## Intramolecular recombination between partially homologous sequences in *Escherichia coli* and *Xenopus laevis* oocytes

(branch migration/histocompatibility antigens/DNA end)

Jean-Pierre Abastado, Sylvie Darche, François Godeau, Brigitte Cami, and Philippe Kourilsky

Unité de Biologie Moléculaire du Gène U. 277 Institut National de la Santé et de la Recherche Médicale, UAC 115, Centre National de la Recherche Scientifique, Institut Pasteur, 25, rue du Dr. Roux, 75724 Paris Cedex 15, France

Communicated by Marianne Grunberg-Manago, April 14, 1987

ABSTRACT We describe a system to analyze the individual contribution of a single physical DNA end on intramolecular recombination between partially homologous sequences. We took advantage of this partial sequence divergence to measure the distance separating the DNA end from the final recombination event. We show that a single physical DNA end stimulates recombination when located in a region of homology. Recombination frequency decreases gradually with the distance from the DNA end. A recombinational hot spot is found at the end of the region of homology. A large insertion of unrelated DNA interferes asymmetrically with this process, suggesting that a recombinogenic signal propagates along the region of homology.

The molecular details of the recombination processes are still far from being completely deciphered. Intensive genetic studies in Escherichia coli and Saccharomyces cerevisiae have allowed the identification of a variety of genes that participate in several pathways (1-4). Due to the development of in vitro systems, the pivotal role of RecA protein is well documented (5-7). However, the structure of recombination intermediates in vivo is still a matter of controversy. Results of gene conversion experiments, particularly in Ascobolus (8) and yeast (9-11), led to the formulation of hypotheses on the structures of DNA intermediates possibly involved in recombination. Thus, it is often assumed that correction of heteroduplex DNA is a major source of conversion events (8, 12), but the existence of heteroduplex DNA (13-16) and its involvement in conversion and recombination are difficult to assess in vivo.

Most of these studies have involved homologous DNA sequences displaying minor differences. Little is known about the recombination between partially homologous sequences. The effects of heterology have been mainly studied from a quantitative point of view and we wanted to know how mismatches or gaps affect the type of recombinants that are recovered.

In S. cerevisiae, recombination between a chromosomal gene and a transfected homologue was found to be strongly stimulated when the latter was carried on a linear piece of homologous DNA rather than on a circular plasmid (17). This observation provided the basis for the double-strand break recombination model of Szostak *et al.* (18). Evidence that double-strand breaks stimulate recombination has also been obtained in mammalian cells (19, 20) and in *E. coli* with plasmid (4, 21) or phage systems (22).

When linear DNA recombines, both ends interact with the homologous duplex, resulting in branch migration in opposite directions and the formation of two Holliday junctions. A detailed analysis of double-strand break-induced recombination might be easier if one dealt with a single class of events initiated from a single end.

To study this question both in *E. coli* cells and in *Xenopus laevis* oocytes, we devised an experimental system in which only one of two ends can initiate recombination, while the other is blocked by a heterologous sequence. The sequences that recombine were chosen to display enough divergence to allow rapid analysis of recombinant plasmids by colony hybridization using several oligonucleotide probes specific for one or the other parental sequence. Our studies show that a recombination signal propagates from the break and that the introduction of a piece of heterologous DNA in the homology region creates a hot spot of recombination nearby. These results are discussed in the context of the current recombination models (18, 20, 23).

## MATERIALS AND METHODS

**Bacterial Strains.** Pop 2556, kindly provided by O. Raibaud, is MM 294 ( $hrs^-$ ,  $hrm^+$ ,  $thi^-$ ,  $endoI^-$ ,  $recA^+$ ) (24) carrying an additional MalT 250 (25) mutation. The presence in this strain of a plasmid carrying a *supF* gene is monitored by plating onto EMB-maltose, where  $mal^+$  and  $mal^-$  colonies display a black or a white color, respectively. 1046 ( $met^-$ ,  $hsd_KS^-R^+M^+$ , *supE*, *supF*, *recA56*) is our standard *recA^-* strain (26).

