Breakage—reunion and copy choice mechanisms of recombination between short homologous sequences

Dominique Brunier, Ben P.H. Peeters, Sierd Bron and S. Dusko Ehrlich

Laboratoire de Génétique Microbiennne, Institut de Biotechnologie, INRA—Domaine de Vilvert, 78350 Jouy en Josas, France and
Department of Genetics, Center of Biological Sciences, Kerklaan 30, 9751 NN Haren (Gn), The Netherlands

Communicated by S. D. Ehrlich

To study recombination between short homologous sequences in *Escherichia coli* we constructed plasmids composed of the pBR322 replicon, M13 replication origin and a recombination unit inserted within and inactivating a gene encoding chloramphenicol resistance. The unit was composed of short direct repeats (9, 18 or 27 bp) which flanked inverted repeats (0, 8 or 308 bp) and a gene encoding kanamycin resistance. Recombination between direct repeats restored a functional chloramphenicol resistance gene, and could be detected by a simple phenotype test. The plasmids replicated in a double-stranded form, using the pBR322 replicon, and generated single-stranded DNA when the M13 replication origin was activated. The frequency of chloramphenicol-resistant cells was low (10⁻⁸–10⁻⁴) when no single-stranded DNA was synthesized but increased greatly (to 100%) after induction of single-stranded DNA synthesis. Recombination between 9 bp direct repeats entailed no transfer of DNA from parental to recombinant plasmids, whereas recombination between 18 or 27 bp repeats entailed massive transfer. The presence or length of inverted repeats did not alter the pattern of DNA transfer. From these results we propose that direct repeats of 9 bp recombine by a copy choice process, while those ≥ 18 bp can recombine by a breakage—reunion process. Genome rearrangements detected in many organisms often occur by recombination between sequences < 18 bp, which suggests that they may result from copy choice recombination.

Key words: recombination/Escherichia coli/chloramphenicol resistance/breakage—reunion/copy choice process

**Introduction**

Genome rearrangements play an important role in evolution (Ohno, 1970; Anderson and Roth, 1977), medicine (Yunis, 1983; Croce, 1987; Koenig *et al.*, 1987; Yen *et al.*, 1987) and biotechnology (Ehrlich *et al.*, 1986). They are probably ubiquitous since they have been found in all organisms in which they were sought. However, the molecular events that underlie genome rearrangements are not well understood. Several studies indicate that rearrangements often result from recombination between short homologous sequences, both in prokaryotes (Farabaugh *et al.*, 1978; Albertini *et al.*, 1982; Jones *et al.*, 1982; Lopez *et al.*, 1984; reviewed by Ehrlich, 1989), and eukaryotes, including man (Efstratiadis *et al.*, 1980; de Zamaroczy *et al.*, 1983; Hogan and Faust, 1984; Nalbantoglu *et al.*, 1986; Ribas-Aparicio *et al.*, 1987; reviewed by Meuth, 1989). It was suggested that such

