysis and the maps compared to the maps of the cellular insertions (Figure 2, Appendix A). There was perfect

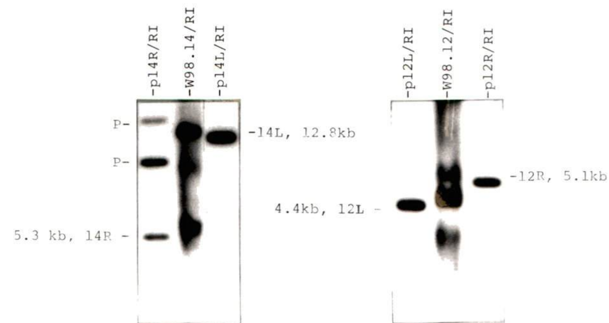


Figure 3 Comigration of the cloned junctions with the corresponding cellular fragments.

Plasmid and cellular DNAs were digested with EcoRI, then migrated through an agarose gel in adjacent wells, Southern blotted, (Southern, 1975) and hybridized with a radioactive Py probe. The junctions cloned into plasmids comigrate with the corresponding junctions in the cellular DNAs. (The slight retardation of migration of the cellular DNAs has been observed before; Wallenburg, 1984) P, fragments resulting from the partial digestion of p 14R.

agreement between the two maps in all cases. The next step was to produce a more detailed restriction map of the junctions (Figure 4). This served two purposes. First, this allowed me to localize, with more precision than was possible when analyzing total cellular DNA, the crossing-over point at the junctions. Secondly, it allowed me to choose restriction enzymes which would later serve as start points for sequencing the junctions by Maxam and Gilbert sequencing.

Localization of repetitive DNA and INS similarities.

Since the B2 and MT repetitive sequences should be present once in every 15-30 Kbp of FR3T3 DNA, it seemed likely that RmI could have integrated into or near homologous sequences. I therefore determined the localization of cellular sequences flanking the junctions that were either highly repetitive or homologous to Ins. This was done by blotting the restriction enzyme analysis gels of the four plasmids as well as those of a plasmid containing the right junction of clone 11.6 (p11.6R) and hybridizing to probes of (i) total FR3T3 DNA or (ii) the RmI/HincII-L fragment (which contains all of Ins as well as 683 bp of flanking Py sequences) or (iii) with an "Ins-only" probe which consisted of the Hind III-SphI fragment of Ins. This 728 bp fragment contains all of the B2 and MT sequences of Ins and 165 bp and 35 bp of unique sequence DNA on either side. Thus the FR3T3 probe would

Figure 4 The physical maps of the cloned RmI-cellular junctions.

Each junction was cloned as an EcoRI fragment as described in Materials and Methods and Results. Thus both EcoRI junctions of a single clone are shown in juxtaposition separated by the EcoRI site of RmI. The name of the clone from which they were derived is given on the extreme left. The inserted sequences of RmI have been aligned with the linearized representation of RmI given at the top of the figure. (Ins has been duplicated for the purpose of illustration.) The legend for the RmI sequences is the same as in Figure 1. The line represents the cellular sequences flanking each insertion. IS is an EcoRI cellular fragment that represents the arrangement of the cellular sequences at the left side of the RmI insertion in clone 11.6 before integration. (Plasmid p2.6 contains the IS fragments cloned into pBR325 digested by EcoRI.) It has been aligned with the corresponding cellular sequences in 11.6 directly above. S1 through S7 and S9 delineate the probes used for the restriction enzyme mapping of the initial sites. S8 is the probe used to determine the fate of the initial sequences after integration (see Figure 14). Refer to the text and figure legends for the uses of S10 and S11.

Restriction enzymes:

A, HpaI; B, BamHI; E, EcoRI; F, HinfI; G, BglII; H, HindIII; I, BglI; J, AvaI; K, KpnI; M, SmaI; P, PstI; S, SalI; V, PvuII; X, XbaI. Not all restriction sites are shown. Sites with an asterisk were used to sequence the junctions.

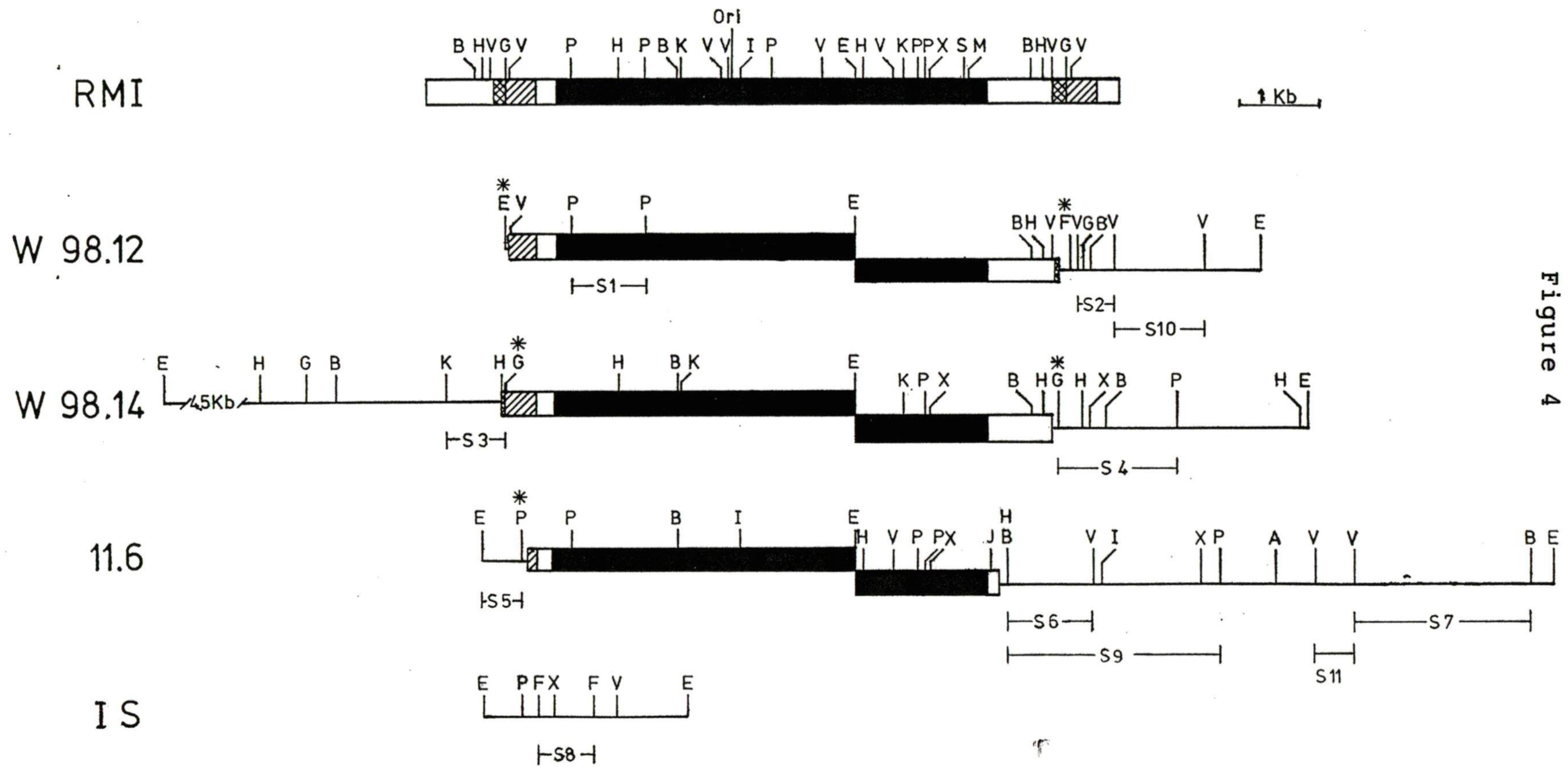


Figure 4

detect all repetitive host sequences while the HincII-L and Ins-only probes would detect sequences which were homologous with Ins, probably B2 and MT. The post-hybridization washes were performed at reduced stringency (Materials and Methods) because the populations of repetitive DNA share less than 100% homology. Since the HincII-L probe was isolated from RmI prepared by Hirt extraction (and therefore contained 5% Py molecules), and the Ins-only probe had been isolated from a plasmid, some contamination of the probes by Py or pBR322 sequences was considered to be inevitable. Thus fragments containing only Py or only plasmid sequences served as negative controls; fragments known to contain homologous or repetitive sequences from Ins served as positive controls. It should be noted that a much weaker signal is expected from the FR3T3 probe since it is total cellular DNA.

During these experiments I noticed that fragments of RmI which contained the right RmI Py/Ins junction frequently hybridized to FR3T3 probes (Figures 5, 6, and 7). However no highly repetitive mouse sequences have been identified in this region (Bourgaux et al, 1982). Since I was able to detect sequences repeated on the order of 5-10 copies per genome using the FR3T3 probe (see below), perhaps there are Ins (mouse) sequences near this junction which are homologous to rat sequences repeated a small number of times in FR3T3.

11.6R

Fragments of p11.6R which are specific to the adjacent cellular sequences hybridized with the FR3T3 probe but not with the Ins-only probe (Figure 5). 11.6R mapped close to the right RmI Py/Ins junction and no highly repetitive sequences are present there in RmI (see above). Furthermore, host sequences downstream from the junction also hybridized with the FR3T3 probe but not with the Ins probe. This suggested that there were repetitive sequences present in the host DNA flanking the insertion, but that they were not homologous to B2 or MT. Since they hybridized with the 6.8 and 6.1 Kbp BamHI fragments, the 4.7 and 2.6 Kbp PvuII and the 5.1 and 1.9 HindIII fragments, but not with the 2.9 Kbp PvuII fragment, the repetitive sequences could be localized to two non-contiguous regions (Figure 5).

14R

Fragments containing the last 100 bp (approximately) of cellular sequences at the distal end of the 14R junction fragment would sometimes hybridize weakly with the FR3T3 probe but never to INS containing probes (data not shown). This suggests that there may be some repetitive DNA 2.5 Kbp from the 14R junction but that these sequences are not homologous to B2 or MT. Other than the aforementioned Py-INS junction fragments (e.g. the 2.2 Kbp p14R/PvuII

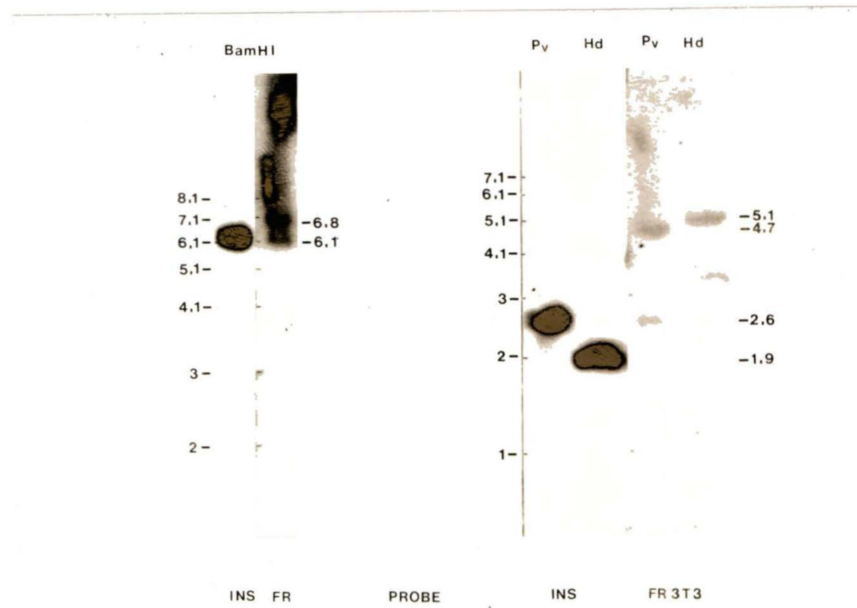
Figure 5 Repetitive sequences in p11.6R.

A. p11.6R was digested with BamHI, PvuII (Pv) or HindIII (Hd), migrated through agarose gels beside molecular weight markers, then blotted bidirectionally onto nylon membranes. The bottom blots were hybridized with an INS specific probe (INS, the RmI/HincII L fragment - see text). The top blots were hybridized with total FR3T3 DNA which was sheared, then radioactively labelled (FR). The shown autoradiograms were exposed either overnight (INS) or for 6 days (FR) at -70° C. The molecular weights of fragments are shown in Kbp.

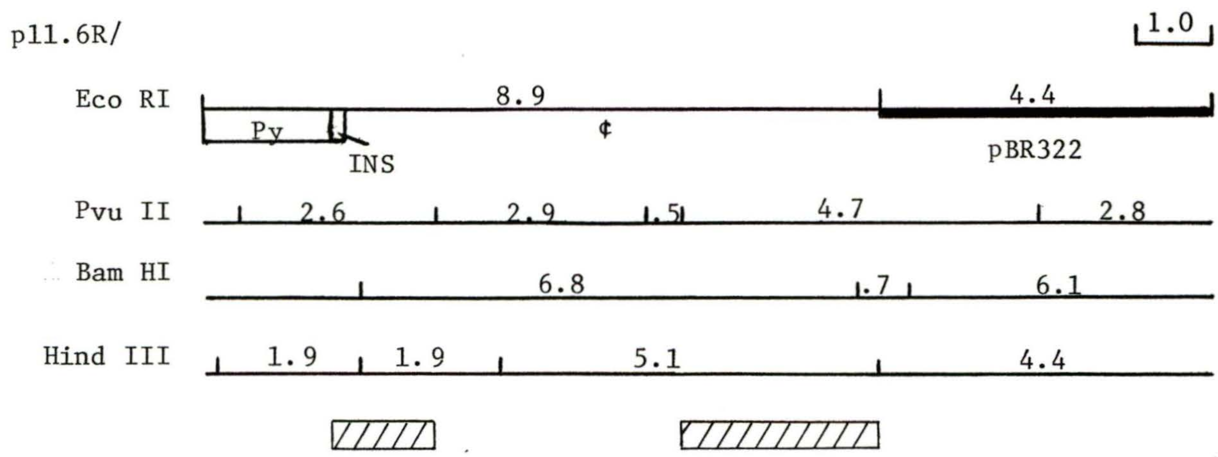
B. Graphic restriction map of p11.6R cut with the above restriction enzymes. The first line shows the p11.6R EcoRI cloning sites as well as the position of the RmI(Py), RmI(INS), and flanking cellular (c) and pBR322 sequences. The map has been linearized at one of the EcoRI sites for graphic purposes. Fragment lengths are given in Kbp. The hatched boxes at the bottom of the figure delimit the regions containing the repetitive DNA.

Figure 5

A



B



fragment) no other cellular fragments consistently hybridized to either the FR3T3 or Ins probes (Figure 6).

14L

The 2.4 Kbp BglII-KpnI fragment of p14L (which contains the MT sequences of RmI) hybridized to FR3T3 and INS as expected (Figure 6). The 3.9 Kbp BglII fragment of host DNA flanking the insertion in 14L (see Figure 6) also hybridized strongly with the FR3T3 probe. The .6 Kbp, and .8 Kbp BglII and the 1.8 Kbp BglII-KpnI fragments which flank the 3.9 Kbp fragment also hybridized but only weakly. Neither the 0.8 Kbp nor the 3.9 Kbp fragments hybridized with the Ins probe; therefore, these repetitive sequences are not B2 or MT since Ins contains both of these SINE elements. Since no single SINE element could account for the >4 Kbp of repetitive DNA, one possible interpretation of these results is that the 3.9 Kbp fragment contains a (partial) LINE element that terminates in the flanking 0.8 Kbp fragment. The 0.6 Kbp fragment and the 1.8 BglII-KpnI fragment both hybridized with the INS probe, but only weakly, especially the 1.8 Kbp fragment. Finally, the 0.7 Kbp BglII+KpnI junction-containing fragment hybridized with the INS probe but not with the FR3T3 probe. Since the BglII site which is the right limit of this fragment is the RmI-BglII site which is in the B2 sequences, it might be expected that an FR3T3 probe should hybridize to it. The

Figure 6 Repetitive sequences in p14R and p14L.

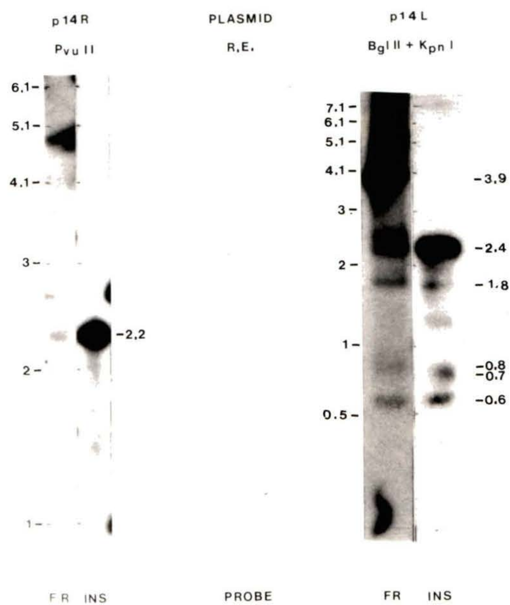
A. p14R was digested with PvuII, p14L was doubly digested with BglII and KpnI. They were migrated through agarose gels beside molecular weight markers, then blotted bidirectionally onto nylon membranes. The bottom blots were hybridized with an INS specific probe (INS, the RmI/HincII L fragment - see text). The top blots were hybridized with total FR3T3 DNA which was sheared, then radioactively labelled (FR). The shown autoradiograms were exposed either overnight (INS) or for 6 days (FR) at -70° C. Fragment lengths are shown in Kbp.

(Note that hybridization of p14L/BglII + KpnI with the FR3T3 probe illuminates the 0.8 Kbp fragment but not the 0.7 Kbp fragment, whereas hybridization of p14L/BglII + KpnI with the Ins probe illuminates the 0.7 Kbp fragment but not the 0.8 Kbp fragment.)

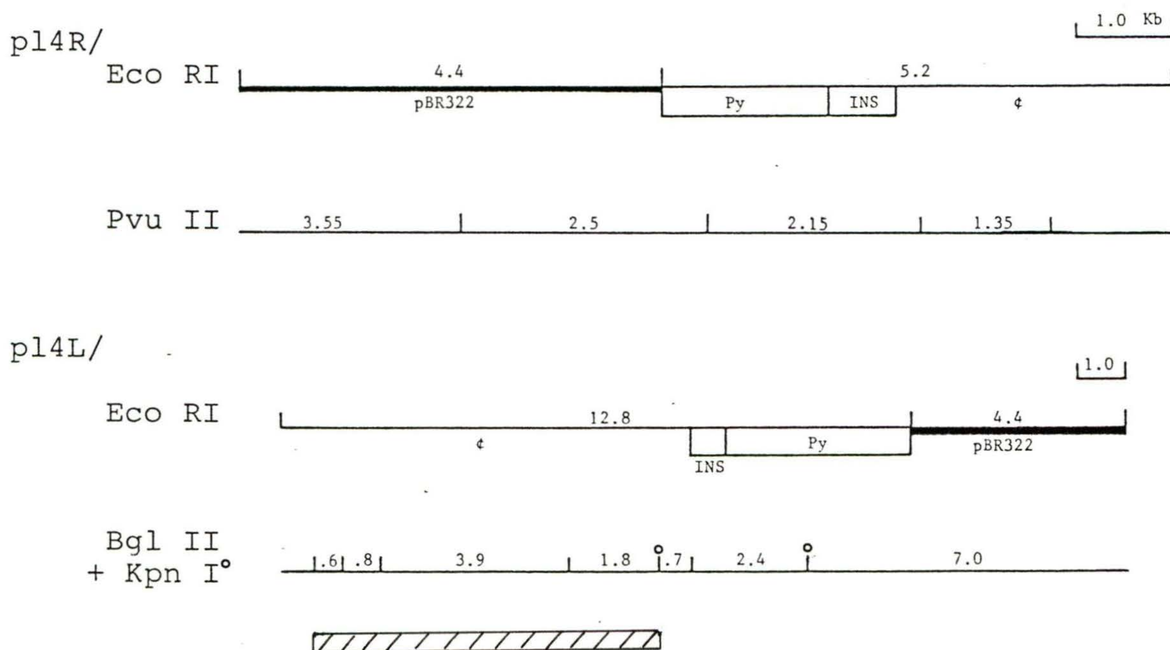
B. Graphic restriction map of p14R and p14L cut with the above RE. The top lines show the p14R and p14L EcoRI cloning sites as well as the position of the RmI(Py), RmI(INS), and flanking cellular (c) and pBR322 sequences. The maps have been linearized at one of the EcoRI sites for graphic purposes. Fragment lengths are shown in Kbp. The hatched boxes at the bottom of the figure delimit the regions containing the repetitive DNA.

FIGURE 6

A



B



lack of hybridization indicates that the actual crossing-over is situated very near the BglII site, and that the B2 sequences remaining at the junction are insufficient for hybridization to the FR3T3 probe. This was confirmed by sequencing (see below).

12R

Fragments containing the 12R junction hybridized to FR3T3 and Ins (Figure 7). Since 12R had been mapped to the repetitive sequences of RmI, this result was not unexpected. No other cellular fragments hybridized with the FR3T3 or Ins probes.

Both p2.6 (IS, Figure 4) and p6.4 (Materials and Methods) had several bands which hybridized to FR3T3 but not to Ins probes (data not shown). These hybridizing sequences were not mapped further. Plasmid p12L contained too few host DNA sequences to be screened for repetitive DNA by hybridization. The cellular DNA was sequenced however and compared to known repetitive sequences (see below).

In summary, a total of ~27.35 Kbp of cellular DNA was analyzed. Highly repetitive sequences hybridizing to FR3T3 were found in p14L at two locations, in p14R, in p11.6R and in p2.6 and p6.4. Of these, only the short repetitive sequence in p14L hybridized with sequences in Ins.

Figure 7 Repetitive sequences in p12R.

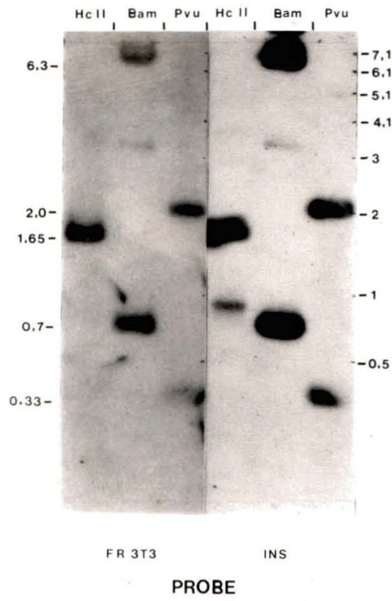
A. p12R was digested with BamHI (Bam), PvuII (Pvu) or HincII (HcII), migrated through agarose gels beside molecular weight markers, then blotted bidirectionally onto nylon membranes. The bottom blots were hybridized with an INS specific probe (INS, the RmI/HincII L fragment - see text). The top blots were hybridized with total FR3T3 DNA which was sheared, then radioactively labelled (FR). The shown autoradiograms were exposed either overnight (INS) or for 6 days (FR) at -70° C. Fragment lengths are given in Kbp.

(Note that the BamHI digestion was not complete. Therefore there are intermediate sized bands which hybridize with both the FR3T3 and INS probes; eg. p12R/BamHI vs FR3T3 illuminates bands at 7.0 (6.3 + 0.7) and 3.2 (2.5 + 0.7) Kbp and p12R/BamHI vs INS illuminates the 3.2 Kbp fragment. Furthermore, the 1 Kbp HincII band detected with the INS probe does not exist in p12R. It is not visible in the ethidium bromide stained gel [data not shown], nor does it hybridize with the FR3T3 probe [all repetitive DNAs detected by the INS must also hybridize with the FR3T3 probe]. I beleive it to be the result of hybridization between pBR322 sequences contaminating the plasmid-derived INS probe and a minor plasmid contaminant present in the minipreparation of p12R DNA.)

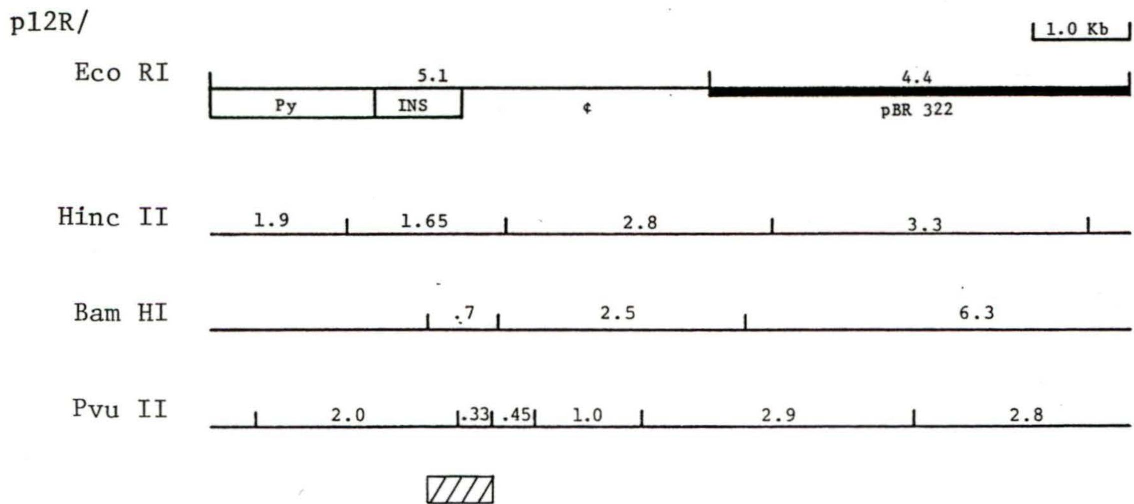
B. Graphic restriction map of p12R cut with the above REs. The first line shows the p12R EcoRI cloning sites as well as the position of the RmI(Py), RmI(INS), and flanking cellular (c) and pBR322 sequences. The map has been linearized at one of the EcoRI sites for graphic purposes. Fragment lengths are given in Kbp. The hatched boxes at the bottom of the figure delimit the regions containing the repetitive DNA.

Figure 7

A



B



Detailed maps of the junctions.

Once the junctions were cloned into plasmids I wanted to map the crossover sites to the nearest RmI RE sites in order to precisely localize the crossover with respect to the RmI sequences, but I also wanted to map the nearby cellular RE sites. The closest convenient RE in either the cellular or RmI sequences would serve as a start point to sequence the junction. The cellular sites would also serve to isolate unique sequence DNA as close as possible to the junction to be used as probes when analyzing the rearrangement of the cellular DNA.

12R

The junction had been mapped to a region bounded on one side by the presence of an RmI PvuII site (bp 4076) and on the other by the absence of a BglII site (bp 4236), a fragment of 160 bp (Figure 4). This analysis also revealed the presence of a cellular BglII site ~ 400 bp from the PvuII. In the plasmid, the junction was found to be isolated to a 330 bp PvuII fragment, which was cut into 200 and 100 bp by a cellular HinfI site. This HinfI site was later used to sequence the junction.

12L

Southern analysis of the cellular DNA had mapped the

crossover between the *MboI* (4650, present) and *PvuII* (4331, absent) sites (319 bp). In the plasmid the *EcoRI* site used to clone the junction was found to be unexpectedly close to the junction. The distance between the cellular *EcoRI* cloning site and the RmI *MboI* site was about 200 bp. Thus the crossing-over had occurred within 200 bp of the *MboI* site. The *EcoRI* site was used to sequence the junction (Figure 4).

14R

The crossover had already been mapped to a fragment of only 121 bp (*HindIII* (3955) to *PvuII* (4076), Figure 4) by Southern analysis of the cellular DNA. Analysis of the plasmid revealed the presence of a cellular *BglII* and *PvuII* site ~200 and ~250 bp away from the *HindIII* site respectively. The sequencing was done from the *BglII* site.

14L

The initial Southern analysis had located the crossover to a 376 bp *HindIII* (3955)-*PvuII* (4331) fragment (Figure 4). The presence of the RmI *BglII* site in the plasmid and the absence of the 255 *PvuII* fragment of Ins, more narrowly defined this region to a 160 bp *BglII* (4236)-*PvuII* (4076) fragment. Subsequent analysis of the cellular DNA with the enzymes *AvaII* (4171) and *MboII* (4162) suggested that the crossover had in fact occurred within

the 9 bp between the two sites (data not shown). The <u>BglII</u> site was used to sequence the junction.

Preparation of unique sequence probes.

From the analysis of the repetitive DNA present in the plasmids, I was able to localize the repetitive and unique sequence host DNA on the restriction enzyme maps of the plasmids, and was thus able to select unique sequence fragments to be used as probes to analyze the host sequences at the site of integration. These probes were initially chosen to be as close as possible to the junctions to ensure that a maximum number of restriction enzymes would be informative. The farther the probes were from the junctions the greater the probability that a given restriction enzyme would cut between the junction and the probe, preventing me from using that enzyme to map the host site.