**Plasmids.** pKKD and pKSD are derivatives of pBR322 carrying the H-2K<sup>d</sup> sequence of pH-2<sup>d</sup>-33 (27) and the H-2D<sup>d</sup> sequences of pAG 64 (28) cloned in the same orientation in the tetracycline-resistance gene. Between these two H-2 cDNA sequences, we inserted the *supF* gene in pKSD and the kanamycin-resistance ( $Km^R$ ) gene in pKKD. The maps of these two constructs are presented in Fig. 3.

pKKDg has the same structure as pKKD except that the two H-2 sequences are derived from the genomic clones 2.14 (29) and c49-2 (30).

pKKDins was derived from pKKD by inserting 550 base pairs (bp) from M13mp8 into the Kpn I site located in the H-2K<sup>d</sup> sequence.

**Microinjection.** Fifteen nanoliters of DNA (0.5 mg/ml) was microinjected into the nuclei of X. *laevis* oocytes. Incubation was for 5 hr and DNA was recovered as described (31).

Isolation of Spontaneous Recombinants. Ninety-six individual colonies of pop 2556 carrying plasmid pKSD (which includes a suppressor) were inoculated in 5 ml and grown to saturation. After three sequential passages in which 10  $\mu$ l were inoculated into 2 ml of broth, nonsuppressor colonies were reisolated on EMB-maltose plates.

In Situ Hybridization with Oligonucleotide. The sequences of the 12 oligonucleotides are given in Fig. 1. In situ hybridization was performed as described (26) except for the temperature of the last wash, which was 45°C (probes 3K,

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. \$1734 solely to indicate this fact.

Abbreviations: Km<sup>R</sup> and Km<sup>S</sup>, kanamycin-resistant and kanamycinsensitive; Amp<sup>R</sup>, ampicillin-resistant.

4K, 4D, 5K, 5D), 50°C (probes OK, OD, 2K, 3K), or 53°C (probes 1K, 1D, 2K).

## RESULTS

**Experimental Design.** Linearized plasmids containing homologous sequences at both ends yield a very high proportion of recombinant plasmids when introduced into *E. coli* cells (4, 32) or *X. laevis* oocytes (31). Accordingly, we constructed the following plasmids.

(i) In a pBR322 background, we cloned two cDNA sequences encoding the mouse  $H-2K^d$  and  $H-2D^d$  antigens, which display  $\approx 10\%$  heterology distributed as shown in Fig. 1. The two cDNAs were inserted in the same orientation with appropriate linkers (Fig. 1).

(*ii*) Between the two H-2 sequences, we inserted a genetic marker, either supF or  $Km^R$ . The resulting plasmids pKSD and pKKD both carry the ampicillin-resistance  $(Amp^R)$  marker (Fig. 1). The presence of supF was monitored as described by transforming a strain carrying an amber mutation.

(*iii*) pKKD was linearized by restriction enzyme digestion and used to transform *E. coli* to ampicillin resistance. Transformants were examined for the presence of the  $Km^R$ gene, which is lost upon homologous recombination between the K and D sequences (Fig. 2). Alternatively, linearized pKKD DNA was microinjected into oocyte nuclei and incubated for 5 hr. Recovered DNA was used to transform *E*.