![Fig. 1. Models of recombination between directly repeated sequences carried on ssDNA plasmids. Continuous and interrupted lines represent pre-existing and newly synthesized DNA strands respectively, direction of replication is indicated by arrows, repeats are indicated by letters, complementary sequences by primed letters. (A) Copy choice recombination. (1) Rolling circle replication of the plasmid generates circular ssDNA molecule; (2) DNA synthesis is initiated on the ss template and proceeds into one of the repeats; (3) the repeat that has not been replicated anneals with the newly synthesized complementary sequence; (4) a recombination intermediate is generated by continuation of DNA synthesis; (5a) a recombinant molecule is generated by copying the short strand of the recombination intermediate. This recombinant is composed of two newly synthesized DNA strands; (5b) a recombinant molecule is generated by cleaving away the ss loop of the recombination intermediate. This recombinant is composed of one newly synthesized and one parental DNA strand. (B) Break—join recombination. (1) Rolling circle replication initiated at the replication origin (the arrow) progresses and displaces a DNA strand; (2) the repeat carried on the displaced strand anneals with the non-replicated complementary sequence; (3) a circular recombinant DNA molecule, annealed to the parental plasmid, is generated by a break—join step (such a step was postulated by Meselson and Radding, 1975); (4) the ss recombinant molecule is separated from the parental plasmid by the progression of DNA replication; (5) the ss recombinant is converted into a ds form by complementary strand DNA synthesis. This recombinant is composed of one newly synthesized and one parental DNA strand.](image-url)
recombination may be due to DNA replication errors, similar to those invoked to explain frame-shift mutations (Streisinger et al., 1966). The key step of this process, conceptually related to copy choice recombination (Lederberg, 1955) is slippage of the DNA replication machinery on a single-stranded (ss) template from one repeat to another (Efstratiadis et al., 1980; Egner and Berg, 1981; Foster et al., 1981; Glickman and Ripley, 1984; Schaaper et al., 1986) and Bacillus subtilis (Brunier et al., 1988). We recently showed that in Escherichia coli precise excision of a Tn10-related transposon occurs by such slippage (Brunier et al., 1988). The slippage takes place between 9 bp direct repeats which flank the 78 bp inverted repeats comprising the transposon ends. Other studies indicated that the proximity of inverted repeats stimulates recombination between direct repeats in E.coli (Egner and Berg, 1981; Foster et al., 1981; Glickman and Ripley, 1984; Schaaper et al., 1986) and Bacillus subtilis (Peeters et al., 1988), suggesting that inverted repeats may be crucial for the slippage and that only a limited class of genome rearrangements (such as transposon excision) may proceed by copy choice recombination. To test this possibility we used plasmids carrying either direct repeats only or direct repeats flanking inverted repeats. Direct repeats of 9 bp recombined by a copy choice process, irrespective of the presence or the length of inverted repeats (8 and 308 bp were tested). In contrast, direct repeats of 18 or 27 bp recombined predominantly by a breakage—reunion process. This shows that the mechanism of recombination depends on the length of direct repeats rather than on the proximity of inverted repeats. Since genome rearrangements often result from recombination between direct repeats much shorter than 18 bp, copy choice recombination may well be a general mechanism for the generation of such rearrangements.

Results

Experimental approach

Copy choice recombination can take place without any transfer of DNA from the parental to the progeny molecules, whereas break—join recombination cannot (Stahl, 1986).

This is illustrated in Figure 1, using models applicable to plasmids that replicate in a rolling circle mode (cf. Gruss and Ehrlich, 1989, for review). A similar copy choice model was presented previously (Ehrlich et al., 1986); the break—join model is related to that proposed by Meselson and Radding (1975). In order to distinguish between the two types of models, we measured the transfer of DNA during recombination between short, directly repeated sequences carried on E.coli plasmids, using a recently developed methodology (Brunier et al., 1988, see below). For that purpose a family of plasmids was constructed (Table I). Each plasmid consisted of the following three elements (Figure 2): (1) plasmid pHV60, a hybrid between pBR322 and the chloramphenicol resistance (CmR) gene of plasmid pC194 (Michel and Ehrlich, 1984); (2) the replication origin of phage M13; (3) a recombination unit inserted between the Cm gene. The recombination units, unique for each plasmid, were all composed of short direct repeats flanking a kanamycin resistance (KmR) gene. The length of the direct repeats was 9, 18 or 27 bp; they corresponded to a part of the CmR gene sequence (Figure 2). In some units direct repeats flanked inverted repeats of either 8 or 308 bp, while in others no inverted repeats were present. The units with inverted repeats contained the KmR gene of plasmid pHJ1 on a 1.4 kb segment; those without inverted repeats contained the KmR gene of Tn5 on a 1.2 kb segment.

Recombination between short direct repeats present within a unit restored the activity of the CmR gene and excised the KmR gene from the plasmid, which allowed detection by a simple phenotypic test. Cells of the recombinant phenotype (several hundred CmR clones were analyzed) invariably harbored plasmids identical in size and restriction pattern to a recombinant plasmid standard.