There were not enough cellular sequences in the plasmid p12L to permit an analysis of their nature (i.e. repetitive or unique). I therefore proceeded directly to determine the sequence.

The crossover at 12R had been mapped to within 160 bp of the (4076) RmI <u>PvuII</u> site, and was contained within a 330 bp <u>PvuII</u> fragment in the plasmid. Immediately flanking this fragment are two <u>PvuII</u> fragments of 450 bp and 950 bp which contain only unique sequence DNA. The 450 bp

fragment was used to analyze the rearranged site (S2, Figure 4). S10, the 950 bp PvuII fragment immediately adjacent to S2, was used once (see Figure 8).

The .7 Kbp BglII-KpnI fragment (S3, Figure 4) of p14L which contained the junction did not hybridize to the FR3T3 probe although it did hybridize to the Ins-only probe as already discussed. In spite of the fact that it may contain B2 sequences this fragment was used to probe the initial site because 1. it did not hybridize to the FR3T3 probe and 2. it contained the cellular sequences immediately flanking the junction.

From p14R a 1.35 Kbp PvuII fragment of cellular DNA which was ~250 bp from the RmI HindIII site and which contained only unique sequence DNA was used as a probe (S4, Figure 4).

Four probes were eventually chosen from p11.6R (Figure 4 and see below). The first was a 2.7 Kbp BamHI-PstI fragment (S9) from the BamHI site ~200-300 bp from the crossover. The second was a 1.1 Kbp BamHI-PvuII subfragment (S6) of S9. The third was a 2.3 Kbp PvuII-BamHI fragment (S7) from the PvuII site at ~4.5 Kbp from the crossover. Finally, S11 (Figure 4) was used only once (Figure 8).

The probe (S5) used to analyze the rearrangement in

11.6L had originally been prepared by A. Nepveu (1984). It was a 570 bp fragment produced by digestion of p11.6L (p4.7 in the reference) with PstI and EcoRI and which is situated ~50 bp from the junction (S5, Figure 4).

Source of the sequences flanking the junctions.

The clones used in this study had been produced by transfecting FR3T3 cells with circular RmI, in the presence of a large excess of carrier DNA, by the calcium-phosphate precipitation technique. Previous studies had already shown that linear DNA transfected in the presence of carrier could form high molecular weight complexes of several hundred Kbp by ligation to that carrier, and that these complexes, termed transgenomes (Scangos and Ruddle, 1981) or pekelosomes (Perucho et al, 1981), subsequently integrated into the host chromosomes (see Introduction). Since I wanted to study the interaction of RmI with the host DNA, it was primordial that I determine the source of the flanking sequences into which it had integrated: was it host FR3T3 DNA or carrier DNA? As I had used mouse (LTA) carrier DNA in the transfections of the W-clones and salmon sperm (s.s.) carrier DNA for 11.6, it was possible to determine the source of the flanking DNA by hybridization of the unique sequence probes to the clone, FR3T3, LTA and salmon sperm DNAs (Figure 8).

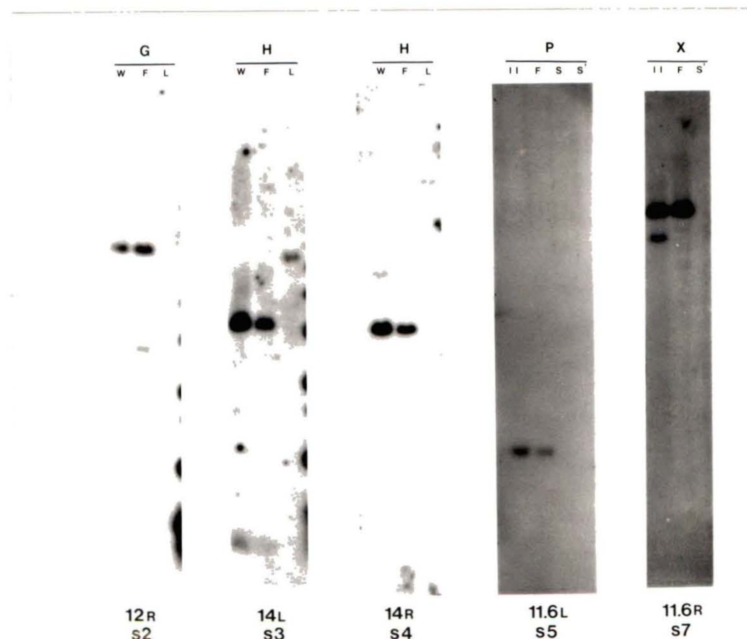


Figure 8. RmI integrated directly into host cellular DNA and not into the carrier DNA.

Five micrograms of clone (W or 11), host FR3T3 (F), and carrier LTA (L) and 5 (S) or 10 (S') micrograms of carrier salmon sperm DNAs were digested with restriction enzymes, migrated beside each other on agarose gels, blotted and hybridized to probes of cellular DNA. The probes (S2, S3, S4 and S5 shown in Figure 4), were chosen to contain only unique sequence DNA and to cover the region immediately flanking the site of integration. Probe S7 (Figure 4) hybridized to sequences present on the order of 5 to 10 copies per genome. The junctions analysed and the probes used are indicated beneath each autoradiogram.

The host DNA immediately flanking 11.6R (which includes the Xba site) has undergone a rearrangement (see text), and therefore this XbaI fragment (lower band in 11.6) does not comigrate with the unrearraned fragments (upper band in 11.6 and FR3T3). Note that the single copy rearranged band is several times less intense than the unrearranged bands.

Restriction enzymes:

G, BglII; H, HindIII; P, PstI; X, XbaI.

The DNA from each clone was cleaved so as to "separate" the probe DNA from the RmI insertion. The FR3T3, LTA and salmon sperm DNAs were digested with the same enzyme. In this way the fragment flanking the RmI insertion in the clone DNA would be expected to comigrate with its allele and with the corresponding fragment in the source DNA. In those cases where the enzyme did not separate the entire probe from the insert, an extra band appears in the digestion of the clone DNA which corresponds to the rearranged junction-containing fragment, and its allele is consequently reduced in intensity by a one copy equivalent (since the rearranged copy no longer comigrates with it). W98.12L could not be analyzed because no probe was available.

Junction 12R was analyzed with BglII and EcoRI; 14L with HindIII, BamHI, and PvuII; 14R with BglII, EcoRI, HindIII, BamHI, and PvuII; 11.6L with XbaI and PstI; and 11.6R with BamHI, EcoRI, and XbaI. In all cases the probes hybridized with the host FR3T3 DNA (Figure 8, not all restriction enzyme digestions are shown). In 11.6L, and 11.6R there was no hybridization to the carrier DNA. In 12R S10 hybridized weakly with a non-comigrating band in the carrier LTA DNA (data not shown), whereas S2 did not. Probes S3 and S4 (from clone W98.14) did hybridize weakly to non-comigrating bands in the carrier LTA DNA. I believe this to be due to cross-hybridization between rat and mouse DNA.

In summary, these results show that RmI had interacted directly with the host DNA in all cases and not by the intermediary of a transgenome in spite of the fact that it had been transfected in the presence of carrier. Thus the junctions studied here are the consequence of the mechanisms present in the cell for the integration of exogenous (circular) DNA into its chromosomes and not the products of simple end-to-end ligation with a mass of "free" DNA.

Sequences.

In order for me to be able to determine the nature of the recombination event (site-specific, homologous, or illegitimate) that led to the integration of RmI in each of the clones it was necessary to determine the nucleotide sequence across each of the junctions. The junctions were sequenced using the chemical cleavage method of Maxam and Gilbert (1977, 1980). The restriction enzymes used are identified with asterisks in Figure 4 and have been previously described. The 3′ recessed ends were labelled with αP^{32}-dNTPs using Klenow fragment DNA polymerase (as described in Materials and Methods). The sequence of 12R was read from 3 different gels resulting from two different sets of reactions. 12L was read from 3 different gels of 3 different sets of reactions. Sequence 14L was read from two gels. The sequence for 14R was read from one gel. This sequence does not cover the junction but is the

sequence of host cellular DNA from the BglII site situated ~100 bp from the junction and read going away from the junction (Figure 4). Multiple attempts to determine the sequence across the junction were unfruitful. However, since the repetitive sequences of INS had been mapped to sequences 4122-4240 (B2, Sylla et al, 1984) and 4277-4650 (MT, Bastien and Bourgaux, 1987), and the crossover at 14R had been mapped between sites 3955 and 4076 (see above), it was clear, without obtaining the sequence across the junction, that the crossover had not occurred in the repetitive sequences of RmI.

The sequences across the junctions 12L, 12R, and 14L have been aligned with the corresponding RmI sequences (Figure 9). The crossover at all of the junctions mapped to the repetitive sequences of RmI, either B2 (12R and 14L) or MT (12L). If RmI integrated by site-specific recombination I would expect to find specific nucleotide sequences or structures at the junctions. Or, if the preferential involvement of the repetitive sequences of Ins was the result of homologous recombination with endogenous repetitive sequences then I would expect to find at the junctions complete chimaeric repetitive elements originating in part from RmI and in part from the host. Neither of these is seen. Instead the junctions are homologous with RmI up to the presumed crossover point beyond which there is no significant homology. What is found at all the junctions are truncated repetitive

```
                                     ▼
RmI    CATGGACTGCTCTTCCAGAGGACCTGGGTTCAATTCCCAGCACCCACATGGCGGCTCACAGTCATCTGTAA
       ||||||||||||||||||||||| ||||||                        ooo oo ooooooooooooo ooooooooooo ooo   oooo o o oo oooo
12R    CATGGACTGCTCTTCCAGAGGACTTGGGTTGTGGCTAAGTTCCCTTTATTTTGCTTTGCTAAGGAACCTCCAGATTGATTTCCAGAGTGGGCAGACCAATATCCACTCCCCAAA

                                   ▼
RmI    AGGNGGGCCCAGCCCATTGTAAGTGGTGCCATCCCTGGGCTGGTAGTCCTGATTCTATAAGAAATCAGGCGGAGCAAGCCAAGGGAAGCAAGCCAGTAAGCAGCACCCCTTCACGGCCTCTGCATCAG
                                   |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||| |||||||||||
12L    TAGACTGGTCCTACTTTAATATCTATGCCCATCCCTGG-CTGGTAGTCCTGATTCTATAAGAAATCAGGCGGAGCAAGCCAAGGGA-GCAAGCCAGTAAGCAGCACCCCTTCACGGYCTCTGCATCAG

                                                                      ▼
RmI    GGCTCAGGAGTTAGGAGCATGGACTGCTCTTCCAGAGGACCTGGGTTCAATTCCCAGCACCCACATGGCGGCTCACAGTC
                                                                     ||| ||| ||||||||||
14L    AGAGAGACAAACAGACAGAGACAAAGAGACAGATAGAGAAGCTTGGCAGTGAGGAGGTAGAGACA-GGCAGCTCACAGTC

                                                                  ▼
RmI    TCAGGCGGAGCAAGCCAAGGGAAGCAAGCCAGTAAGCAGCACCCCTTCACGGCCTCTGCATCAGCTCCTGCCTTCCAAGTTCCTGCCCTGTGTG
                                                                  ||||||||||||| |||||||||||||||||||||
11.6L  GATCTCTAAAAGATTTCAAGAGTTTGTCTTTGAGGTGCCACAAAGTTCTAT-C-TGTTAATCAGCTCCTGCC-TCCAAGTTCCTGCCCTGTGTG
       ||||||||||||||||||||||||||||||||||||| ||||||||||||| | ||||||
IS     GATCTCTAAAAGATTTCAAGAGTTTGTCTTTGAGGTGTCACAAAGTTCTATTCGTGTTAAGAAGGAAGAATTTCCTTAATATAAGCTATGTCAT


14R    TATGTCTCACCAAGACACACTGCTGGGGTTGCCATGGGGACGGGCAGGGGGGAGCAGGCGCG
```

Figure 9 The sequences across and near the junctions.

The sequences across the junctions shown in comparison to the corresponding sequences of RmI. 11.6L is also compared to the initial site (IS, Figure 4). Arrowheads indicate the crossover point in each junction. Vertical lines are drawn between homologous bases. Dashes have been inserted to allow alignment of homologous sequences. Bases with a halo (in the sequence of 12-R) are homologous with L1Rn (see text). The sequence 14R does not cover the junction but is the sequence of host cellular DNA read from the *BglII* site situated ~100 bp from the junction (see Figure 4) and going away from the junction. All sequences are shown in the 5′ to 3′ orientation. The sequences of 11.6L and IS had been previously determined by A. Nepveu (Nepveu, 1984)

elements linked to sequences unrelated to the repetitive element. Thus we concluded that in all three cases the junctions were the result of illegitimate integration.

It has been proposed (Kato _et al_, 1986) that integration of exogenous DNA occurs preferentially in host cellular repetitive sequences. The sequences across all of the junctions, and of the cellular sequences flanking 14R, and of the original site of 11.6 (Figures 4 and 9) were screened by computer (see Materials and Methods) for the presence of highly repetitive DNA. None of the _junctions_ had significant homologies with the major families of rodent repetitive sequences, including B1 (Krayev _et al_, 1980), B2 (Krayev _et al_, 1982), MT (Heinlein _et al_, 1986), or L1Rn (D'Ambrosio _et al_, 1986), with the exception of a short sequence close to the junction 12R (Figure 9). A stretch of 56 bp had a similarity of 78% with a middle segment of the LINE repetitive element L1Rn (33), including 36 bp which were 90% similar. I also found that the sequence of a region of the plasmid p2.6 (that contains the original host site of 11.6, see Figure 4) is that of a rat LINEs, L1Rn (Figure 10). These results were to be expected since the probes derived from the cellular DNA at (S3) or near the junctions (S2, S4, S5, S6, S7, S8, and S9) all hybridized only with unique host DNA, with the exceptions of S6, S7, and S9 which hybridized to sequences present in about 5-10 copies per genome. Thus, my results do not support the hypothesis that integration occurs preferentially in host repetitive sequences.

Figure 10 p2.6 contains a rat LINE (L1Rn) sequence.

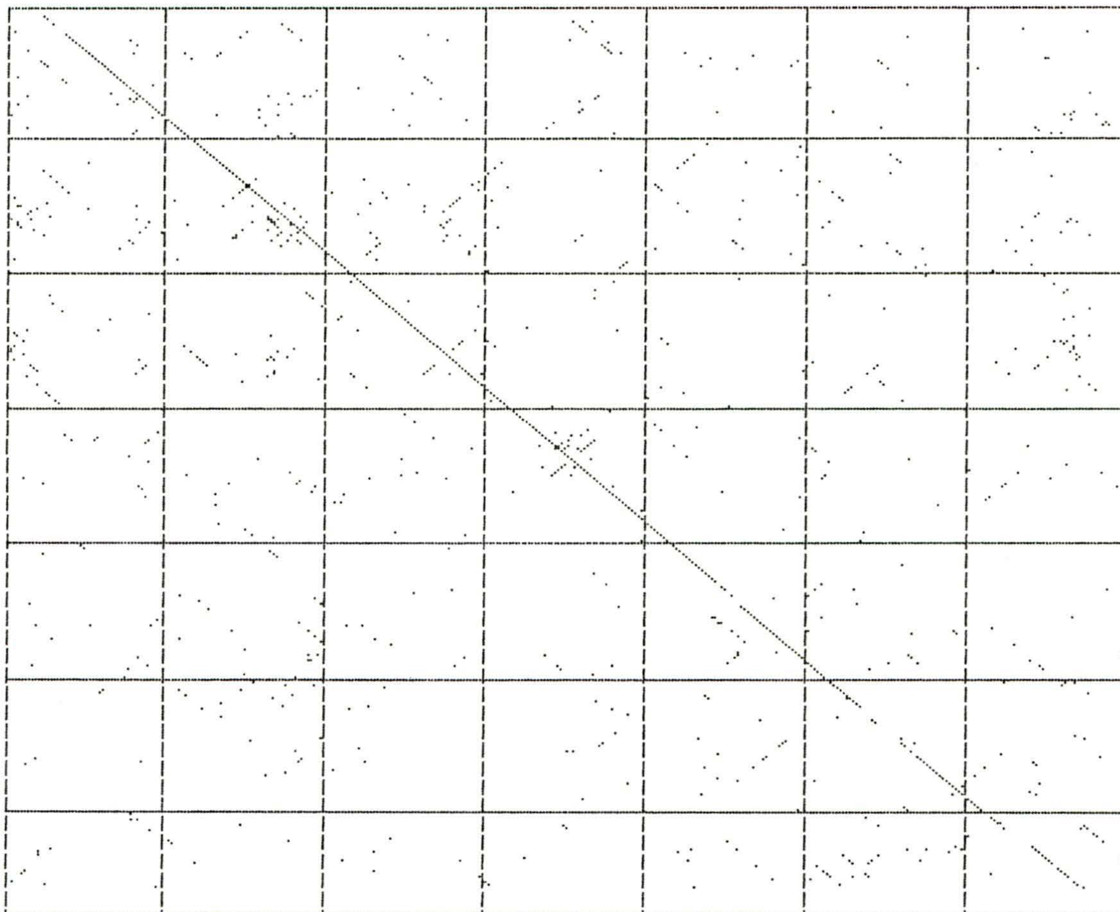
The sequence of a 339 bp region of p2.6 between, but not encompassing, the unique PvuII and the right EcoRI site (Figure 4) of the cloned IS fragment was determined (A. Nepveu and P. Chartrand, unpublished data). This sequence was compared to the sequence of L1Rn (D'Ambrosio et al, 1986) as described in Materials and Methods. The dot matrix comparison shown here was produced using the "Matrix" program of Zweig (1984). The horizontal sequence is that of L1Rn from bp 5750 to 6100, the vertical sequence is that of p2.6 (p2.8 in the Figure) as previously determined. The stringency was for sequence homology of 9 bases out of a window of 10 bases.

Figure 10

Horizontal ("A") axis:
L1Rn sequence.

Vertical ("B") axis:
p2.8 sequence.

Partial Matrix from A= 5750 to 6100,and B=1 to 339
Plot for Homology over 9 out of 10 bases. compression 1.



Rearrangements of the host sequences.

The cellular probes (S2, S3, S4, S5, S6, S7 and S9) previously described and used to determine the provenance of the cellular sequences flanking the insertions were used to deduce the physical maps of the host sequences contiguous to the integration sites before integration.

The restriction maps in the parental cell lines were established as follows. Total cellular DNA from the clone was digested with a restriction enzyme, the fragments separated by gel electrophoresis, blotted to membranes and then hybridized with one of the probes flanking the insertion in that clone. If the probe does not contain a restriction site for the enzyme used then it should detect two fragments: the fragment which has been rearranged by the RmI insertion and the unrearranged allele. These fragments can be distinguished by either of two methods. The unrearranged allele can be positively identified by comparison to the pattern produced by the same RE in FR3T3, the parental cell line, DNA in which only the unrearranged allele should be present, or the RE-digested clone DNA can be probed with a Py probe which hybridizes with the rearranged fragment only. The length of this rearranged fragment is used to map the restriction enzyme site in the flanking cellular DNA by measuring the distance from the known site in RmI. Then, in turn this site is used to establish the restriction map of the unrearranged allele.

In this way restriction maps of the unoccupied integration sites were constructed for both sides of W98.14 and 11.6 and for the right side of W98.12. As explained previously I could not map the unoccupied site for W98.12L since I did not have sufficient flanking cellular sequences in the fragment I had cloned in p12L. I could however establish a restriction enzyme map of the flanking cellular sequences by using a Py probe specific for the left side of the RmI insertion.

These maps are presented in Figure 11 along with the physical map of the flanking host sequences of the integration sites. In none of the three clones could I match the restriction maps established with the probe from one side with the restriction maps established for the other side.

W98.12

The restriction map for the unoccupied host sequences to the left of S2 was determined and compared to the map of the cellular sequences actually flanking W98.12 on the left (Figure 11). It is evident that the two maps do not coincide. The *KpnI* site nearest to the left of W98.12 is 15 Kbp away from the insertion site. The most distant *KpnI* site mapped to the left of S2 is 12 Kbp away. This cannot be the same *KpnI* site because of the differences in length and in the positions of surrounding restriction sites.

Figure 11 Maps of the cellular DNAs rearranged by the integration of RmI.

Beside the name of each clone (given on the extreme left) is the map of the cellular DNA after integration. (The integrated sequences of RmI are represented by wavy lines). Aligned above and below each map are the maps of the initial sites before integration as determined by Southern analysis using probes of cellular DNA, shown as solid bars. (These correspond to S2 in W98.12; S3 and S4 in W98.14; and S5 and S7 in 11.6; see Figure 4). Open bars: cellular DNA flanking the insertions. Solid lines: cellular DNA before integration. Hatched bars: cellular DNA flanking the insertion in 11.6 which has been reorganized (see text). -R- : initial site as mapped by a probe from the right flanking sequences. -L- : initial site as mapped by a probe from the left flanking sequences.

Restriction enzymes:

A, HpaI; B, BamHI; E, EcoRI; G, BglII; H, HindIII; I, BglI; J, AvaI; K, KpnI; M, SmaI; P, PstI; V, PvuII; X, XbaI. Not all restriction sites are shown.

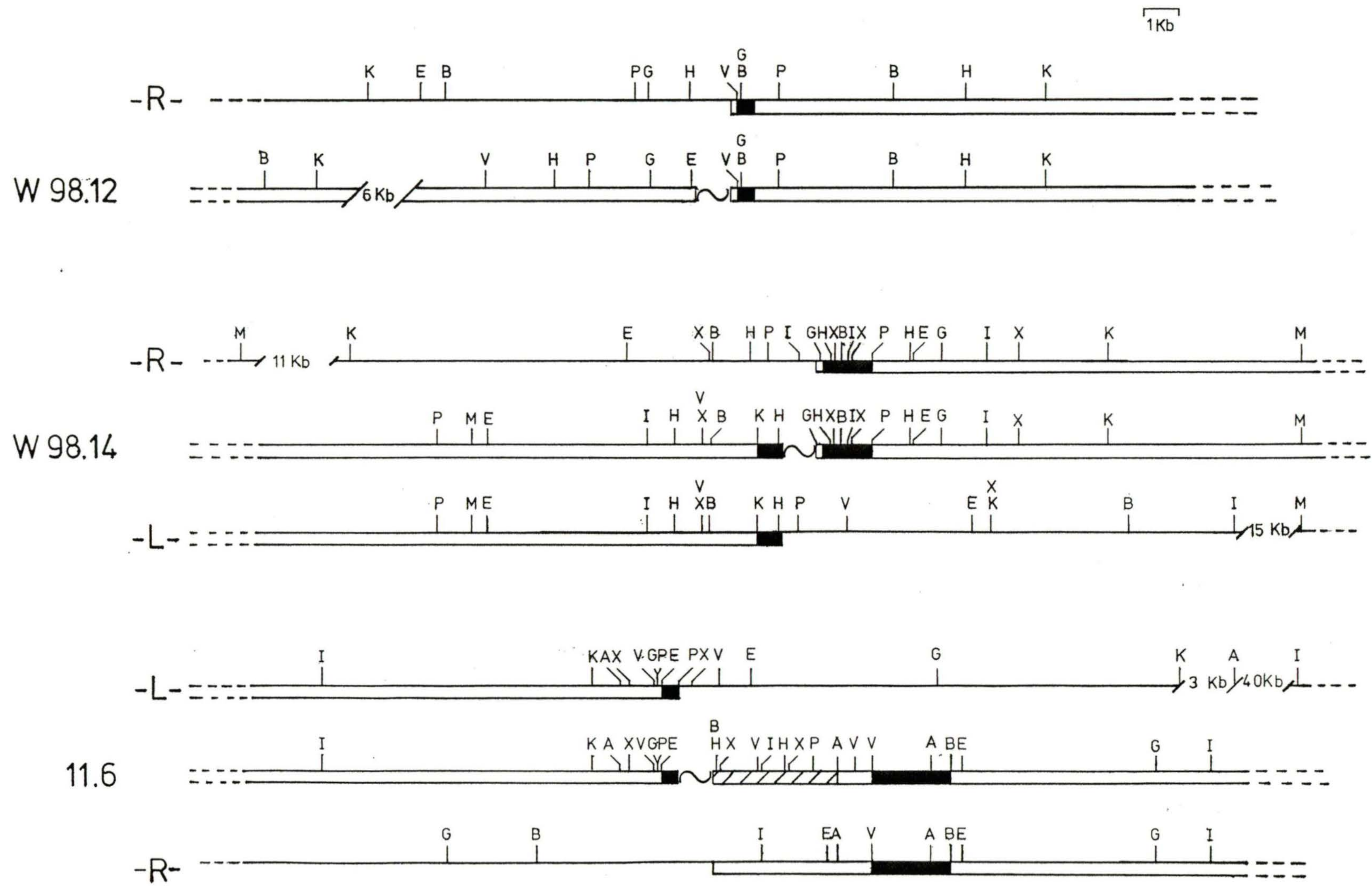


Figure 11

Therefore the minimum length of the rearrangement of host sequences in W98.12 is 12 Kbp.

W98.14

Our initial analysis of W98.14 led us to conclude that a rearrangement of at least 55 Kbp had been produced (Wallenburg et al, 1987). This conclusion was based on the fact that the restriction maps to the right of S3 and to the left of S4 were distinct and non-overlapping. Once the Field Inversion Gel Electrophoresis (FIGE) technique was developed we used it to determine more precisely the length of the rearrangement in W98.14. This showed us that S3, S4 and Py probes all hybridized to the same 59 Kbp HpaI fragment from W98.14 which represented the rearranged site, but also that S3 and S4 hybridized to the same 48 Kbp HpaI fragment in W98.14 (Figure 12) and FR3T3 (data not shown). Furthermore, it was found that S3 and S4 hybridized to the same 33 Kbp SmaI fragment in FR3T3, (but not in W98.14) and also hybridized to the same 40 Kbp fragment in W98.14/SmaI as did a Py probe (data not shown). The fact that S3 and S4 both hybridized to the same unrearranged allelic fragment with two different restriction enzymes (in FR3T3) proved that the DNAs flanking both sides of an insertion were linked before the integration event. It also obliged me to reevaluate the size and nature of the rearrangement in W98.14.

Any hypothesis for the rearrangement in W98.14 would have to explain the increase in size of the HpaI and SmaI fragments following integration: the unrearranged HpaI fragment measures 48 Kbp in both W98.14 and FR3T3, but following integration a 59 Kbp HpaI fragment is detected. In the case of SmaI the picture is more complex; before integration in FR3T3, S3 and S4 detect a 33 Kbp fragment, following integration in W98.14 they detect a 40 kbp fragment. However the 33 kbp fragment is no longer present in W98.14. What has happened to the unrearranged allele? Furthermore we have to take into account the fact that the restriction map determined from each side of W98.14 is totally different (Figure 11), as illustrated by the KpnI digest. W98.14 DNA digested with KpnI produces two fragments which hybridize with probe S3, one fragment at 3.1 Kbp and another at 6.5 Kbp, which comigrates with the unique FR3T3/KpnI fragment (Figure 12). Thus the KpnI site flanking RmI on the left is 3.1 Kbp from the RmI KpnI site and 0.7 Kbp from the end of the RmI insertion, (this is confirmed in the map of p14L; see Figure 4), and the S3 sequences originate from a 6.5 Kbp fragment. W98.14/KpnI also gives 2 fragments which hybridize with S4, one at 14 Kbp and another at 28 Kbp which comigrates with the unique FR3T3/KpnI fragment (Figure 12). Since S4 is 2.4 Kbp to the right of the RmI KpnI site, it is situated 10 Kbp from the right KpnI restriction site and 17 Kbp from the left restriction site of the 28 Kbp fragment (see Figures 4 and 13).

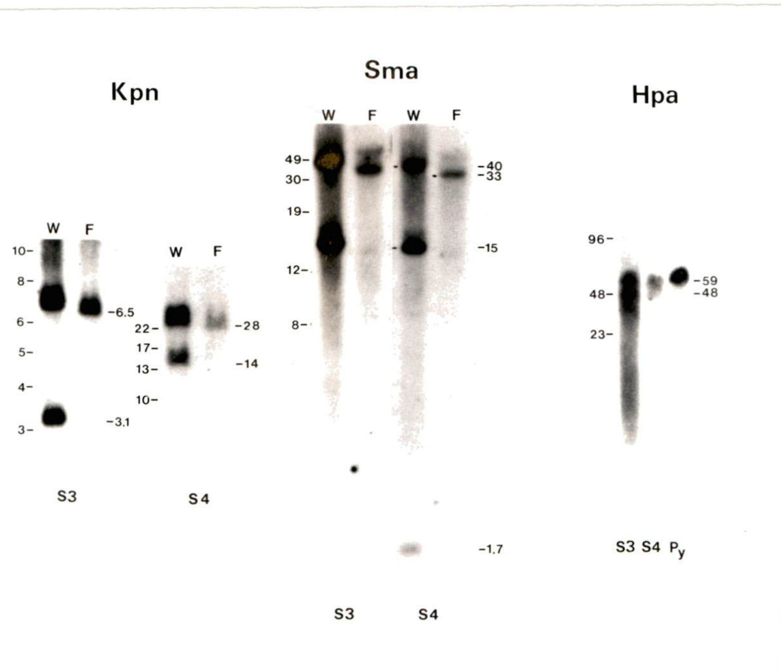


Figure 12. Restriction enzyme analysis of the rearrangement in W98.14.

