MOLECULAR AND CELLULAR BIOLOGY, Feb. 1994, p. 888–895
0270-7306/94/$04.00+0
Copyright © 1994, American Society for Microbiology
Vol. 14, No. 2

Mechanisms of Overlap Formation in Nonhomologous DNA End Joining

PETRA PFEIFFER,[1*] SILKE THODE,[2] JULIANE HANCKE,[1] AND WALTER VIELMETTER[1]

Institut für Genetik der Universität zu Köln, D-50931 Cologne, Federal Republic of Germany,[1] and Department of Neuroscience, University of California San Diego, La Jolla, California 92093-0624[2]

Received 5 August 1993/Returned for modification 5 October 1993/Accepted 5 November 1993

Rejoining of nonhomologous DNA termini plays a central role in processes of illegitimate recombination. In *Xenopus* egg extracts, DNA ends with noncomplementary 4-nucleotide antiparallel single-strand protrusions are assumed to be joined by formation of short mismatched overlap intermediates. The extents of these overlaps may be set by single fortuitously matching base pairs and determine the patterns of subsequent gap filling and nick ligation. Under conditions of alternative overlap settings, rules for the most probable joining pathway and the effects of mismatches on junction formation were analyzed. We show that in certain cases, fill-in and ligation converting overlap intermediates into covalently closed junctions may proceed in the presence of unrepaired mismatches, whereas in other cases, completion of junction formation is preceded by removal of mismatches. Results are discussed in relation with "alignment" proteins postulated to structurally support overlap heteroduplexes during junction formation.

Eucaryotic cells share DNA repair mechanisms that can join unrelated ends of broken duplex DNA (21). On the molecular level, these end-joining mechanisms convert nonhomologous terminal DNA structures, often ending as protruding single strands (PSS), into seamless junctions by using reaction schemes necessarily more complex than mere ligation. These processes may be thought of as part of the repair pathways that eliminate highly lethal DNA double-strand breaks from genomic DNA (26) and sometimes generate chromosomal rearrangements (11). However, end-joining mechanisms are also involved in various processes of illegitimate recombination including VDJ joining of immunoglobulin genes (3, 9), integration of foreign (e.g., viral) DNA into chromosomes (2), and transposition events (1).

Inspection of junctions produced during illegitimate rearrangements of mammalian chromosomes provided the first indications that junctional sequences often contain base exchanges and nucleotide losses or gains within a narrow range of roughly 20 bp around junctional breakpoints. Furthermore, it was suspected that small patches of homology might contribute to junction formation (21, 23). These characteristic features found in naturally occurring junctions were verified and extended in studies employing artificial DNA substrates with defined nonhomologous restriction ends. Such substrates with various terminal configurations were offered either in vivo to cultured mammalian cells (13, 19, 20, 27) or fission yeast cells (4) or in vitro to extracts derived from human cells (14) or *Xenopus laevis* eggs (16). These widely differing systems yielded remarkably similar junctional patterns, suggesting the existence of comparable systems of nonhomologous DNA end joining, at least within the vertebrate domain.

In the *Xenopus* egg extract system, junctions form by two reaction pathways designated as fill-in and overlap modes. The mode of joining is determined by the terminal configurations of PSS. In the fill-in mode, PSS abutting blunt ends or pairs of oppositely directed PSS tails (blunt/PSS or 5′PSS/3′PSS) are consistently preserved by DNA fill-in synthesis (25) (Fig. 1). In the overlap mode, nonhomologous antiparallel PSS tails (5′PSS/5′PSS or 3′PSS/3′PSS) are assumed to transiently form short overlap heteroduplexes. The extent of overlap is defined by microhomology patterns of fortuitously matching base pairs (Fig. 1). In this case, gap filling and, to some extent, mismatch removal precede completion of junction formation by ligation (16).

Fill-in and degradation of PSS tails and final nick ligation, however, do not fully describe the capacity of joining systems. For instance, fill-in of 3′PSS tails, whose recessed 5′ ends cannot prime DNA synthesis, is primed at the partner terminus before the ends have been linked by ligation (Fig. 1). This finding led us to postulate terminal DNA-binding proteins that align juxtaposed partner termini (25). These "alignment" proteins probably play an equally indispensable role in the formation of overlap junctions, where they guarantee the proper settings of energetically weak single-base matches (16).