The plasmids were maintained in E.coli by pBR322 replication functions, which generate no detectable ssDNA (te Riele et al., 1986a,b). Synthesis of plasmid ssDNA was induced by activating the M13 replication origin. The thermosensitive replication protein of the phage fd (a close relative of M13) was routinely used for that purpose; plasmid
Km

Fig. 2. Structure of plasmids used to study recombination mechanisms. (Top) The vector plasmid (circle) carrying the recombination unit (insert). The vector sequences derive from pBR322 (thin line), pC194 (dashed line) and M13 (hatched line); Km, Ap and Tc refer to genes encoding chloramphenicol, ampicillin and tetracycline resistance, oriM to the M13 replication origin. Direct and inverted repeats are shown as triangles and arrows respectively. The numbers refer to the length of inverted repeats, Km_p and Km_T to the kanamycin resistance gene of pH1 or Tn5, carried on 1.4 and 1.2 kb segments respectively. (Bottom) Part of the Cm^R gene. Sequences repeated in the recombination unit are indicated by arrows, the numbers indicate the length of the duplication.

Table II. Proportion of recombinant plasmids in overnight cultures of E.coli cells

<table>
<thead>
<tr>
<th>Length of inverted repeats (bp)</th>
<th>Length of direct repeats (bp)</th>
<th>9</th>
<th>18</th>
<th>27</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>1.1 x 10^{-5}</td>
<td>1.4 x 10^{-5}</td>
<td>1 x 10^{-4}</td>
<td></td>
</tr>
<tr>
<td>308</td>
<td>1.5 x 10^{-5}</td>
<td>1.4 x 10^{-5}</td>
<td>9 x 10^{-5}</td>
<td></td>
</tr>
</tbody>
</table>

ssDNA synthesis was repressed at 42°C, and was induced at 30°C. The fd gene II, encoding the replication protein, was present in the plasmid-carrying cells on a lambda-derived prophage (Geider et al., 1985). Alternatively, in some experiments, ssDNA synthesis was induced by infecting the plasmid-carrying cells with phage M13.

Fig. 3. Stimulation of recombination by ssDNA synthesis. HVC1801 cells, carrying the fd gene II and harboring various plasmids, were grown for specified times at 42 or 30°C (open and closed symbols respectively). The proportion of recombinants was determined by plating appropriately diluted cultures on plates supplemented with Ap (100 µg/ml) or Cm (8 µg/ml) and incubating the plates at 42°C for 12 or 36 h respectively (the time necessary for full colony development on the two types of plates was different). IR and DR refer to inverted and direct repeats respectively, the numbers indicate the repeat length. Direct repeats of 9 and 18 bp present in units carrying inverted repeats recombined with indistinguishable kinetics; for clarity only the former is shown.

Synthesis of ssDNA stimulates recombination between short direct repeats

Recombination within the different recombination units was infrequent in the absence of ssDNA synthesis. This was reflected in the proportion of cells having the recombinant phenotype (Cm^R) in overnight cultures of cells harboring the plasmids described above (Table II). The lowest values, 10^{-8}—10^{-7}, were observed for recombination units lacking inverted repeats; 100—1000 times higher values were observed for units containing either short or long inverted repeats. A similar stimulatory effect of inverted repeats was reported previously in several studies in E.coli (Foster et al., 1981; Egner and Berg, 1981) and B.subtilis (Peeters et al., 1988). The exact extent of this effect cannot be estimated from our experiments, since the recombination units with
determined by the length of the direct and inverted repeats.

A small but reproducible effect of the length of direct repeats was also detected: (i) with the recombination units lacking inverted repeats, 10 times less CmR cells were generated in overnight cultures when the direct repeats were 9 bp rather than 18 bp (Table II). With the same recombination units, the recombination frequency between the 9 bp direct repeats was 3–4 times lower than between the 18 bp, as measured on exponentially growing cells by a Luria–Delbrück fluctuation test (not shown). (ii) With recombination units containing inverted repeats, recombination between 18 bp direct repeats was less frequent than that between 27 bp direct repeats, as estimated from the proportion of CmR cells in overnight culture (Table II) and from the fluctuation test (a 3- to 4-fold difference was observed). A correlation between the recombination efficiency and the length of short repeats has been observed previously (Albertini et al., 1982; Watt et al., 1985; Shen and Huang, 1986; Peeters et al., 1988). We have no explanation at present for the lack of such a correlation for 9 and 18 bp direct repeats within recombination units containing inverted repeats (Table II).