Five micrograms of genomic DNAs (W=W98.14, F=FR3T3) were restricted with KpnI, SmaI or HpaI, electophoresed through 0.7% (KpnI) or 0.35% (SmaI) unique field agarose gels, or a 0.8% agarose FIGE gel (HpaI), transferred to nylon membranes and hybridized with S3, S4, or Py probes as indicated below each autoradiogram. Only W98.14 DNA was digested with HpaI. In the case of SmaI, the same membrane was first hybridized with S3, stripped, then rehybridized with S4. The positions of the molecular weight markers migrated through the same gels are indicated on the left of each autoradiogram and the sizes of the hybridizing fragments are indicated on the right. Sizes are given in Kbp.

The length of the RmI insertion is 7 Kbp. The increase in size of the _SmaI_ fragment is 7 kbp and of the _HpaI_ fragment is 11 Kbp. Considering the limits of sensitivity of the FIGE technique in molecular weight determination it is conceivable that in the case of _HpaI_ the real increase is 7 kbp. If this is true it would suggest that RmI conceivably integrated into the 33 kbp _SmaI_ fragment and the 48 kbp _HpaI_ fragment with little or no loss of cellular sequence. However, to then be able to explain the absence of an overlapping RE map, especially in the case of _KpnI_, we must invoke some rearrangement, which may be totally internal to the _SmaI_ and _HpaI_ fragments, but which does not result in a net gain or loss of cellular DNA, (i.e.) an inversion.

A working model, (Figure 13) which can explain only some of the data but which facilitates conceptualization by graphically displaying much of the data, has RmI integrating 9.5 Kbp from one end of the 28 Kbp _KpnI_ fragment with an inversion of the 20.5 Kbp of DNA immediately flanking the breakpoint. This brings the left-end _KpnI_ site to within .7 Kbp of the left end of RmI but in an inverted orientation. In the "normal" unrearranged position this DNA makes up part of a 6.5 Kbp _KpnI_ fragment adjacent to the 28 Kbp fragment. Thus S3 would originate from that 6.5 Kbp fragment. Obviously the RE pattern would be sufficiently changed to prevent recognition of the map.

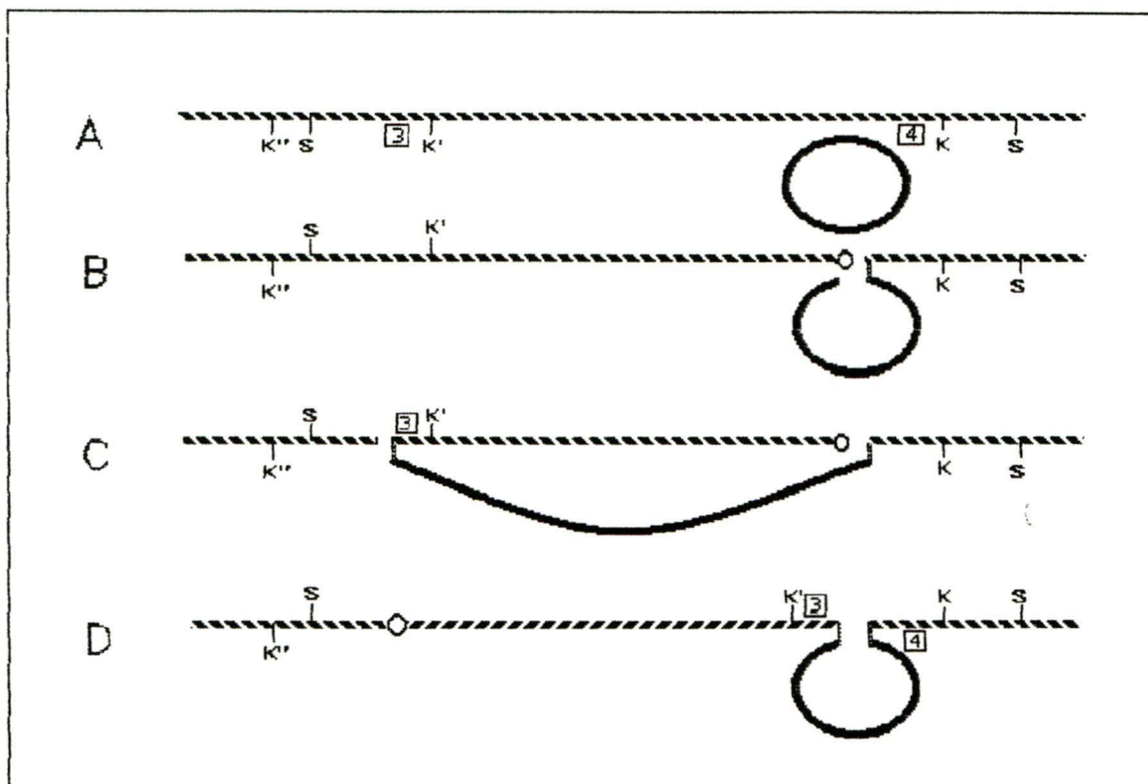


Figure 13

<u>Figure</u> <u>13</u> Potential model of the rearrangement in W98.14.

A graphic model which explains some of the observed rearrangements in W98.14 (see text). RmI is represented by a thick solid line or circle. The hatched line represents the genomic DNA, the inverted sense of the hatching indicates the inverted genomic fragment. The open boxes labelled '3' and '4' identify the sequences from which probe S3 and S4 originate. The distance in "A" between the <u>SmaI</u> sites is 33 Kbp; the distances between the <u>KpnI</u> sites <u>(K)</u> are 28 Kbp (K to K'), and 6.5 Kbp (K' to K"). The map is not drawn to scale.

The RmI recombines with one genomic sequence 9.5 Kbp from the right <u>KpnI</u> site (K) and another genomic sequence 20.5 Kbp to the <u>left</u> of the first crossover (and 0.7 Kbp to the left of K'). Following integration of RmI the intervening sequences are found in an inverted orientation. Probe S3, which originates from a 6.5 Kbp fragment and is 20 Kbp from S4 before integration is found separated from S4 only by the 7 Kbp of RmI following integration. The length of the <u>SmaI</u> fragment is unchanged. The <u>HpaI</u> sites (see text) are <u>external</u> to the crossover sites (<u>not</u> shown).

The SmaI and HpaI restriction sites are placed external to the inversion in such a model so that the respective increases in MW (approximately 7 Kbp) observed with these REs are respected (Figure 13).

However, neither this model, nor any "simple" model can explain all of the changes in W98.14. First, RmI contains a SmaI site within the Py sequences (Bourgaux et al, 1982). However S3, S4, and Py all hybridize to the same 40 Kbp fragment. This suggests that the SmaI site has been inactivated. If we presume that SmaI **does** cut between S3 and S4 within the Py sequences then (by coincidence) both flanking fragments measure 40 Kbp, for a total of 73 Kbp (80 Kbp minus the 7 Kbp RmI insertion) of cellular DNA between the two SmaI sites. This is more than the double of the original 33 Kbp SmaI fragment which contains both S3 and S4 and cannot be "simply" explained.

Secondly, hybridization of W98.14/SmaI with S3 and S4 reveals two fragments, one of which is the 40 Kbp rearranged fragment (Figure 12), which also hybridizes to a Py probe (not shown). However, we do not find the unrearranged 33 Kbp fragment identified in FR3T3. Both S3 and S4, but not Py, hybridize in W98.14 to a new 15 Kbp fragment, and S4 (but not S3) further hybridizes to a 1.5 Kbp fragment, none of these are present in FR3T3. The absence of an unrearranged SmaI allele is all the more surprising in light of the results presented below.

In W98.14 the sequences from both sides of the integration site appear to have been amplified to an equal extent, compared to the parental FR3T3 cell line, in an event unlinked to the integration. This is illustrated in Figure 14. DNA from W98.14 and FR3T3 was restricted with BglII and hybridized to probes S3, S4, and S5. The relative intensity of the resulting bands was determined by densitometry scanning (see Materials and Methods). S5 hybridized to a cellular band unrelated to the integration site in clone W98.14 and it was used to correct for small variations of intensity due to variation of DNA content in each lane. When the band intensities are corrected, we find that the S4 bands and the total of the S3 bands are twice as intense in the W98.14 clone than in the parental FR3T3. Thus there are twice as many copies of the flanking cellular sequences in the W98.14 clone than in its parent. But when we look at the integration site as illustrated by the lowest S3 band for which there is no corresponding band in FR3T3, we find it to be only half as intense as the upper S3 band in FR3T3. This tells us that the amplified copies are the allelic sequences and not the integration site. This amplification of the flanking sequences is also evident in Figure 8. A simple explanation for these results is that integration occurred fortuitously at a site that had been amplified compared to the parental sequences. Note that the amplification did not affect unrelated sequences (as judged by the S5 bands that are of equal intensities in both lanes), nor the integration site itself

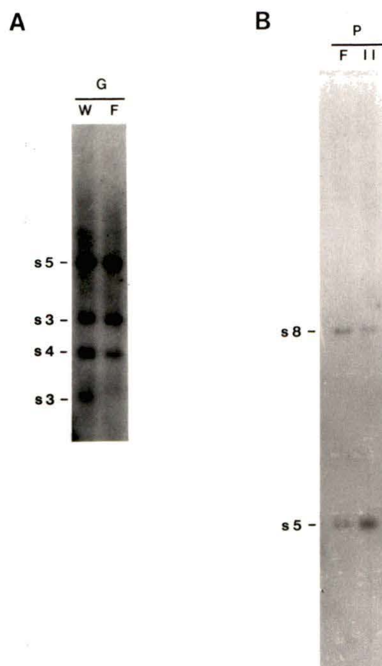


Figure 14. Duplication in W98.14 and deletion in 11.6.

A. Autoradiogram of cellular DNAs digested with BglII (G), run through an agarose gel, transferred to a nylon membrane and hybridized with probes S3, S4, and S5.
W, 5 micrograms W98.14 DNA; F, 5 micrograms FR3T3 DNA. The bands hybridizing with each probe are indicated. S3 hybridizes with two bands in W98.14. The upper band corresponds to unrearranged DNA and comigrates with an FR3T3 band, the lower band corresponds to the sequences flanking the RmI insertion.

B. Integration of RmI results in a deletion of cellular DNA. Five micrograms of 11.6 (11) and FR3T3 (FR) DNAs were digested with PstI (P) and migrated through an agarose gel, blotted and hybridized against probes S5 and S8 (Figure 4). S8 is a probe of cellular DNA from the initial site of 11.6.

judging by the intensity of the lower S3 band in Figure 14. Thus it is the allelic site that is amplified, either *in situ* or by chromosome polyploidy.

11.6

In clone 11.6 I was unable to find any unrearranged fragments which would hybridize to probes from both sides of the insertion. Digestion of 11.6 and FR3T3 DNAs with *BglI* generated a fragment that hybridized with S5 and that in low concentration agarose (.35%) unique field gels migrated much more slowly than 50 Kbp. I initially estimated that this fragment extended 90 Kbp to the right of S5 (Wallenburg *et al*, 1987). Since it did not hybridize with either S6, S9 or S7 the minimum rearrangement in 11.6 was estimated to be 90 Kbp. The use of FIGE to measure this *BglI* fragment more precisely, revealed that it was closer to 65 Kbp in length (Figure 15) and hence the minimum length of the rearrangement was recalculated to be at least 55 Kbp as follows:
S5 and Py probes both hybridize to a fragment of 13 Kbp in 11.6/*BglI* (Figure 15). This places the left flanking cellular *BglI* site 13 Kbp to the left of the RmI *BglI* site, or 10 Kbp to the left of S5. The 65 Kbp unrearranged *BglI* fragment to which S5 hybridizes thus extends 55 Kbp to the right of the insertion point. Since none of S6, S7 or S9 hybridize to this fragment, it follows that at least 55 kbp of DNA have been rearranged.

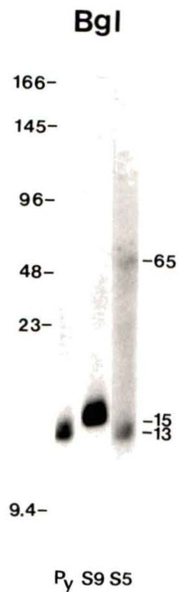


Figure 15. Rearrangement in 11.6.

Fifteen ug of 11.6 DNA were digested with BglI, separated into 3 aliquots of 5 ug each, migrated through a 0.8% agarose FIGE gel, transferred to a nylon membrane and hybridized to S5, S9, or Py probes as indicated below each autoradiogram. The position of the molecular weight markers (lambda DNA/HindIII, lambda concatamers, and phage T4 DNA) migrated through the same gel are indicated on the left, the sizes of the hybridizing fragments are indicated on the right . Fragment sizes are given in Kbp. S9 hybridizes with another fragment of 4.6 Kbp which had migrated off this gel (data not shown).

Fate of the rearranged DNA in 11.6.

Did integration result in deletion of host sequences? To answer this question DNAs from clone 11.6 and FR3T3 were digested with PstI, which separates the flanking unrearranged host DNA from the host DNA rearranged by the integration of RmI (Figure 4). I then hybridized with the S5 probe homologous to the unrearranged host DNA and the S8 probe (Figure 4) homologous to the rearranged host DNA (Figure 14). The autoradiogram was scanned using an LKB scanning densitometer and the S5 band was used to correct for variation of DNA content in each lane. After correction, the band hybridizing to the S8 probe was found to be only half as intense in 11.6 as in FR3T3. Since no other band hybridizing to the S8 probe is found in 11.6 DNA, the rearranged host DNA must have been deleted. This was confirmed by analyses with other restriction enzymes (data not shown).

The special case of 11.6R.

During the analyses of the host sequences flanking the insertion a number of observations served to distinguish the sequences flanking 11.6R. They are described here.

During the construction of the restriction map of 11.6R I discovered that the cellular sequences immediately flanking the insertion on the right had been rearranged

when compared to the host sequences before integration. The extent of the rearrangement was determined as follows. The BamHI site immediately flanking the junction should separate S6 and S9 from the RmI insertion, and therefore produce a fragment that comigrates with its allele and the FR3T3 BamHI fragment. The map of the cloned junction further shows no BamHI sites between S6 and S7 (Figure 4), so these probes should hybridize with the same BamHI fragments. In fact digestion with BamHI produces fragments in FR3T3 and 11.6 of 5.3 Kbp and a rearranged fragment at 7.0 Kbp in 11.6 when hybridized with S6 (data not shown). S7 hybridizes with this same 7.0 Kbp fragment but with 12 Kbp fragments in FR3T3/BamHI and 11.6/BamHI instead of the 5.3 Kbp fragment (data not shown) indicating the presence of a BamHI site between S6 and S7 in the unrearranged sequence which is absent from the flanking DNA at the rearranged site. Hybridization of 11.6/PvuII with S9 (Figure 16) detects two rearranged fragments only one of which hybridizes with Py (data not shown) as well as two fragments comigrating in FR3T3/PvuII. Although PvuII cuts the probe once this can only be explained by a rearrangement of the flanking host sequences in 11.6 compared to FR3T3. In the cloned junction (Figure 4) the single flanking PstI site which defines the distal end of S9, seperates the S7 probe from the RmI insertion. Thus S7 should not hybridize with a rearranged fragment in 11.6/PstI. As shown in Figure 16, S7 in fact hybridizes with a 13 Kbp rearranged fragment in 11.6/PstI, as well as

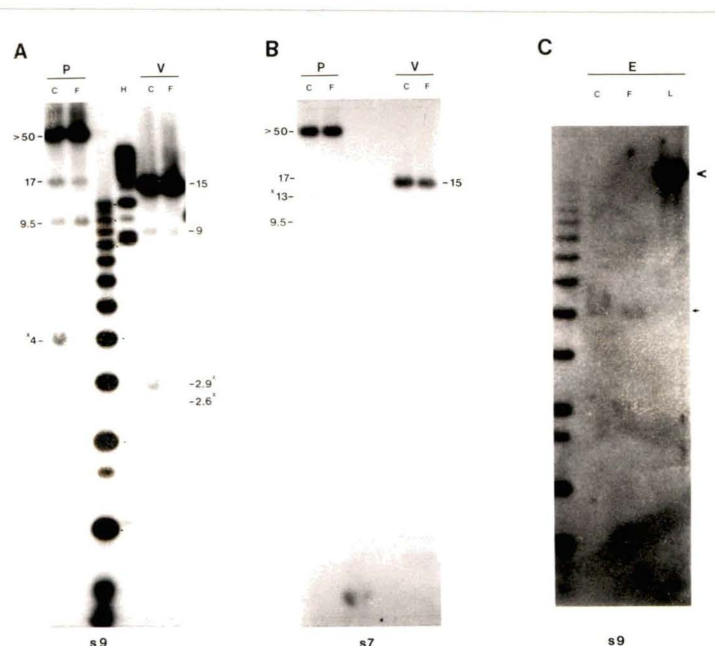


Figure 16. 11.6R

Five micrograms of genomic DNA from clone 11.6 (C) or from FR3T3 (F) or LTA (L) cells ere restricted with PstI (P), PvuII (V), or EcoRI (E), electrophoresed through agarose gels, tranferred to nylon membranes and hybridized with probes S7 or S9 (see Figure 4) as indicated below each autoradiogram. Fragments which have been rearranged by the RmI insertion in 11.6 are indicated with a superscript x.

A: S9 hybridizes with two rearranged fragments in 11.6/PvuII, at 2.6 and 2.9 kbp. Only the 2.6 kbp fagment hybridizes with a Py probe (data not shown). S9 also hybridizes with two unrearranged fragments of 9 and 15 kbp that are common to 11.6/PvuII, and FR3T3/PvuII. S9 hybridizes with a 4 kbp rearranged fragment in 11.6/PstI, but also with 3 unrearranged fragments also detected in FR3T3/PstI at 9.5, 17 and >50 kbp. Note that the 15 kbp PvuII and the >50 kbp PstI fragments are much more intense than the single copy rearranged fragments in 11.6.

B: The same membrane as in A: was stripped of the S9 probe and rehybridized with probe S7. S7 detects the same unrearranged fragments as S9 in both the PvuII and PstI digestions, however it also detects a 13 kbp rearranged fragment in 11.6/PstI. (The 9 kbp PvuII fragment is not visible in this exposition). Note that the relative intensity of hybridization of the unrearranged fragments is the same as with S9.

C: S9 hybridizes much more intensely with an EcoRI fragment in mouse LTA DNA (indicated by the arrowhead) than in either 11.6 or FR3T3 rat DNAs (indicated by the small arrow).

with 3 fragments also found in FR3T3/PstI (see below). Proceeding with the same type of analysis for subsequent restriction enzymes, moving away from the junction, it was evident that both the S6 and S9 probes were fragments which had been rearranged. However using S7 I mapped an HpaI site ~3.5 Kbp from the junction which was outside the rearrangement (data not shown). Furthermore, probes S6 and S9 hybridized with the same 16 Kbp FR3T3/PvuII, and the same FR3T3/PstI fragments as S7 (Figure 16) as well as the same 14 Kbp FR3T3/BglII, 17 Kbp FR3T3/ClaI and 16 Kbp FR3T3/SstII fragments as did S7 (data not shown). The latter 3 enzymes do not cut in the flanking sequences. This shows that S6 and S9 were contiguous with S7 before integration. Thus as a result of the integration (less than) 3.5 Kbp of host DNA immediately flanking the integration site were rearranged. The exact nature of the rearrangement has not been elucidated.

Several lines of evidence lead me to believe that the sequences flanking 11.6R are present in more than two copies per diploid genome. First, it should be noted that probes S6, S9,and S7 hybridize with sequences that appear to be present in greater than two copies per genome (Figures 8 and 16). When the enzymes used do not cut inside the probe the intensity of the rearranged band in 11.6 corresponds to a single gene copy. From the relative intensities of these single copy bands compared to their alleles, I estimate that there are approximately 5 to 10

copies of this DNA per FR3T3 genome. Secondly, hybridizations of 11.6 and FR3T3 DNAs digested with restriction enzymes PstI and PvuII (Figure 16), and BglII, EcoRI, KpnI, and HpaI with S6, S9 and S7 consistently led to the illumination of more bands than expected. The intensity of hybridization with these bands was often weak. This sits well with the hypothesis that they are related loci sharing degenerated homology with the probes. This would also explain the differences in the restriction enzyme positions. Finally, S7 and S11 hybridize more strongly to sequences present in mouse (LTA) DNA than to fragments of 11.6 and FR3T3 DNA (Figure 16), yet S7 and S11 are rat DNA (remember that the carrier DNA used in the generation of 11.6 was salmon sperm DNA). The fact that "unique" sequence DNA from the rat is highly conserved and amplified in the mouse further supports the hypothesis that the sequences flanking 11.6R are present in several (degenerate?) copies in the rat.

In retrospect, the search for repetitive sequences may have detected the presence of this low copy repeat DNA, since the fragments containing the 11.6R junction (and therefore parts of S9 and S6) or S7 had occasionally hybridized with the FR3T3 probe, yet none of the three 11.6R probes (S9, S6, or S7) produced a smear when hybridized to digested FR3T3 DNA, indicating that they do not contain highly repetitive DNA. Another possible explanation lies in the different conditions used for the

hybridizations. If S6, S7, and S9 contain highly repetitive sequences which are highly divergent from the prototype sequence, then they may have been detectable in the conditions of lower stringency used to screen for unique and repetitive sequences but would not produce a smear in the conditions of high stringency used to analyze the source of the flanking DNA. In this case I have to presume that at least five copies of this DNA in the rat genome have retained a high degree of similarity.

DISCUSSION

I had previously shown that FR3T3 cells transformed by RmI at 39°C (W9 clones) have a nonrandom pattern of integration involving preferentially the repetitive sequences of RmI and that integration had probably involved circular molecules (Wallenburg et al, 1984). The objective of this study was to examine the viral-cellular junctions at the molecular level in the hopes of determining the cause for the involvement of the RmI repetitive sequences and also to probe further the integration mechanism.

It is clear that the integration events looked at (Figure 9) were not the result of homologous recombination, nor site-specific recombination but rather are the product of illegitimate recombination. This leads me to postulate that the overrepresention of the repetitive sequences in the crossover events is the result of their acting as "hotspots" for illegitimate recombination. This is not the first time that SINEs sequences have been proposed as hotspots for illegitimate recombination (Rogers, 1981 and 1985) and could explain why repetitive sequences are often found at illegitimate recombination junctions such as translocation breakpoints (Heisterkamp et al, 1985; Lehrman et al, 1985; Vanin et al, 1983) or in small polydispersed circular (spc) DNA (cellular excision products, Hollis and Hindley, 1986; Krolewski et al, 1984; Riabowol et al, 1985). It also raises the interesting possibility that if

reintegration of spc DNA in the host genome were to occur it would preferentially occur via the repetitive sequences. This would preserve the integrity of the unique DNA present in these spc molecules.

Kato *et al* (1986) suggested that integration into the genome happened preferentially in genomic repetitive sequences. This can be interpreted as indicating that genomic repetitive sequences can also act as hotspots for illegitimate recombination. In that study Camerini-Otero and co-workers define repetitive DNA as any sequence with greater than 5-10 copies per genome, and not only highly repetitive families. In clone 11.6 the cellular sequences flanking the right junction into which RmI integrated are present in 5-10 copies per genome, and thus qualify as repetitive DNA by those standards, and would seem to support the hypothesis of Kato *et al* (1986). However, the sequences flanking the other 4 junctions were unique. Since the recombination event in each of those cases had occurred in the repetitive sequences of RmI it is possible that this obviated the need for the integrating DNA (RmI in this case) to "seek out" repetitive sequences on the chromosome. In other words, the hotspot effect would be mediated by repetitive sequences in either the donor or the recipient.

Illegitimate recombination breakpoints have been previously associated with short "patchy" homologies

between the recombining sequences, and both topoisomerase I and topoisomerase II recognition sites. None of the cross-overs mapped here was found to have significant homology between the recombining sequences nor to be near topoisomerase I or topoisomerase II recognition sites (data not shown). In fact, the region of RmI the richest in topoisomerase II sites is the region the most under-represented in crossover events (data not shown).

Recombination hotspots previously described include the prokaryotic chi site in lambda phage (Smith, 1983; Lam et al, 1974), and sites in fungi (Angel et al, 1970; Macdonald and Whitehouse, 1979; Gutz, 1971), at the major histocompatibility (MHC) locus in mice (Kobori et al, 1986; Steinmetz et al, 1986) and a fragment from the human B-globin locus that has been shown to stimulate recombination in S. cerivisiae (Treco et al, 1985). These hotspot sequences enhance recombination between homologous sequences by a mechanism which appears to be mediated by the cutting action of a recombination endonuclease which specifically recognizes those (hotspot) sequences (Ponticelli et al, 1985; Taylor et al, 1985; Markham and Whitehouse, 1982). A mitotic hotspot in yeast (HOT 1) has also been localized (Voelkel-Meiman et al, 1987), it too stimulates homologous recombination, although probably mediated by a transcription stimulating mechanism.

These recombination hotspots stimulate recombination not necessarily at the site itself but in the region in which they are found, in a directional or non-directional manner. The hotspot activity of RmI does not increase the overall rate of integration (see below). However, as described in the introduction, more than one region of RmI is overrepresented in the crossings-over: the region containing the repetitive sequences and the region immediately flanking that region from Py m.u. 30 to 40. (Wallenburg, 1984; Wallenburg et al, 1984). I do not know why the latter region was overrepresented in recombination events, however one possible explanation is that the hotspot activity of the repetitive sequences of INS extends directionally into the flanking Py sequences.

Even though the host genome contained upwards of 10^5 endogenous repetitive sequences homologous to the ones present in RmI available as potential targets for homologous integration, we found no evidence that RmI integrated into these sites by homologous recombination. Obviously the small number of events (6) looked at does not exclude that such events possibly occurred. However the transforming efficiencies of RmI and tsP155 are the same (N. Gusew, unpublished results). If the presence of the repetitive sequences procured an independent efficient integration pathway to RmI via homologous recombination, we might expect RmI to be a more efficient transformer than tsP155. This presumes of course that the two integration/

recombination pathways are independent and do not share a rate-limiting step. If however the pathways (illegitimate and homologous) do compete for a rate-limiting step, then the end result rate of integration (and therefore transformation) would be unchanged. However the nature of the integration events should reflect that of the most efficient pathway. Since RmI does not transform more efficiently than tsP155 and all of the junctions examined were the result of illegitimate recombination it can be concluded that homologous integration is not an efficient procedure in the case of RmI.

As described in the introduction, mammalian cells have been shown to be quite proficient in homologous recombination, the chromosomal sequences have been shown to be available targets and RmI has over 10^5 homologous endogenous targets. Why then is there no evidence of homologous integration? Several groups using intramolecular homologous recombination assays have examined the role of homology length on homologous recombination, and have observed recombination between sequences sharing from 5 Kbp (Subramani and Berg, 1983) to as little as 46 bp (Lin et al, 1984), however the efficiency of the recombination decreases rapidly when the total length of homology between the recombining partners drops below 200-300 bp (Chang and Wilson, 1987; Rubnitz and Subramani, 1984; Subramani and Berg, 1983). The lengths of the repetitive DNA sequences in RmI are 120 bp of B2 and 370 bp of MT. Furthermore the

individual members of the B2 family share only 70% homology, while members of the MT family share 90% homology (Bastien and Bourgaux, 1987), and a recent study has indicated that stretches of uninterrupted homology are necessary for efficient (intrachromosomal) recombination (Waldman and Liskay, 1988). Thus, it is possible that the length and degree of homology between the repetitive sequences of RmI and those of the host DNA were insufficient for homologous integration.

