Nucleic Acids Research, Vol. 20, No. 16 4269–4274

Molecular analysis of DNA junctions produced by illegitimate recombination in human cells

Anne Stary and Alain Sarasin
Laboratory of Molecular Genetics, Institut de Recherches Scientifiques sur le Cancer, BP no. 8, 94801 Villejuif, France

Received April 30, 1992; Revised and Accepted June 30, 1992

ABSTRACT

In a human HeLa derived-cell line carrying permanently a single integrated copy of an SV40 shuttle vector, the transient expression of the SV40 T-antigen led to the production of heterogeneous populations of circular DNA molecules which retained both integrated vector and its surrounding cellular sequences. Comparison between the integrated copy and the linear maps of 80 different plasmids rescued in bacteria suggested that the formation of circular DNA was the result of bidirectionnal replication from the SV40 origin of replication followed by a single intramolecular joining leading to the cyclization of the replicated molecules. Sequence analysis of 45 recombinational junctions demonstrated that the cyclization occurred *via* illegitimate recombination process which did not require preferential nucleotide sequence at the joining sites. However, extensive characterization of recombination junctions revealed that the sequences involved in the recombination at each side of the SV40 origin of replication were not randomly distributed, suggesting the presence of regions which were more prone to be involved in the illegitimate recombination process in human cells. Search of common features usually implied in illegitimate recombination in mammalian cells revealed some association of these regions with palindromes, A + T-rich DNA segments, alternating purine/pyrimidine sequences and *Alu* family repeats.

INTRODUCTION

Illegitimate recombination, defined operationally as the cellular process leading to the joining of DNA sequences with little or no homology, often occurs in mammalian cells. This type of recombination process is involved in a variety of rearrangements including deletions, insertions, chromosomal translocations and gene amplification (1–3). These DNA modifications can produce several cellular abnormalities eventually leading to specific genetic diseases or to carcinogenesis (2, 4). Much of the information on the molecular mechanisms of the recombination process has been provided from studies in rodent or monkey cells (1, 2) whereas the illegitimate recombination processing in human cells is still poorly understood. The study of eukaryotic illegitimate recombination is restricted by the difficulty in isolating the parental DNA molecules implied in a given recombinational event. An experimental and reproducible system is therefore necessary to analyze illegitimate recombination in human cells, to characterize the sequences involved in this process and to determine whether some sequences are more prone to illegitimate recombination. Since these rearrangements affect both genomic DNA and exogenous DNA molecules transferred into mammalian cells, plasmid DNAs represent an easier target for the molecular analysis of this type of process.

As previously described in simian virus 40 (SV40) transformed rat cells, illegitimate recombination junctions have been isolated on circular DNA produced by chromosomal excision of viral DNA after cell fusion with monkey cells expressing the SV40 T-antigen (5, 6). In this paper we present an extensive study of illegitimate recombination junctions during viral gene excision from the chromosome of a human cell line, H-G1 cells, a HeLa-derivative cell line harboring a single copy of an SV40 shuttle vector (p205-GTI) covalently integrated in the genome. The transient expression of the SV40 T-antigen led to the production of an heterogeneous population of excised plasmids which retained sequences of both integrated shuttle vector and its surrounding cellular DNA. Precise studies on the excision process carried out in human cells were performed on individual circular DNAs which have conserved sequences required for an efficient recovery in bacteria. This system offers several advantages for studying illegitimate recombination. First, since a unique and well characterized integrated copy of the shuttle vector serves as the progenitor for all the recombinational events, the parental DNA sequences involved in the recombination junctions can be precisely defined. Second, these recombination events can be induced repeatedly and the comparison of the structure of independent recombined plasmids can be systematically carried out. Finally, since the integrated copy carries all sequences required for selection and maintenance in bacteria, the recovery of the recombination products do not require extensive cloning protocols and the analysis of multiple illegitimate recombination junctions could be then performed. This experimental approach therefore enables one to easily determine the nature of the recombination cross-over events in human cells.

In the present report, we characterize the structure of 80 recombined plasmids obtained after transient T-antigen dependent excision of the SV40 shuttle vector integrated into the human H-G1 cell genome. We show that all excised molecules remained colinear to the integrated copy and were circularized by

illegitimate recombination. DNA sequences of 45 illegitimate recombination junctions are presented and the possible nature of sequences involved in this process is analysed.