Recombination between short direct repeats was greatly stimulated by the induction of ssDNA synthesis. This was shown by a rapid increase in the proportion of cells having the recombination phenotype upon activation of the M13 replication origin (Figure 3). With all plasmids the population of recombinant cells increased to 100%, but the rate of increase was 2–3 times faster for units carrying inverted repeats than for units lacking such repeats. These observations indicate that inverted repeats are not required for stimulation of recombination by ssDNA synthesis, but that they may be able to enhance it.

The mechanism of recombination depends on the length of direct repeats

A breakage–reunion mechanism of recombination always entails some transfer of DNA from parental to recombinant molecules, while a copy choice mechanism does not need to (Figure 1; Stahl, 1986). The very high efficiency of recombination upon induction of ssDNA synthesis allowed us to measure this transfer as follows (cf. Brunier et al., 1988). Total cell DNA was labeled with radioactive thymidine under conditions restrictive for recombination (42°C), labeling was discontinued and recombination was induced by a temperature shift (to 30°C). Recombination was allowed to proceed long enough to generate 10–100%
of recombinant cells. Plasmid DNA was then extracted and analyzed by gel electrophoresis. Parental and recombinant DNA, which differed in size, could be detected by fluorescence in all cases. The gel was cut in slices, plasmid DNA was recovered and used to transform competent E. coli cells to ampicillin resistance (Ap^R; the genetic marker present on both parental and recombinant plasmids) and to measure the radioactivity. Representative results are shown in Figure 4; these have been reproduced at least twice for each plasmid shown in Figure 2.

Two peaks of Ap^R transformants were observed in each case, one at the position of the parental plasmid, the other at the position of the recombinant plasmid. The latter represented between 30 and 90% of the total number of transformants. Only one peak of radioactivity, co-migrating with the parental plasmid, was observed when the recombinant units contained direct repeats of 9 bp. In contrast, two peaks of radioactivity, one co-migrating with the parental, the other with the recombinant plasmid, were detected when the units contained direct repeats of 18 bp (and 27 bp, not shown). The specific radioactivities of the recombinant plasmids, relative to that of parental plasmids, are given in Table III (limits of detection are indicated when no radioactivity was observed in the recombinant plasmid peak). These results suggest that direct repeats of 9 bp recombine by a copy choice mechanism, irrespective of the presence or the length of adjacent inverted repeats. In contrast, at least some of the direct repeats >18 bp appear to recombine by a breakage—reunion mechanism.

The specific activity of the plasmids resulting from recombination between direct repeats >18 bp was 2-4 times lower than that of the parental plasmids (Table III). Among the possible explanations is that they may be generated by two different mechanisms, one that transfers DNA from parental to recombinant plasmids and one that does not. To test the ability of different repeats to undergo copy choice recombination, which does not transfer any DNA, we prepared parental ssDNA and used it to transform competent E. coli cells lacking the protein necessary to activate the M13 replication origin. In these cells ss plasmids undergo only one round of conversion into a ds form, immediately upon transformation, and then continue to replicate as ds molecules. Slippage of the replication machinery during this round of conversion would give rise to Cm^R transformants. Such transformants were obtained with each of the ss plasmids (Table IV). They did not result from recombination after the plasmids were converted to a ds form, since no Cm^R transformants were obtained in control experiments with ds plasmid DNA (the limit of detection was 1 Cm^R transformant for 10^6 Ap^R transformants). This suggests that copy choice recombination can occur on direct repeats >18 bp. The frequency of the process was, however, <10 times lower than with 9 bp, for reasons presently not understood. Presence of inverted repeats had in all cases a stimulatory effect (Table IV).

