The EMBO Journal vol.11 no.2 pp.725 – 732, 1992

Interstitial telomeres are hotspots for illegitimate recombination with DNA molecules injected into the macronucleus of *Paramecium primaurelia*

Michaël D.Katinka and Florence M.Bourgain

Laboratoire de Génétique Moléculaire, Ecole Normale Supérieure, 46, rue d'Ulm, 75230 Paris cedex, France

Communicated by G.H.Beale

DNA molecules injected into the macronucleus of *Paramecium primaurelia* replicate either as free linear telomerized or chromosome integrated molecules. In the present study we show that when a 1.77 kb *Bam*HI DNA fragment harbouring the *his3* gene of *Saccharomyces cerevisiae* was microinjected into the macronucleus, a fraction of the molecules are integrated into the chromosome via an illegitimate recombination process. The injected molecules were mostly inserted at their extremities at multiple points in the genome by replacing the *Paramecium* sequences. However, insertion sites were not totally at random. Roughly 30% of the molecules were integrated next to or in telomeric repeats. These telomeric repeats were not at the extremities of chromosomes but occupy an internal or interstitial position. We argue that such sites are hotspots for integration as the probability of random insertion near or in an interstitial telomeric site, of which there are 25–60 in a macronucleus is between 5×10^{-4} and 3×10^{-5}.

Key words: illegitimate recombination/interstitial telomeres/ *Paramecium primaurelia*

Introduction

Paramecium, like other ciliates presents the particularity of nuclear dualism. Each species contains at least one small diploid micronucleus, dividing mitotically, weakly transcribed, and as the germinal nucleus, transmitting genetic information through the sexual processes of conjugation and autogamy. Each cell likewise contains a large polyploid (~1000n) macronucleus, dividing amitotically. Because of the high level of transcription in the macronucleus it is usually denoted the somatic or vegetative nucleus. During and following the sexual phases of the cell cycle, the old macronucleus is gradually fragmented and diluted, and a new one develops from the dividing zygotic nucleus. This developmental process is accompanied by a series of complex rearrangements of the genome (for a review on the related holotrichous ciliate *Tetrahymena thermophila* (see Yao, 1989).

When circular or linear DNA molecules from different biological origins (bacteria, yeast, paramecia or polyoma virus) are microinjected directly into the *Paramecium* macronucleus, they are stably replicated as linear telomerized molecules throughout the vegetative cycle (Gilley *et al.*, 1988; Bourgain and Katinka, 1991). We have recently shown (Bourgain and Katinka, 1991) that not all of the injected DNA molecules are maintained as free linear molecules and some are integrated into the macronuclear chromosomes. The proportion of the integrated molecules seems to be independent of their biological origin, the presence or not of pre-existing telomeric repeats at the extremities of the linear molecules injected, or of their size. Insertion frequency of a given injected DNA varied in different injected cells deriving from the same clonal origin. Our preliminary results showed that the various injected molecules were integrated at numerous sites and suggested that insertion was probably mediated through illegitimate recombination (Bourgain and Katinka, 1991).

Illegitimate or non-homologous recombination was defined by Franklin in 1971 as an event of recombination occurring between nucleic sequences that share little or no similarity. However, the term 'little similarity' is somewhat vague and requires a more precise definition. In *Escherichia coli* homologous recombination requires sequences of ~20 bp (e.g. recombination between lambda phage and ColE1 plasmids; Shen and Huang, 1986; Watt *et al.*, 1985). Non-homologous recombination in mammalian somatic cells on the other hand, usually occurs with short direct repeats of up to 7 bp (Nalbantoglu *et al.*, 1986). Thus, similarities of <20 nucleotides can be considered as little or short (Ehrlich, 1989). In integration of retroviral DNA the flanking cellular DNA sequences are arranged as might be expected from a simple insertion event and are readily analysed (Skalka, 1988). However, in studies of integration of foreign DNA into mammalian cells at least two difficulties are encountered. First, there are often repetitive sequences flanking the integration site, and second, integration is usually accompanied by chromosome rearrangements. Thus, there are relatively few random integration sites completely characterized (for a review see Roth and Wilson, 1988). Because integration of foreign DNA into *Paramecium* macronuclear chromosomes seemed random and resembled that of DNA molecules introduced into bacteria or eukaryotic nuclei (for a review see Ehrlich, 1989; Meuth, 1989; Roth and Wilson, 1988), we wanted to gain better insight into the integration phenomenon in *Paramecium* and find out if we could use the *Paramecium* cell as a model system for the study of illegitimate recombination in eukaryotes. To do so we used one of the molecules previously studied (Bourgain and Katinka, 1991): the linear *Bam*HI 1.77 kb DNA fragment, containing the *His3* gene of *Saccharomyces cerevisiae* (BS, Figure 1). This fragment was microinjected into the macronucleus of strain 168 of *P.primaurelia*. We have cloned a number of integrated BS fragments and studied in detail the insertion junctions in four recombination events by sequencing the insertion boundaries followed by analysis of the homologous *Paramecium* sequences prior to integration. Firstly, we established that recombination was mediated by a non-homologous mechanism proceeding via the extremities of the injected BS fragment. Surprisingly, insertion was not completely random but ~30% of the fragments preferentially recombined with interstitial

telomeric repeats or the sequences next to them. Secondly, the *Paramecium* macronuclear sequences left and right of the insertion junction boundaries were sometimes distant by tens of kilobases prior to insertion. The possible mechanism and role of illegitimate recombination in the macronucleus are discussed.

Results

Free and integrated yeast DNA molecules are both found in cells following their injection

We have previously shown that when injected into the macronucleus of *P.primaurelia* DNA molecules from different biological origins, either in linear or circular form, were all maintained as telomerized linear monomers or multimers throughout the vegetative cycle. The efficiency of maintenance varied from one DNA species to another mostly because it was greatly enhanced by the presence of pre-existing terminal telomeric repeats (Bourgain and Katinka, 1991). Furthermore, the presence of such end sequences inhibited concatenation of the injected DNA molecules. Not all of the injected DNA molecules were maintained as pseudochromosomes but part of the material was integrated into the macronuclear chromosomes. The amount of insertion detected varied from one injection experiment to another, and was not dependent on the presence and absence of telomeric repeats at the extremities of the molecules injected.