H-2K4	10 CCGCGGGGGGCGCCG1	20 Iggatggagca	30 GGAGGGGGCCG	40 GAGTATTGGG		60 AGAGAGCCA	70 AGAGCG
H-20°	G	·A			- <u>C-G</u>	-G	GA
H-2X <b>d</b>	80 ATGAGCAGTGGT1	90 ICCGAGTGAGC	100 CTGAGGACCG	110 CACAGAGATAG -G-T-C-C	120 TACAACCAGAG	130 GCAAGGGGGGG	140 SCTCTCA
H-2Kd	CACGTTCCAGCG	<b>1K</b> GATGTTCGGGT	170 GTGACGTGGG	1 80 GTCGGACTGGG	190 GCCTCCTCCG	200 CGGGTACCA	210 TCAGTTC
H-2K <sup>d</sup>	220 GCCTACGACGGC	1D 230 CGCGATTACAT	240 CGCCCTGAAC	250 Gaagacctga	260 NAACGTGGACG	270 GCGGCGGAC	280 ACGGCGG
H-2K <sup>d</sup>	290 CGCTGATCACCA	300 Gacgcaagtgg	310 GAGCAGGCTG	320 GTGATGCAGA	330 GTATTACAGGG	340 CCTACCTAG	350 Agggcga
H-2Kd	360 GTGCGTGGAGTG	370 GCTCCGCAGAT	3 80 ACCTGG/GC1	2 K	400 CGCTGCTGCG	410 CACAGATTC	420 CCCAAAG
H-2Kd	430 GCCCATGTGACC	440 TATCACCCCAC	450	2 D 460 GATGTCACCC	470 TGAGGTGCTGG	a 80 GC CCT GG GC	490 TTCTACC
H-2Dd	500	510	-CGG- 520	530	540	550	560
H-2Dd	C 570	580	590	600	<b>A</b> -	620	-G 6 3 0
H-2D4	640	GGATGGAACCI	K	670	680	1GGGAAGGA	GC AGAAT
H-2Kd	TACACATGCCAT	GTOCACCATA	GGGGCTGCC	IGAGCCTCTCA	CCCTGAGATGO	GGCAAGGAG	AAGCTTC
H-2Kd H-2Dd	710 CTCCATCCACT		730 GTAATCATT	7 40 SCTGTTCTGG1	750 TGTCCTTGGAC	760 SCTGCAATAG TGG-CA	770 TCACTGG
H-2Kg	780 AGCTGTGGTGGG 	790 TTTTGTGATG		GAAC: CAGGI	820 GGAAAAGGAG1	830 IGAACTATGO 	840 TCTGGCT
H-2K <sup>d</sup>	850 CCAGGCTCCCAC	860 ACCTCTGATC	B70 IGTCTCTCCC	860 Agatggtaaac	890 TGATGGTTCA1	900 Igaccorent	910 TOTOTIC
11-2K <sup>d</sup> H-2D <sup>d</sup>	920 CGTGAAGACAGO T	:TG					

FIG. 1. Localization of the oligonucleotide probes. The alignment of  $H-2K^d$  and  $H-2D^d$  is presented in this figure. There are 85 nucleotide differences and 2 gaps within 924 bases. Twelve oligonucleotides, each 14 bases long, were synthesized, representing 6 areas of substantial heterology dispersed over this entire region (boxed areas). These were used as specific probes for either K or D sequences in the analysis of recombinant clones.



FIG. 2. The "snail" technique. pKKD (*Left*) DNA was linearized in either the H-2K<sup>d</sup> or H-2D<sup>d</sup> region and used to transform *E. coli* cells to ampicillin resistance. Kanamycin-sensitive colonies were subsequently analyzed using oligonucleotide probes and shown to contain recombinant plasmids (*Right*). The "snail" (*Center*) represents the putative structure of the recombination substrate.

coli, and  $Amp^R Km^S$  (Km-sensitive) colonies were further studied.

(iv) To characterize recombinants, we performed colony hybridization with a set of 12 oligonucleotides (Fig. 3), of which 6 matched the K sequence, and 6 matched the D sequence in the same positions. Thus, recombinants should hybridize with only one probe from each pair. This procedure eliminates ambiguities (false positives, deletions) in the interpretation of hybridization data. Finally, the deduced structure could be confirmed by restriction enzyme mapping or DNA sequencing.

**Double-Strand Cuts in the Homologous Regions Stimulate** Recombination in E. coli. pKKD DNA was digested by restriction enzymes in or outside the region of homology and was used to transform E. coli. The yield of  $Amp^{R}$  transformants and the percentage of  $Km^{S}$  clones were measured. As shown in Table 1, the number of  $Km^{S}$  clones was at least 30 to 300-fold higher when pKKD was cut in a homologous rather than a heterologous region. In fact, no  $Km^{S}$  colonies could be detected when a double-strand break was introduced in the nonhomologous region. The structures of several dozen  $Km^{S}$  clones were analyzed. All clones were true recombinants. Cutting pKKD at the Xba I site yielded 10 times less recombinants than cutting in the region of homology. Since 5' or 3' protruding as well as blunt DNA ends stimulated recombination equally well (Table 1), we favor the hypothesis that this intermediate frequency is due to the proximity of this site (17 bp) to the homologous H-2D sequence. In support of this view, subsequent digestion with BAL-31 exonuclease gave a 10-fold increase in recombination frequency.



