MOLECULAR AND CELLULAR BIOLOGY, Oct. 1995, p. 5389–5395 Vol. 15, No. 10
0270-7306/95/$04.00+0

In-Frame Recombination between the Yeast H^+-ATPase Isogenes *PMA1* and *PMA2*: Insights into the Mechanism of Recombination Initiated by a Double-Strand Break

PHILIP SUPPLY,† ALBAN DE KERCHOVE D'EXAERDE, TIZIANA ROGANTI, ANDRÉ GOFFEAU, AND FRANÇOISE FOURY*

Unité de Biochimie Physiologique, Université Catholique de Louvain, B-1348 Louvain-La-Neuve, Belgium

Received 21 June 1995/Returned for modification 12 July 1995/Accepted 20 July 1995

Chimeric *PMA1*::*PMA2* sequences, placed under the control of the *PMA1* promoter, were constructed by in vivo recombination between a gapped linearized plasmid containing the *PMA2* gene and four different fragments of the *PMA1* gene. Correct in-frame assembly of the *PMA* sequences was screened by the expression of the *lacZ* reporter gene fused to the *PMA2* coding region. Restriction and sequencing analysis of 35 chimeras showed that in all cases, the hybrid sequences were obtained as fusions between continuous sequences specific to *PMA1* and *PMA2*, separated by a region of identity. In all but three cases, the junction sequences were not located at regions of greatest identity. Strikingly, depending on the *PMA1* fragment used, junction distribution fell into two categories. In the first, the junctions were scattered over several hundreds of nucleotides upstream of the extremity of the *PMA1* fragment, while in the second, they were concentrated at this extremity. Analysis of the alignment of the *PMA1* and *PMA2* sequences suggests that the distribution is not related to the size of the region of identity at the *PMA1-PMA2* boundary but depends on the degree of identity of the *PMA* genes upstream of the region of identity, the accumulation of successive mismatches leading to a clustered distribution of the junctions. Moreover, the introduction of seven closely spaced mismatches near the end of a *PMA1* segment with an otherwise-high level of identity with *PMA2* led to a significantly increased concentration of the junctions near this end. These data show that a low level of identity in the vicinity of the common boundary stretch is a strong barrier to recombination. In contrast, consecutive mismatches or regions of overall moderate identity which are located several hundreds of nucleotides upstream from the *PMA1* end do not necessarily block recombination.

The *PMA1* and *PMA2* genes in *Saccharomyces cerevisiae* encode homologous P-type ATPases. PMA1, the major plasma membrane proton pump, is essential for growth (22), while PMA2, which is not expressed under normal physiological conditions and whose physiological role is still unknown, appears not to be essential (20). When placed under the control of the *PMA1* promoter, *PMA2* is overexpressed and can partially substitute for *PMA1*; under these conditions, cells grow almost normally at neutral pH but cannot grow under acidic conditions (23). The *PMA1* and *PMA2* genes have 85% sequence identity. Most of the nucleotide substitutions and deletions are concentrated in the N-terminal part of the coding sequence, while the rest of the sequence shows much greater homology.

It has been shown that the homeologous *PMA1* and *PMA2* genes can exchange genetic information by ectopic recombination (8). Analysis of revertants of the *pma1-105* mutation revealed that nearly 90% of the revertants arose by gene conversion between *pma1* and *PMA2*. The gene conversion tracts were continuous and varied in size from 18 to 774 bp, with an average of 237 bp. Their boundaries were not located in the regions of greatest identity and consisted of only short sequences (3 to 26 bp) common to the two genes. However, a bias in the selection of recombination events was introduced by Harris et al. (8), since they selected only those gene conversions that could restore the wild-type *PMA1* phenotype. Moreover, their analysis was restricted to the central part of the *PMA* genes, i.e., the region of highest identity. We have used a different approach in examining the mechanisms of ectopic recombination between the *PMA1* and *PMA2* genes. Our experimental system involves a combination of previously described procedures (14, 15) and is based on the capacity of the yeast to repair gapped linearized plasmids by using information from cotransformed homologous (or homeologous) DNA fragments. In addition, an "in-frame controller" gene was used to select the in-frame recombinants. Thirty-five chimeras generated between the N-terminal region of *PMA1* and the C-terminal region of *PMA2* were subjected to restriction analysis and sequenced. Our data support and extend previous reports on *PMA1* and *PMA2* (8) and cytochrome P-450 cDNAs (15). The chimeras that we obtained corresponded to simple DNA exchanges between *PMA1* and *PMA2* not necessarily occurring in regions of highest identity. Our work also provides new insights into the mechanisms of recombination initiated by a double-strand break and reveals a phenomenon of clustering in the distribution of the junction sequences of the *PMA1*::*PMA2* fusions. Moreover, our procedure can generate an almost indefinite number of fusions which are amenable to biochemical studies, as detailed elsewhere (4).

* Corresponding author. Mailing address: Unité de Biochimie Physiologique, Université Catholique de Louvain, Place Croix du Sud 2-20, B-1348 Louvain-La-Neuve, Belgium. Phone: (32-10)474691. Fax: (32-10)473872.

† Present address: Laboratoire de Microbiologie Génétique et Moléculaire, Institut Pasteur de Lille, F-59019 Lille, France.