In the experiment depicted in Figure 1, *P.primaurelia* strain 168 cells of the same clonal origin were co-injected with both the 1.77 kb *Bam*HI fragment (BS) and the supercoiled 19 kb UXi3 plasmid harbouring the sequences encoding the surface protein G of strain 156 of *P.primaurelia*. The latter DNA was included in the experiment to detect successfully injected clones and as an internal control to monitor maintenance efficiencies (Bourgain and Katinka, 1991). Expression of 168G and 156G surface proteins when cells are cultured at 24°C is readily distinguishable *in vivo* by an immobilization test using specific sera raised against one or other of the proteins. Thus successfully injected 168 cells express on their surface the 156 protein encoded by the injected plasmid. Total cellular DNA from such cells was extracted, electrophoresed on agarose gels, transferred to nylon membranes and hybridized to a BS radiolabelled probe. As shown in Figure 1, for two independently injected clones the ratio of free over integrated BS molecules was quite different. A large proportion of free molecules (up to heptamers) with a low amount of integrated material were detected in clone BS1. These proportions were reversed for clone BS2, free molecules, although detectable, being in much lower amount than the integrated material (Figure 1; the digestion with *Bam*HI generates 1.77 kb monomers of most of the free multimers). The co-injected pUXi3 DNA was maintained as free or integrated species in about the same proportions as the yeast DNA, in BS1 the major part of pUXi3 DNA was maintained as free linear molecules whereas in BS2 it was mostly in the chromosome integrated configuration (Bourgain and Katinka, 1991 and not shown). When BS1 and more clearly BS2 DNA were cut with *Eco*RI (there is no *Eco*RI site in the BS fragment), the hybridization pattern appeared as a practically continuous smear extending from ~1.5 kb to >14 kb, demonstrating the multiplicity of sites into which insertion occurred (lane E, Figure 1).

Cloning and analysis of the integrated BS yeast molecules

To estimate the number of integrated copies in the BS2 clone and to understand the organization at the insertion boundaries, *Eco*RI fragments of *Paramecium* macronuclear DNA were cloned. We took advantage of the absence of an *Eco*RI site in the injected BS fragment and cut total genomic DNA of the BS2 clone with *Eco*RI, to construct a scrambled *Eco*RI library in phage λ EMBL4. Since free linear BS molecules contain telomeres at their ends, only the integrated BS molecules can be cloned into the *Eco*RI library. The haploid genome of *P.primaurelia* has been variously estimated to contain between 45 and 280 Mb (Cummings, 1975; MacTavish and Sommerville, 1980; Soldo *et al.*, 1981). To estimate the number of integrated molecules we screened the recombinant lambda library with three different DNA probes hybridizing to unique sequences in the *Paramecium* genome (one probe specifically hybridizing to the 168G sequence and two unique probes from undetermined non-coding sequences; results not shown). In ~60 000 plaques, transferred to nylon membranes and hybridized, two to five positive phages were detected with the unique sequence probes, whereas 64 positive plaques were revealed by the BS probe. Thus, a fair estimation of the number of integrated BS copies would be between 15 000 and 30 000 copies per macronucleus. Ten BS-containing phages were isolated, grown, and their DNA

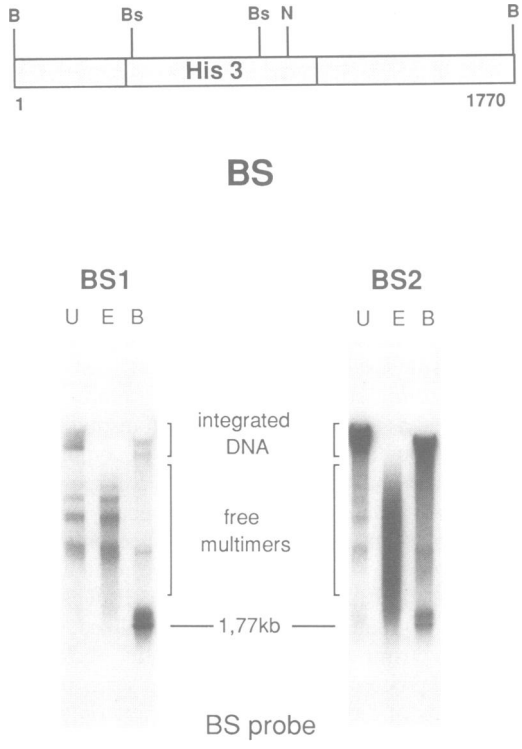

Fig. 1. Free and integrated copies of injected BS fragment. The upper panel depicts the physical map of the 1.77 kb yeast fragment (BS) microinjected into the 168 strain *P.primaurelia* macronucleus. The sequence encoding the *His3* gene is boxed. B, *Bam*HI; Bs, *Bst*BI and N, *Nhe*I. The autoradiograms in the lower panel represent a Southern blot of total DNA extracted from two BS injected clones, electrophoresed on 1% agarose gels and hybridized with the radiolabelled BS probe. The extent of free multimers and integrated material is limited by the brackets. U, Uncut DNA; E, *Eco*RI; B, *Bam*HI. There is no *Eco*RI site in BS.

extracted and cut with *Eco*RI. The autoradiogram in Figure 2 shows that all the lambda phages contained an *Eco*RI fragment of a size >1.7 kb hybridizing with the BS probe.