FIG. 3. Structure of the test plasmids used. pKKD was used as a linear substrate for intramolecular recombination, whereas pKSD was used in the segregation experiment described in the text. RI, *Eco*RI; BH, *Bam*HI; K, *Kpn* I; S, *Sac* I; RV, *Eco*RV; H, *Hind*III; SM, *Sma* I; SP, *Sph* I; X, *Xba* I.

 Table 1.
 Effect of double-strand breaks on recombination

	Amp <sup>R</sup> colonies			
Enzyme(s)	per μg of DNA	Km <sup>s</sup> colonies per μg of DNA		
Uncut	$2 \times 10^{6}$	$2 \times 10^{4}$		
Outside the homology				
region				
BamHI	$6 \times 10^3$	<120 (ND)		
Sph I	$6 \times 10^3$	<480 (ND)		
Sma I	$5 \times 10^{3}$	<100 (ND)		
Xba I	$9 \times 10^{3}$	$1.5 \times 10^{3}$		
In the homology				
region				
Kpn I	$9 \times 10^4$	$2 \times 10^4$		
Sac I	$7 \times 10^4$	$3 \times 10^4$		
Bgl II	$3 \times 10^4$	$1.5 \times 10^{4}$		
<i>Eco</i> RV	$2.2 \times 10^{5}$	$2.2 \times 10^4$		
Xba I + BAL-31	$1.5 \times 10^{4}$	$1.5 \times 10^{4}$		
Double cut				
Sma I + Kpn I	$2.2 \times 10^4$	$2 \times 10^4$		
Sma I + Bgl II	$2 \times 10^4$	$1.5 \times 10^{4}$		
Sma I + EcoRV	$2.3 \times 10^{4}$	$2.2 \times 10^4$		

ND, not detected.

The structures of the plasmid isolated from some  $Amp^R$  $Km^R$  colonies were determined. Most of them (>95%) were identical to pKKD and probably corresponded to plasmid molecules that escaped cutting by the restriction enzyme or were end-to-end ligated after the transformation. To avoid this background, we digested pKKD with two enzymes, creating one end in the homology region (*Kpn* I or *Eco*RV) and another in the *Km<sup>R</sup>* gene (*Sma* I). In comparison with a single cut in the homologous region, the yield of recombinants was unmodified but represented 90–95% of the *Amp<sup>R</sup>* transformants (Table 1).

To measure the rate of recombination of circular DNA, we turned to pKSD, where loss of the supF by homologous recombination is easily monitored in the *malT* amber strain. The presence of the supF gene on a high copy number plasmid is harmful to the host, and bacteria that harbor pKSD grow twice as slowly as those that carry a pKD recombinant derived from pKSD by internal recombination (data not shown). Using this trait, we can estimate the rate of recombination from the instability of  $supF^+$  clones obtained upon transformation with pKSD. Upon serial passages,  $supF^-$  cells invade the bacterial population. This occurred, on average, after 40 generations, from which we calculate a recombination rate of  $10^{-5}$  or less per bacterial generation.

In summary, double-strand breaks stimulate recombination, and they do so only when the break lies in a region of homology (Table 1).

**Recombination Takes Place at a Variable Distance from the Break.** To locate the site of recombination,  $100-200 Amp^R Km^S$  colonies were analyzed with oligonucleotide probes and restriction mapping. Results are shown in Fig. 4 as histograms indicating the percentage of recombinants found in each physical interval.

When pKKD was digested with *Hin*dIII, which cuts once in each homology region (Fig. 3), the histogram was Ushaped (Fig. 4A). This suggested to us that we were observing the superposition of two classes of recombination events proceeding from both ends. To obviate this complexity, pKKD was digested with two enzymes, creating one end in the homology region and another in the nonhomology region. The shape of the histograms was modified: recombination was most frequent near the homologous end, and the frequency decreased with distance; sometimes (D and E) an increase in recombination frequency at the border of the nonhomology region was observed. When double-strand