MATERIALS AND METHODS

Cell cultures

All DNA transfection experiments were performed by the calcium phosphate coprecipitation method (7) and infections were done as previously described (8). The H-G1 human cell line was obtained after transfection of HeLa cells (human epithelioid cervical carcinoma cells) with 10 μg of the intact p205-GTI plasmid (9) as already reported (10). Two days after transfection, cells were plated in selective medium containing 500 μg/ml of G418 (GIBCO Laboratories). Cell cultures were maintained at this drug concentration in Dulbecco's modified Eagle medium supplemented with 10% fetal calf serum, fungizone (2.5 μg/ml) and antibiotics [streptomycin (100 μg/ml) and penicillin (100 units/ml)]. After two months in culture in selective medium, the H-G1 cell line (10) containing the p205-GTI vector as a single integrated copy was isolated (Figure 1A). The induction of the SV40 replication mode was initiated by transfection with 10 μg of the unreplicating pKMT11 plasmid carrying the T-antigen gene (11), per 50-mm dish by calcium phosphate coprecipitation or after infection with SV40 virus (2.5×10^6 PFU/ml).

Plasmid recovery

Low-molecular-weight DNA was purified from cells by a small-scale alkaline lysis method. $10^6 - 10^7$ cells were pelleted in an Eppendorf tube and resuspended in 0.1 ml of 50 mM glucose,