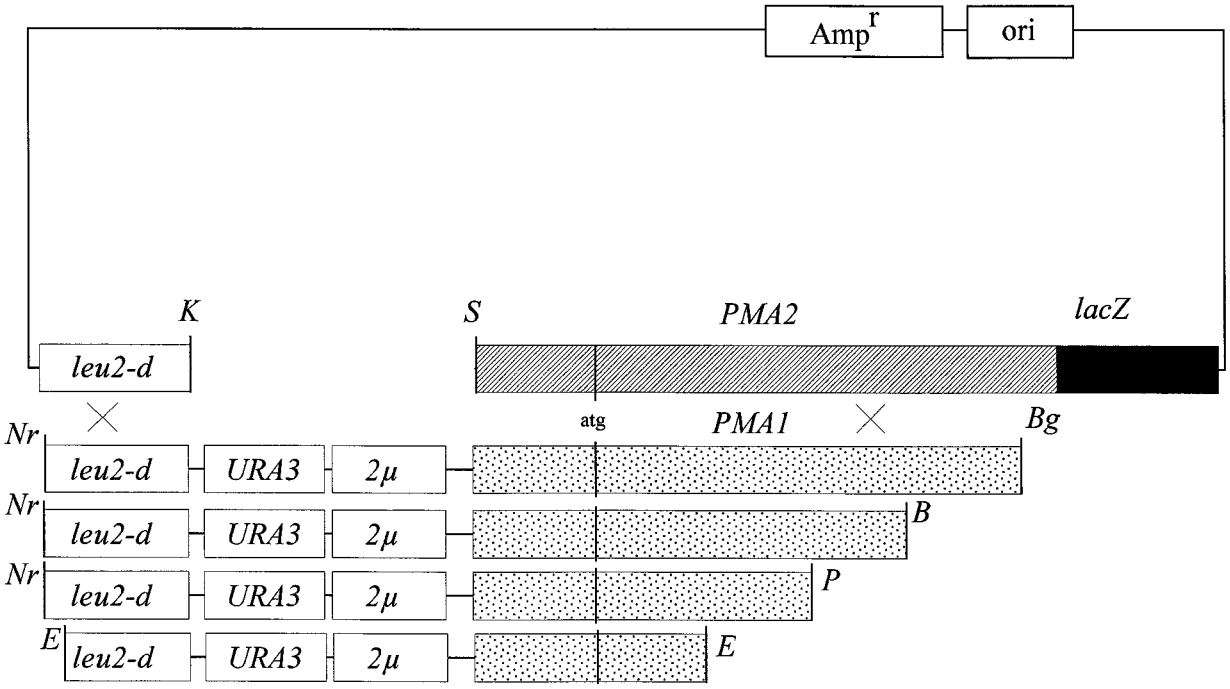

FIG. 1. Generation of *PMA1*::*PMA2* chimeric genes by homeologous intermolecular recombination. Recombination assays were carried out by cotransforming a linearized *PMA2-lacZ*-bearing plasmid derived from plasmid pPSUP2L1 and *PMA1* donor sequences derived from plasmid pPSUP1L. The gapped linearized plasmid was produced from pSPSUP2L1 by removal of the *Kpn*I-*Sal*I fragment containing *URA3* and the 2μm replication origin. ori, pBR322 origin of replication; Amp^r, ampicillin resistance gene; B, *Bam*HI; Bg, *Bgl*II; E, *Eco*RI; Nr, *Nru*I; P, *Pvu*II.

MATERIALS AND METHODS

Strains and media. Haploid strains YPS1-5A and YPS9-7A and diploid strain YPS2-2 have been described previously (23). Strain W303-1B is from R. J. Rothstein (Columbia University, New York, N.Y.). Yeast strains were grown on synthetic glucose medium as previously reported (23). When required, 40 μg of 5-bromo-4-chloro-indolyl-β-D-galactoside (X-Gal) per ml was added to the synthetic medium, which was buffered at pH 7.0 with 0.1 M sodium phosphate.

Plasmid construction. Plasmid pPSUP1L has been described previously (23). The construction of plasmid pPSUP2L1 and chimeric genes has been reported in detail (4). Briefly, pPSUP2L1 was obtained from a tripartite ligation as follows. Plasmid pPTZ-PMA2 (20) was digested with *Bam*HI, which cuts in the polylinker of pTZ, and with *Eco*RV, which cuts at position +2240 of the *PMA2* coding sequence, to isolate a fragment containing the *PMA2* promoter and two-thirds of the *PMA2* coding region. The two other fragments were derived from the 120-U3 vector (12), which contains the *lacZ* gene minus its first ATG. The *Sma*I-*Sac*I fragment contains the 5′ half of the *lacZ* coding region, while the *Sac*I-*Bam*HI fragment contains the 3′ half of the *lacZ* coding region and the rest of p120U3, including the *URA3* and *leu2-d* markers as well as the 2μm replication origin. Ligation of the *Eco*RV end of *PMA2* and the *Sma*I end located upstream of the *lacZ* coding region produces an in-frame fusion between the *lacZ* and *PMA2* genes.

Introduction of mismatches near the extremity of the *Nru*I-*Bgl*II donor. The region upstream of the *Bgl*II site (position +1975 of the *PMA1* coding sequence) in plasmid pECPTZ-PMA1 (23) was subjected to site-directed mutagenesis with the nucleotide 5′GTCGACAATGCGGATGATGCTGCCAGATCTGCTG, which generates a new *Sal*I site (underlined) in the *PMA1* sequence. The mutated plasmids were identified on the basis of this new *Sal*I restriction site and verified by sequence analysis. The mutated *Pvu*II-*Bgl*II fragment (positions 1018 to 1975 of the coding sequence) was substituted for its wild-type homolog in plasmid pPSUP1L.

Yeast transformation and plasmid extraction. The strains were transformed by the lithium acetate method (11), and transformants were selected as Ura$^+$ cells. Each transformation was carried out in the presence of 50 μg of sonicated herring sperm DNA. Plasmids were extracted from *S. cerevisiae* as previously reported (14).

DNA sequencing. The plasmids were extracted from *S. cerevisiae* and amplified in *Escherichia coli*. For several, the region encompassing the junction sequence was cleaved with *Hin*dIII plus *Nru*I and subcloned in the sequencing pTZ18U vector cleaved with *Hin*dIII plus *Sma*I. Single-stranded DNA was sequenced by using the dideoxynucleotide chain termination procedure (2) and synthetic oligonucleotides. The other plasmids were directly sequenced as double-stranded DNA, using the same procedure.

Nomenclature. To simplify this report, the linearized plasmid derived from the pPSUP2L1 plasmid is referred to as the recipient and the linear DNA fragments containing *PMA1* derived from the pPSUPIL plasmid are referred to as donors; the use of these terms is explained in the Discussion. The donor sequences are named by using the names of the restriction enzymes (roman letters) from which they were derived (for example, *Eco*RI-*Eco*RI). The recombinant clones are indicated by the chimeric gene symbol *PMA12* followed by a number and two letters representing the donor sequence from which they were derived (for example, *PMA12*-6EE).