FIG. 4. Distribution of recombination events between H-2K and H-2D sequences along the homology region. In each experiment (A-L), 100 to 200 Km<sup>s</sup> clones were analyzed. For each clone, the recombination point was deduced from the pattern of hybridization with H-2K- or H-2D-specific oligonucleotide probes and from the restriction enzyme profile. The area of each bar in the histograms represents the total number of recombination events observed in the indicated interval, whereas the height of the bar represents the frequency of recombination (number of recombinants per nucleotide). The solid line above each histogram represents 100 bp. Arrowheads mark the position of the homologous end. Numbers under the histogram indicate the position of each respective oligonucleotide probe. (A) pKKD cut by HindIII; (B) cut by Bgl II and Sma I; (C) cut by Kpn I and Sma I; (D) cut by Sac I; (E) cut by EcoRV and Sma I; (F) cut by Xba I and Sma I and treated for 10 min with 5 units of BAL-31 exonuclease. (G) Spontaneous recombinants obtained in independent cultures of pKSD. (H) pKKD-ins cut by EcoRV and Sma I; the position of the insertion of heterologous DNA is denoted by an open triangle. (1) pKKDg cut by EcoRV and Sma I. (J) pKKD cut by EcoRV and Sma I in the RecA<sup>-</sup> host. (K) pKKD cut by EcoRV and Sma I injected in oocytes. (L) pKKD cut by Xba I and Sma I injected in oocytes.

breaks were introduced at different sites in the homology region (*Bgl* II, *Kpn* I, *Sac* I, *Eco*RV, or *Xba* I), a similar type of distribution was observed, with the highest frequency near the homologous end (Fig. 4 B-F).

Thus, the gradient of recombination frequencies is primarily related to the location of the break and not to particular hot spots of recombination found in the sequences themselves. To substantiate this point, we analyzed the distribution of recombinants obtained after transformation with circular DNA. For reasons stated above, we turned to the pKSD plasmid and analyzed the structure of 96 spontaneous independent recombinants that had lost the *supF* marker. The histogram (Fig. 4G) shows an almost uniform distribution, with no indication for hot spots.

**Recombination in X.** *laevis* **Oocytes.** It could be argued that homologous recombination between eukaryotic sequences is somewhat artifactual in *E. coli*. We therefore injected pKKD DNA linearized by *Eco*RV and *Sma* I or *Xba* I and *Sma* I into X. laevis oocyte nuclei. DNA recovered after 5 hr of incubation yielded 20-80 times more  $Km^{S}$  transformants per microgram of DNA than linear pKKD DNA transformed directly into E. coli. This result suggested that the injected DNA recombines in the oocytes. However, we could not rule out the possibility that injection in oocytes only potentiates recombination, which occurs subsequently in E. coli. To exclude this hypothesis, DNA recovered from oocvtes was directly analyzed by Southern blotting and a recombinantspecific fragment was detected (data not shown). This demonstrates that recombination does occur in X. laevis oocvtes. These recombinants cloned in E. coli were further analyzed by colony hybridization with the oligonucleotide probes. Fig. 4 K and L shows the distribution of the recombinants along the sequences. The histograms were qualitatively similar to those obtained in E. coli.

Recombination Associated with a Polar Process. Because the stimulation of recombination by a break in the homologous region is so high, we assume that most recombination events scored in the experiments described above are due to the break, even when they take place at a distance from it. A simple hypothesis, supported by the shape of the histograms, is that a recombination "signal" starts from the break and propagates along the homology region inducing recombination at variable distances. To test this assumption, we introduced a heterologous sequence in the homology region and analyzed the effect on recombination. A 550-bp fragment of M13 phage DNA was thus inserted at the Kpn I site within the K sequence. As shown in Fig. 4H, the distribution of recombinants was significantly affected by the insertion, since recombinants accumulated upstream from the heterology (compare Fig. 4 E and H). More significantly, a hot spot of recombination appeared on the side of the heterology proximal to the break. In this analysis, 5% of the recombinants were found within an 18-bp sequence to the right (Fig. 4H) of the insertion, but none were found in the 18-bp sequence to the left. This experiment establishes the polar character of the process initiated at the break and strongly suggests that heterology creates a recombination hot spot.