On the other hand it is possible that target abundance has no influence on the frequency of homologous integration. When Steele et al (1984) tried to target the 200 copy rDNA sequences of mouse L cells they did not detect homologous integration nor an increase in integration frequency. Zheng and Wilson attempted to target an adenosine deaminase (ADA) gene in a cell line in which this gene was specifically amplified 4,300-fold relative to its parental cell line, Cl-1D (personal communication). They transfected or electroporated cells with substrates with sequence homologies of 1.6, 2.0 and 4.3 Kbp with the ADA gene and which contained the HSV-TK gene as marker. More than 50 HAT resistant clones were screened for a set of restriction fragments diagnostic of targeted recombination, but no targeting events were detected. These authors conclude that targeted integration does not depend upon abundance of the target gene. However neither of these studies is conclusive. Both targeted specifically

amplified genes present at a single locus in the host, and depend absolutely on the efficient expression of the HSV-TK gene. The HSV-TK gene which is normally transcribed by RNA polymerase II may be inactive when integrated into the rDNA sequences which are transcribed by RNA polymerase I, and not all copies of the 4,300 ADA repeats are necessarily active. Moreover a single locus may serve as a single target regardless of the number of copies of the gene present at that locus.

Thomas _et al_ (1986) microinjected varying numbers of defective neo genes into established mouse L-cell lines containing 1, 4 "independent" or 5 head to tail tandem copies of a complementary defective neo gene, integrated at different sites in the genome. They did not observe any change in targeting frequency, and argue therefore that targeting is neither dependent on the concentration of the input plasmid nor on the number of endogenous copies. However the number of clones examined (3) and the number of targeted transformants (1-3 per experiment) were low, and small changes in frequency, within or between cell lines, would have been difficult to detect. Furthermore, it was not established that the four "independent" copies were at distinct loci.

Contrary to the previously described assays, RmI's homologous targets were not confined to a single locus but are highly dispersed throughout the genome and the copy

number was significantly greater (10^5 vs 4, 200, or 4300). The RmI-B2 sequences are relatively short, but the RmI-MT sequences measure 370 bp, well within the range within which intramolecular homologous recombination is efficient. Furthermore recombination between endogenous chromosomal SINEs of similar length has been shown to occur (Lehrman et al, 1985; Lehrman et al, 1986; Krolewski et al, 1984; Shmookler Reis et al, 1983). Based on these observations we suggested (Wallenburg et al, 1987) that integration was independant of target abundance and that the initial step in targeting, that of finding the homologous sequences is not rate-limiting. This would explain why the frequency of homologous integration in mammalian cells is very low compared to bacteria and yeast, even when genome size is taken into account (Lin et al, 1985; Smithe and Berg, 1984; Smithies et al, 1984; Thomas et al, 1986). Thomas et al (1986) came to the same conclusion based on their results, and go on to suggest that later steps such as heteroduplex formation, or Holiday junction resolution must be limiting.

It has been postulated that linear DNA transfected in the presence of carrier forms a high molecular weight structure designated transgenome or pekelosome and that it is these structures that integrate into the host genome (Perucho et al, 1980; Scangos et al, 1981). Transgenomes are the product of ligation between carrier and the transfected DNA and result in molecules of several hundred or thousand Kbp in size. If integration of transfected DNA

was necessarily mediated by transgenomes then this would preclude drawing conclusions on the interaction of specific exogenous DNA with the host genome. My results show that RmI integrated directly into the rat cellular DNA without passing by the intermediary of a transgenome. This (integration of a circular vector directly into the host genome) has also been observed elsewhere (Steele et al, 1984) and shows that transgenomes are not necessary intermediates for the integration of circular DNA even when transfected in the presence of carrier DNA. These results also support the hypothesis presented in the introduction that circular DNA integrates via a different pathway than linear DNA and is not simply converted into linear DNA before integration.

A number of authors have reported that integration of exogenous DNA resulted in genome rearrangements (Allard et al, 1988; Botchan et al, 1980; Hayday et al, 1982; Kato et al, 1986; Neer et al, 1983; Stringer, 1982; Williams and Fried, 1986). However, with one exception (Neer et al, 1983), the extent of these rearrangements had not been determined nor had it been established if host cell sequences were lost as a result of integration. In all cases, (except one; Kato et al, 1986) what has been determined is that the minimum length of the rearranged host sequences is equal to or smaller than the integrated DNA. In the one exception Kato et al (1986) estimated a rearrangement of at least 17 kbp for an insert of about 8

kbp. This was the longest rearrangement thus far described.

As discussed in the introduction initial models of integration of exogenous DNA were based on the papovavirus paradigm and proposed the integration of linear molecules via double crossings-over leading to a "simple" replacement of host sequences by the incoming molecule and creating a deletion of host sequences of similar length as the insertion (Figure 1). With the possible exception of the results of Kato _et al_ (1986), the results described above are not in conflict with this model. My analysis of the integration site of clone 11.6 establishes that indeed integration can result in the deletion of host sequences. However the extent of the rearranged host sequences, which can be more than 55 Kbp for an insert of 6 Kbp (as in 11.6) is incompatible with a linear insertion-replacement model as suggested above. Such extensive rearrangements are more easily explained by an integration mechanism in which the incoming exogenous DNA replaces a looped structure of host cellular DNA (Wallenburg _et al_, 1987).

In support of this model are results which show that eukaryotic chromatin is organized into large loops or domains the bases of which are associated with the nuclear matrix or scaffold (Benyajati and Worcel, 1976; Berezney and Coffey, 1974; Cook and Brazell, 1975).

These matrix association regions (MARs) or scaffold attachment regions (SARs) are evolutionarily conserved, are A+T rich, and have been localized near regions of active chromatin, enhancer regions and, recently, with replication origins in yeast. Significantly, they have also been shown to contain topoisomerase II recognition and putative cleavage sites (Jarman and Higgs, 1988) and Sperry et al (1989) demonstrated that the two MARs they studied specifically bound topoisomerase II. Topoisomerase II is a major structural component of the matrix where it is associated with the bases of the chromatin loops (Berrios et al, 1985; Earnshaw et al, 1985; Earnshaw and Heck, 1985; Gasser et al, 1986; Lewis and Laemmli, 1982). Topoisomerase II protein levels and activity have been shown to fluctuate during the cell cycle being most abundant during DNA replication and low in non-dividing cells (Heck et al, 1986), and it has been shown that the "fixation" of the transformed state in transfected cells requires that the cells undergo at least one division (Topp et al in Tooze, 1980) possibly because high levels of topoisomerase II are required for integration. Furthermore, topoisomerase II is known to create double-strand breaks in DNA in vivo and has been shown to mediate illegitimate recombination in vitro (Bae et al, 1988) and in vivo (Desautels, L., Gusew, N., Ross, W., and P. Chartrand, manuscript in preparation). Finally, their examination of three naturally occurring illegitimate recombination events led Sperry et al (1989) to propose that a dysfunction of MARs is illegitimate

recombination (mediated presumably by topoisomerase II cleavage).

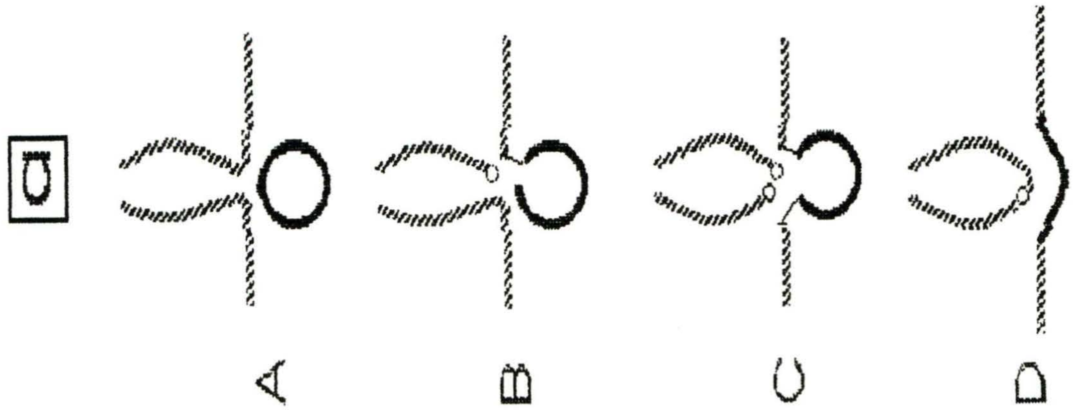
This information immediately suggests a model for integration in which the incoming DNA interacts with the host chromatin at the base of a structural loop, and an exchange reaction mediated by the topoisomerase II present integrates the exogenous DNA into the chromosome (Figure 17). Depending on the sites of the exchanges and the partners involved, this model can easily account for and predicts that deletions and inversions of the host DNA will occur during illegitimate integration, (Figure 17 a and b). It predicts "conservative" integration as well, in which no cellular DNA is lost or rearranged, if the exchanges are limited to the two ends of the cellular DNA generated by the first break (Figure 17 c). Chromatin loops are hypothesized to be formed on the basis of functional or replicative domains and are highly variable in size. Since only the minimum lengths of rearrangements have been determined so far, both here and in the literature, we cannot draw conclusions on the uniformity or non-uniformity of the rearrangement lengths; this model predicts, however, that they reflect the high variability of the chromatin loop sizes.

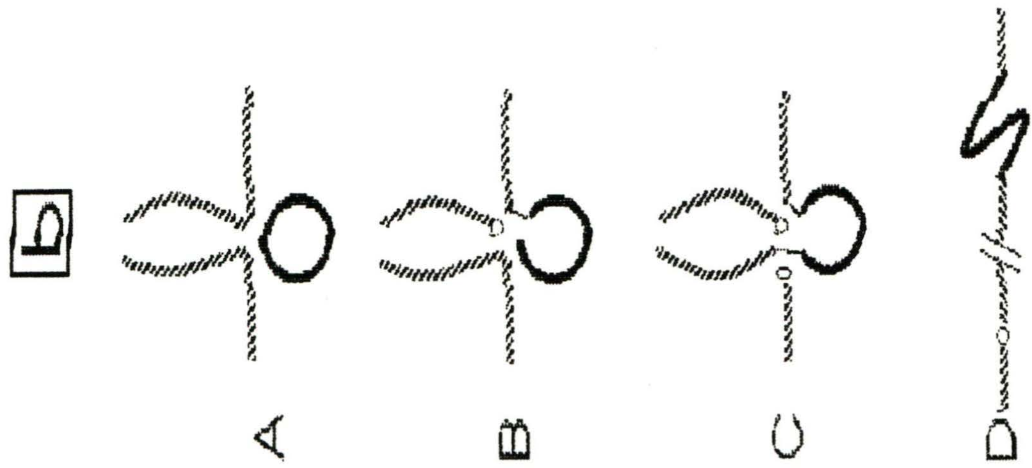
Figure 17. Model for the integration of exogenous DNA.

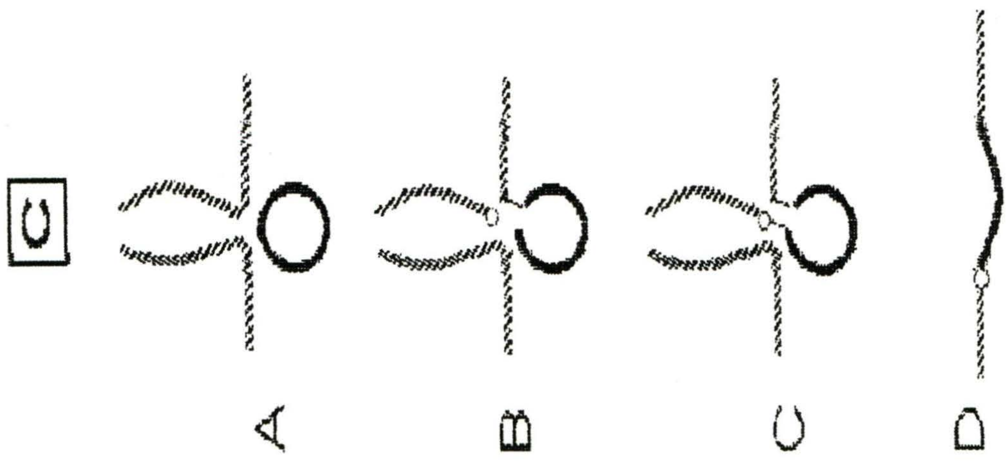
A. The genomic chromatin (represented by a hatched line) forms a loop the bases of which (matrix associated regions, MARs) are associated with the nuclear matrix. The association is mediated by topoisomease II which recognizes and fixes the MARs sequences. The exogenous DNA is represented by a solid circle.
B. The exogenous DNA recombines with the genomic chromatin at the base of a loop through a dysfunction of topoisomerase II (subunit exchange model, Ikeda et al, 1982), resulting in a double-strand break in both the chromatin and the incoming exogenous DNA. (The small open circles serve only to identify "free" ends of chromatin).
C. The "free" end of the incoming exogenous DNA recombines a second time with the genomic DNA.
D. Ligation of "loose" ends and resolution.

The consequences of the integration differ depending on the site of the second recombination in 'C'.
a. If the exogenous DNA recombines with the other base of the chromatin loop, the loop of chromatin is deleted. Ligation of the "free" ends of the liberated chromatin loop (in D) leads to formation of a free circular molecule of chromosomal DNA which is eventually lost from the cell.
b. If the exogenous DNA recombines with the loop of the chromatin at the other base, following ligation and resolution, the DNA within the loop is found in an inverted orientation (shown by the inverted sense of hatching), (D).
c. Finally, if the exogenous DNA recombines with the "free" end of chromatin generated by the first crossover, then integration results without any rearrangement of the genomic DNA.

Figure 17







Consistent with the results described here which show that RmI integrated directly into the host genome, and suggest that it integrated as a circular molecule without passing through a linear intermediate, the integrating substrate is presented in the model as a circular molecule interacting directly with the genome.

The role of topoisomerase II on the integrating molecule is unclear. Since virtually any foreign DNA can be integrated, the presence of topoisomerase II 'consensus' sites is certainly not essential. As mentioned previously the region of RmI the richest in topoisomerase II consensus sequences is the region the most underrepresented in crossover events. However it should be noted that although the topoisomerase II consensus sequence is highly degenerate ($GTN^{A}_{T}A^{C}_{T}ATTNATNN^{A}_{G}$) no perfect matches were mapped in RmI, 15 sites with greater than 86% homology (13/15 bases) were mapped. Furthermore this sequence was established for the *Drosophila melanogaster* topoisomerase II enzyme (Sander and Hsieh, 1985), and although eukaryotic topoisomerases II have very similar activities, some differences in cutting site preference have been noted. Finally, even though the consensus sequence defines preferential topoisomerase II cutting sites determined *in vitro*, topoisomerase II is known to cut DNA at other sites (Sperry *et al*, 1989). Thus it is possible that topoisomerase II also cuts the integrating substrate, even in the absence of apparent topoisomerase II sites.

If the presented model can account for some characteristics of the observed integrations, it is certainly not suggested to be exclusive of other mechanisms. The results presented here for both clones 11.6 and W98.14 indicate that there have been multiple rearrangements of the cellular DNA on one side of the integration site; similar observations have been made before (Allard _et al_, 1988; Botchan _et al_, 1980; Brenner _et al_ 1984; Hayday _et al_, 1982; Ruley and Fried, 1983). Allard _et al_ (1988) found that the cellular sequences flanking the provirus in a Py transformed mouse cell line (Cyp) originated from two cellular sites that are separated in the normal mouse genome by at least 6 Kbp of DNA, and several groups have detected short insertions of "linker" DNA of unknown origin in the sequences across illegitimate recombination junctions (Botchan _et al_, 1980; Brenner _et al_ 1984; Hayday _et al_, 1982; Ruley and Fried, 1983). These results suggest that there is probably added complexity to the mechanism beyond what is proposed in the present model.

ACKNOWLEDGEMENTS

I would like to thank my thesis director Dr. Pierre Chartrand for his constant encouragement and stimulation. I also wish to show my appreciation of his 'grande disponibilité', not to mention his great patience.

I wish to thank Nadine Gusew for her help with a number of techniques especially the laborious preparation of the *in vitro* packaging extracts, and in particular for having cloned, then subcloned, the right junction of 11.6.

I thank Gilles Roy for his constant sound advice as well as for his invaluable help in writing the FIGE program, and for fabricating the switch box.

I would like to thank Maryse Chalifoux, Johanne Proulx, and Roxanne Lussier, for their generous technical assistance.

I thank Dr. Vivianne Bibor-Hardy for her help with the densitometry scanning and Chantal Turmel and Dr. Marc Lalande for the help they gave me in setting up the FIGE technique.

I would like to thank a number of the members of the laboratory, past and present, for their stimulation,

encouragement and friendship; Alain Nepveu, Pierre Trudel, Abdellah Belmazza, and Suzanne Brouillette.

I thank Drs. Ted Bradley, Claude Asselin and Pierre Bourgaux for having accepted to review this thesis. Their comments and criticisms were greatly appreciated.

I thank the 'Fonds pour la Formation de Chercheurs et l'Aide à la Recherche' for their financial support.

Finally, I profoundly thank Marie, Marika, and Catherine, to whom this thesis is dedicated, for their support and tremendous patience.

BIBLIOGRAPHY

Allard, D., L. Delbecchi, D. Bourgaux-Ramoisy, and P. Bourgaux. (1988) MAJOR REARRANGEMENT OF CELLULAR DNA IN THE VICINITY OF INTEGRATED Py DNA. Virology 162; 128-136.

Amati, B.B., and S.M. Gasser. (1988) CHROMOSOMAL ARS AND CEN ELEMENTS BIND SPECIFICALLY TO THE YEAST NUCLEAR SCAFFOLD. Cell 54: 967-978.

Angel, T., B. Austin, and D. G. Catcheside. (1970) REGULATION OF RECOMBINATION AT THE *his-3* LOCUS IN *Neurospora crassa.* Aust. J. Biol. Sci. 23: 1229-1240.

Ayares, D., L. Chekuri, K-Y. Song, and R. Kucherlapati. (1986) SEQUENCE HOMOLOGY REQUIREMENTS FOR INTERMOLECULAR RECOMBINATION IN MAMMALIAN CELLS. Proc. Natl. Acad. Sci. USA 83: 5199-5203.

Bae, Y.-S., I. Kawasaki, H. Ikeda, and L.F. Liu. (1988) ILLEGITIMATE RECOMBINATION MEDIATED BY CALF THYMUS DNA TOPO II in vitro. Proc. Natl. Acad. Sci. USA 85: 2076-2080.

Bandyopadhyay, P.K., S. Watanabe, and H.M. Temin. (1984) RECOMBINATION OF TRANSFECTED DNAs IN VERTEBRATE CELLS IN CULTURE. Proc. Natl. Acad. Sci. USA 81: 3476-3480.

Bastien, L. and P. Bourgaux. (1987) THE MT FAMILY OF MOUSE DNA IS MADE OF SHORT INTERSPERSED REPEATED ELEMENTS. Gene 57: 81-88.

Been, M.D., and J.J. Champoux. (1981) DNA BREAKAGE AND CLOSURE BY RAT LIVER TYPE I TOPOISOMERASE: SEPARATION OF THE HALF-RECTIONS BY USING A SINGLE-STRANDED DNA SUBSTRATE. Proc. Natl. Acad. Sci. USA 78: 2883-2887.

Benton, W.D., and R.W. Davis. (1977) SCREENING gt RECOMBINANT CLONES BY HYBRIDIZATION TO SINGLE PLAQUES *in situ.* Science 196: 180.

Benyajati, C., and A. Worcel. (1976) ISOLATION, CHARACTERIZATION, AND STRUCTURE OF THE FOLDED INTERPHASE GENOME OF *Drosophila melanogaster.* Cell 9: 393-407.

Berezney, R., and D. Coffey (1974) IDENTIFICATION OF A NUCLEAR PROTEIN MATRIX. Biochem. Biophys. Res. Commun. 60: 1410-1419.

Berrios, M., N. Osheroff, and P.A. Fisher. (1985) *In situ* LOCALIZATION OF DNA TOPOISOMERASE II, A MAJOR POLYPEPTIDE

COMPONENT OF THE *Drosophila* NUCLEAR MATRIX FRACTION. Proc. Natl. Acad. Sci. USA 82: 4142-4146.

Birg, F., R. Dulbecco, M. Fried, and R. Kamen. (1979) STATE AND ORGANIZATION OF Py DNA SEQUENCES IN TRANSFORMED RAT CELL LINES. J. Virol. 29: 633-648.

Blackwell, T.K., M.W. Moore, G.D. Yancopoulos, H. Suh, S. Lutzker, E. Selsing, and F.W. Alt. (1986) RECOMBINATION BETWEEN IMMUNOGLOBULIN VARIABLE REGION GENE SEGMENTS IS ENHANCED BY TRANSCRIPTION. Nature 324: 585-589.

Botchan, M., J. Stringer, T. Mitchison, and J. Sambrook. (1980) INTEGRATION AND EXCISION OF SV40 DNA FROM THE CHROMOSOME OF A TRANSFORMED CELL. Cell 20: 143-152.

Botchan, M., W. Topp, and J. Sambrook. (1976) THE ARRANGEMENT OF SV40 SEQUENCES IN THE DNA OF TRANSFORMED CELLS. Cell 9: 269-287.

Bourgaux, P., B. Sylla, and P. Chartrand. (1982) EXCISION OF POLYOMA VIRUS DNA FROM THAT OF A TRANSFORMED MOUSE CELL: IDENTIFICATION OF A HYBRID MOLECULE WITH DIRECT AND INVERTED REPEAT SEQUENCES AT THE VIRAL-CELLULAR JOINTS. Virology 122: 84-97.

Bourgaux-Ramoisy, D., D. Gendron, P. Chartrand, and P. Bourgaux. (1986) AN EXCISION EVENT THAT MAY DEPEND ON PATCHY HOMOLOGY FOR SITE SPECIFICITY. Mol. Cell. Biol. 6: 2727-2730.

Brenner, D.A., S. Kato, R.A. Anderson, A.C. Smigocki, and R.D. Camerini-Otero. (1984) THE RECOMBINATION AND INTEGRATION OF DNAs INTRODUCED INTO MOUSE L CELLS. Cold Spring Harbor Symp. Quant. Biol. 49:151.

Brouillette, S., and P. Chartrand. (1987) INTERMOLECULAR RECOMBINATION ASSAY FOR MAMMALIAN CELLS THAT PRODUCES RECOMBINANTS CARRYING BOTH HOMOLOGOUS AND NONHOMOLOGOUS JUNCTIONS. Mol. Cell. Biol. 7: 2248.

Bullock, P., J.J. Champoux, and M. Botchan. (1985) ASSOCIATION OF CROSSOVER POINTS WITH TOPOISOMERASE I CLEAVAGE SITES:A MODEL FOR NONHOMOLOGOUS RECOMBINATION. Science 230:954.

Bullock, P., W. Forrester, and M. Botchan. (1984) DNA SEQUENCE STUDIES OF SV40 CHROMOSOMAL EXCISION AND INTEGRATION IN RAT CELLS. J. Mol. Biol. 174:55-84.

Carle, G.F., M. Frank, and M.V. Olson. (1986) ELECTROPHORETIC SEPARATIONS OF LARGE DNA MOLECULES BY PERIODIC INVERSION OF THE ELECTRIC FIELD. Science 232:65.

Champoux, J.J., W.K. McCoubrey, and M.D. Been. (1984) DNA STRUCTURAL FEATURES THAT LEAD TO STRAND BREAKAGE BY EUKARYOTIC TYPE-I TOPOISOMERASE. Cold Spring Harbor Symp. Quant. Biol. 49:435.

Chang, X.-B., and J. Wilson. (1987) MODIFICATION OF DNA ENDS CAN DECREASE END JOINING RELATIVE TO HOMOLOGOUS RECOMBINATION IN MAMMALIAN CELLS. Proc. Natl. Acad. Sci. USA 84: 4959-4963.

Chia, W., and P.W.J. Rigby. (1981) FATE OF VIRAL DNA IN NONPERMISSIVE CELLS INFECTED WITH SV40. Proc. Natl. Acad. Sci. USA 78: 6638-6642.

Cook, P.R., and I.A. Brazell. (1975) SUPERCOILS IN HUMAN DNA. J. Cell Sci. 19:261-279.

Croce, C.M. (1977) ASSIGNMENT OF THE INTEGRATION SITE FOR SV40 TO CHROMOSOME 17 IN GM54VA, A HUMAN Cell LINE TRANSFORMED BY SV40. Proc. Natl. Acad. Sci. USA 74:315-318.

Croce, C.M., A.J. Girardi, and H. Koprowski. (1973) ASSIGNMENT OF THE T-ANTIGEN GENE OF SV40 TO HUMAN CHROMOSOME C-7. Proc. Natl. Acad. Sci. USA 70: 3617-3620.

Dailey, L., S. Pellegrini, and C. Basilico. (1984) DELETION OF THE ORIGIN OF REPLICATION IMPAIRS THE ABILITY OF Py DNA TO TRANSFORM CELLS AND TO FORM TANDEM INSERTIONS. J. Virol. 49:984-987.

D'Ambrosio, E., S.D. Waitzkin, F.R. Witney, A. Salemme, and, A.V. Furano. (1986) STRUCTURE OF THE HIGHLY REPEATED, LONG INTERSPERSED DNA FAMILY (LINE OR L1Rn) OF THE RAT. Mol. Cell. Biol. 6:411-424.

Della Valle, G., R.G. Fenton, and C. Basilico. (1981) Py LARGE T ANTIGEN REGULATES THE INTEGRATION OF VIRAL DNA SEQUENCES INTO THE GENOME OF TRANSFORMED RAT CELLS. Cell 23:347-355.

De Lozanne, A., and J.A. Spudich. (1987) DISRUPTION OF THE *Dictyostelium* MYOSIN HEAVY CHAIN GENE BY HOMOLOGOUS RECOMBINATION. Science 236;1086.

Doetschman, T., R.G. Gregg, N. Maeda, M.L. Hooper, D.W. Melton, S. Thompson, and O. Smithies. (1987) TARGETTED CORRECTION OF A MUTANT HPRT GENE IN MOUSE EMBRYONIC STEM CELLS. Nature 330;578-560.

Dorsett, D.L., I. Keshet, and E. Winocour. (1983) QUANTITATION OF A SV40 NONHOMOLOGOUS RECOMBINATION PATHWAY. J. Virol. 48:218-228.

Dretzen, G., M. Bellard, P. Sassone-Corsi, and P. Chambon. (1981) A RELIABLE METHOD FOR THE RECOVERY OF DNA FRAGMENTS FROM AGAROSE AND ACRYLAMIDE GELS. Anal. Biochem. 112:295-298.

Earnshaw, W.C., B. Halligan, C.A. Cooke, M.M.S. Heck, and L.F. Liu. (1985) TOPOISOMERASE II IS A STRUCTURAL COMPONENT OF MITOTIC CHROMOSOME SCAFFOLDS. J. Cell Biol. 100: 1706-1715.

Earnshaw, W.C., and M.M.S. Heck. (1985) LOCALIZATION OF TOPOISOMERASE II IN MITOTIC CHROMOSOMES. J. Cell Biol. 100: 1716-1725.

Eckhart, W. (1974) PROPERTIES OF TEMPERATURE-SENSITIVE MUTANTS OF Py. Cold Spring Harbor Symp. Quant. Biol. 39: 37-40.

Feinberg, A.P., and B. Vogelstein. (1983) A TECHNIQUE FOR RADIOLABELING DNA RESTRICTION ENDONUCLEASE FRAGMENTS TO HIGH SPECIFIC ACTIVITY. Anal. Biochem. 132: 6-13.

Folger, K.R., E.A. Wong, G. Wahl, and M.R. Capecchi. (1982) PATTERNS OF INTEGRATION OF DNA MICROINJECTED INTO CULTURED MAMMALIAN CELLS: EVIDENCE FOR HOMOLOGOUS RECOMBINATION BETWEEN INJECTED PLASMID DNA MOLECULES. Mol. Cell. Biol. 2:1372-1387.

Folger, K., K. Thomas, and M. Capecchi. (1984) ANALYSIS OF HOMOLOGOUS RECOMBINATION IN CULTURED MAMMALIAN CELLS. Cold Spring Harbor Symp. Quant. Biol. 49:123.

Furano, A.V., C.C. Somerville, P.N. Tsichlis, and E. D'Ambrosio. (1986) TARGET SITES FOR THE TRANSPOSITION OF RAT LONG INTERSPERSED REPEATED DNA ELEMENTS (LINEs) ARE NOT RANDOM. Nucleic Acids Res. 14:3717.

Gasser, S.M., T. Laroche, J. Falquet, E. Boy de la Tour, and, U.K. Laemmli. (1986) METAPHASE CHROMOSOME STRUCTURE: INVOLVEMENT OF TOPOISOMERASE II. J. Mol. Biol. 188: 613-629.

Graessmann, A., M. Graessmann, W.C. Topp, and M. Botchan. (1979) RETRANSFORMATION OF A SV40 REVERTANT CELL LINE, WHICH IS RESISTANT TO VIRAL AND DNA INFECTIONS, BY MICROINJECTION OF VIRAL DNA. J. Virol. 32:989-994.