RESULTS

Construction of chimeric *PMA1-PMA2-lacZ* genes. Our strategy, derived from a previously described procedure (14), is based on the construction of a recipient gapped linearized plasmid and donor sequences with various lengths (Fig. 1). The linear plasmid used in this work contains the *leu2-d* selectable marker at one free end (*Kpn*I site) and the promoter and the first 747 codons of the *PMA2* gene fused to the *lacZ* reporter gene at the other free end (*Sal*I site). This plasmid lacks the 2μm replication origin and thus cannot replicate in *S. cerevisiae*. The donor sequences, which include the 2μm replication origin and the *URA3* marker, contain the *leu2-d* sequence at one free end (*Nru*I or *Eco*RI site) and various lengths of the *PMA1* coding sequence at the other. The *leu2-d* and *PMA1* genes are separated by a distance of about 2 kb. The *PMA1* sequences placed under the control of the *PMA1* promoter were obtained as follows. The *PMA1* gene was cleaved at increasing distances from the ATG start codon with *Eco*RI (+484), *Pvu*II (+1018), *Bam*HI (+1599), and *Bgl*II (+1975). The following donor fragments were generated: *Eco*RI-*Eco*RI, *Nru*I-*Pvu*II, *Nru*I-*Bam*HI, and *Nru*I-*Bgl*II, corresponding to *PMA12*-EE, -NP, -NB, and -NG recombinants, respectively. *S. cerevisiae* was cotransformed with the gapped plasmid and each of the four donor sequences. To generate an autonomously replicative plasmid bearing the *URA3* selectable marker, a

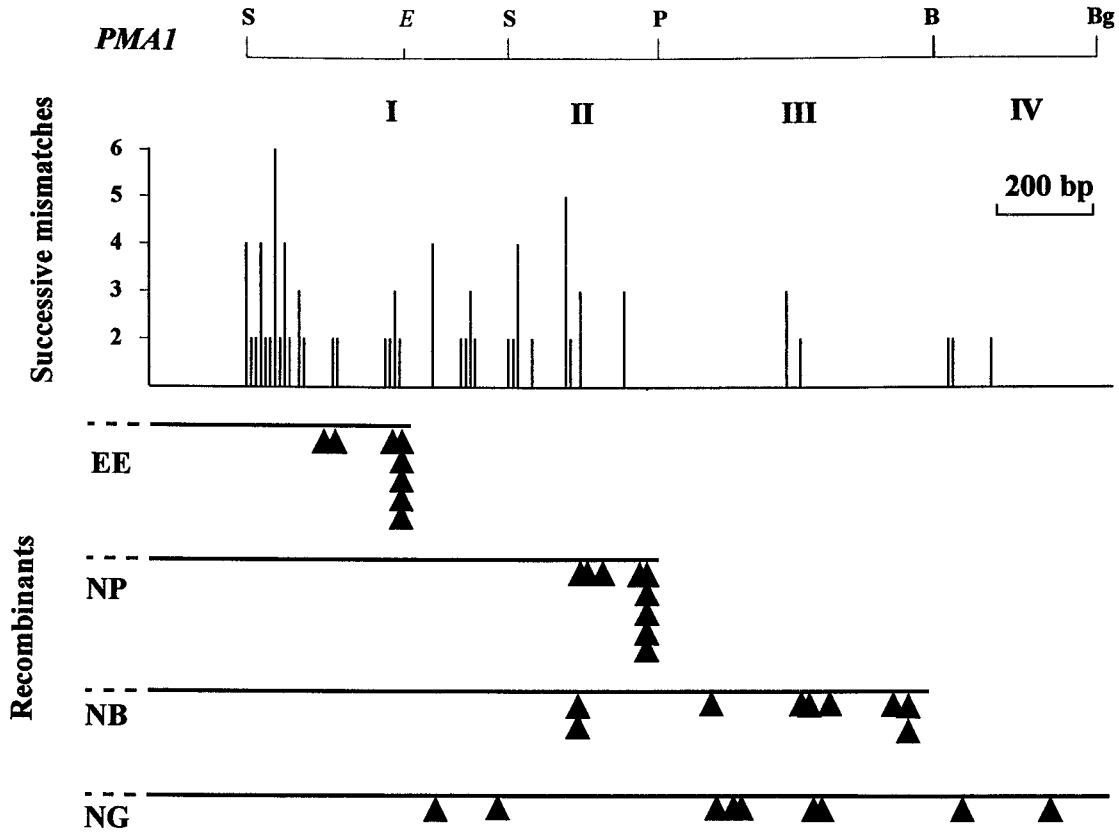

FIG. 2. Localization of the junctions in *PMA1*::*PMA2* chimeric genes. At the top are indicated the restriction sites in the *PMA1* gene used to delineate domains I, II, III, and IV and/or the ends of the *PMA1* gene. In the center, the groups of contiguous mismatches between *PMA1* and *PMA2* are indicated by vertical lines, whose lengths are proportional to the number of contiguous mismatches. Isolated mismatches and mismatches upstream of the first *Sal*I site (no recombinants found in this region) are not shown. The arrowheads at the bottom indicate junctions in the *PMA1* donor sequences. S, *Sal*I; Bg, *Bgl*II; B, *Bam*HI; P, *Pvu*II; E, *Eco*RI.

double crossover must occur between the identical *leu2-d* sequences on one side and the homeologous *PMA1* and *PMA2* sequences on the other side (Fig. 1). The identification of in-frame *PMA1*::*PMA2* fusions under the control of the strong *PMA1* promoter is facilitated by the presence of the *lacZ* reporter gene, which is fused in-frame with *PMA2*; because of β-galactosidase activity, the recombinant clones harboring an in-frame *PMA1*::*PMA2* fusion develop an intense dark blue color on X-Gal-containing medium. About 1,000 transformants of plasmid DNA per μg were obtained, 75% of which were dark blue. The white transformants were not analyzed further. No dark blue colonies were observed in the absence of the donor sequence or recipient plasmid.