After subsequent insertion of four different BS harbouring *Eco*RI fragments into the pUC18 vector, physical maps of these fragments were established as shown in Figure 3a. These clones were chosen from among the 10 isolated because of the relatively small size of the *Eco*RI fragment involved so as to faciliate analysis and cloning. Only the sites used for generating specific probes or for Maxam and Gilbert sequencing purposes are shown. The 5′ and 3′ insertion junctions (5′ and 3′ are defined with respect to the orientation of *His3* gene transcription) of these four clones (In1, In7, In8 and In10) were sequenced using two primers that were complementary to the 5′ and 3′ ends of the 1.77 kb injected fragment (see Materials and methods). The sequences at the insertion boundaries are given in Figure 3b. For seven out of eight junctions the insertion proceeded via the *Bam*HI extremities of the injected BS fragment by blunt end insertion, either after complete or partial removal of the *Bam*HI 4 bp cohesive ends and repair of the remaining protruding nucleotides. A deletion of 255 bp in the 5′ region of the integrated yeast fragment excluded the use of the specific 5′ primer for sequencing the In7−5′ junction and the sequence was established by partial chemical degradation following phosphorylation of the *Bst*BI site at nucleotide 421 (Figure 1). In In8 the 5′ *Paramecium* boundary sequences

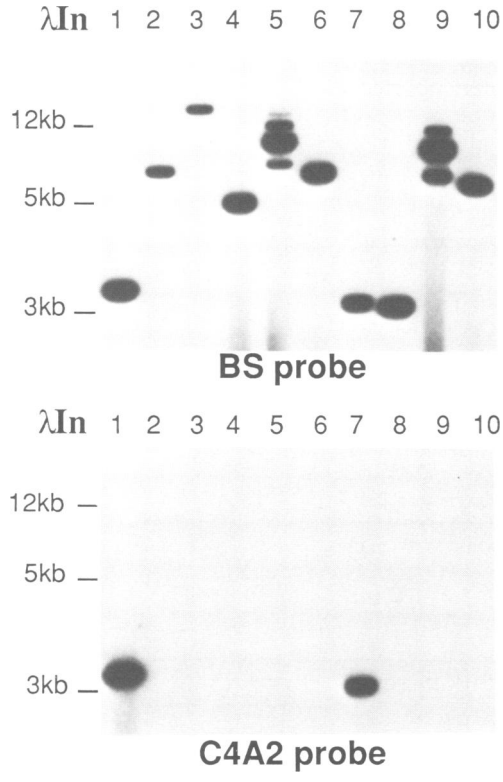

Fig. 2. Southern blots of BS harbouring λ phages. *Eco*RI cut *Paramecium* DNA from the BS2 injected cells was inserted into the EMBL4 λ phage vector and positive plaques were detected with the radiolabelled BS probe. The autoradiogram shows the hybridization pattern of a Southern blot of DNA from 10 positive λ phages, cut with *Eco*RI and electrophoresed on a 0.7% agarose gel with either the BS probe, or the *Tetrahymena* telomeric repeats (C_4A_2) probe. Multiple hybridizing bands in lanes 5 and 9 are due to rearrangements in the phage population.

did not originate from the *P.primaurelia* 168 strain but from recombination between the BS fragment and the co-injected circular cloned 156G gene. The insertion site was located downstream of the gene at nucleotide 14 478 of the 15 917 bp cloned and sequenced 156 DNA (Prat *et al.*, 1986; A.Le Mouël and M.Katinka, unpublished results). We do not know whether recombination between the two co-injected DNAs occurred prior to or following integration. No conserved or repetitive sequence elements other than the ones discussed below were observed. *Paramecium* macronuclear DNA is adenine- and thymine-rich (AT-rich; Gibson and Martin, 1971; MacTavish and Sommerville, 1980; Soldo *et al.*, 1981) and all insertion sites were in AT-rich sequences, without any significant open reading frames. Likewise, no duplication of even short repeats could be detected at the 5′ and 3′ insertion boundaries for any clone.

Insertion frequently occurs in or near internal telomeric repeats

A striking feature revealed by the junction sequences was the presence of telomeric repeats next to two insertion junctions (In1 and In7; see Figure 3b). 28.5 telomeric repeats were found two nucleotides from the In1−5′ junction and 13.5 telomeric repeats at the In7−3′ junction. As described above, using *Eco*RI fragments to construct the lambda recombinant library precludes cloning of the telomerized macronuclear extremities. As shown in Figure 2, only In1 and In7 DNA hybridized to the telomeric probe. To examine if the two clones containing the telomeric repeats were exceptions, the same filters described above, already hybridized with the BS probe (64 positive clones) were rehybridized with the telomeric probes and 118 positive plaques with variable intensities were detected (not shown), intensities that might solely reflect the number of telomeric repeats present in the recombinants. Of these, 20 hybridized with both the BS and telomeric C_4A_2 probes. From this we can conclude first, that there are ~25−60 internal telomeric repeat stretches per genome equivalent (the genome equivalent is defined as the haploid genome), and second, that ~30% of the integrated BS fragments were within, or in the immediate vicinity of internal telomeric repeats. The probability of random insertion at a given site can be calculated roughly by multiplying the number of interstitial telomeric sites per genome equivalent by the mean size in nucleotides of a site divided by the complexity of a haploid genome. The number of sites was found to be between 25 and 60, with an average of ~120−240 bp (20−40 repeats of 6 bp) and the genome complexity according to different authors is of 45−280 Mb (see above). Thus, the probability of insertion at a given site is roughly between 5×10^{-4} and 3×10^{-5}. The experimental value of 0.3 probability for insertion near telomeric repeats is therefore statistically highly significant.

The possibility that telomeric repeats were found next to the BS sequences because the injected BS fragment fused with one chromosome extremity while recombining with internal sequences of the same or another chromosome, is highly improbable and might be expected to be a rare event only. A significant number of the λ *Eco*RI recombinants, however, contained telomeric repeats and of these ~30% also contained the BS sequences (see above). However, to rule out that the telomeric repeats found next to the BS sequence were chromosome ends and not interstitial