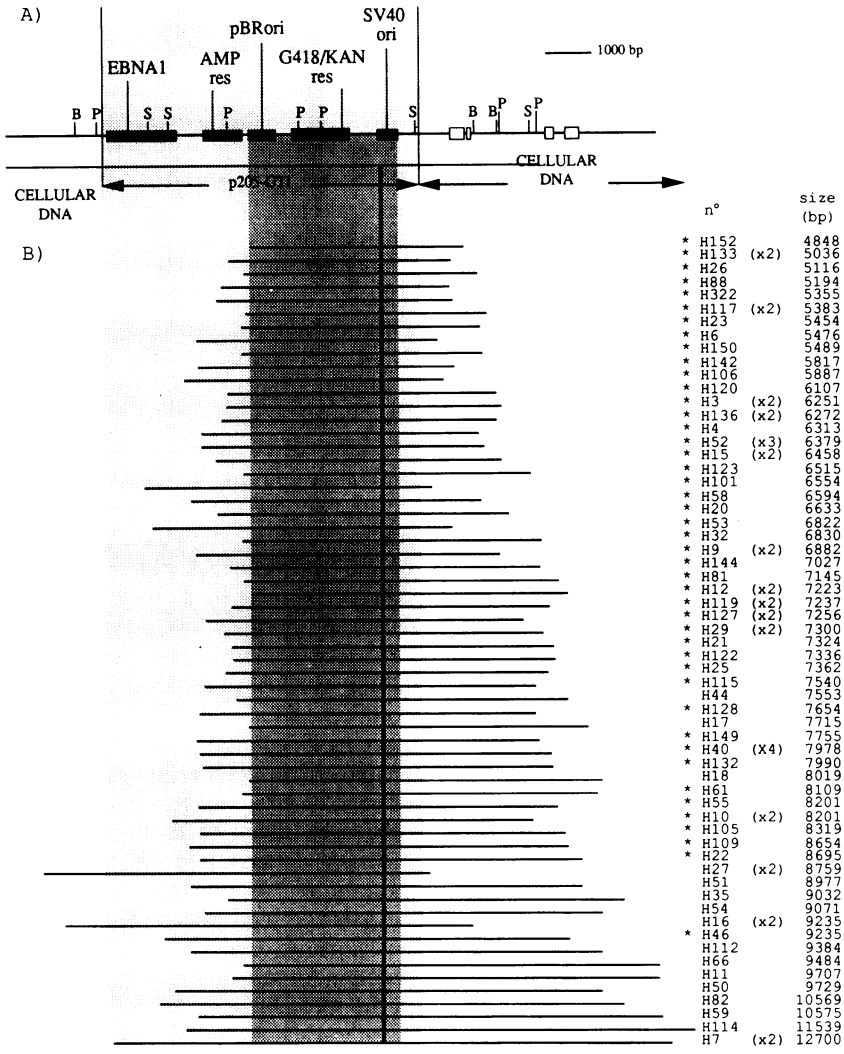

Fig. 1. A) Unique integrated copy of the p205-GTI plasmid in H-G1 cell genome. The light-shaded region indicates the integrated sequence of the p205-GTI plasmid DNA containing SV40 replication origin (SV40 ori), neomycin and kanamycin resistance (G418/KAN res) gene, pBR322 replication origin (pBR ori), ampicillin resistance gene (AMP res) and the EBV nuclear antigen 1 gene (EBNA1). The *Alu* repeats in the adjacent cellular region are shown by open boxes. B, *Bam*HI; P, *Pst*I; S, *Sma*I restriction sites. **B)** Linear maps of 61 different plasmids out of 80 isolated after 11 independent T-antigen boost experiments are aligned with the integrated plasmid DNA. Each circular DNA is represented by its linear map, number and size given in bp. The number in parenthesis indicates the number of identical plasmids isolated in the same T-antigen boost experiment. The light-shaded region indicates the sequence of the p205-GTI vector, the darker region comprises the minimal sequences required to rescue excised plasmids in bacteria and the vertical line indicates the position of the initiation point of the SV40 replication. The extremities of each line correspond to sequences involved in a recombination junction. Stars near plasmid numbers indicate the 45 plasmids in which the recombination junction has been sequenced.

25 mM Tris−HCl (pH 8.0), 10 mM EDTA, followed by addition of 0.2 ml of 0.2 M NaOH, 1% SDS. After 5 min at room temperature, 0.15 ml of 8 M ammonium acetate was added, and the mixture was placed for 15 min at 4°C. After centrifugation for 15 min, the supernatant was collected and combined with 0.6 vol of isopropanol. DNA precipitation was achieved after 5 min at room temperature and DNA was spun down at 4°C for 15 min, washed with cold 70% ethanol, vacuum dried, dissolved in 50 μl of TE (10 mM Tris−HCl, 1 mM EDTA) and finally treated with 10 μg/ml DNAase-free RNAase A for 30 min at 37°C.

Rescue analysis of shuttle vector

Low-molecular-weight DNA isolated from cells by the alkaline lysis method was introduced into the recombination-deficient *E.coli* DH5 α(Bethesda Research Laboratories) by the method of Hanahan (12). Transformants were selected on LB agar medium containing 30 μg/ml of kanamycin (GIBCO Laboratories). Plasmid DNA was purified as described above and analysed using restriction enzymes (New England BioLabs). Digestion products were run on 0.6% agarose gels and analysed by Southern blotting followed by hybridization with specific oligonucleotide probes labelled to a high specific activity using T4 polynucleotide kinase and gamma-^{32}P-ATP (6000 Ci/mM). DNA sequencing was performed by the chain elongation termination method on double-stranded DNA templates using specific oligonucleotides as primers.

RESULTS AND DISCUSSION

Recovery and structure of excised plasmids from the H-G1 cells

Isolation and characterization of the human cell line (H-G1 cells) derived from HeLa cells after transfection with the SV40/EBV hybrid shuttle vector, p205-GTI (9), were carried out as described in Materials and Methods. Genomic DNA of H-G1 cells contained a unique integrated copy of the shuttle vector with no DNA rearrangement except two small deletions at the integration sites which lie in the EBV replication origin (10). Figure 1A presents a map of the chromosomal locus in the human H-G1 cells containing the single integrated copy of the p205-GTI. After integration, the SV40 origin of replication, the G418 eukaryotic