Localization of the recombinant *PMA1*::*PMA2* junctions. Thirty-eight plasmids selected from the *PMA12*-EE, -NP, -NB, and -NG recombinants were analyzed with a combination of *Sal*I, *Bam*HI, *Pvu*II, and *Eco*RI restriction enzymes (Fig. 2). Thirty-five plasmids showed the structure expected for a simple exchange between recipient and donor sequences involving homeologous recombination between the coding sequences of *PMA1* and *PMA2*. Three plasmids with a more complex restriction pattern were not analyzed further. On the basis of restriction analysis, the *PMA1-PMA2* junctions were mapped to four contiguous DNA segments (I, II, III, and IV) delineated by the *Sal*I site on the 5′ side and a *Sal*I, *Pvu*II, *Bam*HI, or *Bgl*II site on the 3′ side of *PMA1* (Fig. 2). A region of 300 to 500 nucleotides encompassing the *PMA1-PMA2* junction was sequenced for each fusion. The sequencing data agreed perfectly with the restriction analysis results and showed that all chimeras correspond to a precise in-frame fusion between the *PMA* genes with a short junction sequence shared by the two genes. Our data support previous reports (8, 15) of the following. (i) The common boundaries are short (Table 1), on average consisting of 10 nucleotides, with six junctions being equal to or shorter than 5 nucleotides (*PMA12*-6EE, -16NP, -19NP, -24NB, -35NG, and -40NG) and four longer than 30 nucleotides (*PMA12*-22NB, -23NB, -30NB, and -39NG). (ii) The boundaries are not located at regions of greatest identity between *PMA1* and *PMA2* except in the case of recombinants *PMA12*-23NB, -30NB, and -39NG. (iii) Sequence analysis showed that the sequences are continuous on both sides of the junction, suggesting overall alignment of the parental sequences, which is in agreement with restriction analysis data. In other words, the chimeric sequences have not been subjected to multiple replacements but result from a simple unique DNA exchange.

The unequal distribution of the junctions is the most striking feature of our work. A highly clustered distribution of the junction sequences was seen in the *PMA12*-EE and -NP recombinants. Five of eight *PMA12*-EE recombinants and five of nine *PMA12*-NP recombinants contain identical junction sequences localized at the extremity of the donor sequence used in the cotransformation experiment (Table 1 and Fig. 2). Moreover, the average distances from the *PMA1* extremity were 47 (±70) and 63 (±53) nucleotides in the *PMA12*-EE and -NP recombinants, respectively. In contrast, in the *PMA12*-NB recombinants, the junctions were dispersed across a region of more than 500 nucleotides, at an average distance of 312

TABLE 1. Localization of junctions between *PMA1* and *PMA2* genes

PMA1 extremity (position[a])	Clone(s)[b]	Last *PMA1*-specific nucleotide[c]	No. of flanking identical nucleotides[d]	Distance from *PMA1* extremity[e]
*Eco*RI (484)	9	306	14	178
	6	342	5	142
	10	456	8	28
	2, 3, 4, 5, 7	478	6	6
*Pvu*II (1018)	19	867	5	140
	13	876	15	131
	15	909	20	109
	16	990	2	28
	12, 14, 17, 18, 20	993	25	25
*Bam*HI (1599)	27, 28	852	14	747
	24	1149	5	450
	29	1332	17	267
	25	1353	11	246
	21	1392	26	207
	22	1527	32	72
	23, 30	1563	36	36
*Bgl*II (1975)	31	567	8	1,408
	36	717	11	1,258
	35	1149	5	826
	37	1194	8	781
	32	1203	17	772
	34	1368	8	607
	40	1380	5	595
	33	1680	11	295
	39	1853	62	122

[a] The position of the *PMA1* extremity (in parentheses) is calculated from the first nucleotide of the *PMA1* coding sequence.

[b] The name of the clone is the number given in the column followed by two letters as defined in "Nomenclature" in Materials and Methods (e.g., 9-EE).

[c] The nucleotide position is calculated from the first nucleotide of the *PMA1* coding sequence.

[d] Number of nucleotides at the junction sequence that are identical between *PMA1* and *PMA2*. When the junction is located at the *PMA1* end of the donor sequence, the nucleotides removed by restriction are subtracted.

[e] Number of nucleotides between the *PMA1* extremity of the donor sequence and the junction of the *PMA* genes calculated from the last *PMA1*-specific nucleotide.

(±279) nucleotides from the *Bam*HI site. This dispersed distribution was even more pronounced in the *PMA12*-NG recombinants, which displayed regularly spaced junction sequences up to a distance of 1,400 nucleotides upstream of the *Bgl*II site (average distance, 740 [±409] nucleotides). Only two of the nine *PMA12*-NB junctions and none of the *PMA12*-NG junctions were located at the extremity of the donor sequence.

What features distinguish the *Eco*RI-*Eco*RI or *Nru*I-*Pvu*II and *Nru*I-*Bam*HI or *Nru*I-*Bgl*II donor sequences? The ends of all four donor sequences were delineated by a region of identity between *PMA1* and *PMA2*, which could be short (6 and 10 nucleotides for *Eco*RI-*Eco*RI and *Nru*I-*Bgl*II DNA fragments, respectively) or long (25 and 36 nucleotides for *Nru*I-*Pvu*II and *Nru*I-*Bam*HI DNA fragments, respectively) (Table 1). However, the *PMA1-PMA2* sequences corresponding to the *Eco*RI-*Eco*RI and *Nru*I-*Pvu*II donors exhibited a much higher degree of heterology over their entire sequences than the sequences corresponding to the *Nru*I-*Bam*HI and *Nru*I-*Bgl*II donors. The level of homology is also much lower near the *PMA1* extremity, with a series of closely spaced mismatches. For instance, the 33-nucleotide *PMA1* tract immediately upstream of the boundary common to *PMA1* and *PMA2* at the *Eco*RI site was interrupted by 13 mismatches (six groups of one, two, or three contiguous mismatches) (Fig. 2 and 3) and had 61% sequence identity with its *PMA2* homolog. The overall identity of the 200 nucleotides upstream of the *Eco*RI site was 82%. Moreover, the first 260 nucleotides of the coding sequences of the *PMA* genes showed only 38% identity and contained numerous deletions or insertions. Homeology at the *Pvu*II site was only slightly better, with 74% identity of the 46 nucleotides upstream of the common boundary (nine groups of one mismatch and one tract of three contiguous mismatches). The next 200 nucleotides upstream showed an overall identity of 78%. In contrast, immediately upstream of the common boundary at the *Bam*HI and *Bgl*II sites, there were only five mismatches for 51 nucleotides (90% identity) and two mismatches for 49 nucleotides (96.5% identity), respectively (Fig. 2 and 3). The overall identities within the first 200 nucleotides upstream of the boundary at the *Bam*HI and *Bgl*II sites were 92.5 and 97.5%, respectively. These data strongly suggest that there is a direct relationship between the clustered distribution of the junctions and the heterology observed in the region upstream of the common boundary. In contrast, to produce recombinants *PMA12*-31NG and -36NG, the heteroduplex must have passed barriers of four, or even five, contiguous mismatches (positions 732 to 735 and 828 to 832 in *PMA1*), as well as a 34-nucleotide region of only 50% identity (positions 819 to 852). However, these stretches of highly divergent DNA are localized more than 1,000 nucleotides upstream of the donor end. The fact that it is possible to form recombinants such as *PMA12*-31NG or -36NG indicates that the requirement for a high degree of homology is less absolute in regions distant from the free end of the donor.