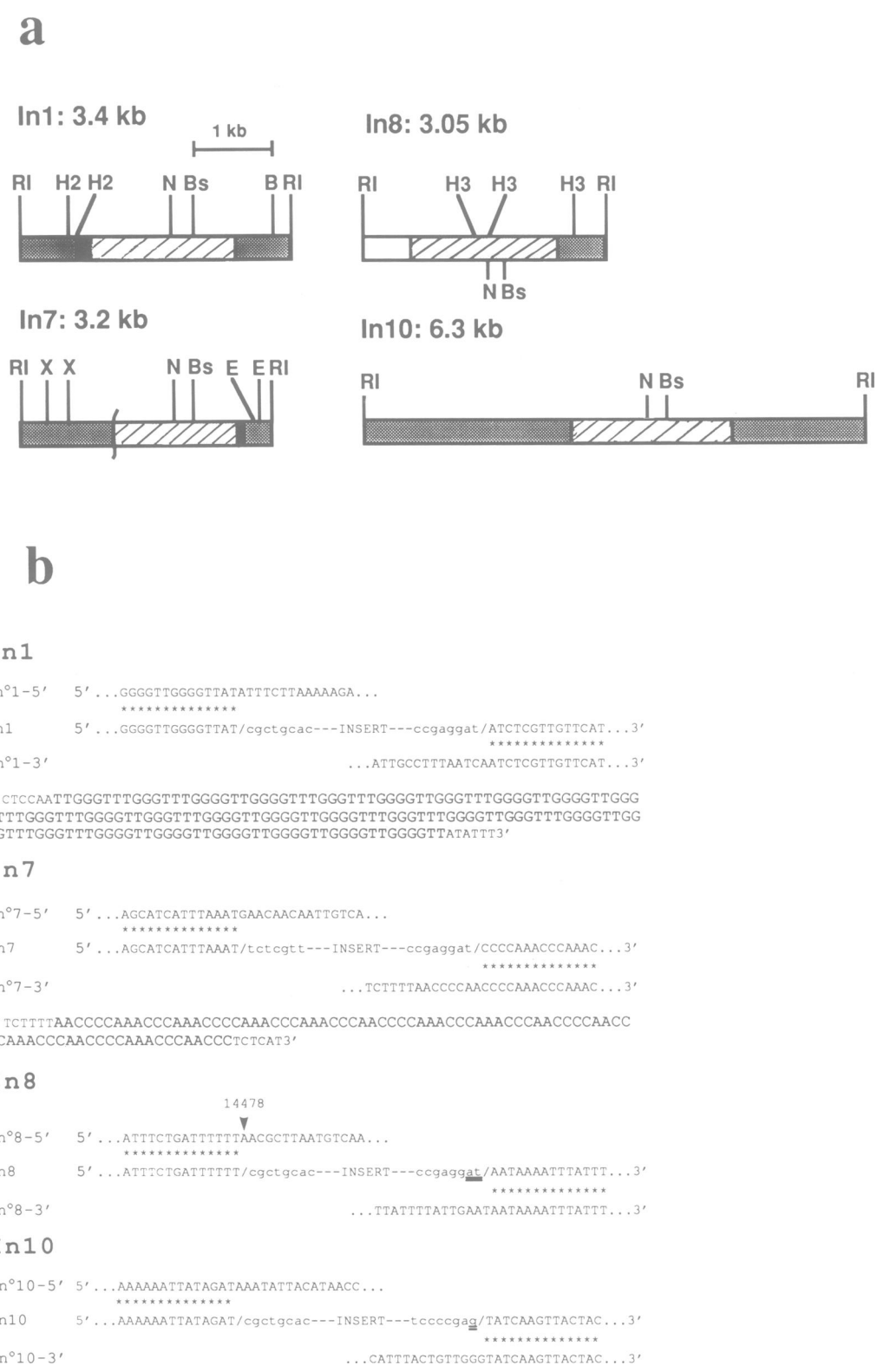

Fig. 3. Analysis of BS insertion boundaries. Physical map of four *Eco*RI fragments harbouring inserted BS sequences (**a**), the light grey boxes represent the BS sequences, the dark grey boxes indicate *Paramecium* sequences and the black boxes show the internal telomeric repeats. The open box in the In8 map represents the co-inserted sequences of the co-injected 156G DNA. The missing 255 bp in the 5′ of BS fragment, in the In7 map are schematized by the wavy vertical line. The sequences of the insertion junctions shown for the four In clones (**b**) with lower case letters representing the yeast BS sequences. The paramecia sequences prior to insertion on the 5′ and 3′ of the insertion junctions (the 5′ and 3′ of an insertion junction is defined in respect to *His3* transcription orientation) are shown above (In° N-5′) and under (In° N-3′) the insertion junctions, respectively. Two (AT) and one (G) nucleotide at the 3′ junctions of respectively In8 and In10, of uncertain origin, are underlined. The whole sequence of the internal telomeric repeats is given below the corresponding In1 and In7 sequences in which only the last repeats are shown. Only the restriction sites used for plasmid orientation, chemical degradation sequencing or for generating specific probes are shown. B, *Bam*HI; Bs, *Bss*HI; E, *Ear*I; H2, *Hinc*II; H3, *Hin*dIII; N, *Nhe*I; RI, *Eco*RI; X, *Xba*I.

telomeres, we performed a number of experiments. Firstly, 168 strain DNA, from uninjected cells, was partially digested by *Sau*3A and the 12−20 kb fragments were inserted into *Bam*HI restriction sites of EMBL3 λ arms. When the encapsidated recombinant phages were screened with a telomeric probe, we could estimate the number of internal telomeric repeats to be comprised between 20 and 40, which is approximately the number of telomeric repeats found in the cloned DNA from the BS-injected 168 cells. Secondly, when a λ EMBL4 library containing DNA, partially digested by *Eco*RI, from the 156 *P.primaurelia* strain (gift from F.Caron), the number of interstitial telomeres found was between 15 and 40. Because strains 156 and 168 were originally isolated in distinct geographical regions (New Haven, Connecticut and Sendai, Japan, respectively), interstitial telomeres appear to be a natural feature of the *P.primaurelia* macronucleus.