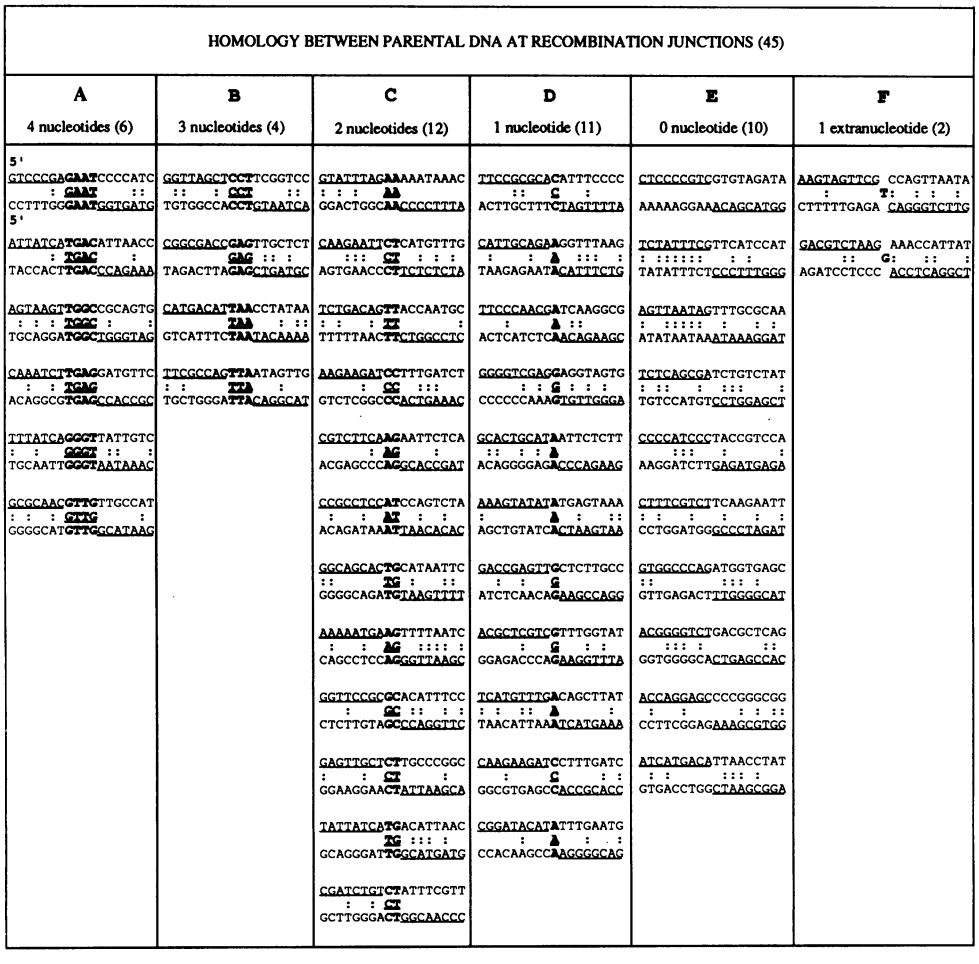

Fig. 2. Nucleotide sequence analysis of 45 recombination junctions. The recombination junctions are classified (columms A through F) according to the decreasing homology between parental DNA at cross-over points. Nucleotides in bold type show nucleotides in common in both parental sequences at the recombination site. Each junction is characterized by the two parental nucleotide sequences spanning from both left (upper sequences) and right (lower sequence) of the SV40 replication origin (Figure 1A). Both sequences are given 5′ (left) to 3′ (right). The underlined sequences correspond to the recombination product and (:) indicates the common nucleotides between parental DNA.

resistance gene and bacterial sequences required to rescue plasmids in bacteria (pBR322 replication origin, ampicillin and kanamycin resistance genes) have been conserved.

Episomal DNA molecules were recovered from three month H-G1 cultures after the induction of the SV40 replication mode carried out either by transient transfection of an unreplicating vector carrying the T-antigen gene or after infection with SV40 virus. This protocol is referred to as a T-antigen boost (9, 13). In the absence of the SV40 T-antigen, episomal molecules were never detected. For each T-antigen boost experiment, low molecular weight DNA was recovered three days after the introduction of the T-antigen into human cells and used to transform bacteria. Plasmids isolated from kanamycin resistant colonies were then digested with various restriction enzymes and analysed on agarose gels. After using Southern blotting following by hybrizations with specific p205-GTI fragments and oligonucleotide probes and then DNA sequencing, the structure of numerous episomes recovered from H-G1 cells submitted to the T-antigen boost has been studied in detail.