The Recombination Signal Can Propagate for Several Kilobases. The histogram shown in Fig. 4F indicates that 8% of the recombination events occur >500 bp from the doublestrand break. This was the maximum distance that could be studied with pKKD. We therefore replaced the cDNA sequences in pKKD with the corresponding H-2 gene sequences. This yielded the plasmid pKKDg, where the homology region spanned 4.5 kilobases. When pKKDg digested by *Eco*RV was transformed into *E. coli*, recombinants were obtained and analyzed with the same set of oligonucleotidic probes (Fig. 4I). This histogram presents a monotonic decrease of the recombination frequency with increasing distance from the break. As many as 1% of the recombination events occurred further than 2.5 kilobases from the break.

## DISCUSSION

We describe an experimental system designed to explore the intramolecular recombination of plasmid DNA in *E. coli* or *X. laevis* oocytes. We constructed plasmids that carry, in the same orientation, two mouse *H-2* sequences (displaying  $\approx 10\%$  divergence) with a marker gene in between to facilitate characterization. A set of 12 synthetic oligonucleotide probes was used to locate exchange events in seven different regions. The exchanges were further defined by restriction enzymes and/or DNA sequencing.

Using this system, we have confirmed previous studies in  $E. \ coli$  (4, 21, 32), showing that a double-strand break enhances the frequency of intramolecular recombinants. This is consistent with the idea that plasmid recombination in wild-type  $E. \ coli$  occurs via a pathway similar to the RecF

pathway (4), since this pathway responds to double-strand cuts. Similarly, Red-dependent recombination of phage  $\lambda$  is stimulated by cuts introduced in vivo at the cos site or by a restriction enzyme (reviewed in ref. 22). In addition, in this system no stimulation is seen when the cut is outside the region of homology. This agrees with our results, showing that altering the position of the double-strand break has a dramatic effect on the yield of recombinants (Table 1). A small sequence of 17 heterologous base pairs is sufficient to reduce the yield of recombinants by a factor of 10 (Table 1). When the break was located further from the homology region, no recombinant could be detected. This indicates that the stimulation factor by a break located in the homology region is at least 30-300. The experiment with pKSD implies that essentially all recombination events observed in our experiments are induced by the break.

When the parental plasmid was cut once in the homology region and once in the  $Km^R$  spacer gene sequence, the histogram was asymmetric. Its shape was not affected by the site of cutting in the  $Km^R$  sequence, but the distribution was strongly dependent on the site cutting in the homologous sequences (K or D). As expected, it was inverted when the D, rather than the K, sequence was cut (Fig. 4F). This observation confirms that all observed recombination events are initiated from the double-strand break in the homology region. Two breaks, one in each homology sequence, result in a U-shaped curve (Fig. 4A), which is, in fact, the superposition of two asymmetric curves each being separately obtainable by blocking one end with a heterologous sequence. Because this asymmetric distribution is observed with plasmid cut at various restriction sites along the homology region, it is not due to a specific sequence in the homologous region. This conclusion is further supported by the analysis of spontaneous recombinants, which shows a uniform distribution (Fig. 4G).

Recombination frequency is maximal in the vicinity of the break and decreases gradually with distance. It often reincreases at the border of the heterology region (i.e., plasmid DNA), which flanks the homologous sequence. It was, therefore, of interest to insert a large stretch of heterologous sequence (550 bp) in one of the homology regions. Recombination was less frequent but not abolished in the region distal to the insertion and a hot spot of recombination was generated on the border of the insertion closest to the break (Fig. 4H). This experiment strongly reinforces the interpretation that a recombination signal is initiated at the homologous end and propagates along the DNA. The insertion of as much as 550 bp of heterologous DNA does not entirely suppress its propagation but apparently favors its arrest, which stimulates recombination on the border of the homology region.

These conclusions agree with other data obtained in different systems. Genetic experiments in *Ascobolus* established that, during meiosis, the frequency of cross-overs also increased in the vicinity of large heterologies (8), suggesting that these create recombination hot spots. In *E. coli, in vitro* studies have demonstrated that the purified RecA protein helps the formation of joint DNA molecules by a polar process initiated at one end. Internal heterologies also slowed or stopped growth of the heteroduplex (7). We therefore favor the interpretation that the distribution of recombinants reflects the size of heteroduplexes initiated by the invasion of one duplex by the homologous free end and subsequent growth by branch migration.