Introduction of closely spaced mismatches at the *PMA1*-bearing end of the *Nru*I-*Bgl*II donor. Our data suggest that the degree of sequence identity upstream of the donor end dictates the distribution of the junctions. In order to test this hypothesis, seven closely spaced mismatches (two groups of four and three adjacent mismatches, respectively, separated by three identical nucleotides) were generated by oligonucleotide site-directed mutagenesis of the *Nru*I-*Bgl*II donor, immediately upstream of the *PMA1-PMA2* boundary (Fig. 4A). The gapped plasmid was cotransformed with the wild-type and mutated donors, and about 80% blue colonies were obtained for each transformation. The junctions of 19 correctly in-frame recombinants from each transformation were determined by restriction analysis, and the precise locations of 24 (13 for the wild type and 11 for the mutated donor) within the *Bam*HI-*Bgl*II fragment were determined by DNA sequencing. While only one junction was found at the extremity of the wild-type donor, about 40% (8 of 19) were clustered downstream of the *Sal*I site included in the mismatches for the mutated donor. Sequencing of six of these eight junctions showed that they were located at the *PMA1-PMA2* boundary (Fig. 4B). The analysis of the proportion of clustered and unclustered junctions by Fisher's exact test gave a value of P equal to 0.013, showing that the mutant and wild-type distributions are significantly different. It must be stressed, however, that the percentage of clustered junctions for the mutant donor is probably underestimated. Indeed, since the mismatches have been generated less than 15 nucleotides upstream of the donor end, a significant percentage of the mismatched extremities might have been eliminated during transformation by the action of cellular exonucleases. A similar degradation process has also certainly shortened a fraction of the ends of the *Eco*RI-*Eco*RI and *Nru*I-*Pvu*II donors; however, because of the high level of heterology upstream of the common boundary between *PMA1* and *PMA2*, recombination

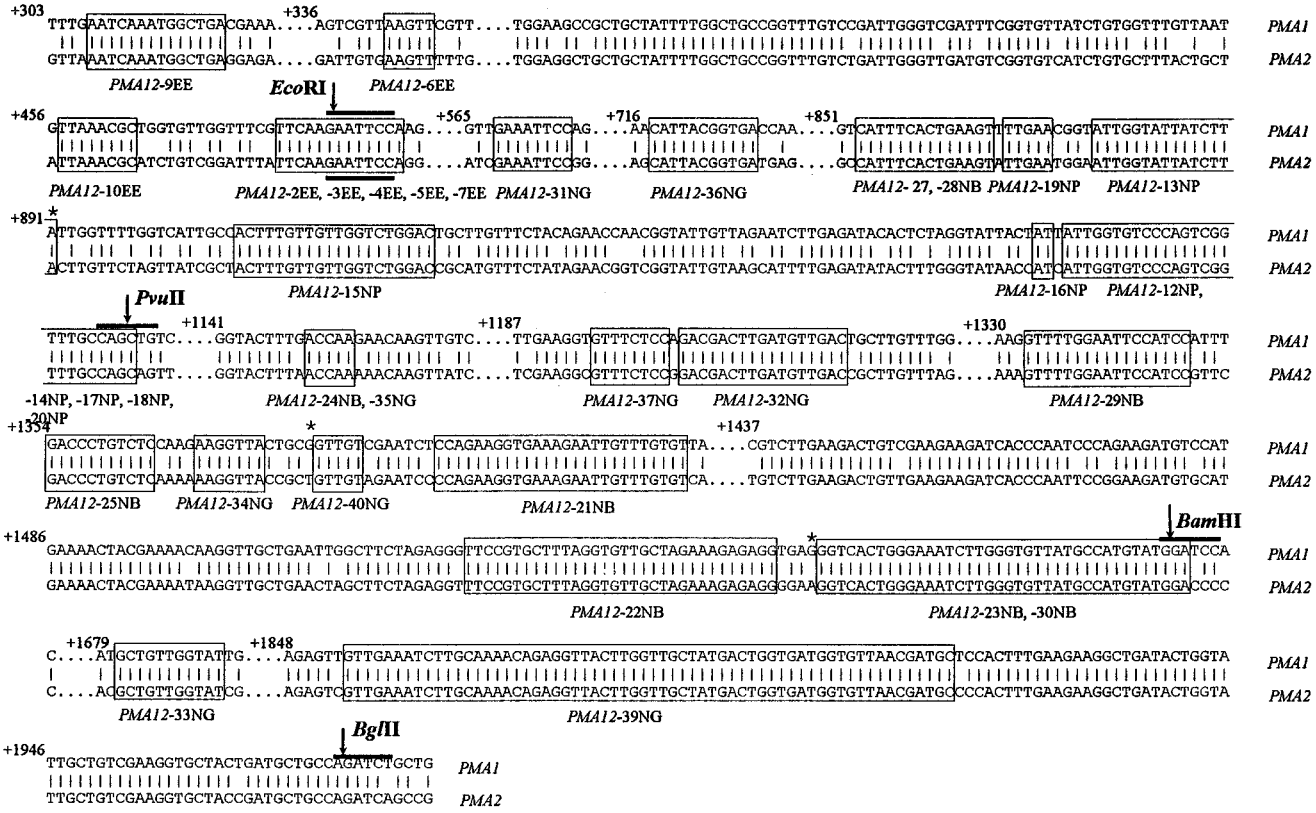

FIG. 3. Alignment of *PMA1* and *PMA2* coding regions (20, 22, 25) and localization of the junction sequences in chimeras. *PMA1* and *PMA2* sequences were aligned by using the NALIGN program (16). Only relevant regions are shown. The *PMA1* sequence (upper sequence) is numbered from the first ATG (+1). Identical residues are indicated by vertical bars. Horizontal bars and arrows above the *PMA1* sequence indicate relevant restriction and cleavage sites, respectively. The recombinant junction sequences are boxed. Three nucleotide substitutions (indicated by a star) were found at positions +891, +1380, and +1563 between the *PMA1* sequences from strains IL125-2B (this study) and Σ1278b (25). The first substitution (G to A) was also found in strain S288C (22), while the other two (C to G and A to G) are specific to strain IL125-2B; all three are silent.

events involving shortened ends of the *Eco*RI-*Eco*RI and *Nru*I-*Pvu*II donors are expected to be rare (*PMA12*-10EE might be such a recombinant).