From 11 independent T-antigen boost experiments, we isolated 116 excised plasmids which exhibited common features independent of the mode of introduction of the T-antigen gene (transfection or infection) and the time in culture of H-G1 cells. Figure 1B shows the linear physical maps of 80 excised episomes compared to the integrated copy. Several points emerged from these data: (i) among 80 plasmids, 61 episomes exhibited different physical maps and a large heterogeneity in size ranging from 4800 to 12700 bp was observed. No predominant class of molecules was obtained, even in the same experiment, suggesting that a random recombination process arose during the formation of these episomes; (ii) although our experimental protocol to isolate these plasmids required only the presence of the bacterial sequences and the SV40 replication origin (darker shadowed region in Figure 1B), all extrachromosomal molecules also retained some cellular sequences located on the 3′ side of the SV40 origin of replication. Moreover most of the episomes retained plasmid and cellular sequences symmetrically distributed around the SV40 origin. This is consistent with a bidirectional and symmetrical activation of the SV40 replication origin in the integrated copy after T-antigen boost and suggests that replication is necessary for excisional recombination; (iii) all episomal molecules were exactly colinear with the parental integrated copy. This indicates the absence of complex rearrangements during viral excision in human cells. The formation of episomes could therefore be explained by a single intramolecular joining of the replicated molecules. Such recombination junctions leading to the formation of circular excision products were investigated.

Illegitimate recombination junctions

Nucleotide sequences of recombination junctions from 45 excised plasmids (stars in Fig. 1B) have been determined (Figure 2). The comparison between the two nucleotide sequences involved in each recombination junction showed the absence of significant homology, indicating that complete nucleotide homology is not required for such recombination. Illegitimate recombination was therefore the process which occurred to circularize the sequences replicated from the SV40 replication origin in the integrated copy. Alignment between parental DNA molecules at the crossover points revealed that 22 joinings occurred between short direct repeat sequences of 2 to 4 bp, leaving one copy of the sequence in the recombined molecules (Figure 2, columns A−C). 11 recombination junctions exhibited only one nucleotide in common between parental DNA and recombined molecules (Figure 2, columm D), while 10 joinings were formed without any common nucleotide (Figure 2, columm E) and two recombination points were associated with the presence of one extra nucleotide (Figure 2, columm F). As was already found for illegitimate recombination junctions in mammalian cells (1, 2), the homology at recombination junction is less than five nucleotides in our system. In addition, the presence of an extra nucleotide in 4% of our recombination junctions is consistent with previous data where comparison of genetic rearrangements in non-immune mammalian cells showed that 10% of illegitimate recombination junctions contain one or several extra nucleotides of unknown origin (14). In some cases, short stretches of homology (3−6 bp) within each parental sequence were noted near the crossover points, especially at recombination junctions where no common nucleotide was found at the breakpoint (Figure 2, columm E). It is possible that, during circularization of excised molecules, the joining of replicated molecules by illegitimate recombination was facilitated by the presence of short direct repeat sequences (2−6 bp) for which one copy was recovered near or directly at the crossover point in the recombination product.

From our study of 90 nucleotide parental sequences involved in the 45 recombination junctions, we observed an extreme diversity of sequences at these crossover points and no consensus sequence for the recombination site was found. Obligatory sites for recombination were not obvious on either the viral or host sequence. The illegitimate recombination process therefore appears to be sequence-independent. The nucleotide sequence analysis also revealed that no further alterations, such as deletions or base substitutions, were detected near any of the recombination joinings. Taken together, our results show that in human cells, the activation of a viral replication origin integrated into a genomic sequence leads to the formation of heterogeneous circular DNA molecules. This process is mediated by a simple joining mechanism which does not require preferential sequences or significant homology between parental DNAs.

Distribution of recombination junction points

In order to determine whether the illegitimate recombination sites were randomly located at each side of the SV40 origin of replication or whether certain sequences were more prone to such recombination, the distribution of target recombination sequences was analysed. On the left side (plasmid DNA) and the right side (cellular DNA) of the SV40 origin of replication in the parental copy (Figure 3A), we located the position of each parental sequence involved in the illegitimate recombination junctions (Figure 3B). The distribution we observed was not random. Indeed, as indicated by shaded bands in Figure 3, recombination points are concentrated at some regions (recombination hot regions) in both plasmid and cellular DNA. This suggests that certain regions were more prone to be involved in the illegitimate recombination process. Search of sequence homology between these recombination hot regions did not reveal any similarity at the nucleotide level. Moreover, recombination sites located in the same hot region on the left side were never joined with recombination sites located in the same hot region on the right side. This suggests that the capacity of one region to recombine is due to a particular property of this region rather than to the mechanism of the recombination process *per se*.