Site of BS insertion

Illegitimate recombination might occur by a number of insertion modes, the simplest one being an insertion between two nucleotides. If this were the case, first, the *Eco*RI fragment that contains the recombination site should be, prior to insertion, of a size inferior by exactly 1.77 kb (the size of integrated BS) and second, the nucleotide sequence established at the 5′ and 3′ insertion junctions should be contiguous. To verify this, *Paramecium* sequences prior to insertion were analysed. The macronucleus being highly polyploid and insertion of BS molecules being random, the lambda recombinant library should also contain the *Paramecium* *Eco*RI fragments of the homologous chromosomes without BS insertion. Thus we screened the same lambda library utilized above with the *Paramecium* sequences on the 3′ side of the insertion junctions. For In7, In8 and In10 none of the *Eco*RI fragments isolated from the library by a *Paramecium* sequence probe from the 3′ side of the insertion site (the 220 bp *Ear*I−*Eco*RI, and the 645 bp *Hin*dIII−*Eco*RI fragments for In7 and In8, respectively, Figure 3a; and a 350 bp *Ssp*I fragment for In10, not shown) had the expected size (not shown). Only the *Eco*RI fragment isolated with the In1−3′ probe (390 bp *Bam*HI−*Eco*RI fragment) had a size of 1.65 kb suggestive of the simple insertion mode. When we screened the library with probes generated on the 5′ side of the BS insertion (a 145 bp *Hin*cII fragment for In1 and a 450 bp *Xba*I fragment for In7 fragment shown in Figure 3a; a 850 bp *Ssp*I fragment for In10, not shown), none of the *Eco*RI fragments revealed by the 5′ probes was of the size expected prior to BS insertion (not shown).

Total *Paramecium* DNA was extracted from uninjected 168 cells, digested with *Eco*RI, electrophoresed, transferred to nylon membranes and hybridized with 5′ or 3′ *Paramecium* sequence probes from one given clone. The 5′ or 3′ probes were generated by digesting the four BS harbouring pUC18 clones by *Nhe*I (position 917 on the *His3* fragment; Figure 1). Because of the telomeric repeats in the *Paramecium* sequences 5′ of In1 and 3′ of In7, the 145 bp *Hin*cII and the 250 bp *Eco*RI−*Ear*I fragment, respectively (Figure 3a), were used because they were devoid of telomeric repeats. The fragments detected by the 5′ versus the 3′ probe were not of the same size (Figure 4a). Nevertheless, their size was identical to the *Eco*RI fragments cloned in the lambda library (these are depicted by the asterisk in Figure 4a). The In8−5′ probe did not hybridize because the sequence originated from the co-injected 156G DNA (see above). The complex hybridization pattern with the In10−5′ probe probably corresponds to insertion into highly repeated DNA sequences.

The lambda recombinant clones isolated from the library constructed with *Eco*RI cut DNA from the BS-injected cellular DNA (mentioned above), were used to establish the nucleotide sequences of the *Paramecium* chromosomes prior to integration of the BS fragments (Figure 3b). The 3′ insertion boundary of In8 and In10 presented an uncertainty of two (AT) or one (G) base pairs, respectively, because they might derive from either *Paramecium* or BS sequences. They are depicted in Figure 3b as belonging to the BS sequences solely for convenience. The 5′ and 3′ *Paramecium* sequence probes of In1 and In7 were used to screen the recombinant EMBL3 lambda library constructed with DNA from the uninjected 168 strain. The In1−5′ and the In7−3′ probes both hybridized to phages that likewise hybridized with a telomeric probe. On the contrary, the In1−3′and the In7−5′ probes did not hybridize to the same phage and moreover, did not hybridize to the telomeric probe. Both the hybridization and the sequencing results described above show that the *Paramecium* sequences at the BS insertion site were not contiguous prior to insertion and thus a replacement and deletion mechanism of *Paramecium* sequences in the process of integration should be envisaged.

Paramecium sequences are eliminated in BS integration

The exact number of macronuclear chromosomes is hard to assess because of the multiple levels of heterogeneity of chromosome ends (Baroin *et al.*, 1987; Forney and Blackburn, 1988; Caron and Meyer, 1989), and the different estimations of the complexity of the haploid genome (Gibson and Martin, 1971; Cummings, 1975; MacTavish and Sommerville, 1980). Nevetheless, the average size of the macronuclear chromosomes being ~300 kb (Preer and Preer, 1979; Caron and Meyer, 1989), an estimation of 300−1000 haploid chromosomes per macronucleus would be fairly accurate. As discussed above, integration possibly proceeds via a replacement−deletion mechanism. We wished to evaluate the amount of *Paramecium* sequence replaced in the insertion process. First, agarose imbedded uninjected *Paramecium* 168 strain chromosomes were submitted to pulse field electrophoresis as previously described (Bourgain and Katinka, 1991) with a commutation period of 80 s, transferred to nylon membranes and hybridized with the probes 5′ or 3′ to the BS insertion site described above. The hybridization pattern was complex with the In7 and In10 probes, hybridization being detected with more than one chromosome (not shown). One should be aware that hybridization of a given probe to more than one chromosome frequently reflects different levels of alternative macronuclear chromosome fragmentation (Caron and Meyer, 1989; F.Caron, unpublished results). The 5′ and 3′ *Paramecium* sequence probes of In1 both hybridized to one chromosome of the same apparent size of ~390−450 kb (Figure 4b). As shown in Figure 4c, restriction mapping revealed that the 5′ and the 3′ BS-flanking *Paramecium* sequences were separated by ~50 kb. Thus, insertion can be associated with long distance sequence replacements. We

do not know if the replaced *Paramecium* sequences were maintained as sub-chromosomic fragments or eliminated.

Discussion

This work shows that a when a 1.77 kb *Bam*HI yeast DNA fragment was directly injected into the macronucleus of a *P.primaurelia* cell it can be integrated at a high copy number (15 000–30 000 copies) into the macronuclear chromosomes. Insertion occurs at multiple sites and is generated by an illegitimate recombination process, mostly directly via the extremities of the injected linear DNA molecules. Integration probably proceeded at a very early stage following injection because no telomeric repeats were added to the integrated BS molecules. Interestingly, ~30% of the recombination events took place in or just next to interstitial telomeric repeats. Moreover, in the four recombination events analysed in detail none appeared as a simple insertion into the macronuclear chromosomes, the In1 recombination event probably proceeding via replacement of *Paramecium* sequences and two other events, In7 and In10, involved even more complex recombination mechanisms (not shown). Illegitimate recombination in the four insertion events studied in this work, does not involve short homologies between the ends of the BS fragments and the insertion site. No tandem or inverted repeats, other than telomeric repeats discussed below, were found either prior to, or following recombination.