DISCUSSION

We have analyzed 35 chimeras generated by in vivo recombination between the homeologous *PMA1* and *PMA2* genes. The mapping of the junction sequences confirms and extends the conclusions drawn from ectopic recombination (8). In addition, our work gives new insights into the mechanism of recombination initiated by a double-strand break.

A new experimental approach. Two types of approaches have been used to study recombination between homeologous genes. One takes advantage of the possibility of selecting recombination events that occur spontaneously between nearly identical chromosomal sequences at ectopic locations in both mitotic and meiotic cells. As early as 1981, it was shown that the *cyc1-11* mutation reverts by recombination with the nonallelic *cyc7* gene encoding iso2-cytochrome *c* (5). More recently, it has been found that almost 90% of the revertants of the *pma1-105* mutation arise by gene conversion between the *PMA1* and *PMA2* genes (8). Although this approach is of great interest, all conversion events which generate a nonfunctional hybrid protein are eliminated. A second method producing plasmid-borne chimeric genes in vitro is based on intermolecular recombination occurring in vivo during yeast cotransformation with two linear DNAs bearing homeologous sequences. This approach was first successfully applied to the production of hybrids between mammalian cytochrome P-450 cDNAs having 73% identity and, less efficiently, between the yeast *ARG4* gene and its human cDNA homolog, which have 52% identity (15). Hybrids between mitochondrial RNA maturase and DNA endonuclease have also been produced by this method (7). Our experimental system, which is derived from the second method, is original in several respects and, most importantly, it makes it possible to define both the recipient plasmid and the donor sequence (see below). We have designed donor sequences in which one free end is identical to one end of the recipient plasmid, while the second free end (*PMA1*) is homeologous to an internal region (*PMA2*) of the recipient plasmid. The donor homeologous sequences are variable in length but are designed in such a way that analysis of the recombination events involved in the production of the observed junctions is simplified. In addition, the in-frame controller *lacZ* gene allows their direct selection as dark blue colonies bearing in-frame *PMA1*::*PMA2* fusions. Even though illegitimate recombination events are rare for highly homologous genes, their proportion rapidly increases when identity is less than 50% (15). The isolation of the correct chimeras is thus facilitated in our system.

Clustering in linear distribution of junction sequences. *PMA1*::*PMA2* hybrids share traits with the *pma1* revertants generated by ectopic recombination and with the P-450 cDNA hybrids. The chimeric genes result from a simple DNA exchange involving fusion between continuous sequences specific

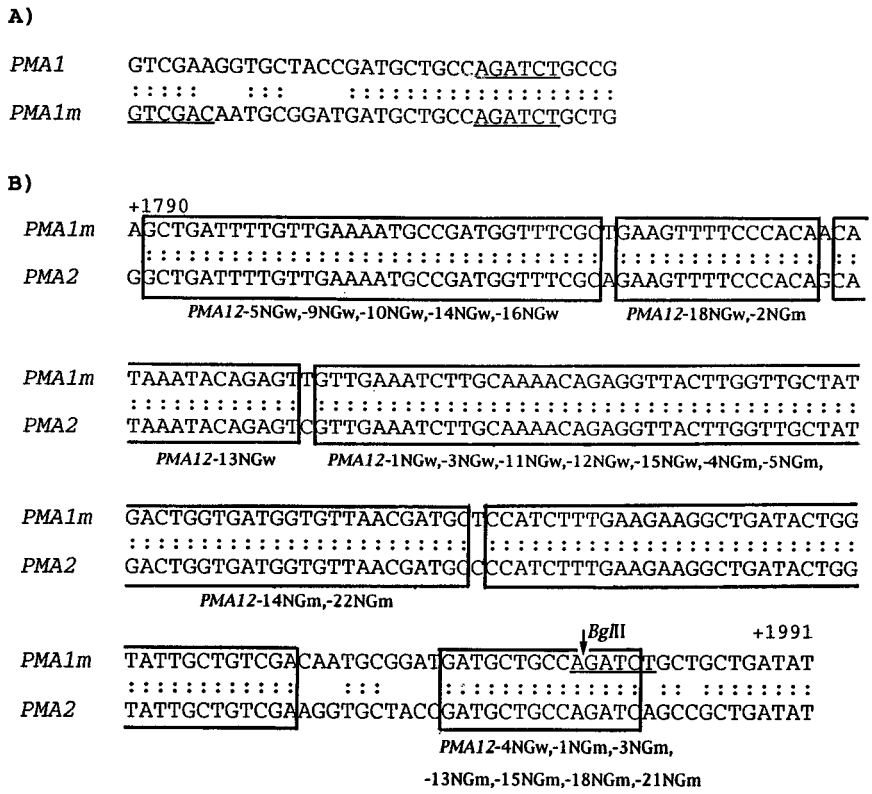

FIG. 4. Effect of consecutive mismatches at the extremity of the *PMA1*-bearing end on the position of junctions. (A) Localization of the mismatches at the *PMA1*-bearing end of the *Nru*I-*Bgl*II donor. The *Sal*I site (on the left), generated by site-directed mutagenesis, and the *Bgl*II site (on the right) are underlined. (B) Localization of the junctions in the *Bam*HI-*Bgl*II fragment. *PMA12*-NGw and *PMA12*-NGm are recombinants issued from transformations with wild-type and mutated *PMA1*-bearing donors, respectively. Alignments were done with the mutated *PMA1* fragment. The recombinant junctions are boxed. Colons indicate identical nucleotides.

to each gene, which are separated by short DNA stretches identical in the two homeologous genes. The junction sequences between the recombined genes are generally not found in regions of highest identity (recombinants *PMA12*-23NB, -30NB, and -39NG in this study being the only exceptions). However, as previously emphasized (8, 15), junction sequences do not necessarily correspond to the initiation site of heteroduplex formation. The final recombination products result from both heteroduplex formation and repair synthesis.