Because the SV40 T-antigen was involved in the excision of integrated SV40 DNA from rat chromosomes (15) and since defined protein-protein interactions between T-antigen molecules

have been demonstrated *in vitro* to physically bind separate DNA sites by the formation of specific DNA loops (16), a possible role of the T-antigen in the recombination processes has been suggested (16). The positions of the recognition sites for T-antigen (17) on plasmid and cellular sequences have been compared to the distribution of the recombination sites. As shown in Figure 3C, some recombination sites and certain hot recombination regions, essentially in the plasmid part, are located near a recognition site for T-antigen but this is not a general feature. Furthermore, no excised plasmid due to recombination involving T-antigen sites in both parental DNAs has been recovered. However, the intervention of the T-antigen in the recombination process cannot be excluded since any DNA region can be attached *in vitro* to another DNA site via a T-antigen bridge (16). Moreover, it has been recently shown (18) that the T-antigen double-hexamer complex binds *in vitro* simultaneously to both forks of unwinding bubbles that are formed during bidirectional DNA unwinding initiated from the SV40 origin. It is therefore possible that *in vivo* promotion of recombination by the T-antigen occurs during the replication process in an independent-sequence manner.

A search for the presence of features commonly identified in illegitimate recombination in non-immune mammalian cells, such as palindromic repeats, members of interspersed repetitive DNA families (2), AT-rich elements (19, 20), alternating purine/pyrimidine runs, polypurine-polypyrimidine tracts (21) and topoisomerase I consensus cleavage sites (21, 22) was carried out. Palindromic repeats (Figure 3D), alternating purine/pyrimidine runs (Figure 3E) and polypurine−polypyrimidine tracts (Figure 3F) were found in the vicinity of several recombination hot regions which were not however always associated with these motifs. Three recombination hot regions in the cellular DNA were located just upstream the *Alu* family of repetitive sequence (open boxes, Figure 3A). Topoisomerase I consensus sequences as determined by Champoux and Bullock (22) were found both on the plasmid and cellular sequences, respectively 176 and 280 motifs on the piece of DNA spanning the two extreme recombination sites (Figure 3G), but no correlation was observed between high concentrations of these motifs and the recombination hot regions. Interestingly, two recombination hot regions (Figure 3B, stars) were located within the two pBR322 regions with high A+T content (Figure 3A, underlined regions a and b) (23). In the pBR322 plasmid, these two regions (Figure 3A, a and b) exhibited a specificity for the insertion of the transposable element IS1 in bacteria (23) which is believed to occur by illegitimate recombination. These two regions were shown to be hypersensitive to a single-strand-specific nuclease at 37°C due to stable DNA unwinding (24). This suggests that these target sites for recombination in bacteria also correspond to recombination hot regions in human cells possibly in relation to a specific property of the DNA sequence as nuclease sensitive site.

CONCLUSIONS

We have analysed here for the first time in human cells a large collection of illegitimate recombination junctions obtained from the same sequence. Most of the recombinational events occurred between short (2−6 bp) direct repeats at or near the cross-over points. Similar short direct repeats were also found in numerous illegitimate recombination junctions in mammalian cells (1, 2)