The exact mechanism of recombination remains to be elucidated but the results indicate that the In1 insertion event probably resulted in deletions of *Paramecium* target DNA. Nevertheless, our results are compatible with general models

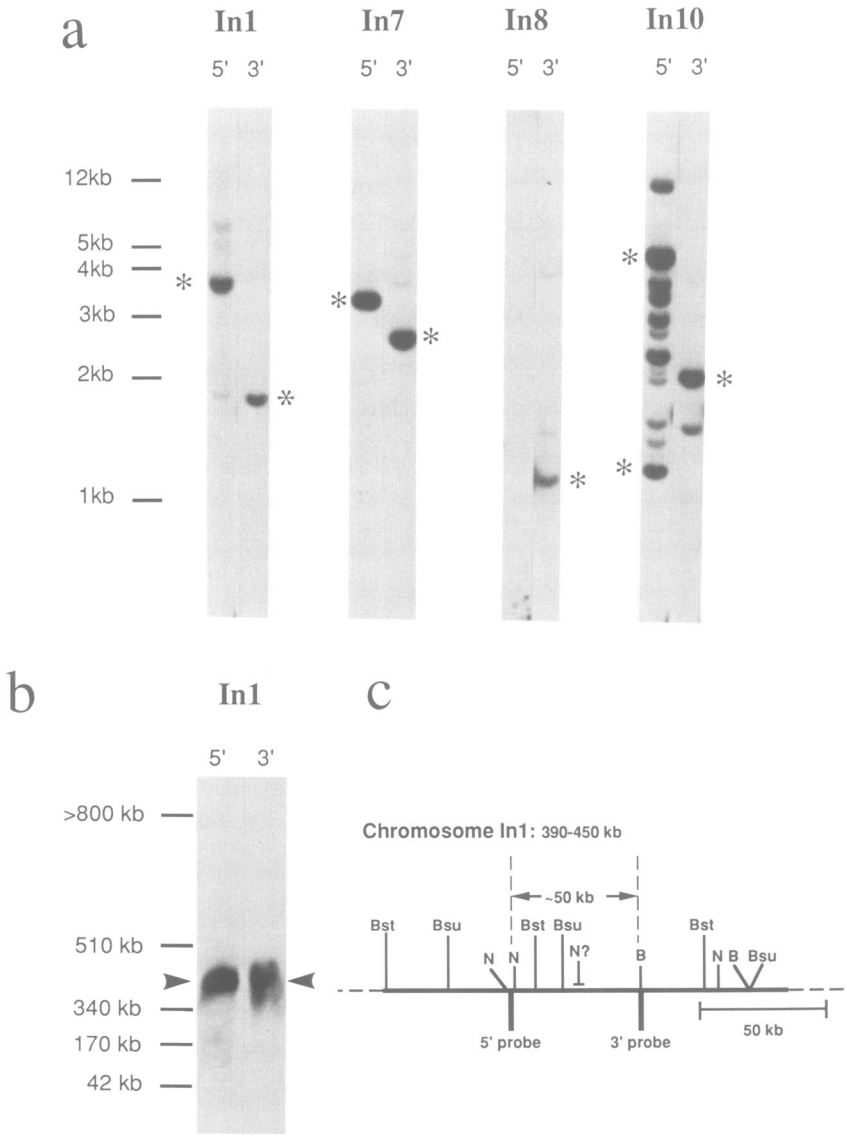

Fig. 4. Southern blot of *Eco*RI cut genomic DNA electrophoresed on a 1% agarose gel (**a**) and macronuclear chromosomes of the 168 *P.primaurelia* strain electrophoresed on a 1.4% agarose CHEF pulse field gel (**b**). The blots in (a) were hybridized with the different 5′ and 3′ flanking *Paramecium* sequence probes derived from In1, In7, In8 and In10 (the 5′ and 3′ orientation of the probes is the same as in Figure 3 and in the text). The *Eco*RI fragments cloned from the uninjected 168 strain DNA using the 5′ and 3′ probes were of the size indicated in (a) by the asterisks. The blot in (b) was hybridized with the 5′ and 3′ *Paramecium* sequence probes from clone In1. (**c**) depicts the physical map of part of the chromosome hybridizing with the In1 probes. The approximate positions of the 5′ and 3′ *Paramecium* sequences flanking the BS insert in clone In1 are shown. B, *Bam*HI; Bst, *Bst*EII; Bsu, *Bsu*36I; N, *Nde*I.

of recombination which propose that breaks or interruptions in the DNA can be converted into recombinogenic sites either by enzymes normally participating in recombination (Cox and Lehman, 1987), or by errors in DNA replication and repair (Lederberg, 1955; Michel and Ehrlich, 1986a,b). A general model for recombination such as the double-strand break repair model proposed by Szostak *et al.* in 1983, might be compatible with our results simply by leaving out the step of partial homologous pairing between the ends of the BS fragment and target DNA. Injected cells contain a large number of either active or potential recombinogenic sites as proved by the large number of BS molecules integrated into the macronuclear DNA. Of the two simple insertion models, either with or without deletions of the target DNA sequence, only the former was observed. However, with only four insertion events analysed the possibility that simple integration does occur cannot be dismissed.

The different number of integration events observed in two cells of the same clonal origin (Figure 1 and data not shown) is difficult to explain at the present time, but this difference might depend either on the phase of the cell cycle in which cells were injected (e.g. the S phase might favour recombination associated with replication), or on the part of the macronucleus into which DNA was injected. Results show however, that a very efficient and highly active enzymatic machinary for recombination is present in the *Paramecium* macronucleus. Recombination might be implicated in housekeeping functions such as chromosome repair, or, DNA content regulation, but another role that can be considered is in control of gene expression. Gene activation following recombination is observed in *Trypanosoma* (Pays and Steinert, 1988). Although the high ploidy ($\sim$1000n) of the macronucleus might be thought of as ruling out recombination as an efficient method for regulating expression, recombination at only a small number of key chromosomes might be sufficient to modulate the expression of certain genes.