The most original finding of our study is the clustered distribution of the junction sequences along the recombination domain. Junction frequency is much greater near the end of the donor fragment in two recombinant families (*PMA12*-EE and -NP). This clustered distribution is similar to that of exchanges of exogenous DNA in *Xenopus* oocytes, which are concentrated near the sites of the original molecular ends (3, 9). In this case, recombination has been shown to occur through single-strand annealing, a process requiring pairing of a single-stranded 3′ tail with a complementary single strand followed by repair synthesis by DNA polymerase and ligation (13). Single-strand annealing is also very efficient in *S. cerevisiae* (6, 19). However, the observation that many junctions in *PMA12*-NB and -NG recombinants occur at a large distance from the *PMA1*-bearing end, as well as the existence of two types of clustering among the recombinants, is not readily explained by this model. Analysis of our results is thus based on the double-strand break repair model (24) with a pair of simple crossovers as in Fig. 1A.

Poles of recombination near the end of *PMA1* may have been generated in our experimental system by the asymmetry in homology between the ends of the DNA fragments bearing *PMA1* and *PMA2*. Indeed, the free extremity of the DNA bearing *PMA1* is delimited by regions of identity with *PMA2*, while, in the gapped linearized plasmid, the free extremity is occupied over 1,000 bp by the *PMA2* promoter that bears no homology to *PMA1*; strand invasion of the gapped linearized plasmid on the *PMA2* side should therefore be promoted by the *PMA1*-bearing ends. As a consequence, the *PMA1*-bearing DNA fragment must be considered the donor and the gapped plasmid must be considered the recipient for the genetic information. Thus, in our experimental system, recombination might be initiated at fixed sites dictated by the sequence of the free donor end. In the system based on intermolecular recombination of homeologous P-450 cDNAs (17), most junction sequences (17 of 20) are also found within a very small region of 100 bp around the break of the linearized gapped plasmid (15), also suggesting clustering in the production of junction sequences; again, the existence of an asymmetric homology at the DNA ends of the gapped duplex and cotransforming DNA fragment suggests that clustering might result from initiation of recombination at fixed sites.

High- and low-level clustering. Depending on the donor sequence, two levels of clustering, high and low, are found. High-level clustering is characterized by concentration of the junctions at the stretch of identical DNA in *PMA2* and the *PMA1*-bearing extremity of the donor sequence. Low-level clustering is characterized by a dispersed distribution of the junctions over a distance of more than 1,000 bp. The degree of

clustering does not correlate with the length of the stretch of identity at the boundary of the fusions. However, it seems to be directly related to the level of identity between *PMA1* and *PMA2* in the DNA segment immediately upstream of the stretch of identity. In the case of the *Eco*RI-*Eco*RI and *Nru*I-*Pvu*II donors, these segments have 60 to 75% identity over a 30- to 50-nucleotide sequence and contain several consecutive mismatches, while in the case of the *Nru*I-*Bam*HI and *Nru*I-*Bgl*II donors, they are more than 90% identical, with only isolated mismatches. Moreover, the introduction of seven closely spaced mismatches in the *Nru*I-*Bgl*II donor immediately upstream of the stretch of identity between *PMA1* and *PMA2* results in a significantly increased clustered distribution of the junctions at the extremity of the donor fragment, as observed with the *Eco*RI-*Eco*RI and *Nru*I-*Pvu*II donors. Thus, a low level of identity in the vicinity of the boundary stretch is a strong barrier to recombination. Our data, however, also suggest that in contrast to the initiation step, once the heteroduplex has been established over several hundreds of nucleotides, the occurrence of isolated groups of two, three, or even five consecutive mismatches as well as regions of overall moderate identity does not necessarily block recombination.

The position of the apparent crossover junction is a complex function of the extent of heteroduplex formation and subsequent mismatch repair. The mismatch repair machinery, several components of which have been conserved through evolution from bacteria to humans, constitutes a block to homeologous recombination in prokaryotes (18). This block may result from the destruction of the homeologous duplex by the complex excision repair reaction. However, it has been shown recently that heteroduplex binding in vitro by MutS and MutL proteins leads to termination of the branch migration step of RecA-catalyzed strand transfer, suggesting that the MutS-MutL complex promotes the disassembly of the homeologous duplex (26). These two processes are not mutually exclusive. A similar control of homeologous recombination may also exist in *S. cerevisiae*. It has been shown that the *MSH2* and *MSH3* gene products significantly decrease homeologous recombination (21), and it has been proposed that mismatch repair proteins regulate the length of symmetric heteroduplex DNA through specific interactions with recombination enzymes (1). In other words, *MSH2* and *MSH3* gene products would block branch migration during homeologous recombination. These hypotheses are supported by the observation that in spontaneous conversion tracts replacing a *pma1* mutation with homeologous *PMA2* sequence, the average length of the recombination tracts is significantly reduced in comparison with the several thousands of nucleotides covered by conversion tracts of quasi-identical sequences (10). Similarly, our own data show that even in the case of the *PMA12*-NG recombinant junctions, which have the least clustered distribution, the average length of the conversion tracts is only 740 nucleotides; this distance is reduced to 312 nucleotides for the *PMA12*-NB recombinants and to 64 and 48 nucleotides for the clustered recombinants *PMA12*-NP and *PMA12*-EE, respectively. It would be of great interest to determine whether the clustering phenomenon described in this report is dependent on the mismatch repair genes *MSH2* and *MSH3*.

ACKNOWLEDGMENTS

We gratefully acknowledge James E. Haber's help in the critical reading of the manuscript and Benoit Moreau for his assistance in statistical analysis.