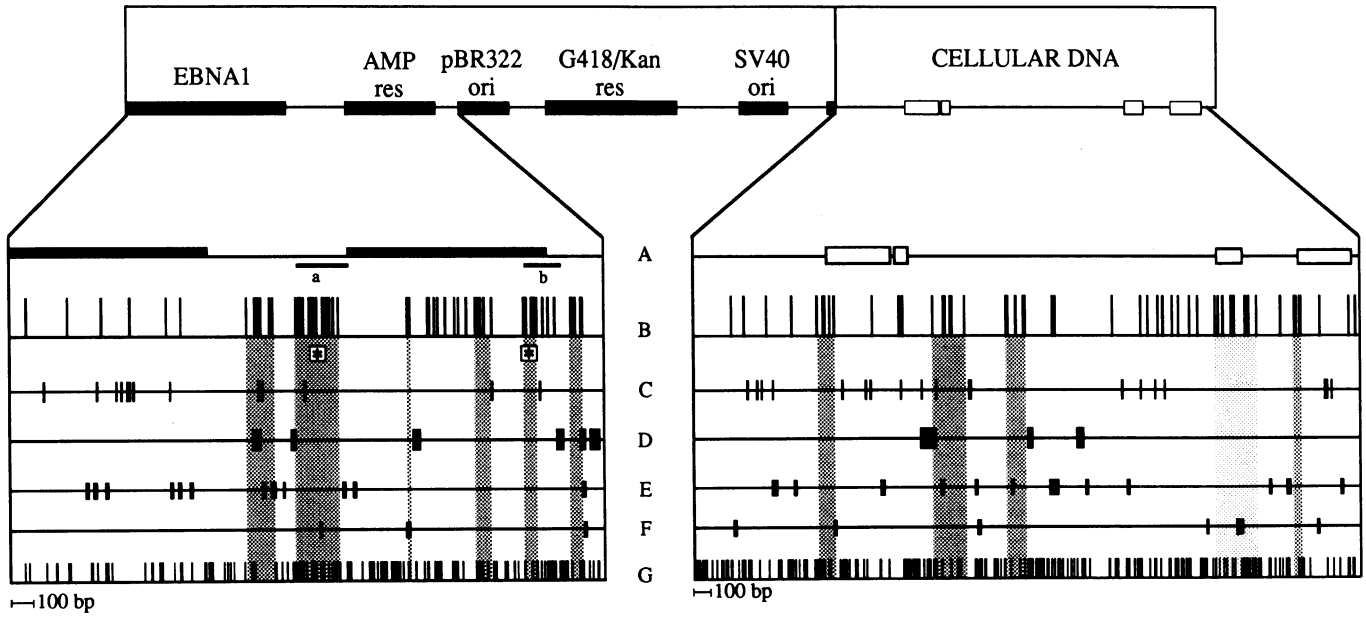

Fig. 3. Distribution of the parental sequences involved in the recombination junctions at each side of the SV40 replication origin. The map of integrated p205-GTI (see legend in Figure 1) is given on the top and its regions implicated in recombination are enlarged in (A). The underlined regions a and b in (A) correspond to the two A+T rich segments in pBR322 plasmid (23). Positions of all recombination junctions (B) have been located according to the maps in (A). The stars in (B) indicate the two recombination hot regions in the the two A+T rich segments in pBR322 plasmid. Positions along the two sequences of several features are also indicated: (C) the SV40 T antigen recognition sites, 5′-GAGGC-3′ (17); (D) the palindromic sequences; (E) the homopyrimidine or homopurine tracts (at least 10 nucleotides); (F) the alternating purine/pyrimidine runs (at least 8 nucleotides); (G) the topoisomerase I cleavage consensus sites determined by Champoux and Bullock (22): 5′-(A or T)-(G or C)-(A or T)-T-3′. The shaded bands depict the regions exhibiting an abnormal concentration of sequences involved in the recombination process.

as exemplified in chromosomal rearrangements produced by deletions in a hamster cellular gene (25). Chromosomal deletions in rodent cells and a circularization process during viral excision in human cells could therefore imply common features of the recombination pathway. We show that there exist DNA regions more prone to be involved in the illegitimate recombination process although no preferential nucleotide sequence at the joining points was found. Most of these regions were associated with features commonly identified with illegitimate recombination in other systems but we did not find a common characteristic for all of them. This is in agreement with all data computed from illegitimate recombination junctions obtained in mammalian cells from very different sources in which no specific DNA structures and sequences were systematically involved (2). Since the recombination junctions recovered in episome structures were obtained immediately after induction of the replication process due to the expression of the T-antigen, this protein could play a role in the recombination process. Indeed, the structures of our circular DNA may be consistent with the model (16) in which the T-antigen favours the association of two replication forks within the same replicating molecules followed by resolution of these DNA structures by suitable enzymes to result in DNA excision. The nature of the illegitimate recombination process we found in particular DNA regions, could also correspond to a simple joining of free ends produced in replicated intermediates. The generation at fragile sequences of breaks due to replication pause sites (26) or nicks in incomplete nascent strands (27) could facilitate these recombination events.