Graham, F., S. Bacchetti, R. McKinnon, C. Stanners, B. Cordell, and, H.M. Goodman. (1980) TRANSFORMATION OF MAMMALIAN CELLS WITH DNA USING THE CALCIUM TECHNIQUE. IN: INTRODUCTION OF MACROMOLECULES INTO VIABLE MAMMALIAN CELLS, Alan Liss Inc., NY NY.

Grunstein, M., and D. S. Hogness. (1975) COLONY HYBRIDIZATION: A METHOD FOR THE ISOLATION OF CLONED DNAs THAT CONTAIN A SPECIFIC GENE. Proc. Natl. Acad. Sci. USA 72: 3961.

Gusew, N., A. Nepveu, and P. Chartrand. (1987) LINEAR DNA MUST HAVE FREE ENDS TO TRANSFORM RAT CELLS EFFICIENTLY. Mol. Gen. Genet. 206:121-125.

Gutz, H. (1971) SITE SPECIFIC INDUCTION OF GENE CONVERSION IN *Schizosaccharomyces pombe.* Genetics 69: 317-337.

Halligan, B.D., J.L. Davis, K.A. Edwards and L.F. Lin. (1982) INTRA- AND INTERMOLECULAR STRAND TRANSFER BY HeLa DNA TOPOISOMERASE I. J. Biol. Chem. 257:3995-4000.

Hanahan, D. (1983) STUDIES ON TRANSFORMATION OF E. COLI WITH PLASMIDS. J. Mol. Biol. 166:557-580.

Hayday, A., H.E. Ruley, and M. Fried. (1982) STRUCTURAL AND BIOLOGICAL ANALYSIS OF INTEGRATED Py DNA AND ITS ADJACENT HOST SEQUENCES CLONED FROM TRANSFORMED RAT CELLS. J.VIROL. 44:67-77.

Heck, M.M.S., W.N. Hittelman, and W.C. Earnshaw. (1988) DIFFERENTIAL EXPRESSION OF DNA TOPISOMERASES I AND II DURING THE EUKARYOTIC CELL CYCLE. Proc. Nat. Acad. Sci. USA 85:1086-1090.

Heinlein, V.R.O., R. Lange-Sablitzky, H. Schoal, and W. Wille. (1986) MOLECULAR CHARACTERIZATION OF THE MT-FAMILY OF DISPERSED MIDDLE-REPETITIVE DNA IN RODENT GENOMES. Nucleic Acids Res. 14:6403.

Heisterkamp, N., K. Stam, J. Groffen, A. de Klein, and G. Grosveld (1985) STRUCTURAL ORGANIZATION OF THE *bcr* GENE AND ITS ROLE IN THE Ph' TRANSLOCATION. Nature 315: 758-761.

Hinnen, A., J. Hicks, and G. Fink. (1978) TRANSFORMATION OF YEAST. Proc. Natl. Acad. Sci. USA 75:1929-1933.

Hollis, M., and J. Hindley. (1986) HUMAN *Sau3A* REPEATED DNA IS ENRICHED IN SMALL POLYDISPERS CIRCULAR DNA FROM NORMAL LYMPHOCYTES. Gene 46: 153-160.

Huberdeau, D., B.S. Sylla, E. Herring-Gillam, D. Bourgaux-Ramoisy, and P. Bourgaux. (1985) ALTERNATIVE EXCISION PRODUCTS ORIGINATING FROM A SINGLE INTEGRATION OF Py DNA. Mol. Cell. Biol. 5: 2608-2612.

Hwang, S.-P., and R. Kucherlapati. (1980) LOCALIZATION AND ORGANIZATION OF INTEGRATED SV40 SEQUENCES IN A HUMAN CELL LINE. Virology 105: 196-204.

Ikeda, H., K. Aoki, and A. Naito. (1982) ILLEGITIMATE RECOMBINATION MEDIATED *in vitro* BY DNA GYRASE OF *E.coli*:

STRUCTURE OF RECOMBINANT DNA MOLECULES. Proc. Natl. Acad. Sci. USA 79: 3724-3728.

Jarman, A.P., and D.R. Higgs. (1988) NUCLEAR SCAFFOLD ATTACHMENT SITES IN THE HUMAN GLOBIN GENE COMPLEXES. EMBO J. 7:3337-3344.

Jasin, M., J. de Villiers, F. Weber, and W. Schaffner. (1985) HIGH FREQUENCY OF HOMOLOGUOS RECOMBINATION IN MAMMALIAN CELLS BETWEEN ENDOGENOUS AND INTRODUCED SV40 GENOMES. Cell 43:695-703

Jayaram, M., and Makkni. (1986) MATING TYPE-LIKE CONVERSION PROMOTED BY THE 2uM CIRCLE SITE-SPECIFIC RECOMBINASE: IMPLICATIONS FOR THE DOUBLE-STRAND-GAP REPAIR MODEL. Mol. Cell. Biol. 6: 3831-3837

Jeanpierre, M. (1987) A RAPID METHOD FOR THE PURIFICATION OF DNA FROM BLOOD. Nucleic Acids Res. 15:9611.

Jelinek, W.R., and C.W. Schmid. (1982) REPETITIVE SEQUENCES IN EUKARYOTIC DNA AND THEIR EXPRESSION. Ann. Rev. Biochem 51: 813-844.

Jurka, J., and T. Smith. (1988) A FUNDAMENTAL DIVISION IN THE Alu FAMILY OF REPEATED SEQUENCES. Proc. Natl. Acad. Sci. USA 85:4775-4778.

Kato, S., R.A. Anderson, and R.D. Camerini-Otero. (1986) FOREIGN DNA INTRODUCED BY CALCIUM PHOSPHATE IS INTEGRATED INTO REPETITIVE DNA ELEMENTS OF THE MOUSE L CELL GENOME. Mol. Cell. Biol. 6: 1787-1795.

Ketner, G., and T.J. Kelly. (1976) INTEGRATED SV40 SEQUENCES IN TRANSFORMED CELL DNA: ANALYSIS USING RESTRICTION ENDONUCLEASES. Proc. Natl. Acad. Sci. USA 73: 1102-1106.

-----------. (1980) STRUCTURE OF INTEGRATED SV40 DNA IN TRANSFORMED MOUSE CELLS. J. Mol. Biol. 144: 163-182.

Kirchhausen, T., J.C. Wang, and S.C. Harrison. (1985) DNA GYRASE AND ITS COMPLEXES WITH DNA: DIRECT OBSERVATION BY ELECTRON MICROSCOPY. Cell 41: 933.

Klar, A.J.S., J.N. Strathern, and J.B. Hicks. (1981) A POSITION EFFECT CONTROL FOR GENE TRANSPOSITION: STATE OF EXPRESSION OF YEAST MATING-TYPE GENES AFFECTS THEIR ABILITY TO SWITCH. Cell 25: 517-524.

Kobori, J.A., E. Strauss, K. Minard, and L. Hood. (1986) MOLECULAR ANALYSIS OF THE HOTSPOT OF RECOMBINATION IN THE

MURINE MAJOR HISTOCOMPATIBILITY COMPLEX. Science 234: 173-179.

Konopka, A.K. (1988) COMPILATION OF DNA STRAND EXCHANGE SITES FOR NON-HOMOLOGOUS RECOMBINATION IN SOMATIC CELLS. Nucleic Acids Res. 16:1739.

Krauss, M.R., B.B.Gourlie, M.L. Bayne, and R.M. Benbow. (1984) Py MINICHROMOSOMES: ASSOCIATED DNA TOPOISOMERASE II AND DNA LIGASE ACTIVITIES. J. Virol. 49: 333-342.

Krayev, A.S., D.A. Kramerov, K.G. Skryabin, A.P. Ryskov, and G.P. Georgiev. (1980) THE NUCLEOTIDE SEQUENCE OF THE UBIQUITOUS REPETITIVE DNA SEQUENCE B1 COMPLEMENTARY TO THE MOST ABUNDANT CLASS OF MOUSE FOLD-BACK RNA. Nucleic Acids Res. 8: 1201-1215.

Krayev, A.S., T.V. Markusheva, D.A. Kramerov, A.P. Ryskov, K.G. Skryabin, A.A. Bayev, and G.P. Georgiev. (1982) UBIQUITOUS TRANSPOSON-LIKE REPEATS B1 AND B2 OF THE MOUSE GENOME: B2 SEQUENCING. Nucleic Acids Res. 10: 7461-7475.

Krolewski, J.J., C.W. Schindler, and M.G. Rush. (1984) STRUCTURE OF EXTRACHROMOSOMAL CIRCULAR DNAs CONTAINING BOTH THE *Alu* FAMILY OF DISPERSED REPETITIVE SEQUENCES AND OTHER REGIONS OF CHROMOSOMAL DNA. J. Mol. Biol. 174: 41-54.

Kucherlapati, R.S., E.M. Eves, K.Y. Song, B.S. Morse, and O. Smithies. (1984) HOMOLOGOUS RECOMBINATION BETWEEN PLASMIDS IN MAMMALIAN CELLS CAN BE ENHANCED BY TREATMENT OF INPUT DNA. Proc. Natl. Acad. Sci. USA 81: 3153-3157.

Kucherlapati, R., S.P. Hwang, N. Shimizu, J.K. McDougall, and M.R., Botchan. (1978) ANOTHER CHROMOSOMAL ASSIGNMENT FOR A SV40 INTEGRATION SITE IN HUMAN CELLS. Proc. Natl. Acad. Sci. USA 75: 4460-4464.

Lalande, M., J. Noolandi, C. Turmel, J. Rousseau, and G.W. Slater. (1987) PFGE: APPLICATION OF A COMPUTER MODEL TO THE SEPARATION OF LARGE DNA MOLECULES. Proc. Natl. Acad. Sci. USA 84:8011-8015.

Lam, S.T., M.M. Stahl, K.D. McMilin, and F.W. Stahl. (1974) *Rec*-MEDIATED RECOMBINATIONAL HOTSPOT ACTIVITY IN BACTERIOPHAGE LAMBDA. II. A MUTATION WHICH CAUSES HOT SPOT ACTIVITY. Genetics 77: 425-433.

Leder, P., D. Tiemeier, and L. Enquist. (1977) EK2 DERIVATIVES OF BACTERIOPHAGE LAMBDA USEFUL IN THE CLONING OF DNA FROM HIGHER ORGANISMS: THE LgtWES SYSTEM. Science 196: 175.

Lehrman, M.A., D.W. Russell, J.L. Goldstein, and M.S. Brown. (1986) EXON-*Alu* RECOMBINATION DELETES 5 Kbp FROM THE LOW DENSITY LIPOPROTEIN RECEPTOR GENE, PRODUCING A NULL PHENOTYPE IN FAMILIAL HYPERCHOLESTEROLEMIA. Proc. Natl. Acad. Sci. USA 83:3679-3683.

Lehrman, M.A., W.J. Schneider, T.C. Sudhof, J.L. Goldstein, and, D.W. Russell. (1985) Science 227: 140-146.

Lewis, C.D., and U.K. Laemmli. (1982) HIGHER ORDER METAPHASE CHROMOSOME STRUCTURE: EVIDENCE FOR METALLOPROTEIN INTERACTIONS. Cell 29: 171-181.

Lin, F.-L., K. Sperle, and N. Sternberg. (1984) HOMOLOGOUS RECOMBINATION IN MOUSE L CELLS. Cold Spring Harbor Symp. Quant. Biol. 49:139.

-----------. (1984) MODEL FOR HOMOLOGOUS RECOMBINATION DURING TRANSFER OF DNA INTO MOUSE L CELLS: ROLE FOR DNA ENDS IN THE RECOMBINATION PROCESS. Mol. Cell. Biol. 4: 1020-1034.

-----------. (1985) RECOMBINATION IN MOUSE L CELLS BETWEEN DNA INTRODUCED INTO CELLS AND HOMOLOGOUS CHROMOSOMAL SEQUENCES. Proc. Natl. Acad. Sci. USA 82:1391-1395.

-----------. (1987) EXTRACHROMOSOMAL RECOMBINATION ON MAMMALIAN CELLS AS STUDIED WITH SINGLE- AND DOUBLE-STRANDED DNA SUBSTRATES. Mol. Cell. Biol. 7:129.

Liskay, R.M., and J.L. Stachelek. (1983) EVIDENCE FOR INTRACHROMOSOMAL GENE CONVERSION IN CULTURED MOUSE CELLS. Cell 35:157-165.

Liskay, R.M., J.L. Stachelek, and A. Letsou. (1984) HOMOLOGOUS RECOMBINATION BETWEEN REPEATED CHROMOSOMAL SEQUENCES IN MOUSE CELLS. Cold Spring Harbor Symp. Quant. Biol. 49:183.

Loyter, A., G.A. Scangos, and F.H. Ruddle. (1982) MECHANISMS OF DNA UPTAKE BY MAMMALIAN CELLS: FATE OF EXOGENOUSLY ADDED DNA MONITORED BY THE USE OF FLUORESCENT DYES. Proc. Natl. Acad. Sci. USA 79:422-426.

Macdonald, M.V., and H.L.K. Whitehouse. (1979) A *buff* SPORE COLOUR MUTANT OF *Sordaria brevicollis* SHOWING HIGH FREQUENCY CONVERSION. 1. CHARACTERISTICS OF THE MUTANT. Genet. Res. 34: 87-119.

Maniatis, T., E.F. Fritsch, J. and Sambrook. (1982) MOLECULAR CLONING: A LABORATORY MANUAL. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Mansour, S.L., K.R. Thomas, and M.R. Capecchi. (1988) DISRUPTION OF THE PROTO-ONCOGENE int-2 IN MOUSE EMBRYO-DERIVED STEM CELLS: A GENERAL STRATEGY FOR TARGETING MUTATIONS TO NON-SELECTABLE GENES. Nature 336:348-352.

Markham, P., and H.L.K. Whitehouse. (1982) A HYPOTHESIS FOR THE INITIATION OF GENETIC RECOMBINATION IN EUKARYOTES. Nature 295: 421-423.

Marvo, S.L., S.R. King, and S.R. Jaskunas. (1983) ROLE OF SHORT REGIONS OF HOMOLOGY IN INTERMOLECULAR ILLEGITIMATE RECOMBINATION EVENTS. Proc. Natl. Acad. Sci. USA 80:2452-2456.

Maxam, A.M., and W. Gilbert. (1977) A NEW METHOD FOR SEQUENCING DNA. Proc. Natl. Acad. Sci. USA 74: 560.

-----------. (1980) SEQUENCING END-LABELED DNA WITH BASE-SPECIFIC CHEMICAL CLEAVAGES. Methods in Enzym. 65:499-560.

McLachlan A.D., and D.R. Boswell. (1985) CONFIDENCE LIMITS FOR HOMOLOGY IN PROTEIN OR GENE SEQUENCES. J. Mol. Biol. 185:39-49.

Miller, B.L., K.Y. Miller, and W.E. Timberlake. (1985) DIRECT AND INDIRECT GENE REPLACEMENTS IN ASPERGILLUS NIDULANS. Mol. Cell. Biol. 5:1714 (-1721).

Neer, A., N. Baran, and H. Manor. (1983) INTEGRATION OF Py DNA INTO CHROMOSOMAL DNA IN TRANSFORMED RAT CELLS CAUSES DELETION OF FLANKING CELL SEQUENCES. J. Gen. Virol. 64:69-82.

Orr-Weaver, T.L., J.W. Szostak, and R.J. Rothstein. (1981) YEAST TRANSFORMATION: A MODEL SYSTEM FOR THE STUDY OF RECOMBINATION. Proc. Natl. Acad. Sci. USA 78:6354-6358.

Perucho, M., D. Hanahan, and M. Wigler. (1980) GENETIC AND PHYSICAL LINKAGE OF EXOGENOUS SEQUENCES IN TRANSFORMED CELLS. Cell 22: 309-317.

Piche, A., and P. Bourgaux. (1987b) RESOLUTION OF A Py-MOUSE HYBRID REPLICON: VIRAL FUNCTION REQUIRED FOR RECOMBINATION. J.Virol. 61:845-850.

-----------. (1987a) RESOLUTION OF A Py-MOUSE HYBRID REPLICON: RELEASE OF GENOMIC VIRAL DNA. J.Virol. 61:840.

Pomerantz, B.J., M. Naujokas, and J. Hassell. (1983) HOMOLOGOUS RECOMBINATION BETWEEN TRANSFECTED DNAs. Mol. Cell. Biol. 3:1680-1685.

Ponticelli, A.S., D.W. Schultz, A.F. Taylor, and G.R. Smith. (1985) Chi-DEPENDENT DNA STRAND CLEAVAGE BY RecBC ENZYME. Cell 41: 145-151.

Reed, K.C., and D.A. Mann. (1985) RAPID TRANSFER OF DNA FROM AGAROSE GELS TO NYLON MEMBRANES. Nucleic Acids Res. 13:7207-.

Riabowol, K., R.J. Shmookler Reis, and S. Goldstein. (1985) INTERSPERSED REPETITIVE AND TANDEMLY REPETITIVE SEQUENCES ARE DIFFERENTIALLY REPRESENTED IN EXTRACHROMOSOMAL COVALENTLY CLOSED CIRCULAR DNA OF HUMAN DIPLOID FIBROBLASTS. Nucleic Acids Res. 13: 5563-5584.

Rigby, P.W.J., M. Dieckmann, C. Rhodes, and P. Berg. (1977) LABELLING DNA TO HIGH SPECIFIC ACTIVITY in vitro BY NICK-TRANSLATION WITH DNA POLYMERASE I. J. Mol. Biol. 113: 237-251.

Rogers, J. (1985) ONCOGENE CHROMOSOME BREAKPOINTS AND Alu SEQUENCES. Nature 317:559.

-----------. (1985) ORIGINS OF REPEATED DNA. Nature 317:765.

Roth, D.B., and J.H. Wilson. (1985) RELATIVE RATES OF HOMOLOGOUS AND NONHOMOLOGOUS RECOMBINATION IN TRANSFECTED DNA. Proc. Natl. Acad. Sci. USA 82: 3355-3359.

Roy, G., J.C. Wallenburg, and P. Chartrand. (1988) INEXPENSIVE AND SIMPLE SET-UP FOR FIGE. Nucleic Acids Res. 16:768.

Rubnitz, J., and S. Subramani. (1984) THE MINIMUM AMOUNT OF HOMOLOGY REQUIRED FOR HOMOLOGOUS RECOMBINATION IN MAMMALIAN CELLS. Mol. Cell. Biol. 4:2253-2258.

-----------. (1985) RAPID ASSAY FOR EXTRACHROMOSOMAL HOMOLOGOUS RECOMBINATION IN MONKEY CELLS. Mol. Cell. Biol. 5:529-537.

-----------. (1986) EXTRACHROMOSOMAL AND CHROMOSOMAL GENE CONVERSION IN MAMMALIAN CELLS. Mol. Cell. Biol. 6: 1608.

Ruley, H.E., L. Lania, F. Chaudry, and M. Fried. (1982) USE OF A CELLULAR POLY-A SIGNAL BY VIRAL TRANSCRIPTS IN Py TRANSFORMED CELLS. Nucleic Acids Res. 10:4515.

Ruley, H.R., and M. Fried. (1983) CLUSTERED ILLEGITIMATE RECOMBINATION EVENTS IN MAMMALIAN CELLS INVOLVING VERY SHORT SEQUENCE HOMOLOGIES. Nature 304:181.

Sander, M. and T.S. Hsieh. (1985) *Drosophila* TOPOISOMERASE II DOUBLE-STRAND DNA CLEAVAGE: ANALYSIS OF DNA SEQUENCE HOMOLOGY AT THE CLEAVAGE SITE. Nucleic Acids Res. 13:1057-1072.

Savageau, M.A., R. Metter, and W.W. Brockman. (1983) STATISTICAL SIGNIFICANCE OF PARTIAL BASE-PAIRING POTENTIAL: IMPLICATIONS FOR RECOMBINATION OF SV40 DNA IN EUKARYOTIC CELLS. Nucleic Acids Res. 11:6559.

Scangos, G., and F.H. Ruddle. (1981) MECHANISMS AND APPLICATIONS OF DNA-MEDIATED GENE TRANSFER IN MAMMALIAN CELLS - A REVIEW. Gene 14:1-10.

Scangos, G., K.M. Huttner, D.K. Jurcicek, and F.H. Ruddle. (1981) DNA-MEDIATED GENE TRANSFER IN MAMMALIAN CELLS: MOLECULAR ANALYSIS OF UNSTABLE TRANSFORMANTS AND THEIR PROGRESSION TO STABILITY. Mol. Cell. Biol. 1: 111-120.

Schwartzberg, Colicelli, and S. Goff. (1985) RECOMBINATION BETWEEN A DEFECTIVE RETROVIRUS AND HOMOLOGOUS SEQUENCES IN HOST DNA: REVERSION BY PATCH REPAIR. J.Virol 53:719-726.

Shapira, G., J.L. Stachelek, A. Letsou, L.K. Soodak, and R.M. Liskay. (1983) NOVEL USE OF SYNTHETIC OLIGONUCLEOTIDE INSERTION MUTANTS FOR THE STUDY OF HOMOLOGOUS RECOMBINATION IN MAMMALIAN CELLS. Proc. Natl. Acad. Sci. USA 80:4827-4831.

Shaul, Laub, Walker, and Rutter. (1985) HOMOLOGOUS RECOMBINATION BETWEEN A DEFECTIVE VIRUS AND A CHROMOSOMAL SEQUENCE IN MAMMALIAN CELLS. Proc. Natl. Acad. Sci. USA 82:3781-3784.

Shmookler Reis, R.J., C.K.J. Lumpkin, K.T. Riabowol, and, S. Goldstein. (1983) EXTRA CHROMOSOMAL CIRCULAR COPIES OF AN '*inter-Alu*' UNSTABLE SEQUENCE IN HUMAN DNA ARE AMPLIFIED DURING *in vitro* AND *in vivo* AGEING. Nature 301: 394-398.

Singh, L., and K.W. Jones. (1984) THE USE OF HEPARIN AS A SIMPLE COST-EFFECTIVE MEANS OF CONTROLLING BACKGROUND IN NUCLEIC ACID HYBRIDIZATION PROCEDURES. Nucleic Acids Res. 12: 5627.

Small, J., and G. Scangos. (1983) RECOMBINATION DURING GENE TRANSFER INTO MOUSE CELLS CAN RESTORE THE FUNCTION OF DELETED GENES. Science 219:174.

Smith, A.J.H., and P. Berg. (1984) HOMOLOGOUS RECOMBINATION BETWEEN DEFECTIVE NEO GENES IN MOUSE 3T6 CELLS. Cold Spring Harbor Symp. Quant. Biol. 49: 171-181.

Smith, G.R. (1983) CHI HOTSPOTS OF GENERALIZED RECOMBINATION. Cell 34: 709-710.

Smith, T.F., M.S. Waterman, and C. Burks. (1985) THE STATISTICAL DISTRIBUTION OF NUCLEIC ACID SIMILARITIES. Nucleic Acids Res. 13:645.

Smithies, O., M.A. Koralewski, K.-Y. Song, and R. Kucherlapati. (1984) HOMOLOGOUS RECOMBINATION WITH DNA INTRODUCED INTO MAMMALIAN CELLS. Cold Spring Harbor Symp. Quant. Biol. 49:161.

Smithies, O., R. Gregg, S.G. Boggs, M.A. Koralewski, and, R.S. Kucherlapati. (1985) INSERTION OF DNA SEQUENCES INTO THE HUMAN CHROMOSOMAL B-GLOBIN LOCUS BY HOMOLOGOUS RECOMBINATION. Nature 317:230.

Southern, E.M. (1975) DETECTION OF SPECIFIC SEQUENCES AMONG DNA FRAGMENTS SEPARATED BY GEL ELECTROPHORESIS. J. Mol. Biol. 98:503-517.

Sperry, A.O., V.C. Blasquez, and W.T. Garrard. (1989) DYSFUNCTION OF CHROMOSOMAL LOOP ATTACHMENT SITES: ILLEGITIMATE RECOMBINATION LINKED TO MATRIX ASSOCIATION REGIONS AND TOPOISOMERASE II. Proc. Natl. Acad. Sci. USA 86: 5497-5501.

Steele, R.E., A.H. Bakken, and R.H. Reeder. (1984) PLASMIDS CONTAINING MOUSE rDNA DO NOT RECOMBINE WITH CELLULAR RIBOSOMAL GENES WHEN INTRODUCED INTO CULTURED MOUSE CELLS. Mol. Cell. Biol. 4:576-582.

Stephens, R.M. (1985) A SEQUENCERS' SEQUENCE ANALYSIS PACKAGE FOR THE IBM PC. Gene Anal. Tech. 2: 67-75.

Stringer, J.R. (1981) INTEGRATED SV40 DNA: NUCLEOTIDE SEQUENCES AT CELL-VIRUS RECOMBINANT JUNCTIONS. J. Virol. 38:671-679.

-----------. (1982) DNA SEQUENCE HOMOLOGY AND CHROMOSOMAL DELETION AT A SITE OF SV40 DNA INTEGRATION. Nature 296:363.

Subramani, S., and J. Rubnitz. (1985) RECOMBINATION EVENTS AFTER TRANSIENT INFECTION AND STABLE INTEGRATION OF DNA INTO MOUSE CELLS. Mol. Cell. Biol. 5:659-666.

Subramani, S., and P. Berg. (1983) HOMOLOGOUS AND NONHOMOLOGOUS RECOMBINATION IN MONKEY CELLS. Mol. Cell. Biol. 3:1040-1052.

Sylla, B.S., D. Allard, G. Roy, D. Bourgaux-Ramoisy, and P. Bourgaux. (1984) A MOUSE DNA SEQUENCE THAT MEDIATES INTEGRATION AND EXCISION OF Py DNA. Gene 29: 343-350.

Sylla, B.S., D. Bourgaux-Ramoisy, and P. Bourgaux. (1980) INDUCTION OF VIRAL DNA SYNTHESIS IN CLONAL DERIVATIVES OF A

PERMISSIVE CELL LINE TRANSFORMED BY A ts Py. Virology 100:357-369.

Sylla, B.S., D. Huberdeau, D. Bourgaux-Ramoisy, and P. Bourgaux. (1984) SITE-SPECIFIC EXCISION OF INTEGRATED Py DNA. Cell 37: 661-667.

Taylor, A.F., D.W. Schultz, A.S. Ponticelli, and G.R. Smith. (1985) RecBC ENZYME NICKING AT Chi SITES DURING DNA UNWINDING: LOCATION AND ORIENTATION-DEPENDENCE OF THE CUTTING. Cell 41: 153-163.

Thomas, K.R., and M.R. Capecchi. (1987) SITE-DIRECTED MUTAGENESIS BY GENE TARGETING IN MOUSE EMBRYO-DERIVED STEM CELLS. Cell 51;503-512.

Thomas, K.R., K.R. Folger, and M.R. Capecchi. (1986) HIGH FREQUENCY TARGETING OF GENES TO SPECIFIC SITES IN THE MAMMALIAN GENOME. Cell 44:419-428.

Topp, W.C., D. Lane, and R. Pollack. (1980) TRANSFORMATION BY SV40 AND Py VIRUS. p. 205-296. In J. Tooze (ed.) DNA Tumor Viruses part 2, Cold Spring Harbour Laboratory, NY.

Treco, D., B. Thomas, and N. Arnheim. (1985) RECOMBINATION HOT SPOT IN THE HUMAN ß-GLOBIN GENE CLUSTER: MEIOTIC RECOMBINATION OF HUMAN DNA FRAGMENTS IN *Saccharomyces cerevisiae*. Mol. Cell. Biol. 5: 2029-2038.

Vanin, E.F., P.S. Henthorn, D. Kioussis, F. Grosveld, and, O. Smithies. (1983) UNEXPECTED RELATIONSHIPS BETWEEN FOUR LARGE DELETIONS IN THE HUMAN ß-GLOBIN GENE CLUSTER. Cell 35: 701-709.

Voelkel-Meiman, K., R.L. Keil, and G.S. Roeder. (1987) RECOMBINATION-STIMULATING SEQUENCES IN YEAST RIBOSOMAL DNA CORRESPOND TO SEQUENCES REGULATING TRANSCRIPTION BY RNA POLYMERASE I. Cell 48:1071-1079.

Vogelstein, B., and D. Gillespie. (1979) PREPARATIVE AND ANALYTICAL PURIFICATION OF DNA FROM AGAROSE. Proc. Natl. Acad. Sci. USA 76;615-619.

Voliva, C.F., S.L. Martin, C.A. Hutchison III, and M.H. Edgell. (1984) DISPERSAL PROCESS ASSOCIATED WITH THE L1 FAMILY OF INTERSPERSED REPETITIVE DNA SEQUENCES. J. Mol. Biol. 178:795-813.

Waldeck, W., H. Zentgraf, and G. Sauer. (1982) in vitro FORMATION OF MULTIMERIC DNA STRUCTURES MEDIATED BY PURIFIED SV40 CHROMATIN. Proc. Natl. Acad. Sci. USA 79: 1027-1031.