This work was supported by grants from the Services de la Politique Scientifique: Action Sciences de la Vie and the Fonds National de la Recherche Scientifique, Belgium. P.S. and A.D.K.D. were supported by fellowships from the Institut pour l'Encouragement de la Recherche Scientifique. F.F. is a Senior Research Associate of the Fonds National de la Recherche Scientifique Belge.

REFERENCES

1. **Alani, E., R. A. G. Reenan, and R. D. Kolodner.** 1994. Interaction between mismatch repair and genetic recombination in *Saccharomyces cerevisiae*. Genetics **137:**19–39.
2. **Biggin, M. D., T. J. Gibson, and G. F. Hong.** 1983. Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. Proc. Natl. Acad. Sci. USA **80:**3963–3965.
3. **Carroll, D., C. W. Lehman, S. Jeong-Yu, P. Dohrman, R. J. Dawson, and J. K. Trautman.** 1994. Distribution of exchanges upon homologous recombination of exogenous DNA in *Xenopus laevis* oocytes. Genetics **138:**445–457.
4. **de Kerchove d'Exaerde, A., P. Supply, and A. Goffeau.** 1995. Generation of chimeric proteins by *in vivo* homologous gene recombination in yeast: example of an essential P-type ATPase. *In* K. W. Adolph (ed.), Molecular microbiology techniques, in press. Academic Press, San Diego, Calif.
5. **Ernst, J. F., J. W. Steward, and F. Sherman.** 1981. The *cyc1-11* mutation in yeast reverts by recombination with non-allelic gene: composite genes determining the iso-cytochrome *c*. Proc. Natl. Acad. Sci. USA **78:**6334–6338.
6. **Fishman-Lobell, J., N. Rudin, and J. E. Haber.** 1992. Two alternative pathways of double-strand break repair that are kinetically separable and independently modulated. Mol. Cell. Biol. **12:**1292–1303.
7. **Goguel, V., A. Delahodde, and C. Jacq.** 1992. Connections between RNA splicing and DNA intron mobility in yeast mitochondria: RNA maturase and DNA endonuclease switching experiments. Mol. Cell. Biol. **12:**696–705.
8. **Harris, S., K. S. Rudnicki, and J. E. Haber.** 1993. Gene conversions and crossing over during homologous and homeologous ectopic recombination in *Saccharomyces cerevisiae*. Genetics **135:**5–16.
9. **Jeong-Yu, S., and D. Carroll.** 1992. Effect of terminal nonhomologies on homologous recombination in *Xenopus laevis* oocytes. Mol. Cell. Biol. **12:** 5426–5437.
10. **Judd, S. R., and T. D. Petes.** 1988. Physical lengths of meiotic and mitotic gene conversion tracts in *Saccharomyces cerevisiae*. Genetics **118:**401–410.
11. **Kuo, C.-L., and J. L. Campbell.** 1983. Cloning of *Saccharomyces cerevisiae* DNA replication genes: isolation of the *CDC8* gene and two genes that compensate for the *cdc8-1* mutation. Mol. Cell. Biol. **3:**1730–1737.
12. **Lahaye, A., H. Stahl, D. Thinès-Sempoux, and F. Foury.** 1991. PIF1: a DNA helicase in yeast mitochondria. EMBO J. **10:**997–1007.
13. **Lin, F.-L., K. Sperle, and N. Sternberg.** 1984. Model for homologous recombination during transfer of DNA into mouse L cells: role for DNA ends in the recombination process. Mol. Cell. Biol. **4:**1020–1034.
14. **Ma, H., S. Kunes, P. J. Schatz, and D. Botstein.** 1987. Plasmid construction by homologous recombination in yeast. Gene **58:**201–216.
15. **Mézard, C., D. Pompon, and A. Nicolas.** 1992. Recombination between similar but not identical DNA sequences during yeast transformation occurs within short stretches of identity. Cell **70:**659–670.
16. **Myers, E. W., and W. Miller.** 1988. Optimal alignments in linear space. Comput. Appl. Biosci. **4:**11–17.
17. **Pompon, D., and A. Nicolas.** 1989. Protein engineering by cDNA recombination in yeasts: shuffling of mammalian cytochrome P450 functions. Gene **83:**15–24.
18. **Rayssiguier, C., D. S. Thaler, and M. Radman.** 1989. Interspecific recombination between *Escherichia coli* and *Salmonella typhimurium* is disrupted in mismatch-repair mutants. Nature (London) **342:**396–401.
19. **Rudin, N., E. Sugarman, and J. E. Haber.** 1989. Genetic and physical analysis of double-strand break repair and recombination in *Saccharomyces cerevisiae*. Genetics **122:**514–534.
20. **Schlesser, A., S. Ulazewski, M. Ghislain, and A. Goffeau.** 1988. A second transport ATPase gene in *Saccharomyces cerevisiae*. J. Biol. Chem. **263:** 19840–19847.
21. **Selva, E. M., L. New, G. F. Crouse, and R. S. Lahue.** 1995. Mismatch correction acts as a barrier to homeologous recombination in Saccharomyces cerevisiae. Genetics **139:**1175–1188.
22. **Serrano, R., M. C. Kielland-Brandt, and G. R. Fink.** 1986. Yeast plasma membrane ATPase is essential for growth and has homology with (Na^+ + K^+) and Ca^{2+}-ATPases. Nature (London) **319:**689–693.
23. **Supply, P., A. Wach, D. Thinès-Sempoux, and A. Goffeau.** 1993. Proliferation of intracellular structures upon overexpression of the PMA2 ATPase in *Saccharomyces cerevisiae*. J. Biol. Chem. **268:**19744–19752.
24. **Szostak, J. W., T. L. Orr-Weaver, and R. J. Rothstein.** 1983. The double-strand break repair model for recombination. Cell **33:**25–35.
25. **Van Dyck, L., J. H. Petretski, H. Wolosker, G. Rodrigues, A. Schlesser, M. Ghislain, and A. Goffeau.** 1990. Molecular and biochemical characterization of the Dio-9-resistant *pma1-1* mutation of the H^+-ATPase from *Saccharomyces cerevisiae*. Eur. J. Biochem. **194:**785–790.
26. **Worth, L., S. Clark, M. Radman, and P. Modrich.** 1994. Mismatch repair proteins MutS and MutL inhibit RecA-catalyzed strand transfer between diverged DNAs. Proc. Natl. Acad. Sci. USA **91:**3238–3241.