Our data, however, do not necessarily imply that the recombination "signal" is heteroduplex DNA. We have performed similar experiments in a  $recA^{-}$  strain of *E. coli* and found no qualitative differences (Fig. 4J). Although surprising, this observation is in agreement with the finding that transformation with linearized dimeric plasmids is RecA

independent (21). Also, in our experiments, the recombination signal can apparently bypass more heterologies (trinucleotide mismatches, regions of dense dispersed single nucleotide mismatches, and even the 550-bp insertion) than *recA*mediated heteroduplex growth *in vitro* (6, 7).

Whatever the molecular nature of the recombination signal, we have found that it can propagate along at least 2.5 kilobases of DNA. It must be noted that our "homologous" sequences differ by  $\approx 10\%$  of their nucleotides—a feature that facilitated the analysis of recombinants. We do not know whether this degree of heterology significantly impairs the migration of the recombination signal, which might otherwise traverse longer distances of DNA.

Even though most of our experiments have been performed in *E. coli*, we have obtained similar results in *X. laevis* oocytes. Thus, the conclusions obtained in *E. coli* are probably of general value. In addition, since double-strand DNA is not replicated in oocytes (33), the observed recombination process seems to be independent of replicative DNA synthesis.

In E. coli, homologous recombination between partially divergent sequences (as described here) could be considered somewhat artifactual (although recombination between plasmids or phages, such as  $\lambda$ , 434, and 80, offers biological grounds for such situations). However, it is of major importance in eukaryotes, where it plays a significant role in the evolution of multigene families (34). The use, in the present studies, of two sequences that encode highly polymorphic mouse class I transplantation antigens is irrelevant. But we (34), and others (35, 36), propose that recombination and conversion in the large mouse class I H-2 multigene family can explain this polymorphism through combinatorial processes, which shuffle divergent nucleotides and thereby create diversity. Stimulation of recombination by large heterologies may account for the clustering of recombination events in the major histocompatibility complex of the mouse (37, 38).

On practical grounds, our system is more convenient than the original one of Weber and Weissman (32) for generating hybrid DNA molecules sometimes difficult to obtain by standard *in vitro* genetic engineering techniques. The recombination pattern, although random, is predictable, and recombination can be favored in a given region by varying the location of the break. In 2000 recombinants analyzed, out of which 22 have been sequenced for other purposes (unpublished data), we did not observe any aberrant event that had not been readily detected by the oligonucleotide probes. Our constructs bear polylinkers, which make them available as convenient vectors for the engineering of DNA sequences by *in vivo* recombination.

We are grateful to O. Raibaud for constructing the bacterial strain Pop 2556, to P. W. Rigby for the gift of the plasmid pAG64, and to C. Jaulin for stimulating discussions and critical reading of the manuscript. We are indebted to O. Noël and to C. Verna for their help in the preparation of the manuscript.

- Clark, A. J., Sandler, S. J., Willis, D. K., Chu, C. C., Blanar, M. A. & Lovett, S. T. (1984) Cold Spring Harbor Symp. Quant. Biol. 49, 453-462.
- 2. Cohen, A. & Laban, A. (1983) Mol. Gen. Genet. 189, 471-474.
- Matfield, M., Badawi, R. & Brammar, W. J. (1985) Mol. Gen. Genet. 199, 518-523.