The more interesting feature of the present work is the high number ($\sim$30%) of insertion events occurring near or in internal telomeric repeats (interstitial telomeres). Otherwise, insertion was located in AT-rich sequences, which is not unexpected in an organism with a AT content of $\sim$75% (MacTavish and Sommerville, 1980; Soldo *et al.*, 1981). We show here that the number of interstitial telomeres per macronuclear genome equivalent is $\sim$40, and because we estimate the number of chromosomes to be of a few hundreds there is roughly one interstitial telomere per 10 chromosomes. Thus, interstial telomeres are not present in every macronuclear chromosome. There is to our knowledge no report describing interstitial telomeres in *Paramecium* or *Tetrahymena* macronuclear DNA. On the other hand Cherry and Blackburn (1985) and Herrick *et al.* (1985) described internally located telomeric repeats in the micronuclear DNA of the oligohymenophora species *Tetrahymena thermophila* and the polyhymenophora species *Oxytricha falax*, respectively. Their studies showed that the telomeric repeats were located at the ends of transposon-like elements. Because we analysed only short sequences near the interstitial telomeres of In1 and In7 we cannot confirm or infer the linkage of these sequences to transposon-like elements in *Paramecium*, however, no sequence similarity was found between them. Experiments are now in progress in both the 168 and the 156 strains of *P.primaurelia* to assess the presence of transposon-like elements near interstitial telomeric repeats.

The presence of interstitial telomeric sites in mammalian chromosomes is now fairly well documented (for a review see Hastie and Allshire, 1989). Not only were these sequences present, but a number of recombination events, either meiotic or somatic, could be associated with the presence of telomeric repeats. Such recombinational events occurring at random produce 20–30% of human neoplastic states. Sen and Gilbert in 1988, have shown that single-stranded oligonucleotides with sequences characteristic of the guanine-rich strand in the immunoglobulin heavy chain switch region can self-associate to form four-stranded structures of the G4 type (Sundquist and Klug, 1989; Williamson *et al.*, 1989). Moreover, fragile sites prone to recombination are frequently monotonous stretches of either cytosine and thymine or guanine and thymine residues on one strand (Sutherland *et al.*, 1985; Nussbaum *et al.*, 1986). Furthermore, DNA minisatellites that are potential candidates as mammalian recombinational hot spots, characterized by a number of authors (see Hastie and Allshire, 1989 for a review), present a high guanine content ($\sim$60%). One of the hot spots for recombination in the mouse major histocompatibility complex also has a >50% content of guanine on one strand (Steinmetz *et al.*, 1986). Finally, the instability in the progeny of certain restriction fragments following meiosis in *Plasmodium falciparum*, enhanced $\sim$100-fold, could be correlated with the presence in these DNA fragments of internal telomeric repeats (Vernick *et al.*, 1988). The data summarized above implies that repetitive monotonous sequences with a high guanine content are prone to high efficiency of recombination. Thus, a general explanation for the high frequency of recombination at the sites of interstitial telomeres observed by us might be that proposed by Hastie and Allshire (1989). They suggest that in eukaryotic organisms, repeated sequences are preferential sites for errors during DNA replication because of the possibility of slippage during replication inducing in the process recombinogenic hot spots. Another explanation would be that the interstitial telomeres are recognized by the enzymatic machinary of telomere replication and because the interstitial telomeres cannot be elongated strand separation and eventual breakage lead to recombinogenic hot spots. Finally, the data presented show that *Paramecium* might reveal itself to be an interesting model organism for studies on illegitimate recombination in eukaryotes because of: (i) the ease of obtaining a high number of integrated molecules in the macronuclear chromosomes, and (ii) the polyploidy of the macronucleus allowing a relatively straightforward study in the same cell of the sequences at the insertion sites prior to or following recombination.

Materials and methods

Paramecia cell culture and DNA microinjection

Paramecium primaurelia strain 168 was grown as described elsewhere (Bourgain and Katinka, 1991). The injected *Bam*HI 1.77 kb fragment (BS) contained the *Saccharomyces cerevisiae his3* gene that codes for imidazoleglycerophosphate dehydratase (EMBL accession number X03245). The fragment was purified twice on low melting agarose gels from a pYAC3 vector, and co-injected with the UXi3 recombinant plasmid harbouring the *P.primaurelia* 156G surface protein gene (Prat *et al.*, 1986). pUXi3 was used to monitor successfuly injected clones and to compare replication

maintenance efficiencies. The final concentration of total injected DNA was 5 mg/ml (the BS/pUXi3 ratio was ~11). Microinjection was previously described (Bourgain and Katinka, 1991).

DNA extraction, electrophoresis and Southern hybridization

Total *Paramecium* DNA was extracted as described before (Bourgain and Katinka, 1991). Standard or pulse field CHEF electrophoresis as well as Southern hybridization were performed according to usual methods (Sambrook *et al.*, 1989) or as previously described (Bourgain and Katinka, 1991). Pulse field commutation time used was 80 s for *Paramecium* macronuclear chromosome separation (50 up to 800 kb), or 5 s for restriction fragment analysis (3 up to 150 kb). The λ EMBL4 vector used in this work was constructed by Frischauf *et al.* (1983).

Cloning of paramecia genomic DNA

DNA was extracted from cells 20 generations following microinjection and digested with *Eco*RI (no *Eco*RI site is present in the 1.7 kb BS fragment). The *Eco*RI fragments were then ligated to the *Bam*HI and *Eco*RI digested arms of bateriophage λ EMBL4 (Frischauf *et al.*, 1983) and packaged *in vitro* (Amersham kit reference N334). Packaged recombinant phages were plated on *E.coli* C600P2 cells. Positive plaques were detected after transfer to nylon Hybond$^+$ membranes (Amersham) and hybridization with a BS ^{32}P-labelled probe. λ phages with a BS harbouring fragment were grown on the *E.coli* Q358 or 1106 strains, the phage DNA extracted and the fragment that hybridized to the BS probe was then subsequently inserted into pUC18 according to standard techniques (Sambrook *et al.*,1989). Restriction fragments containing *Paramecium* sequences at the insertion boundaries were used as probes to isolate from the same lambda library the *Paramecium* DNA fragments prior to BS insertion. The recombinant EMBL3 (Frischauf *et al.*, 1983) library was constructed with DNA extracted from uninjected 168 cells. DNA was partially digested with *Sau*3A and after electrophoresis on a low melting agarose gel, DNA molecules of sizes between 12 and 20 kb were agarase purified as described above. Otherwise, ligation, encapsidation, infection and screening were performed essentially according to the same protocol used for the construction of the EMBL4 *Eco*RI library.