ACKNOWLEDGEMENTS

We are grateful to Drs F.Bourre, L.Daya-Grosjean, M.Mezzina and C.Sarasin for critical reading of the manuscript. This work was supported by grants from the Association pour la Recherche sur le Cancer (Villejuif, France), the Fondation pour la Recherche Médicale (Paris, France) and the Commission of the European Communities (contract B17-0034, Brussels, Belgium). A.Stary held a fellowship from the Association pour la Recherche sur le Cancer (Villejuif, France).

REFERENCES

1. Roth,D. and Wilson,J. (1988) In Kucherlapati,R. and Smith,G.R. (eds), *Genetic recombination*. Am. Soc. Microbiol., Washington, DC, 621-653.
2. Meuth,M. (1989) In Berg,D.E. and Howe,H.H. (eds), *Mobile DNA*. Am. Soc. Microbiol., Washington, DC, 833-860.
3. Stark,G.R., Debatisse,M., Giulotto,E. and Wahl,G.M. (1989) *Cell* **57**, 901-908.
4. Bishop,J.M. (1991) *Cell* **64**, 235-248.
5. Botchan,M., Stringer,J., Mitchison,T. and Sambrook,J. (1980) *Cell* **20**, 143-152.
6. Bullock,P., Forrester,W. and Botchan,M. (1984) *J. Mol. Biol.* **174**, 55-84.
7. Wigler,M., Pellicer,A., Silverstein,S. and Axel,R. (1978) *Cell* **14**, 725-731.
8. Menck,C.F.M., Sarasin,A. and James,M.R. (1987) *Gene* **53**, 21-29.
9. Stary,A., James,M.R. and Sarasin,A. (1989) *J. Virol.* **63**, 3837-3843.
10. Stary,A. and Sarasin,A. (1992) *J. Gen. Virol.* **73**, in press.
11. Gerard,R.D. and Gluzman,Y. (1985) *Mol. Cell. Biol.* **5**, 3231-3240.
12. Hanahan,D. (1983) *J. Mol. Biol.* **166**, 557-580.
13. Heinzel,S.S., Krysan,P.J., Calos,M.P. and DuBridge,R.B. (1988) *J. Virol.* **62**, 3738-3746.
14. Roth,D.B., Chang,X.-B. and Wilson,J.H. (1989) *Mol. Cell. Biol.* **9**, 3049-3057.
15. Miller,J., Bullock,P. and Botchan,M. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 7534-7538.
16. Schiedner,G., Wessel,R., Scheffner,M. and Stahl,H. (1990) *EMBO J.* **9**, 2937-2943.
17. Borowiec,J.A., Dean,F.B., Bullock,P.A. and Hurwitz,J. (1990) *Cell* **60**, 181-184.
18. Wessel,R., Schweizer,J. and Stahl,H. (1992) *J. Virol.* **66**, 804-815.
19. Woodworth-Gutai,M. and Nathans,D. (1978) *J. Mol. Biol.* **126**, 275-288.
20. Hyrien,O., Debatisse,M., Buttin,G. and Robert de Saint Vincent,B. (1987) *EMBO J.* **6**, 2401-2408.
21. Konopka,A.K. (1988) *Nucleic Acids Res.* **16**, 1739-1758.
22. Champoux,J.J. and Bullock,P.A. (1988) In Kucherlapati,R. and Smith,G.R. (eds), *Genetic recombination*. Am. Soc. Microbiol., Washington, DC, 655-666.
23. Zerbid,D., Gamas,P., Chandler,M., Prentki,P., Bass,S. and Galas,D. (1985) *J. Mol. Biol.* **185**, 517-524.
24. Kowalski,D., Natale,D.A. and Eddy,M.J. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 9464-9468.
25. Nalbantoglu,J., Hartley,D., Phear,G., Tear,G. and Meuth,M. (1986) *EMBO J.* **5**, 1199-1204.
26. Schimke,R.T. (1988) *J. Biol. Chem.* **263**, 5989-5992.
27. Wahl,G.M. (1989) *Cancer Res.* **49**, 1333-1340.