Discussion

Genome rearrangements often result from recombination between short homologous sequences (cf. Ehrlich, 1989, for review). A model system to study such recombination is the precise excision of transposons Tn5 and Tn10 (Egner and Berg, 1981; Foster et al., 1981; DasGupta et al., 1987; Jannière and Ehrlich, 1987). We have recently shown that the excision of a transposon derived from Tn10 occurs by copy choice recombination (Brunier et al., 1988). This raised a question whether other recombination events between short direct repeats can occur by the same mechanism, as was previously suggested (Efstratiadis et al., 1981; Albertini et al., 1982). Transposon excision corresponds to recombination between 9 bp direct repeats adjacent to transposon ends, which are long inverted repeats (78 bp in the above study; Brunier et al., 1988). In contrast, most other instances of genome rearrangements occur by recombination between short homologous sequences that are not adjacent to long inverted repeats and are only occasionally adjacent to short inverted repeats (Glickman and Ripley, 1984). Since the presence and length of inverted repeats affect the frequency of transposon excision (Egner and Berg, 1981; Foster et al., 1981), it was possible that they also affect the mechanism of recombination between short direct repeats. To test this possibility we used a series of E. coli plasmids that carried short direct repeats (9, 18 or 27 bp) adjacent to long (308 bp), short (8 bp) or no inverted repeats. Recombination between the direct repeats was highly efficient when ssDNA synthesis was induced in these plasmids. This allowed us to test the mechanism of recombination by following transfer of DNA from parental to recombinant plasmids. A copy choice process should entail no such transfer, while a breakage—reunion process should. No transfer was detected for recombination between 9 bp repeats, irrespective of the presence or the length of adjacent inverted repeats. In contrast, massive transfer of material was detected with only slightly longer repeats, 18 or 27 bp.

Could the lack of detectable transfer be due to a preferential accumulation of recombinant plasmids in the bacterial population rather than to copy choice recombination? For example, the recombinant plasmids could replicate faster than the parental plasmids, or even completely inhibit replication of the parental plasmids. Multiple copies of recombinant plasmid could therefore be generated from a single recombination event; the parental material transferred to the original recombinant could thus be too dilute to be detectable. This hypothesis is ruled out by the following evidence. The recombinants were generated with the same kinetics from parental plasmids carrying 9 or 18 bp direct repeats (Figure 3). The putative preferential accumulation of recombinants would therefore have been comparable in the two experiments. This is not surprising, since the two types of parental plasmids differed by 9 bp only (the difference in the length of direct repeats) out of a total of ~8 kb, and gave rise to identical recombinant plasmids. Since a massive transfer of material was detected with plasmids carrying 18 bp direct repeats, the lack of transfer with plasmids carrying 9 bp direct repeats could not have been due to preferential accumulation of recombinants. We therefore conclude that the direct repeats of 9 bp recombine by a copy choice process, possibly as indicated in Figure 1A (steps 1-5a). In contrast, at least a significant proportion of the direct repeats >18 bp recombine by a breakage—reunion process, irrespective of the presence or the length of adjacent inverted repeats. These results indicate that the mechanism of recombination is determined by the length of the recombining direct repeats, and not by the presence of inverted repeats.

The details of the breakage—reunion recombination are
not known at present. Two alternatives may be considered. The first is that the process involves slippage of the replication machinery on the ssDNA template (Figure 1A, steps 1–4), and that the break–join step corresponds to cleavage of the ssDNA loop carried by the recombination intermediate and by the ligation of the ends (Figure 1A, step 5b; cf. Albertini et al., 1982). An argument against this hypothesis is that such cleavage is not detected during recombination of the 9 bp repeats, which presumably proceeds via a similar intermediate. Another alternative is that the process is a bona fide homologous recombination. A model of such a process is presented in Figure 1B, but other models were also proposed (cf. Conley et al., 1986; DasGupta et al., 1987). An appropriate genetic analysis may enable us to distinguish between these alternatives. Interestingly, a rather sharp cut-off exists for break–join recombination, which is efficient on repeats of 18 bp, but not detectable on repeats of 9 bp. Experiments with repeats of intermediate length may help define the dependence of breakage–reunion recombination on the length of the recombining sequences, and could give insights into mechanisms of homologous recombination. Some copy choice recombination between direct repeats of 18 and 27 bp seems to occur (Table IV), which suggests that these repeats may recombine by the two pathways, breakage–reunion and copy choice.