DNA sequencing

Insertion junctions were sequenced by the dideoxynucleotide termination method (Sanger *et al.*, 1980) using two primers specific for the two ends of the BS sequence:5'-GCTAGGGAACAGGACCGTGC-3' and 5'-CCGTCGTGGTGGTTACGGCAAC-3'. *Paramecium* sequences of the chromosomes prior to insertion were sequenced either by the dideoxynucleotide termination technique using specific oligonucleotide primers (not shown) or by the partial chemical degradation technique (Maxam and Gilbert, 1980).

Acknowledgements

We wish to thank François Caron for stimulating discussions throughout this work; Linda Martin, Eric Meyer and Anne Le Mouël for critically reading the manuscript. This work was supported by grant no. 88/1470 from the Direction des Recherches, Etudes et Techniques, Ministère de la Défense, and by a grant from the Ministère de la Recherche et de l'Education Supérieure (Structure et Fonction des Macromolécules et systèmes Intégrés).

References

Baroin,A., Prat,A. and Caron,F. (1987) *Nucleic Acids Res.*, **15**, 1717–1728.
Bourgain,F. and Katinka,M. (1991) *Nucleic Acids Res.*, **19**, 1541–1547.
Caron,F. and Meyer,E. (1989) *Annu. Rev. Microbiol.*, **43**, 23–42.
Cherry,J.M. and Blackburn,E.H. (1985) *Cell*, **43**, 747–758.
Cox,M.M. and Lehman,I.R. (1987) *Annu. Rev. Biochem.*, **56**, 229–262.
Cummings,D.J. (1975) *Chromosoma*, 53, 191–208.
Ehrlich,S.D. (1989) In Berg,D.E. and Howe,M.H. (eds), *Mobile DNA*. American Society for Microbiology, Washington DC, pp. 799–832.
Forney,J.D. and Blackburn,E.H. (1988) *Mol. Cell. Biol.*, **8**, 251–258.
Franklin,N. (1971) In Hershey,A.D.(ed.), *The Bacteriophage Lambda*. Cold Sring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 175–194.
Frischauf,A.M., Lehrach,H., Poustka,A. and Murray,N. (1983) *J. Mol. Biol.*, **170**, 827–842.
Gibson,I. and Martin,N. (1971) *Chromosoma*, **35**, 374–382.
Gilley,D., Preer Jr,J.R., Aufderheide,K.J. and Polisky,B. (1988) *Mol. Cell. Biol.*, **8**, 4765–4772.
Hastie,N.D. and Allshire,R.C. (1989) *Trends Genet.*, **5**, 326–331.
Herrick,G., Cartinhour,S.W., Dowson,D., Ang,D., Sheets,R., Lee,A. and Williams,K. (1985) *Cell*, **43**, 759–768.
Lederberg,J. (1955) *J. Cell. Comp. Physiol.*, **45**, 75–107.
MacTavish,C. and Sommerville,J. (1980) *Chromosoma*, **78**, 147–164.
Maxam,A. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499–550.
Meuth,M. (1989) In Berg,D.E. and Howe,M.H. (eds), *Mobile DNA*. American Society for Microbiolgy, Washington DC, pp. 833–860.
Michel,B. and Ehrlich,S.D. (1986a) *Proc. Natl. Acad. Sci. USA*, **83**, 3386–3390.
Michel,B. and Ehrlich,S.D. (1986b) *EMBO J.*, **5**, 3691–3696.
Nalbantoglu,J., Hartley,D., Phear,G. and Meuth,M. (1986) *EMBO J.*, **5**, 1199–1204.
Nussbaum,R.L., Airhart,S.D. and Ledbetter,D.H. (1986) *Am. J. Med. Genet.*, **23**, 715–721.
Pays,E. and Steinert,M. (1988) *Annu. Rev. Genet.*, **22**, 107–126.
Prat,A., Katinka,M., Caron,F. and Meyer,E. (1986) *J. Mol. Biol.*, **189**, 47–60.
Preer Jr,J.R and Preer,L.B. (1979) *J. Protozool.*, **26**, 14–18.
Roth,D. and Wilson,J. (1988) In Kucherlapati,R. and Smith,R. (eds), *Genetic Recombination*. American Society for Microbiology, Washington DC, pp. 621–653.
Sambrook,J., Fritch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*. 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Sanger,F., Coulson,A.R., Barrell,B.G., Smith,A.J.H. and Roe,B.A. (1980) *J. Mol. Biol.*, **143**, 161–178.
Sen,D. and Gilbert,W. (1988) *Nature*, **334**, 364–366.
Shen,P. and Huang,H.V (1986) *Genetics*, **112**, 441–457.
Skalka,A.N. (1988) In Kucherlapati,R. and Smith,R. (eds), *Genetic Recombination*. American Society for Microbiology, Washington DC, pp. 701–724.
Soldo,A.T., Brickson,S.A. and Larin,F. (1981) *J. Protozool.*, **28**, 377–383.
Sundquist,W.I. and Klug,A. (1989) *Nature*, **342**, 825–829.
Steinmetz,M., Stephan,D. and Fischer Lindahl,K. (1986) *Cell*, **44**, 895–904.
Sutherland,G.R., Baker,E. and Fratini,A. (1985) *Am. J. Med. Genet.*, **23**, 715–721.
Szostak,J.W., Orr-Weaver,T.L., Roththstein,R.J. and Stahl, F.W. (1983) *Cell*, **33**, 25–35.
Vernick,K.D., Walliker,D. and MacCutchan,T.F. (1988) *Nucleic Acids Res.*, **16**, 6973–6985.
Watt,V.M., Ingles,C.J., Urdea,M.S. and Rutter,W.J. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 4768–4772.
Williamson,J.R., Raghuraman,M.K. and Cech,T.R. (1989) *Cell*, **59**, 871–880.
Yao,M.-C. (1989) In Berg,D.E. and Howe,M.H. (eds), *Mobile DNA*. American Society for Microbiology, Washington DC, pp. 715–734.

Received on July 18, 1991; revised on October 8, 1991