Waldman, A.S., and R.M. Liskay. (1988) DEPENDENCE OF INTRACHROMOSOMAL RECOMBINATION IN MAMMALIAN CELLS ON UNINTERRUPTED HOMOLOGY. Mol. Cell. Biol. 8: 5350-5357.

Wallenburg, J.C., A. Nepveu, and P. Chartrand. (1984) RANDOM AND NONRANDOM INTEGRATION OF A POLYOMAVIRUS DNA MOLECULE CONTAINING HIGHLY REPETITIVE CELLULAR SEQUENCES. J. Virol. 50: 678-683.

-----------. (1987) INTEGRATION OF A VECTOR CONTAINING RODENT REPETITIVE ELEMENTS IN THE RAT GENOME. Nucleic Acids Res. 115:7849.

Weiner, A.M., P.L. Deininger, and A. Efstratiadis. (1986) NONVIRAL RETROPOSONS: GENES, PSEUDOGENES, AND TRANSPOSABLE ELEMENTS GENERATED BY THE REVERSE FLOW OF GENETIC INFORMATION. Ann. Rev. Biochem. 55:631-661.

Williams, T.J., M. and Fried. (1986) INVERTED DUPLICATION-TRANSPOSITION EVENT IN MAMMALIAN CELLS AT AN ILLEGITIMATE RECOMBINATION JOIN. Mol. Cell. Biol. 6: 2179-2184.

Witke, W., W. Nellen, and A. Noegel. (1987) HOMOLOGOUS RECOMBINATION IN THE *Dictyostelium* α-ACTININ GENE LEADS TO AN ALTERED mRNA AND LACK OF PROTEIN. EMBO 6:4143-4148.

Zweig, S.E. (1984) ANALYSIS OF LARGE DOT NUCLEIC ACID MATRICES ON SMALL COMPUTERS. Nucleic Acids Res. 12: 767-776.

APPENDIX A

JOURNAL OF VIROLOGY, June 1984, p. 678–683
0022-538X/84/060678-06$02.00/0
Copyright © 1984, American Society for Microbiology
Vol. 50, No. 3

Random and Nonrandom Integration of a Polyomavirus DNA Molecule Containing Highly Repetitive Cellular Sequences

JOHN C. WALLENBURG, ALAIN NEPVEU, AND PIERRE CHARTRAND*

Département de Microbiologie, Faculté de Médecine, Centre Hospitalier Universitaire, Sherbrooke, Québec, Canada J1H 5N4

Received 18 October 1983/Accepted 14 February 1984

RmI is a circular DNA molecule that consists of a complete polyomavirus genome with an insertion (Ins) of mouse cellular DNA. This polyomavirus genome carries a mutation which renders its replication, but not its transforming ability, temperature sensitive. Ins contains both unique and repetitive cellular DNA sequences. We transfected RmI into rat cells at the permissive and nonpermissive temperatures for replication and isolated clones that had integrated RmI in their genomes. In this paper, we describe detailed mapping of the integrated RmI sequences present in 37 different cell clones. Our results indicated that transfection at the permissive temperature resulted in a random integration pattern, whereas transfection at the nonpermissive temperature resulted in a nonrandom integration pattern. The nonrandom insertions had a preferential length and preferential endpoints. We argue from these results that the nonrandom integration pattern is related to the presence of Ins and that the switch between nonrandom integration and random integration reflects a modification of the integrating substrate. When both are active, the random mechanism dominates the nonrandom mechanism.

The initial analyses of integrated simian virus 40 (SV40) (3, 18) and polyomavirus (Py) (1) sequences using restriction enzymes showed that integration is not a site-specific process. These results were confirmed by many authors, with one exception (22). It was also shown that the integrated viral sequences could be associated with different chromosomes (8, 9, 16, 21). Finally, detailed analyses of the viral cellular junctions indicated the absence of specific sequences at the crossover sites (2, 15, 25, 28, 29). From these observations, it was concluded that the integration of Py and SV40 genomes was due to a random process. Such a process is expected to generate viral insertions of random length with random endpoints. Relatively few insertions have been precisely mapped and compared. Ketner and Kelly (19) looked at the lengths and the endpoints of viral insertions present in four clones, but the number of insertions involved and the precision of mapping were not sufficient to permit a statistical analysis.

Is the random integration process mentioned above the sole mechanism of integration for Py and SV40 genomes? The only results pertinent to this question are the results of Botchan and collaborators (4, 12), who have shown that retransformation of a cell line already carrying an SV40 insertion always result in distinct insertions. This could be due to a very low probability that the donor molecules find their homologous copies in the cellular genome. It could also mean that the illegitimate integration process is favored over the homologous process.

To study some of the aspects of integration mentioned above, we used a spontaneous recombinant DNA molecule (RmI) as the donor in an integration assay. Because of its particular structure and composition (Fig. 1), we felt that RmI had the potential to integrate via different types of recombination mechanisms. RmI is produced upon induction of C12a1, a mouse cell line carrying integrated sequences from ts-P155, a Py mutant (30). ts-P155 is thermosensitive for replication because of a mutation in gene *a*, but its transformation efficiency is the same at both temperatures (11). When shifted down from the nonpermissive temperature to the permissive temperature, C12a1 cells produce RmI and ts-P155 genomes in a 20:1 ratio. RmI (Fig. 1) is a circular, 7.1-kilobase pair DNA molecule that consists of 1.03 copies of the ts-P155 genome and a 1.6-kilobase pair insertion (Ins) of mouse cellular DNA (30). Ins is part of the flanking cellular sequences at the viral integration site in C12a1 (7). RmI has a complete uninterrupted Py early region, as well as a Py origin of replication. Ins contains both unique and repetitive cellular sequences (5). The repetitive sequences are homologous to disperse repetitive sequences of rodents, and their copy number is on the order of 10^5 copies per haploid genome (Nepveu, unpublished data). The two viral-cellular junctions have a particular structure. The viral sequences on one side of Ins form a direct 182-base pair repeat with the viral sequences on the other side, and the first cellular 7 base pairs at one junction form an imperfect inverted repeat with the first cellular 7 base pairs at the other junction (5). As part of the inverted repeat, Ins starts with the nucleotides TG and ends with CA. These nucleotides are the first and last nucleotides of retrovirus genomes, as well as repetitive elements of *Drosophila* and yeast (17). Thus, RmI is a circular molecule containing (i) a complete Py genome that is thermosensitive for replication, (ii) both unique and repetitive mouse genomic DNA, and (iii) two viral-cellular DNA junctions with direct and inverted repeats.

Like ts-P155, RmI can replicate at 33°C but not at 39°C, and it can transform rat cells at 39°C (D. Huberdeau, B. S. Sylla, D. Bourgaux-Ramoisy, and P. Bourgaux, manuscript in preparation). We transfected RmI into rat cells at the permissive and nonpermissive temperatures and isolated clones that had integrated RmI in their genomes. In this paper, we present an analysis of the integrated RmI sequences present in 37 different cell clones. Our conclusions are (i) that the model system which we used can generate from the same substrate random and nonrandom integration patterns, (ii) that these patterns are linked to the replicative state of the integrating molecules and the presence of cellular

* Corresponding author.

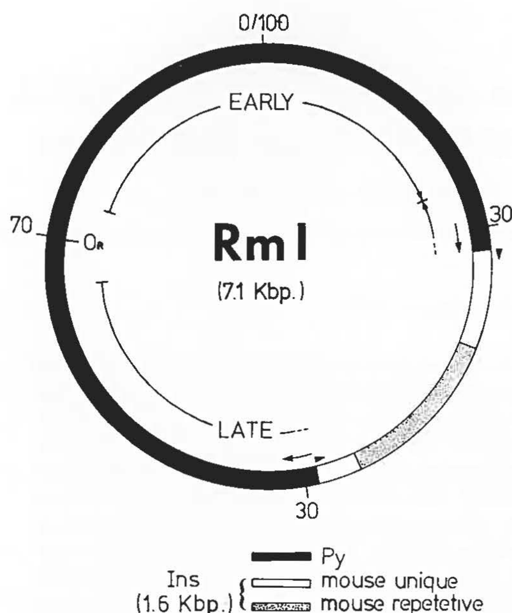


FIG. 1. Physical structure of RmI, as determined previously (5, 30). The numbers refer to standard Py map units. The arrows indicate the positions of the viral direct repeats, and the arrowheads indicate the positions of the cellular inverted repeats, whose sizes are exaggerated for the purpose of illustration. Ins is divided into three segments; the middle segment containing repetitive cellular DNA sequences is flanked on either side by segments consisting mainly of unique cellular DNA sequences (5; B. S. Sylla and P. Bourgaux, unpublished data). Kbp, Kilobase pairs.

sequences, and (iii) that the random integration mechanism is dominant over the nonrandom integration mechanism.

MATERIALS AND METHODS

Preparation of RmI. C12a1 cells grown at 39°C were shifted down to 33°C, and after 40 to 48 h, RmI was collected by Hirt extraction and purified as previously described (5).

Transfection. Fischer rat 3T3 cells (26) passaged beforehand at 39°C were seeded 24 h before transfection at a density of 2.5×10^5 cells per 90-mm petri dish. The transfection protocol used was that of Graham and collaborators (13, 14), with minor modifications. The carrier was high-molecular-weight DNA extracted from mouse cells. After addition of the calcium phosphate DNA precipitate, the cells were incubated at either 33 or 39°C for 20 h. The medium was then replaced with fresh medium, and all cells were incubated at 39°C for an additional 28 h before passing. All subsequent incubations were done at 39°C. The medium used at all times was Dulbecco modified Eagle medium (Flow Laboratories, Inc.) supplemented with 5.8×10^{-6} g of L-glutamine (Sigma Chemical Co.) per ml, 10% fetal bovine serum (Flow), 0.0004% streptomycin (Glaxo Laboratories), 1 U of penicillin (Glaxo) per ml, and 50 U of nystatin (E. R. Squibb & Sons) per ml.

Selection of clones. At 48 h after addition of the calcium phosphate precipitate, the cells were trypsinized, duplicate petri dishes were pooled, and the cells were seeded at a density of 5×10^4 cells per 60-mm petri dish.

After 1 to 2 weeks well-isolated foci of densely growing cells were picked and transferred to Linbro dishes or 35-mm petri dishes. When confluent, the clones were reseeded at a very low density (less than 50 cells per 60-mm petri dish), and after an additional 1 to 2 weeks of growth, one isolated colony per clone was picked and expanded. Although it remains possible that two clones isolated from the same experiment are related, the uniqueness of each clone was shown by the individuality of the patterns of integration.

Preparation and endonuclease digestion of cellular DNA. Confluent monolayers of cells in 90-mm petri dishes were washed with phosphate-buffered saline and then lysed with 1 ml of buffer A (10 mM Tris, pH 7.9, 10 mM EDTA, 10 mM NaCl, 0.5% sodium dodecyl sulfate, 50 μg of predigested proteinase K [Sigma] per ml). The lysate was extracted two or three times with phenol saturated with TE (10 mM Tris, pH 7.9, 1 mM EDTA), once with phenol-chloroform-isoamyl alcohol (25:24:1), and once with chloroform-isoamyl alcohol (24:1). The salt concentration was adjusted to 400 mM NaCl, and 2 volumes of ethanol (−20°C) was layered gently on top. The DNA was spooled on glass rods, washed in consecutive baths of 70, 80, 90, and 100% ethanol, dried, and dissolved in TE to a final concentration of 0.5 to 1.0 mg of DNA per ml. For each restriction endonuclease analysis, 10 μg of DNA was digested by using a 5× excess of enzyme in a total volume of 35 μl.

Electrophoresis and blotting. Cellular DNA was electrophoresed as previously described (7). The DNA was then transferred to nitrocellulose sheets (type BA85; Schleicher & Schuell Inc.) by the technique of Southern (27).

Nick translation and hybridization. The nick-translation procedure used was that of Rigby et al. (24), and the hybridization procedure used was that of van der Ploeg and Flavell (32), with minor modifications (7).

Goodness-of-fit test. For statistical analysis of our results we used the Kolmogorov-Smirnov goodness-of-fit test (20), which is considered to be superior to the chi-square test (6). This test allowed us to compare an observed distribution with any hypothetical distribution and also to compare two observed distributions with one another. The hypothetical distributions to which we compared our data were random distributions (also referred to as uniform or rectangular distributions), in which all insertion lengths (or all endpoints, depending on the distribution compared) had an equal probability of being represented.

Since the goodness-of-fit test requires that distributions be divided into continuous nonoverlapping intervals, the distribution of the lengths of the insertions (see Fig. 3) was arbitrarilly divided into the following four intervals: 20 to 40, 40 to 60, 60 to 80, and 80 to 100%. For the endpoint distribution (see Fig. 4), we used the following five intervals: Py map units 0 to 27, map unit 27 to the repetitive sequences of Ins, the repetitive sequences of Ins, the rest of Ins to map unit 39, and map units 39 to 70 (Fig. 1).

Because we selected for an intact transforming region, we did not include in the distribution analysis the length interval from 0 to 20% and the endpoint interval between map units 70 and 100. We made the assumption that the distributions of the insertions which we selected for had the same characteristics as the distributions of unselected insertions. In other words, we assumed that the randomness or specificity of the observed distributions were independent of the selection procedure.

As shown below (see Fig. 3 and 4), there was a certain uncertainty in the exact lengths and endpoints of the insertions. In the few cases where this resulted in an observation spanning two intervals, the observation was assigned to the interval in such a way as to favor the random distribution hypothesis. This means that our analysis is biased in favor of the hypothesis that our observed distributions were random distributions.

RESULTS

Number of insertions and tandems. RmI induces the formation of foci in rat cells (see above). We used this assay to isolate cells which integrated RmI. Transfected cells were either maintained continuously at 39°C or incubated at 33°C for 20 h immediately after the addition of the DNA precipitate and then returned to 39°C. RmI is thermosensitive for replication, but not for transformation, because of the ts-P155 mutation in large T (see above). Incubation at 33°C would be expected to activate the temperature-sensitive function of large T and permit initiation of replication. The efficiency of focus formation was similar in both cases, approximately 100 foci per μg of RmI per 10^6 cells.

A total of 74 foci were isolated and cloned; the 49 clones originating from the petri dishes which had been shifted to 33°C for 20 h were designated the W3 series, and the 25 clones originating from the petri dishes maintained at 39°C were designated the W9 series. To evaluate the complexity of the integration pattern, high-molecular-weight DNA from each of the clones was digested with *Eco*RI, *Bgl*II and *Bam*HI. There are single restriction sites for these enzymes in either RmI (*Eco*RI, *Bgl*II) or Py (*Eco*RI, *Bam*HI). As mentioned above, the Hirt preparation of RmI used for transfection contained RmI and Py molecules in a ratio of 20:1. Screening the clones with the restriction enzymes indicated which clones had only a few distinct insertions, thus permitting detailed physical mapping. All clones displaying simple restriction patterns were then analyzed with *Hpa*I, which does not cut RmI or Py and thus produces a single band for each distinct integration site. The results obtained with *Hpa*I confirmed the results obtained with *Eco*RI, *Bam*HI, and *Bgl*II. Furthermore, subsequent fine mapping of the insertions in these clones was also in agreement with this interpretation. The majority of the W9 clones (72%) had only one insertion, whereas the majority of the W3 clones (69%) had multiple insertions (Table 1).

The analysis with the single-cut enzymes *Eco*RI, *Bgl*II, and *Bam*HI also told us which clones did not have full copies of either RmI or Py in head-to-tail tandem. These are the clones that did not produce a fragment which comigrated with unit length RmI or Py after digestion with these enzymes. The vast majority of the W3 clones (88%) and W9 clones (96%) did not have full head-to-tail tandems (Table 1).

Physical maps of the insertions. Using the approach described previously (7), we determined physical maps of the insertions in some clones with two insertions and in all of the clones with single insertions except clones W38.23 and W98.11. Clones W38.23 and W98.11 had single insertions with full head-to-tail tandems, which made it difficult to map with precision the endpoints. The maps of 18 W3 clones and 19 W9 clones are shown in Fig. 2; these maps were constructed from an analysis of the restriction patterns of the insertions by using 11 different restriction enzymes (data not shown). In all cases, the insertions are of RmI unit length or less.

TABLE 1. Number of insertions and presence of tandems

Clones	No. of clones without full head-to-tail tandems/total no. of clones analyzed	No. of clones with one insertion/total no. of clones analyzed
W3	43/49 (88)[a]	15/49 (31)
W9	24/25 (96)	18/25 (72)

[a] The values in parentheses are percentages.

Analysis of the lengths and the endpoints of the insertions. Illegitimate recombination is expected to generate insertions having random lengths and random endpoints. Figures 3 and 4 show analyses of the insertions shown in Fig. 2 according to length and to the distribution of the endpoints, respectively.

In Fig. 3 the length of each insertion is depicted as a percentage of the RmI unit length. Because of our isolation procedure, the minimum length of the insertions was expected to correspond to the sequences necessary for the formation of foci. This was the transforming region of Py, which corresponds to 20% of the RmI length. A goodness-of-fit analysis (see above) of the W3 and W9 length distributions revealed that we can reject the hypothesis that the W3 and W9 length distributions are samples from the same distribution at a level of significance of 0.005. We cannot reject the hypothesis that the W3 length distribution is a sample of a random distribution. However, the hypothesis of a random distribution of the W9 lengths can be rejected at a level of significance of 0.005. This is because in the W9 clones there was a preferential length, close to unit length, for the insertions.

Figure 4 shows the distribution of the endpoints of the insertions in relation to RmI. Again, because of our isolation procedure, endpoints were not expected to occur in the transforming region. A goodness-of-fit analysis (see above) of the W3 and W9 endpoint distributions revealed that we can reject the hypothesis that the W3 and W9 endpoint distributions are from the same distribution at a level of significance of 0.025. We cannot reject the hypothesis that the W3 endpoint distribution is a sample of a random distribution. In the case of the W9 distribution, the random distribution hypothesis can be rejected at a level of significance of 0.05. This is due to the high occurrence of endpoints in the repetitive region of Ins and the Py sequences between 30 and 40 map units and also the low occurrence of endpoints in the Py sequences between 40 and 70 map units.

DISCUSSION

In this study we analyzed the integration pattern of a Py genome containing highly repetitive cellular sequences (RmI). Our results indicate that the pattern of integration of RmI can be random or nonrandom depending on the conditions of the assay.

When the cells were shifted down to 33°C for 20 h immediately after transfection and then back to 39°C (W3 clones), we observed that the distribution of the lengths and the endpoints of the insertions was random. Because of the large number of individual clones compared, these results give a statistical basis for the hypothesis that the integration of papovaviruses is mediated by a random process.

When the assay was done entirely at 39°C (W9 clones), the distribution of the lengths and the endpoints of the insertions was not random. There was a preferential length which was close to unit length. This raises the possibility that the state of the integrating substrate is different in the W9 and W3 clones. The illegitimate integration of Py or SV40 is thought of as the insertion of a linear segment that involves two crossing-over events, and this would explain the random length insertions in the W3 clones (31). The kind of integration process that would result in unit length insertions is the bacteriophage lambda type, where the donor is circular and only one crossing-over event is necessary. The integration of a circular molecule via a single crossing-over event has not been reported in mammalian cells. In the case of the W9 clones, even though most of the insertions are close to unit

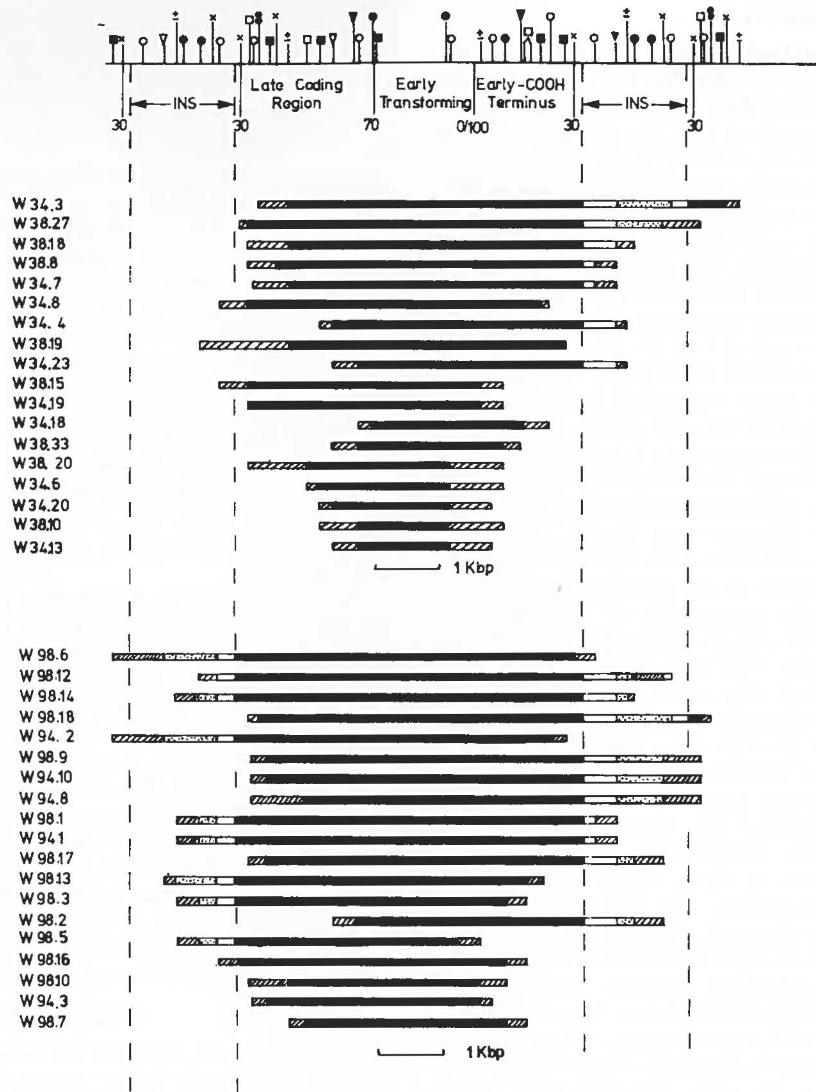


FIG. 2. Physical map of the insertions in W3 and W9 clones. The code for the RmI sequences is the same as that shown in Fig. 1. The cross-hatched boxes on each end define the region of uncertainty of the insertions where the recombination events occurred. At the top there is a linearized representation of RmI divided into the following four continuous regions: the cellular DNA (INS), the Py late coding region, the Py transforming region, and the carboxy terminus of the Py early region. The numbers refer to standard Py map units. Also indicated are the positions of the following key restrictions sites used for mapping: *Hpa*II (■), *Pst*I (□), *Pvu*II (●), *Mbo*I (○), *Hha*I (▼), *Bam*HI (▽), *Hin*cII (⧫), *Hin*dIII (±), and *Hph*I (×). Clones W38.18, W38.19, W34.6, W94.8, and W98.2 had two distinct insertions, only one of which was mapped precisely. Clone 34.7 also had two distinct insertions; we cannot be absolutely sure whether the two crossover events shown here are from the same insertion or represent two ends from distinct insertions. Kbp, Kilobase pair.

length, they nevertheless have two distinct endpoints. Thus, if the insertion in the W9 clones was the result of the integration of a circular molecule via a single crossing-over event, we have to postulate that integration was followed by rearrangements. These rearrangements could be necessary to stabilize the integrated structure. One difference between the integration of a linear molecule versus a circular molecule is the effect on the recipient molecule. The insertion of a linear segment, as in the case of Py, results in a deletion of a similar length in the recipient molecule (23). On the other hand, the insertion of a circular molecule via a single crossing-over event does not produce a deletion in the recipient molecule. Analysis of the cellular genome rearrangements at the sites of integration might give us some clue as to the integration mechanisms. This work is in progress.

The nonrandom distribution of the endpoints in the W9 clones is not the result of a single effect. There are two regions overrepresented; the major one is the repetitive region of Ins, and the other one occurs near the right viral-cellular junction in the 30- to 40-map unit Py fragment. There is also a low occurrence of endpoints in the Py sequences between 40 and 70 map units. The high occurrence of endpoints in the 30- to 40-map unit Py fragment is hard to explain, unless this is the result of postintegration rearrangements that delete certain Ins sequences. The high occurrence of endpoints in the region containing the repetitive sequences of Ins raises the possibility of a specific interac-

tion with the homologous cellular counterparts. This interpretation can be assayed by analysis of the recipient cellular site before integration. As mentioned above, this work is in progress.

One striking conclusion from our results is that the random mechanism of integration is dominant over the nonrandom mechanism. In other words, when the random mechanism is activated by a temperature shift down, it supersedes the nonrandom mechanism. As mentioned above, the Py genome in RmI originates from a thermosensitive mutant (ts-P155) that carries a mutation in the gene coding for the early protein large T, such that RmI replicates at 33°C but not at 39°C. Thus, the shiftdown of 20 h probably initiates RmI replication. The resulting modification of the substrate is sufficient to favor a random integration process. Della-Valle and co-workers (10) have suggested that an important step in the process of Py integration could be viral DNA replication, which would explain the effect of large T on transformation frequencies.

In conclusion, our results give a statistical basis for the hypothesis that the integration of papovaviruses is a random process resulting in random distribution of length and endpoints, as the result of nonspecific double crossing-over events. However, in the absence of certain viral functions and with the addition of a cellular insertion, the integration process becomes nonrandom. We believe that this is due to the presence of highly repetitive cellular sequences. In this case unit length molecules are preferentially inserted. We believe that this could be the result of the integration of a circular molecule. Finally, the random integration process is dominant over the nonrandom process. Botchan and co-workers (4, 12) did not observe specific integration of SV40 molecules in the cellular DNA of a cell line already containing an insertion of SV40 sequences. The results obtained with the W3 clones argue that a much higher copy number of insertions would not have necessarily changed the results of these workers.

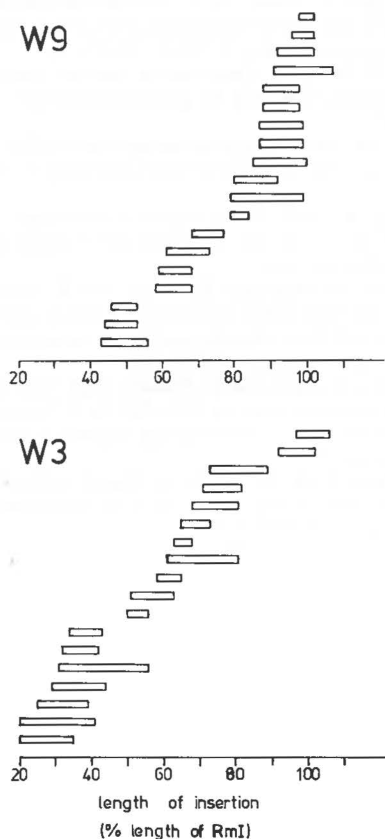


FIG. 3. Lengths of the insertions in the W3 and W9 clones. The scale indicates percentages of the length of the full RmI molecule. Each box represents one insertion shown in Fig. 2 and defines a minimum length and a maximum length for each insertion. The uncertainty in the actual length is due to the uncertainty in the precise location of the endpoints (Fig. 2, cross-hatched boxes).

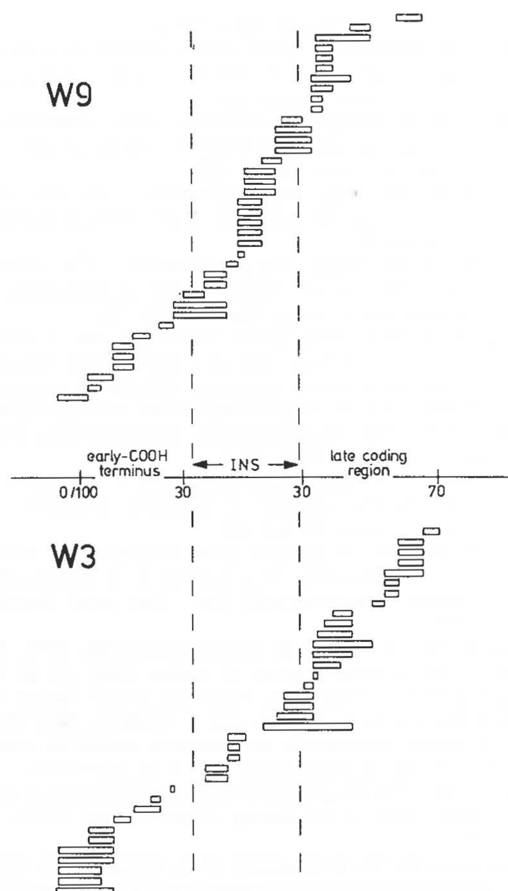


FIG. 4. Distribution of the endpoints of the insertions in the W3 and W9 clones in relation to RmI. Each box represents one endpoint (of which there are two per insertion), and these boxes correspond exactly to the cross-hatched boxes in Fig. 2. The RmI regions shown in the middle are the same as those shown in Fig. 2; they represent a linearization of the circular RmI molecule.