- Symington, L. S., Morrison, P. & Kolodner, R. (1985) J. Mol. Biol. 186, 515-525.
- Cassuto, E., West, S. C., Mursalim, J., Conlon, S. & Howard-Flanders, P. (1980) Proc. Natl. Acad. Sci. USA 77, 3962–3966.
- DasGupta, C. & Radding, C. M. (1982) Proc. Natl. Acad. Sci. USA 79, 762–766.
- 7. Bianchi, M. E. & Radding, C. M. (1983) Cell 35, 511-520.
- Rossignol, J. L., Nicolas, A., Hamza, H. & Langin, T. (1984) Cold Spring Harbor Symp. Quant. Biol. 49, 13-21.
- Esposito, M. S. & Wagstaff, J. E. (1981) in *The Molecular Biology of the Yeast Saccharomyces: Life Cycle and Inheritance*, eds. Strathern, J. N., Jones, E. W. & Broach, J. R. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 341-370.
- Sugarawa, N. & Szostak, J. W. (1983) Proc. Natl. Acad. Sci. USA 80, 5675-5679.
- 11. Klein, H. L. (1984) Nature (London) 310, 748-753.
- 12. Holliday, R. (1984) Genet. Res. 5, 282.
- 13. Fox, M. S. (1978) Annu. Rev. Genet. 12, 47-68.
- 14. Mejean, V. & Claverys, J. P. (1984) J. Bacteriol. 158, 1175-1178.
- 15. Lichten, M. & Fox, M. (1984) Proc. Natl. Acad. Sci. USA 81, 7180-7184.
- Symington, L. S., Morrison, P. & Kolodner, R. (1985) Mol. Cell. Biol. 5, 2361–2368.
- 17. Orr-Weaver, T., Szostak, J. & Rothstein, R. (1981) Proc. Natl. Acad. Sci. USA 78, 6354-6358.
- Szostak, J. W., Orr-Weaver, T. L., Rothstein, R. J. & Stahl, F. W. (1983) Cell 33, 25-35.
- Brenner, D. A., Smigocki, A. C. & Camerini-Otero, R. D. (1986) Proc. Natl. Acad. Sci. USA 83, 1762–1766.
- Lin, F. L., Sperle, K. & Sternberg, N. (1984) Mol. Cell. Biol. 4, 1020-1034.
- 21. Conley, E. C. & Saunders, J. R. (1984) Mol. Gen. Genet. 194, 211-218.
- 22. Stahl, F. W. (1986) Prog. Nucleic Acid Res. Mol. Biol. 33, 169-194.
- 23. Meselson, M. & Radding, C. (1975) Proc. Natl. Acad. Sci. USA 72, 358-361.
- 24. Talmadge, K. & Gilbert, W. (1980) Gene 12, 235-241.
- 25. Debarbouille, M. & Schwartz, M. (1979) J. Mol. Biol. 132, 521-534.
- Abastado, J. P., Cami, B., Huynh Dinh, T., Igolen, J. & Kourilsky, P. (1984) Proc. Natl. Acad. Sci. USA 81, 5792– 5796.
- Lalanne, J. L., Delarbre, C., Gachelin, G. & Kourilsky, P. (1983) Nucleic Acids Res. 11, 1567-1577.
- Brickell, P. M., Latchman, D. S., Murphy, D., Willison, K. & Rigby, P. W. J. (1983) Nature (London) 306, 756-760.
- 29. Kvist, S., Roberts, L. & Dobberstein, B. (1983) EMBO J. 2, 245-254.
- 30. Sher, B. T., Nairn, R., Coligan, J. E. & Hood, L. E. (1985) Proc. Natl. Acad. Sci. USA 82, 1175-1179.
- Carroll, D., Wright, S. H., Wolff, R. K., Grzesivk, E. & Maryon, E. B. (1986) Mol. Cell. Biol. 6, 2053-2061.
- 32. Weber, H. & Weissmann, C. (1983) Nucleic Acids Res. 11, 5661-5669.
- Gurdon, J. B. & Wickens, M. P. (1983) Methods Enzymol. 101, 20-36.
- 34. Kourilsky, P. (1983) Biochimie 65, 85-93.
- Schulze, D. H., Pease, L. R., Geier, S. S., Reyes, A. A., Sarmiento, L. A., Wallace, R. B. & Nathenson, S. G. (1983) *Proc. Natl. Acad. Sci. USA* 80, 2007–2011.
- Weiss, E., Golden, L., Zakut, R., Mellor, A., Fahrner, K., Kvist, S. & Flavell, R. A. (1983) *EMBO J.* 2, 453-463.
- Steinmetz, M., Minard, K., Horvath, S., McNicholas, J., Frelinger, J., Wake, C., Long, E., Mach, B. & Hood, L. (1982) *Nature (London)* 300, 35-42.
- Uematsu, Y., Kiefer, H., Schulze, R., Fischer-Lindalh, K. & Steinmetz, M. (1986) EMBO J. 5, 2123-2129.