The length of the homologous sequences observed to recombine in various instances of genome rearrangements (cf. Ehrlich, 1989) was almost invariably <18 bp, which suggests that these sequences could have undergone a copy choice recombination. Much further work is, however, needed to ascertain that copy choice recombination can take place in genomes other than E. coli plasmids, or in organisms other than E. coli, and yield rearrangements other than deletions. Significant differences between two bacteria, E. coli and B. subtilis, were already observed. The direct repeats which recombined with the same efficiency in the former organism (9 and 18 bp, neighboring either short or long inverted repeats; Table II) recombined with vastly different efficiencies in the latter (a 1000-fold difference was observed; Peeters et al., 1988). This may be due to different mechanisms of recombination in the two hosts. However, since DNA replication always occurs on ss templates, and since many processes may generate ssDNA, there is ample opportunity in any cell for slippage of the DNA synthesizing machinery. This suggests that copy choice recombination may underlie numerous instances of genome rearrangements occurring in various organisms.

**Materials and methods**

Bacterial strains, plagues and plasmids are listed in Table I. Standard rich media (LB and LB Thy; Miller, 1972), supplemented, when appropriate, with ampicillin (100 μg/ml), kanamycin (15 μg/ml) or chloramphenicol (8 μg/ml), were used for cell growth. Transformation and plasmid transduction were previously described (Dagert and Ehrlich, 1979, 1983). Plasmid DNA was extracted from E. coli by the clear lysate method (Clewell and Helinski, 1969) and purified by chromatography on hydroxyapatite columns (Colman et al., 1978). For analytical purposes, plasmid DNA was prepared by the rapid extraction method (Bimboim and Doly, 1979). ssDNA was extracted from phage particles as described by Zinder and Boeke (1982). For cloning and crude structural analysis, restriction enzymes BglII, EcoRI, Ncol, Smal and XmnI were from Boehringer Mannheim and were used according to the supplier’s instructions. Nucleotide sequence analysis was performed using the dyeodeoxy method described by Sanger et al. (1977) on ds plasmids. The required primers, and other oligonucleotides were synthesized on an Applied Biosystems 380-B DNA synthesizer.

Induction of ssDNA synthesis in plasmids and the transfer of radioactivity from parental to recombinant molecules were previously described (Brunier et al., 1988). The fluctuation test (Luria and Delbrück, 1943) was performed as follows. For each determination 40 independent 1 ml cultures were inoculated with 10–105 exponentially growing CmR cells carrying parental plasmids. The cultures were grown to a density which varied between 106 and 108 cells/ml in different experiments, and chloramphenicol was added to a concentration of 8 μg/ml. A sample of each culture was used to determine the total cell number; the remainder was incubated for 12 h at 37°C. Some cultures were fully grown while in others no growth was observed. The average number (k) of recombinant CmR cells among all the cultures was calculated from the proportion of cultures not grown (P0) by applying the Poisson distribution law (P0 = e–k). The recombination frequency was estimated by dividing the average number of CmR cells by the average number of viable cells.

The plasmids used are shown in Figure 1. They were derived from plasmid pCM (previously called pHV606; Brunier et al., 1988), pCM9X, pCM18X, pCM27X, pCM9K, pCM18K and pCM27K were constructed by transferring the recombinant units as Ncol segments from the pHF3 plasmid series (Peerers et al., 1988) onto pCM. A 1.4 kb HaeII segment of plasmid pHH1 (Trieu-Cuot and Courvalin, 1985) which encodes a KmR gene was carried between the repeats in these plasmids. In the plasmids of the X series the KmR gene was flanked by inverted repeats of 308 bp, in those of the K series it was flanked by inverted repeats of 8 bp. The 308 bp inverted repeats were composed of the 8 bp inverted repeats of the K series to which the 300 bp fragment of IS50 was adjoined (cf. Peerers et al., 1988). Duplications in the recombinant units devoid of any inverted repeats were generated in pCM by the method of Peerers et al. (1988), using synthetic oligonucleotides that contained no palindromic sequences and carried a BglII and a Smal restriction site. A 1.2 kb HindIII–BglII fragment of the transposon Tn5 encoding a KmR gene was carried between the repeats in these plasmids.

**Acknowledgements**

This work was supported by a grant from the Commission des Communautés Européennes (no. BAP-0141-F) to S.D.E. and by Gist-Brocades NV (Delft, The Netherlands) to B.P.H.P.

**References**


Lucia, S.E. and Delbrück, M. (1943) Genetics, 28, 491–511.

Received on April 17, 1989; revised on June 16, 1989