ACKNOWLEDGMENTS

We are grateful to P. Bourgaux, D. Bourgaux-Ramoisy, and colleagues for permission to refer to unpublished data and for access to the C12a1 cell line. Furthermore, we thank P. Bourgaux for critical reading of the manuscript and L. Tétreault and R. Black for statistical analysis of the results. We also thank S. Brouillette for valuable assistance, M. Chalifoux for drawing some of the figures, and N. Bernard and D. Laliberté for typing the manuscript.

This work was supported by grants from the National Cancer Institute of Canada and the Medical Research Council of Canada. P.C. is the recipient of a scholarship from the Medical Research Council of Canada. A.N. is the recipient of a studentship from the Medical Research Council of Canada.

LITERATURE CITED

1. **Birg, F., R. Dulbecco, M. Fried, and R. Kamen.** 1979. State and organization of polyoma virus DNA sequences in transformed rat cell lines. J. Virol. **29**:633–648.
2. **Botchan, M., J. Stringer, T. Mitchison, and J. Sambrook.** 1980. Integration and excision of SV40 DNA from the chromosome of a transformed cell. Cell. **20**:143–152.
3. **Botchan, M., W. Topp, and J. Sambrook.** 1976. The arrangement of simian virus 40 sequences in the DNA of transformed cells. Cell **9**:269–287.
4. **Botchan, M., W. Topp, and J. Sambrook.** 1978. Studies on simian virus 40 excision from cellular chromosomes. Cold Spring Harbor Symp. Quant. Biol. **43**:709–719.
5. **Bourgaux, P., B. S. Sylla, and P. Chartrand.** 1982. Excision of polyoma virus DNA from that of a transformed mouse cell. Identification of a hybrid molecule with direct and inverted repeat sequences at the viral-cellular joints. Virology **122**:84–97.
6. **Bradley, J. V.** 1968. Distribution-free statistical tests. Prentice-Hall, Inc., Englewood Cliffs, N.J.
7. **Chartrand, P., N. Gusew-Chartrand, and P. Bourgaux.** 1981. Integrated polyoma genomes in inducible permissive transformed cells. J. Virol. **39**:185–195.
8. **Croce, C. M.** 1977. Assignment of the integration site for simian virus 40 to chromosome 17 in GM54VA, a human cell line transformed by simian virus 40. Proc. Natl. Acad. Sci. U.S.A. **74**:315–318.
9. **Croce, C. M., A. J. Girardi, and H. Koprowski.** 1973. Assignment of the T-antigen gene of simian virus 40 to human chromosome C-7. Proc. Natl. Acad. Sci. U.S.A. **70**:3617–3620.
10. **Della-Valle, G., R. G. Fenton, and C. Basilico.** 1981. Polyoma large T antigen regulates the integration of viral DNA sequences into the genome of transformed cells. Cell **23**:347–355.
11. **Eckhart, W.** 1974. Properties of temperature-sensitive mutants of polyoma virus. Cold Spring Harbor Symp. Quant. Biol. **39**:37–40.
12. **Graessmann, A., M. Graessmann, W. C. Topp, and M. Botchan.** 1979. Retransformation of a simian virus 40 revertant cell line, which is resistant to viral and DNA infections, by microinjection of viral DNA. J. Virol. **32**:989–994.
13. **Graham, F. L., S. Bacchetti, R. McKinnon, C. Stanners, B. Cordell, and H. M. Goodman.** 1980. Transformation of mammalian cells with DNA using the calcium technique, p. 3–25. *In* R. Baserga, C. Croce, and G. Rovera (ed.), Introduction of macromolecules into viable mammalian cells. A. R. Liss Inc., New York.
14. **Graham, F. L., and A. J. van der Eb.** 1973. A new technique for the assay of infectivity of human adenovirus 5 DNA. Virology **52**:456–467.
15. **Hayday, A., H. E. Ruley, and M. Fried.** 1982. Structural and biological analysis of integrated polyoma virus DNA and its adjacent host sequences cloned from transformed rat cells. J. Virol. **44**:67–77.
16. **Hwang, S.-P., and R. Kucherlapati.** 1980. Localization and organization of integrated simian virus 40 sequences in a human cell line. Virology **105**:196–204.
17. **Jelinek, W. R., and C. W. Schmid.** 1982. Repetitive sequences in eukaryotic DNA and their expression. Annu. Rev. Biochem. **51**:813–844.
18. **Ketner, G., and T. J. Kelly, Jr.** 1976. Integrated simian virus 40 sequences in transformed cell DNA: analysis using restriction endonucleases. Proc. Natl. Acad. Sci. U.S.A. **73**:1102–1106.
19. **Ketner, G., and T. J. Kelly, Jr.** 1980. Structure of integrated simian virus 40 DNA in transformed mouse cells. J. Mol. Biol. **144**:163–182.
20. **Kraft, C. H., and C. van Eeden.** 1968. A nonparametric introduction to statistics. The Macmillan Co., New York.
21. **Kucherlapati, R., S.-P. Hwang, N. Shimizu, J. K. McDougall, and M. R. Botchan.** 1978. Another chromosomal assignment for a simian virus 40 integration site in human cells. Proc. Natl. Acad. Sci. U.S.A. **75**:4460–4464.
22. **Mougneau, E., F. Birg, M. Rassoulzadegan, and F. Cuzin.** 1980. Integration sites and sequence arrangement of SV40 DNA in a homogeneous series of transformed rat fibroblast lines. Cell **22**:917–927.
23. **Neer, A., N. Baran, and H. Manor.** 1983. Integration of polyoma virus DNA into chromosomal DNA in transformed rat cells causes deletion of flanking cell sequences. J. Gen. Virol. **64**:69–82.
24. **Rigby, P. W. J., M. Dieckmann, C. Rhodes, and P. Berg.** 1977. Labelling deoxyribonucleic acid to high specific activity *in vitro* by nick-translation with DNA polymerase I. J. Mol. Biol. **113**:237–251.
25. **Ruley, H. E., L. Lania, F. Chaudry, and M. Fried.** 1982. Use of cellular polyadenylation signal by viral transcripts in polyoma virus transformed cells. Nucleic Acids Res. **10**:4515–4524.
26. **Seif, R., and F. Cuzin.** 1977. Temperature-sensitive growth regulation in one type of transformed rat cells induced by the *tsa* mutant of polyoma virus. J. Virol. **24**:721–728.
27. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
28. **Stringer, J. R.** 1981. Integrated simian virus 40 DNA: nucleotide sequences at cell-virus recombinant junctions. J. Virol. **38**:671–679.
29. **Stringer, J. R.** 1982. DNA sequence homology and chromosomal deletion at a site of SV40 DNA integration. Nature (London) **296**:363–366.
30. **Sylla, B. S., D. Bourgaux-Ramoisy, and P. Bourgaux.** 1980. Induction of viral DNA synthesis in clonal derivatives of a permissive cell line transformed by a temperature-sensitive polyoma virus. Virology **100**:357–369.
31. **Topp, W. C., D. Lane, and R. Pollack.** 1980. Transformation by SV40 and polyoma virus, p. 205–296. *In* J. Tooze (ed.), DNA tumor viruses, part 2. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
32. **van der Ploeg, L. H. T., and R. A. Flavell.** 1980. DNA methylation in the human γδβ-globin locus in erythroid and nonerythroid tissues. Cell **19**:947–958.

APPENDIX B

Volume 15 Number 19 1987 Nucleic Acids Research

Integration of a vector containing rodent repetitive elements in the rat genome

John C.Wallenburg[1], Alain Nepveu[2] and Pierre Chartrand[1]*

Département de microbiologie, Faculté de médecine, Centre Hospitalier Universitaire, Sherbrooke, Québec J1H 5N4, Canada

Received June 26, 1987; Revised and Accepted August 30, 1987

ABSTRACT

We have previously shown that integration of a polyoma vector containing rodent repetitive elements into rat cellular DNA is non-random (Wallenburg et al. J. Virol. 50: 678-683). Junctions between the polyoma vector and the host DNA occur in the repetitive sequences of the vector about ten times more frequently than would be expected if sequences from the vector were used randomly for integration. In this paper we looked at the host sequences involved in these junctions. Our analysis did not reveal any repetitive or specific sequences and we presume therefore that the repetitive sequences of the vector acted as hot spots for illegitimate recombination. We also analysed the integration mechanism and found that: First, even though the polyoma vector was transfected in the presence of carrier DNA, integration did not involve the formation of a transgenome. Second, in at least one of the clones analysed, integration resulted in deletion of host DNA sequences. Third, the host DNA displaced at the integration site was considerably longer than the integrated segment.

INTRODUCTION

Highly repeated short and long interspersed sequences (SINES and LINES) appear to be an ubiquitous component of mammalian genomes (1, 2, 3). Amongst the better characterized are the Alu (4) and KpnI (5) repetitive families in human, and B1 (6), B2 (7) and MIF-1 (8) repetitive families in mouse. These elements have copy numbers of about 10^5 that are scattered through their host genome. The role, if any, of these repetitive units remains unknown. They would appear however, to be capable of various recombination processes such as retroposition, gene conversion, translocation, excision, homologous and non-homologous recombination (reviewed in 2, 3).

We previously investigated the integration in the rat genome of a polyoma derived vector (RmI) containing rodent repetitive sequences (9). RmI (10, 11, Fig. 1) consists of a complete polyoma genome with an insertion of mouse cellular sequences (Ins), that contains copies of two rodent repetitive elements, B2 (7) and MT (12). We observed that integration of RmI in rat cells occurred preferentially via the repetitive sequences present in Ins. These results were subsequently confirmed in a separate study

© IRL Press Limited, Oxford, England.

(13). In fact, junctions between host cellular DNA and RmI occurred about ten times more frequently in the portion of Ins containing the repetitive sequences than would have been expected for a random process. To find out more about the recombination hot spot behaviour of the repetitive sequences of RmI, we cloned and sequenced four RmI-cellular junctions occurring in the repetitive sequences of RmI. The detailed analyses of these junctions are presented in this paper. We could not identify host sequences at the junctions that were repetitive or specific and presume therefore that the repetitive sequences of RmI acted as hot spots for illegitimate recombination events.

In studying the integration mechanism, we also made the following observations: first, even though RmI was transfected into rat cells with carrier DNA, it integrated directly into the host genome and not via the formation of a transgenome (14, 15, 16). Second, in at least one of the clones analysed, integration was accompanied by deletion of host DNA sequences and finally the host DNA sequences displaced at the site of integration were considerably longer than the ones being inserted.

MATERIALS AND METHODS

Cell lines

Cell lines W98.12 and W98.14 have been described previously (9). They were produced by transfecting Fisher rat 3T3 (FR3T3) cells (17) with RmI using the calcium phosphate technique (18, 19). The carrier DNA used was high molecular weight DNA isolated from mouse LTA cells (19). Clone 11.6 was produced by using the same protocol with minor modifications (56). Briefly, 5×10^5 FR3T3 cells were seeded on a 90 mm petri dish 24 hours prior to transfection with 0.4 micrograms RmI and 13 micrograms of salmon sperm carrier DNA. The medium was replaced with fresh medium 20 hours post transfection, and 44 hours post transfection, cells were trypsinized and reseeded at 5×10^4 cells/60 mm petri dish. After 14 to 17 days, foci of densely growing cells were isolated, subcloned and grown up for DNA extraction. Clone 11.6 contains a single insertion of RmI as determined by detailed restriction enzyme analysis (56).

DNA preparation and restriction enzyme analysis

Total cellular DNA was prepared as previously described (9). Restriction enzymes were used as suggested by the manufacturer with a 5x excess of enzyme. For analyses of cellular DNA, 5 to 10 micrograms of high molecular weight DNA were digested. RmI was prepared as previously described (10).

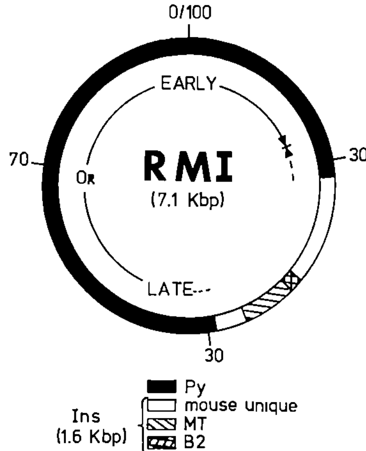


Figure 1
The physical structure of RmI as previously determined (9, 10). The numbers refer to standard polyoma map units and OR identifies the polyoma origin of replication. The early and late coding regions have been indicated with the late region interrupted by Ins. Ins is an insertion of mouse cellular DNA. B2 and MT refer to the B2 (11) and MT (12) families of rodent repetitive DNA.

Probes and hybridization

For Southern analyses the DNA was transferred to nitrocellulose (type BA85, Schleicher & Schwell Inc.) or nylon (Hybond, Amersham) membranes by the technique of Southern (20), with the modifications suggested by the manufacturer. The hybridization procedure was that of van der Ploeg and Flavell (21) with minor modifications (22). Probes were prepared by the nick translation procedure of Rigby *et al* (23) or by using the random hexamer-primed synthesis technique (24).

Cloning the RmI-cellular DNA junctions

High molecular weight cellular DNA was digested with *Eco*RI and the lengths of the RmI-cellular DNA junctions determined. The appropriate fragments were enriched on sucrose gradients (25) and recovered by two rounds of ethanol precipitation. The cloning was done in vectors λgtWES. λB (26) or λL47.1 (27) followed by *in vitro* packaging. The junction-containing recombinants were identified by plaque hybridization (28) against a polyoma probe, and the inserts subcloned into pBR322 at the unique *Eco*RI site, except 11.6L which was cloned in pAT153/*Eco*RI. The initial site of 11.6 (IS) was cloned by a similar procedure as a 9 Kbp *Bgl*II fragment into λL47.1 digested with *Bam*HI. An *Eco*RI subfragment of this recombinant, a fragment of 2.6 Kbp, was subcloned into pBR322.

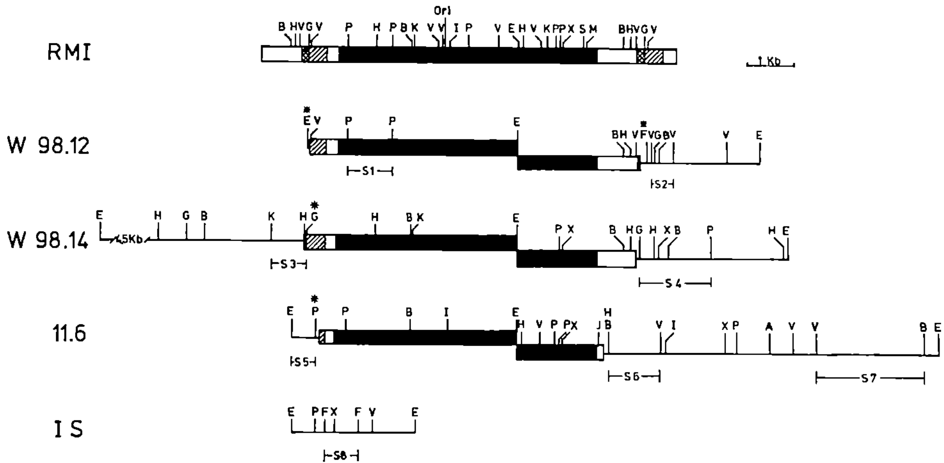


Figure 2
The physical maps of the cloned RmI-cellular junctions. Each junction was cloned as an EcoRI fragment as described in Materials and Methods. Thus both EcoRI junctions of a single clone are shown in juxtaposition separated by the EcoRI site of RmI. The name of the clone from which they were derived is given on the extreme left. The inserted sequences of RmI have been aligned with the linearized representation of RmI given at the top of the figure. (Ins has been duplicated for the purpose of illustration.) The legend for the RmI sequences is the same as in figure 1. The line represents the cellular sequences flanking each insertion. IS is an EcoRI cellular fragment that represents the arrangement of the cellular sequences at the left side of the RmI insertion in clone 11.6 before integration. It has been aligned with the corresponding cellular sequences in 11.6 directly above. S1 through S7 delineate the probes used for the restriction enzyme mapping of the initial sites. S8 is the probe used to determine the fate of the initial sequences after integration (see figure 6A). Restriction enzymes : A, HpaI; B, BamHI; E, EcoRI; F, HinfI; G, BglII; H, HindIII; I, BglI; J, AvaI; K, KpnI; M, SmaI; P, PstI; S, SalI; V, PvuII; X, XbaI. Not all restriction sites are shown. Sites with an asterisk were used to sequence the junctions.

Sequencing

The sequences across the junctions were determined using the chemical cleavage procedure of Maxam and Gilbert (29, 30).

Sequence comparisons

The determined sequences were compared to published sequences for the B1 (6), B2 (7), R (31) and MT (12) consensus sequences and Bam5 (32), and L1Rn, (33) families of repetitive DNA, using the DOTMATRIX (stringency=60, 70 or 80% over 10 bases) and HOMOLOGY (stringency=65% over 20 bases) programs of Stephens (34).

Copy number

Relative copy number of DNA fragments from Southern analysis were

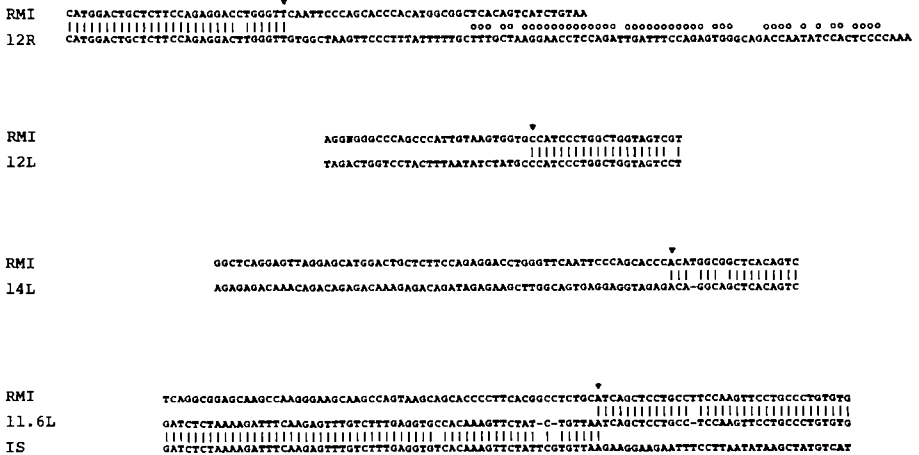


Figure 3
The sequences across the junctions shown in comparison to the corresponding sequences of RmI. 11.6L is also compared to the initial site (IS). Arrow heads indicate the cross-over point in each junction. Vertical lines are drawn between homologous bases. Dashes have been inserted to allow alignment of homologous sequences. Bases with a halo (in the sequence of 12-R) are homologous with L1Rn (see text).

determined by densitometry scanning of autoradiograms using an LKB Ultroscan XL and accompanying Gelscan XL software.

RESULTS

Cloning of the RmI insertions in clones W98.12, W98.14 and 11.6

Clones W98.12, W98.14 and 11.6 were selected for analysis of the role played by repetitive sequences in the integration of RmI in the rat cellular genome. Each of these clones contained only one RmI insertion and in each case at least one RmI-cellular DNA junction had occurred in the repetitive sequences present in RmI (Fig. 2). Restriction enzyme mapping located the right junction of insertion W98.12 (12R) and the left junction of insertion W98.14 (14L) in the B2 repetitive sequences present in RmI, while the left junction of W98.12 (12L) and the left junction of 11.6 (11.6L) mapped in the MT repetitive sequences present in RmI. Total cellular DNA from each clone was digested with EcoRI and each junction cloned separately (see Material and Methods).

Nature of the endogenous nucleotide sequences at the junctions

The restriction sites designated by asterisks in Fig. 2 were used to determine the nucleotide sequences of RmI-cellular junctions 12R, 12L, 14L and 11.6L (Fig. 3). These have been aligned with the corresponding RmI

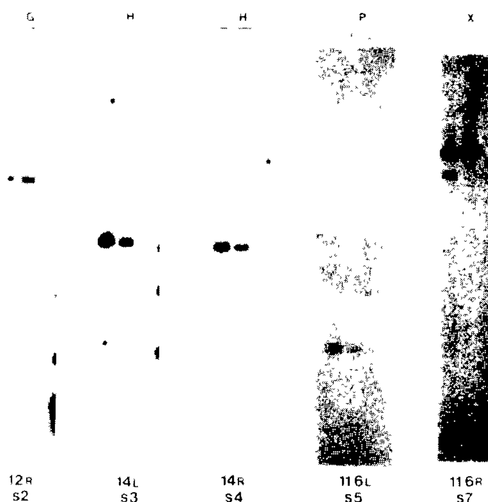


Figure 4
RmI integrated directly into host cellular DNA and not into the carrier DNA. Five micrograms of clone (W or 11), host FR3T3 (F), and carrier LTA (L) and 5 (S) or 10 micrograms (S′) of carrier salmon sperm. DNAs were digested with restriction enzymes, migrated beside each other on agarose gels, blotted and hybridized to probes of cellular DNA. The probes (S2, S3, S4 and S5 shown in Fig. 2), were chosen to contain only unique sequence DNA and to cover the region immediately flanking the site of integration. Probe S7 (Fig. 2) hybridized to sequences present in the order of 5 to 10 copies per genome. The junctions analysed and the probes used are indicated beneath each autoradiogram. Restriction enzymes: G, BglII; H, HindIII; P, PstI; X, XbaI. The host DNA immediately flanking 11.6R (which includes the Xba site) has undergone a rearrangement, and therefore this Xba fragment (lower band in 11.6) does not comigrate with the unrearranged fragments (upper band in 11.6 and FR3T3). Note that the single copy rearranged band is several times less intense than the unrearranged bands.

sequences. The junctions are homologous with RmI up to a certain point (see arrowhead in Fig. 3) beyond which there is no significant homology. In all four cases the break point occurred in repetitive sequences of RmI, either B2 (12R, 14L) or MT (12L, 11.6L). The preferential involvement of RmI repetitive sequences in integration could have been the result of homologous recombination with endogenous repetitive sequences of the host genome. However, if this had been the case, we would have expected to find at the junctions, complete chimaeric repetitive elements originating in part from RmI and in part from the host. What we found at the junctions in all cases were truncated repetitive elements linked to sequences unrelated to the repetitive element. There is some divergence between the repetitive sequences at the junctions and the repetitive sequences present in RmI. We do not know how these originated, but it should be noted that RmI is

produced by amplification in mouse cells and these base differences could represent variations between RmI molecules. They could also be mutations induced by transfection or by integration (35, 36). To confirm that homologous endogenous repetitive sequences were not present at the junctions, we used the flanking cellular sequences of junction 11.6L (Fig. 2, fragment S5) as a probe to clone the cellular sequences of 11.6L in their original arrangement before integration. This segment is designated IS in Fig. 2. We determined the nucleotide sequence across the site where RmI had integrated. (IS in Fig. 3). This sequence was found to be homologous to the cellular sequence found at the 11.6L junction but had no significant homology with the repetitive sequences of RmI present at the junction. Thus junction 11.6L occurred in repetitive sequences of RmI but did not involve homologous repetitive sequences of the rat cellular genome. Although as mentioned above, the structures for junctions 12R, 12L and 14L make it highly unlikely that homologous cellular sequences were involved in integration, we cannot exclude this possibility since the nature of the original host sequences was not determined.

It has been proposed (37) that integration of exogenous DNA occurs preferentially in host cellular repetitive sequences. We did a computer search of the sequences for the presence of highly repetitive sequences in the cellular host DNA at or near the junctions (see Material and Methods). None of the four cellular junctions had significant homologies with the major families of rodent repetitive sequences, including B1 (6), B2 (10), R (31), Bam5 (32), MT (12) or L1Rn (33), with the exception of a sequence close to junction 12R (Fig. 3). This sequence had a homology of 78% over 56 bp, including a 36 bp stretch of 90% homology, with a middle segment of the LINE repetitive element L1Rn (33). Furthermore, probes derived from the cellular DNA at (Fig. 2, fragment S3) or near the junctions (Fig. 2, fragments S2, S4, S5, S6, S7, S8) hybridized with only one band of host DNA (see Fig. 4 and 6). Probes S6 and S7 hybridized with sequences present in the order of 5 to 10 copies per genome (Fig. 4). Thus our results do not indicate that integration would occur preferentially in host highly repetitive DNA. Finally, in comparing the host sequences at the junction, we did not find any common sequence or structure at or near the junction.

Analysis of the integration mechanism

It has been shown that when DNA is transfected into mammalian cells with carrier DNA, it can form structures designated transgenomes or pekelosomes that can integrate into the host DNA (14, 15, 16). Transgenomes or pekelosomes are high molecular weight DNA structures of several hundred Kbp that result from ligation between the transfected DNA and the carrier

DNA. Since we wished to examine the interaction between RmI and endogenous host sequences, it was necessary to determine whether RmI had integrated directly into the host DNA or via a transgenome. As we had used salmon sperm or mouse DNA as carrier, cellular DNA segments near the junctions were used as probes in Southern analysis to determine if the DNA flanking the RmI insertions originated from the host or carrier DNA (Fig. 4). For each junction we hybridized the corresponding cellular probes (S2, S3, S4, S5, S7 in Fig. 2) to DNAs originating from the clones (lanes W or 11), from the parental FR3T3 cell line (lanes F) and from the carrier mouse LTA (lanes L) or salmon sperm (lanes S and S'). The left junction of clone 98.12 (12L) could not be analysed because too few non-RmI sequences had been cloned to make a probe (Fig. 2). The restriction enzymes used to digest the DNA introduced a cut in the cellular sequences near the cross-over point (see Fig. 2). Thus if the cellular sequences at the junction originated from the host, we would expect the probe to hybridize with bands co-migrating in the clone and the parental cell line and not present in the carrier. This is what we found in all cases (Fig. 4). With clones 98.12 and 11.6 there was no hybridization between the probes and the carrier. Probes S3 and S4 (from clone 98.14) did hybridize weakly to non co-migrating bands in the carrier DNA. We believe this was due to cross-hybridization between rat and mouse DNA. For junctions 12R and 14R we see more than one band in both the clones and the parental line. This is because the probe contains a site for the restriction enzyme used in the analysis and thus the probe hybridizes to more than one fragment of cellular DNA. We repeated such analyses with a number of restriction enzymes (data not shown) and the results confirmed that the flanking cellular sequences of all the junctions analysed originated from the parental FR3T3 DNA and that carrier was not involved in the integration of RmI. Furthermore in the case of junction 11.6L we have already shown that the nucleotide sequences up to the RmI insert originated from the parental FR3T3 genome (Fig. 3). Thus we conclude that integration of RmI did not involve the formation of transgenomes or pekelosomes, even though RmI was transfected in the presence of carrier DNA.

Rearrangement of host sequences at the site of integration of RmI

The cellular probes described above , S2, S3, S4, S5, S6 and S7 could be used to deduce the physical map of the host sequences contiguous to the integration sites before integration. This was done by establishing their restriction maps in the parental cell line. These maps are presented in Fig. 5 along with the physical map of the flanking host sequences at the integration sites. In none of the three clones were we able to establish that the host sequences flanking each side of the integration site were

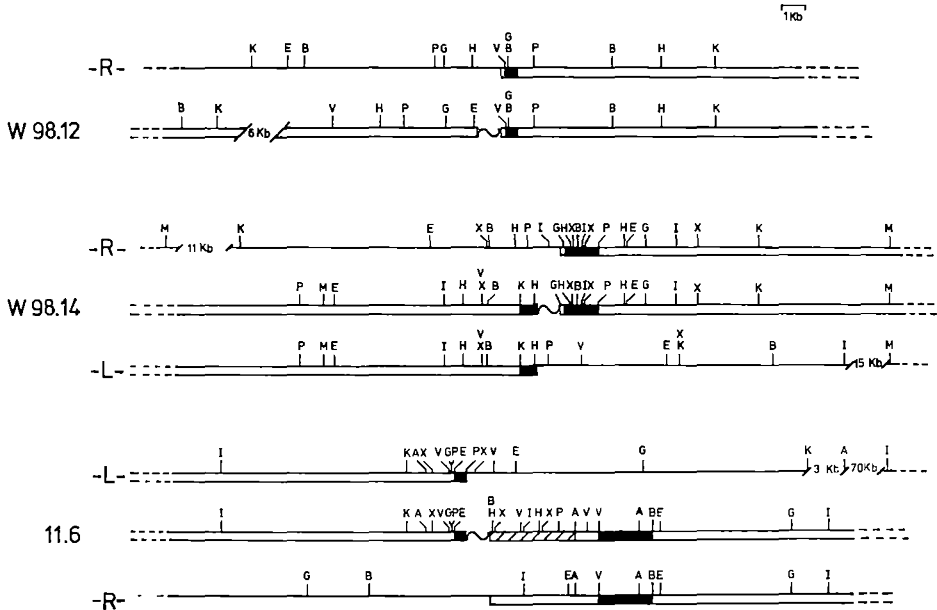


Figure 5
Maps of the rearrangements of the cellular DNA associated with the integration of RmI. Beside the name of each clone (given on the extreme left) is the map of the cellular DNA after integration. (The integrated sequences of RmI are represented by wavy lines.) Aligned above and below each map are the maps of the initial sites before integration as determined by Southern analysis using probes of cellular DNA, shown as solid bars. These correspond to probes S2 to S7 in figure 2. Open bars: cellular DNA flanking the insertions. Solid lines: cellular DNA before integration. Hatched bars: cellular DNA flanking the insertion in 11.6 which has been reorganized (see text). -R- : initial site as mapped by a probe from the right flanking sequences. -L-: initial site as mapped by a probe from the left flanking sequences. Restriction enzymes: A, HpaI; B, BamHI; E, EcoRI; G, BglII; H, HindIII; I, BglI; K, KpnI; M, SmaI; P, PstI; V, PvuII; X, XbaI. Not all restriction enzyme sites are shown.

linked before the integration event. That is, in none of the three cases did we find a host fragment that would hybridize with probes from both sides of the integration site nor could we match the restriction maps established with the probe from one side with the restriction maps established for the other side. What we can deduce from these maps however, is that before integration the flanking cellular sequences of clones W98.12, W98.14 and 11.6 were separated by at least 12 Kbp, 55 Kbp and 90 Kbp respectively. These numbers were obtained as follows: in the case of clone W98.12 we could establish the restriction map of the cellular site before integration up to 12 Kbp to the left of the S2 probe and it differed totally from the

restriction map of the cellular sequences present at the left of the integration site (see Fig.5). In the case of clone W98.14, before integration we found that the restriction map for at least 29 Kbp to the right of the S3 fragment and 26 Kbp to the left of the S4 fragment were distinct. Thus S3 and S4 were originally separated by at least 55 Kbp. Finally in the case of clone 11.6, we found that digestion of the parental DNA with BglI generated a fragment that in low concentration agarose (.35%) migrated much higher in the gels than 50 kb. We estimate that this fragment extends 90 Kbp to the right of S5. It did not hybridize with either S7 or S6. We also found that the cellular DNA immediately flanking the 11.6R junction (Fig. 5, hatched region) had undergone rearrangement as compared to the host sequences from the right hand side before integration (see Fig. 4 legend). However, using the S6 and S7 probes (Fig. 2) we could establish that these rearranged flanking sequences were linked with the unrearranged distal cellular sequences before integration but in a different configuration (data not shown).

The above results raised the question: were the host flanking sequences linked before integration or did they originate from distinct chromosomes? A fortuitous event indicates that at least in clone W98.14 the flanking cellular sequences appeared to be linked together before integration. This is illustrated in Fig. 6A. We have put side by side DNA from W98.14 and FR3T3 and hybridized them with probes S3, S4 and S5. The relative intensity of the resulting bands was determined by densitometry scanning (see Material and Methods). S5 hybridizes to a cellular band unrelated to the integration site in clone W98.14 and it was used to correct for small variations of intensity due to variation of DNA content in each lane. When the band intensities are corrected, we find that the S4 bands and the total of the S3 bands are twice as intense in the W98.14 clone than in the parental FR3T3. Thus there are twice as many copies of the flanking cellular sequences in the W98.14 clone than in its parental clone. But when we look at the integration site as illustrated by the lowest S3 band for which there is no corresponding band in FR3T3, we find it to be only half as intense as the upper S3 band in FR3T3. This tells us that the amplified copies are the allelic sequences and not the integration site. This amplification of the flanking sequences is also evident in Fig. 4. A simple explanation for these results is that integration occurred fortuitously in sequences that were amplified compared to the parental sequences. The fact that the sequences from both sides of the integration site have been amplified to an equal extent in an event unlinked to integration, suggests that these sequences were linked prior to integration. Note that the amplification did not affect

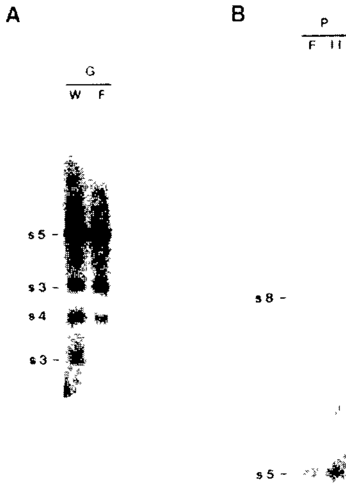


Figure 6
A. Autoradiogram of cellular DNAs digested with BglII (G), run through an agarose gel, transferred to a nylon membrane and hybridized with probes S3, S4 and S5. W, 5 micrograms W98.14 DNA; F, 5 micrograms FR3T3 DNA. The bands hybridizing with each probe are indicated. S3 hybridizes with two bands in W98.14. The upper band corresponds to unrearranged DNA and comigrates with an FR3T3 band, the lower band corresponds to the sequences flanking the RmI insertion.
B. Integration of RmI results in a deletion of cellular DNA. Five micrograms of 11.6 (11) and FR3T3 (F) DNA were digested with PstI (P) and migrated through an agarose gel, blotted, and hybridized against probes S5 and S8 (figure 2). S8 is a probe of cellular DNA from the initial site of 11.6.

unrelated sequences as judged by the S5 bands that are of equal intensities in both lanes.

Did integration result in deletion of host sequences? To answer this question we digested 11.6 and FR3T3 DNA with PstI, which separates the flanking unrearranged host DNA from the host DNA rearranged by the integration of RmI (Fig. 2, 11.6 and I.S.). We then hybridized with the S5 probe homologous to the unrearranged host DNA and the S8 probe homologous to the rearranged host DNA. The results are shown in Fig. 6B. We did a densitometric scanning of the autoradiogram and used the S5 band to correct for variation of DNA content in each lane. After correction, we find that the band hybridizing to the S8 probe is only half as intense in 11.6 as in FR3T3. Since no other band hybridizing to the S8 probe is found in 11.6 DNA, we conclude that the rearranged host DNA has been deleted. This was confirmed by analyses with other restriction enzymes (data not shown).

DISCUSSION

We had previously established that exogenous repetitive sequences were hot spots for integration in the rat genome (9, 13). In this paper, we examined the host DNA into which the exogenous DNA had integrated. Our analysis failed to reveal the presence of repetitive or specific host sequences. Thus it would appear that integration occurred through

illegitimate recombination and that repetitive sequences can act as hot spots for such events. This would explain, as has been hypothesized by Rogers (38), why repetitive sequences are often found at illegitimate junctions such as translocation break points (39, 40, 41) or in cellular excision product referred to as small polydispersed circular DNA (spc) (42, 43, 44). It also raises the interesting possibility that if reintegration of spc DNA in the host genome was to occur it would preferentially occur via the repetitive sequences. This would preserve the integrity of the unique DNA present in these spc molecules.

Another point to be made is that even though the host genome contained upwards of 10^5 endogenous repetitive sequences homologous to the ones present in RmI, there is no indication that integration occurred in these sequences. Obviously the small number of events that we looked at does not exclude that such events would occur at a lower frequency than illegitimate integration. It is possible that the length and degree of homology between the repetitive sequences of RmI and that of the host DNA were not sufficient to favor homologous integration. It has been shown however that homologous recombination between endogenous chromosomal repetitive sequences can occur (41, 42, 45, 46). It is also possible that homology searching is not the rate limiting step that will determine if integration is to occur by homologous or illegitimate recombination. This would explain why the frequency of homologous integration in mammalian cells is very low compared to bacteria and yeast, even when taking into account genome size (47, 48, 49, 50).

It has been postulated that DNA transfected in the presence of carrier forms a high molecular weight structure designated transgenome or pekelosome and that it is these structures that integrate into the host genome (14, 15, 16). Transgenomes are the product of ligation between carrier and the transfected DNA and result in molecules of several hundred Kbp in size. If integration of transfected DNA was necessarily mediated by transgenomes then this would preclude drawing conclusions on the interaction of specific exogenous DNA with the host genome. In this paper we have demonstrated that RmI integrated on its own, without the formation of transgenomes. Thus transgenomes are not a necessary intermediate for the integration of transfected DNA.

A number of authors have reported that integration of exogenous DNA resulted in genome rearrangements (15, 37, 51, 52, 53, 54, 55). However the nature and the extent of these rearrangements were undetermined nor had it been established if host sequences were lost as a result of integration. In all cases, what had been determined as the minimum length of the rearranged

host sequences was equal or smaller to what had been integrated, except in one instance where the rearrangement was of at least 17 Kbp for an insertion of about 8 Kbp (37). This represented the longest rearrangement so far determined. Thus from these results it could be hypothesized that integration involved a simple replacement of host sequences by an incoming linear molecule causing a deletion of similar length as the insertion. Our analysis of the integration site of clone 11.6 establishes that indeed integration can result in the deletion of host sequences. However the extent of the rearranged host sequences, which can be more than 90 Kbp for an insert of 6 Kbp is incompatible with a linear insertion-replacement model as suggested above. Such extensive rearrangements could be explained by an integration mechanism in which the incoming exogenous DNA replaces a looped structure of host cellular DNA.

ACKNOWLEDGEMENTS

We are grateful to Gilles Roy for his help in the computer searches for sequence homologies, to Viviane Bibor-Hardy for her help with densitometry scanning and to Nadine Gusew for her help in cloning in the lambda vectors and for the initial cloning of 11.6R. We thank Johanne Proulx for excellent technical assistance, Maryse Chalifoux for excellent technical assistance as well as for drawing some of the figures, Suzanne Brouillette, Gilles Roy and Pierre Trudel for valuable discussions and Nicole Langlois for typing the manuscript.

This work was supported by a grant from the Medical Research Council of Canada. P. C. is a Medical Research Council of Canada scientist. J. C. W. is the recipient of a studentship from the Fonds pour la formation de chercheurs et l'aide à la recherche.

*To whom correspondence should be addressed

Present addresses: [1]Canadian Red Cross, 3131 Sherbrooke Street East, Montreal, Québec H1W 1B2, Canada and [2]Biochemistry Department, State University of New York, Stony Brook, NY 11794, USA

LITERATURE CITED

1. Singer, M.F. (1982) Cell 28, 433-434.
2. Jelinek, W.R. and Schmid, C.W. (1982) Ann. Rev. Biochem. 51, 813-844.
3. Rogers, J.H. (1985) Int. Rev. Cytol. 93, 187-279.
4. Houck, C.M., Rinehart, F.P. and Schmid, C.W. (1979) J. Mol. Biol. 132, 289-306.
5. Adams, J.W., Kaufman, R.E., Kretschmer, P.J., Harrison, M. and Nienhuis, A.W. (1980) Nucl. Acid Res. 8, 6113-6128.
6. Krayev, A.S., Kramerov, D.A., Skryabin, K.G., Ryskov, A.P., Bayev, A.A. and Georgiev, G.P. (1980) Nucl. Acids Res., 8, 1201-1215.
7. Krayev, A.S., Markusheva, T.V., Kramerov, D.A., Ryskov, A.P., Skryabin,

K.G., Bayev, A.A. and Georgiev, G.P. (1982) Nucl. Acids Res., 10, 7461-7475.
8. Brown, S.D.M. and Dover, G. (1981) J. Mol. Biol. 150, 441-466.
9. Wallenburg, J.C., Nepveu, A. and Chartrand, P. (1984) J. Virol., 50, 678-683.
10. Bourgaux, P., Sylla, B.S. and Chartrand, P. (1982) Virology, 122, 84-97.
11. Sylla, B.S., Allard, D., Roy, G., Bourgaux-Ramoisy, D. and Bourgaux, P. (1984) Gene, 29, 343-350.
12. Heinlein, V.R.O., Lange-Sablitzky, R., Schoal, H. and Wille, W. (1986) Nucleic Acids Res., 14, 6403-6416.
13. Gusew, N., Nepveu, A. and Chartrand, P. (1987) Mol. and Gen. Genet. 206, 121-125.
14. Perucho, M., Hanahan, D. and Wigler, M. (1980) Cell, 22, 309-317.
15. Robins, D.M., Ripley, S., Henderson, A.S. and Axel, R. (1981) Cell, 23, 29-39.
16. Scangos, G.A., Huttner, K.M., Juricek, D.K. and Ruddle, F.H. (1981) Mol. Cell. Biol., 1, 111-120.
17. Seif, R. and Cuzin, F. (1977) J. Virol., 24, 721-728.
18. Graham, F.L., and van der Eb, E.J. (1973) Virology, 52, 456-467.
19. Graham, F.L., Bachetti, S., McKinnon, R., Stanners, C., Cordell, B. and Goodman, H.M. (1980) In R. Baserga, C. Croce and G. Rovera (ed.). Introduction of macromolecules into viable mammalian cells. A.R. Liss Inc., New York, p. 3-25.
20. Southern, E.M. (1975) J. Mol. Biol., 98, 503-517.
21. van der Ploeg, L., Flavell, H.T. and R.A. (1980) Cell, 19, 947-958.
22. Chartrand, P., Gusew-Chartrand, N. and Bourgaux, P. (1981) J. Virol., 39, 185-195.
23. Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Berg, P. (1977) Mol. Biol., 113, 237-251.
24. Feinberg, A.P. and Vogelstein, B. (1983) Anal. Biochem., 132, 6-13.
25. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Cold Spring Harbor Laboratory, Cold Spring Harbor, USA.
26. Leder, P., Tiemeier, D. and Enquist, L. (1977) Science, 196, 175-177.
27. Loenen, W.A. and Brammar, W.J. (1980) Gene, 10, 249-259.
28. Benton, W.D. and Davis, R.W. (1977) Science, 196, 180-182.
29. Maxam, A.M. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 560-564.
30. Maxam, A.M. and Gilbert, W. (1980) Methods in Enzymology, 65, 499.
31. Gebhard, W., Meitinger, T., Höchtl, J. and Zachau, H.G. (1982) J. Mol. Biol., 157, 453-471.
32. Fanning, T.G. (1982) Nucl. Acids Res., 10, 5003-5013.
33. D'Ambrosio, E., Waitzkin, S.D., Witney, F.R., Salemme, A. and Furano, A.V. (1986) Mol. Cell. Biol., 6, 411-424.
34. Stephens, R.M. (1985) Gene Anal. Tech., 2, 67-75.
35. Calos, M.P., Lebkowski, J.S. and Botchan, M.R. (1983) Proc. Natl. Acad. Sci. USA 80, 3015-3019.
36. Razzaque, A., Mizusawa, H. and Seidman, M.M. (1983) Proc. Natl. Acad. Sci. USA 80, 3010-3014.
37. Kato, S., Anderson, R.A. and Camerini-Otero, R.D. (1986) Mol. Cell. Biol., 6, 1787-1795.
38. Rogers, J. (1985) Nature, 317, 559.
39. Vanin, E.F., Henthorn, P.S., Kioussis, D., Grosveld, F. and Smithies, O. (1983) Cell, 35, 701-709.
40. Heisterkamp, N., Stam, K. and Groffen, J. (1985) Nature, 315, 758-761.
41. Lehrman, M.A., Schneider, W.J., Südhof, T.C., Goldstein, J.L. and Russell, D.W. (1985) Science, 227, 140-146.
42. Krolewski, J.J., Schindler, C.W. and Rush, M.G. (1984) J. Mol. Biol., 174, 41-54.
43. Riabowol, K., Shmookler, R.J. and Goldstein, S. (1985) Nucl. Acids Res. 13, 5563-5584.

44. Hollis, M. and Hindley, J. (1986) Gene 46, 153-160.
45. Shmokler Reis, R.J., Lumpkin, C.K. Jr., Riabowol, K.T. and Goldstein, S. (1983) Nature, 301, 394-398.
46. Lehrman, M.A., Russell, D.W., Goldstein, J.L. and Brown, M.S. (1986) Proc. Natl. Acad. Sci. USA 83, 3679-3683.
47. Smith, A.J.H. and Berg, P. (1984) Cold Spring Harbor Symp. Quant. Biol., 49, 171-181.
48. Smithies, O., Koralewski, M.A., Song, K.-Y. and Kucherlapati, R.S. (1984) Cold Spring Harbor Symp. Quant. Biol., 49, 161-170.
49. Lin, F.-L., Sperle, K. and Sternberg, N. (1985) Proc. Natl. Acad. Sci. USA 82, 1391-1395.
50. Thomas, K.R., Folger, K.R. and Capecchi, M.R. (1986) Cell, 44, 419-428.
51. Botchan, M., Stringer, J., Mitchison, T. and Sambrook, J. (1980) Cell, 20, 143-152.
52. Hayday, A., Ruley, H.E. and Fried, M. (1982) J. Virol., 44, 67-77.
53. Stringer, J.R. (1982) Nature, 296, 363-366.
54. Neer, A., Baran, N. and Manor, H. (1983) J. Gen. Virol., 64, 69-82.
55. Williams, T.J. and Fried, M. (1986) Mol. Cell. Biol., 6, 2179-2184.
56. Nepveu, A. (1985) Ph.D. Thesis, Université de Sherbrooke, Sherbrooke, Canada.

APPENDIX C

Volume 16 Number 2 1988 Nucleic Acids Research

Inexpensive and simple set-up for field inversion gel electrophoresis

Gilles Roy, John C.Wallenburg and Pierre Chartrand*

Canadian Red Cross, BTS, 3131 Sherbrooke St. East, Montreal, Quebec H1W 1B2, Canada
Submitted December 11, 1987

The introduction of field inversion gel electrophoresis now permits the easy characterization of very long fragments of DNA (1). This is in the process of revolutionizing mammalian genetics by enabling the mapping of genetic markers on fragments much longer than was possible by conventional gel electrophoresis (2). We have devised a simple and inexpensive set up for the use of this technique based on equipment already available in most laboratories, namely: 1. horizontal gel apparatus; 2. home made switch box; 3. power supply; 4. microcomputer (IBM PC or compatible). The switch box (Fig. 1) is a very simple electronic design made of components readily available at electronic stores. The terminals of the gel apparatus are connected to the switch box through the output terminals (P2). The input terminals (P1) are used to connect the switch box to one or two power packs. The use of two power packs permits the use of two different voltages for forward and reverse electrophoresis. If the same voltage is applied forward and reverse, the same power pack is connected to both pairs of input terminals on the switch box. The switch box is interfaced with a microcomputer (P4). The output signal from the computer is zero or 5 volts for the forward and reverse pulses, respectively. The computer is used to control the forward and reverse pulses as a function of time. The program written in IBM Basic, asks for the gel duration and number of periods. The pulse times are then entered manually, allowing great flexibility, or calculated automatically. The operational status is continuously displayed on the screen. We routinely use this set up to run FIGE gels as described by Carle et al (1) and Lalande et al (3) and separate, on the same gels, DNAs ranging from 2 kbp to 2000 kbp. Any further information can be obtained from the authors (514) 527-1501. To obtain a copy of the program, send a blank 5 1/4" floppy diskette to the authors.

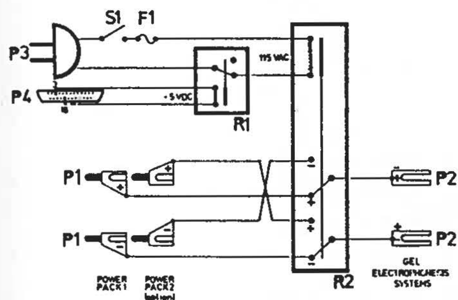


Fig. 1 Schematic diagram of the switch box. S1: switch (unipolar, 115 VDC); F1: fuse (115 VAC, 0.2 A); R1: Relay (unipolar, miniature 5 VDC); R2: Relay (bipolar, 115 VDC). P1: banana plugs ♂; P2: banana plugs ♀; P3: outlet plug (115 VAC); P4: IBM line printer connector (25 positions). Connections to the printer port are made to pins 3 and 18, as shown. Use AWG22 insulated wires.

*To whom correspondence should be addressed

References

1. Carle, G.F. et al. (1986) Science, 232, 65-68.
2. Barlow, D.P. and Lehrach, H. (1987) Trends in Genetics, 3, 167-171.
3. Lalande, M. et al. (1987) Proc. Natl. Acad. Sci. USA, 84, in press.

© IRL Press Limited, Oxford, England.

APPENDIX D

FIGE program: "Cycler"

This program, in conjunction with the following program, allows the user to calculate and to program a set of parameters (number of periods in a ramp, period lengths, and number of cycles to repeat the ramp) for a FIGE gel, automatically. The following program ensures the continuity in the case of a power failure.

```
20 COLOR 14,1,12
30 REM PROGRAMME PULSATEUR-INVERSEUR: initialisation automatique
40 REM****************INITIALISATION DES PARAMETRES DE PULSATION************
50 DIM S(100), N(100)
60 CLS
70 PRINT "        P R O G R A M M E  P U L S A T E U R - I N V E R T E U R"
80 PRINT: PRINT " ELECTROPHORESE DU:" ; DATE$
90 PRINT: PRINT
100 PRINT " Quel est le voltage initial";: INPUT VS
110 PRINT "                             inverse";: INPUT VN
140 PRINT " Temps total de migration (en heures)";: INPUT RUNTIME
150 PRINT " Pour combien de périodes différentes";: INPUT C
155 PRINT " How many cycles to repeat these periods?";: INPUT CYCLENO
160 CYCLETIME=RUNTIME*3600/CYCLENO
165 PERIODE=CYCLETIME/C
180 PRINT: PRINT " RAMPE DISCONTINUE DE";C;"PERIODES DE
";PERIODE;"SECONDES"
190 PRINT: PRINT " REPEATED";CYCLENO;"CYCLES. CYCLE TIME
=";(CYCLETIME/60);"MIN"
200 PRINT: PRINT " DUREE TOTALE=";RUNTIME;"hrs."
210 PRINT: PRINT
230 PRINT " Nombre de secondes pour la migration initiale=";: INPUT S(1)
240 PRINT "                                              inverse =";: INPUT N(1)
250 PRINT: PRINT " Nombre de secondes additionnelles  chaque nouvelle
période."
260 PRINT "                                    Migration initiale=";: INPUT SINC
270 INPUT "                                              inverse ="; NINC
280 CLS: PRINT "PERIODE=";1;"MIGRATION INITIALE=";S(1);"MIGRATION
INVERSE=";N(1)
300 FOR X=2 TO C
310 S(X)=S(X-1)+SINC
320 N(X)=N(X-1)+NINC
330 PRINT "PERIODE=";X;"MIGRATION INITIALE=";S(X);"MIGRATION
INVERSE=";N(X)
340 FOR W=0 TO 500: NEXT
350 NEXT X
360 BEEP: PRINT: PRINT "                 --------> Vérifiez les paramètres
<--------"
370 PRINT " -------> Vérifier vos installations et la boîte de contrôle
<------"
380 PRINT: PRINT "                       APPUYEZ SUR UNE TOUCHE POUR
DEMARRER";: INPUT W
385 OPEN "a:PULSE.DAT" FOR OUTPUT AS #1
386 PRINT #1,VS,VN,RUNTIME,C,S(1),N(1),SINC,NINC,CYCLENO,CYCLETIME,PERIODE
387 CLOSE #1
```

```
390 REM *************INSTRUCTION DE L'INTERFACE DE
PULSATION****************
400 CLS: X=1 : V=1 : LOCATE 1,5: PRINT "DUREE TOTALE  =";RUNTIME;"hrs."
410 GOSUB 620
415 FOR V=1 TO CYCLENO
420 FOR X=1 TO C
425 OPEN "A:PULSEU.DAT" FOR OUTPUT AS #1
426 PRINT #1,X,V
427 CLOSE #1
428 LOCATE 4,5 : PRINT "CYCLE #";V;"OF";CYCLENO
430 LOCATE 5,5 : PRINT "PERIODE #";X;"SUR";C
440 TIME$="00:00:00"
450 LOCATE 2,5: PRINT "TOTAL TIME
REMAINING:";(RUNTIME-(CYCLETIME*(V-1)/3600))
455 LOCATE 3,5: PRINT "MIN. REMAINING IN
CYCLE:";((CYCLETIME-(PERIODE*(X-1)))/60)
460 REM boucle de la pulsation S>N
470 ON TIMER(S(X)) GOSUB 660
480 TIMER ON
490 LOCATE 6,5: PRINT "MINUTES RESTANTES DANS CETTE
PERIODE:";CINT((PERIODE-TIMER)/60)
500 IF POINTEUR =0 GOTO 500 ELSE TIMER OFF
510 REM boucle de la pulsation N>S
520 ON TIMER(N(X)) GOSUB 620
530 TIMER ON
540 LOCATE 6,5: PRINT "MINUTES RESTANTES DANS CETTE
PERIODE:";CINT((PERIODE-TIMER)/60)
550 IF POINTEUR = 5 GOTO 550 ELSE TIMER OFF
560 IF PERIODE>TIMER GOTO 460 ELSE NEXT X
565 NEXT V
570 REM ********** procedure d'avertissement de fin de programme
**************
575 PRINT
"****************************************************************"
576 PRINT " Electrophorèse termineé.......!"
577 PRINT " Couper l'alimentation de l'ordinateur"
578 KILL "pulse.dat":KILL "PULSEU.DAT"
580 FOR Z=1 TO 100000!
590 BEEP : NEXT Z
610 END : REM
***************************fin*********************************
620 REM soubroutine IBM  0 volt
630 MOTOR 0 : POINTEUR=0
640 LOCATE 10,15: PRINT " MIGRATION INITIALE A";VS;"VOLTS
POUR";S(X);"SECONDES"
650 RETURN
660 REM soubroutine IMB  5 volts; applique +5 volts sur la pin 3 et 5
670 MOTOR 1 : POINTEUR=5
680 LOCATE 10,15: PRINT " MIGRATION INVERSE  A";VN;"VOLTS
POUR";N(X);"SECONDES"
690 RETURN
```

FIGE program: "Recycler"

This program, in conjunction with the previous program, ensures the continuity of a FIGE gel in the event of a power failure.

```
10 COLOR 10,12,14
30 REM PROGRAMME PULSATEUR-INVERSEUR: initialisation automatique
40 REM***************INITIALISATION DES PARAMETRES DE PULSATION************
50 DIM S(100), N(100)
60 CLS
95 OPEN "a:pulse.dat" FOR INPUT AS #1
96 INPUT #1, VS,VN,RUNTIME,C,S(1),N(1),SINC,NINC,CYCLENO,CYCLETIME,PERIODE
97 CLOSE #1
165 OPEN "a:pulseu.dat" FOR INPUT AS #1
166 INPUT #1,Y,U
167 CLOSE #1
280 CLS
300 FOR X=2 TO C
310 S(X)=S(X-1)+SINC
320 N(X)=N(X-1)+NINC
350 NEXT X
360 BEEP
400 CLS: X=Y : V=U : LOCATE 1,5: PRINT "DUREE TOTALE  =";RUNTIME;"hrs."
410 GOSUB 620
415 FOR V=U TO CYCLENO
420 FOR X=Y TO C
425 OPEN "a:pulseu.dat" FOR OUTPUT AS #1
426 PRINT #1,X,V
427 CLOSE #1
428 LOCATE 4,5 : PRINT "cycle #";V;"of";CYCLENO
430 LOCATE 5,5 : PRINT "PERIODE #";X;"SUR";C
435 LOCATE 23,15: PRINT " Il y a eu panne de courant à la période #"; Y
;",cycle" ; V
440 TIME$="00:00:00"
450 LOCATE 2,5: PRINT "HEURES TOTALES
RESTANTES:";(RUNTIME-(CYCLETIME*(V-1)/3600))
455 LOCATE 3,5: PRINT "MIN. REMAINING IN
CYCLE:";((CYCLETIME-(PERIODE*(X-1)))/60)
460 REM boucle de la pulsation S>N
470 ON TIMER(S(X)) GOSUB 660
480 TIMER ON
490 LOCATE 6,5: PRINT "MINUTES RESTANTES DANS CETTE
PERIODE:";CINT((PERIODE-TIMER)/60)
500 IF POINTEUR =0 GOTO 500 ELSE TIMER OFF
510 REM boucle de la pulsation N>S
520 ON TIMER(N(X)) GOSUB 620
530 TIMER ON
540 LOCATE 6,5: PRINT "MINUTES RESTANTES DANS CETTE
PERIODE:";CINT((PERIODE-TIMER)/60)
550 IF POINTEUR = 5 GOTO 550 ELSE TIMER OFF
560 IF PERIODE>TIMER GOTO 460 ELSE NEXT X
565 NEXT V
570 REM ********** procedure d'avertissement de fin de
programme**************
```