In this study, we present a detailed analysis of the joining patterns assigned to the overlap mode. Our first aim is to define the principles that determine the most probable overlap setting between two PSS tails under conditions of alternative base match formation. The second aim is to elucidate the problem of mismatch correction in overlap heteroduplexes of junctional intermediates. The results will be discussed in relation to alignment functions and their influence on the final step of ligation.

MATERIALS AND METHODS

Preparation of egg extracts. Extracts from Ca ionophore-stimulated *X. laevis* eggs were prepared by the modified method described recently (22).

Preparation of DNA substrates for nonhomologous DNA end joining. All DNA substrate types used in this study were derived from modified pSP65 plasmids (Promega) containing insertions of a 1.2-kb *Bam*HI-*Sal*I and a 1.8-kb *Sma*I-*Bam*HI fragment of lambda DNA cloned in the equivalent sites of

* Corresponding author. Mailing address: Institut für Genetik der Universität zu Köln, Weyertal 121, D-50931 Cologne, Federal Republic of Germany. Phone: 49-221-470-3403. Fax: 49-221-470-5170. Electronic mail address: PFEIFFER@GENETIK.UNI-KOELN.DE.

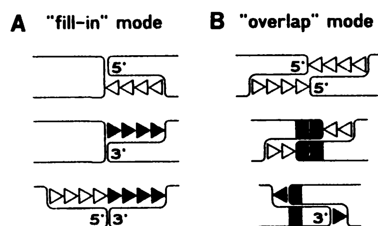


FIG. 1. Models for the modes of nonhomologous DNA end joining. (A) Abutting terminal configurations are joined by fill-in of PSS tails (open triangles mark filling in from the recessed 3′ end; solid triangles symbolize filling in primed at the unlinked partner terminus). (B) Pairs of antiparallel PSS may form short overlaps whose extent is set by single-base matches (solid boxes).

the original pSP65 polylinker. Preparation of 3-kb linear plasmid substrates carrying two nonhomologous restriction ends was controlled by complete excision of the diagnostic lambda insert from the polylinker by using two different restriction enzymes (Boehringer, Mannheim, Germany). To verify successful double digestion, we subjected gel-purified DNA substrates to a T4 DNA ligation assay in which truly nonhomologous substrate molecules cannot be recircularized (16). The spectrum of available unique restriction sites was extended by inserting additional self-complementary linkers (sequences are given in the 5′→3′ direction) into the *Sma*I or *Hin*dII site of the polylinker of the pSP65 derivatives: *Nco*I (GC/CATGGC), *Apa*I (CGGGCC/CG), *Kpn*I (C GGTAC/CG), *Asp*-718 (CG/GTACCG), and *Bst*XI [CCAN (N)$_4$/NTGG: CCATACGT/ATGG; linkers for nonmatching 3′PSS: CCATAATT/ATGG and CCATTTAA/ATGG]).

Cloning of joined products and analysis of junctional sequences. Cloning of circular head-to-tail-joined products in *Escherichia coli recA* strains HB101 and DH5α, miniscale preparation of clonal plasmid DNA, and sequencing of the 5′-3′ strands of cloned junctions with an SP6-specific primer by the dideoxy-chain termination method were performed as described previously (16).

Inhibition of the joining reaction by ddNTPs. Assay conditions for the end joining activity in the presence and absence of dideoxynucleoside triphosphates (ddNTPs) and the method used for visualization of reaction products by in situ hybridization of dried agarose gels were exactly as described previously (25).

Ligation assays with partially purified ligase. Partially purified ligase (PPL) from complete *Xenopus* egg extracts was prepared by the method of Thode et al. (25) with the following modifications. For dialysis of complete undiluted extract and equilibration of DEAE-cellulose columns, buffer A (10 mM Tris · HCl [pH 8.3], 6 mM magnesium acetate, 10% (vol/vol) glycerol, 5 mM β-mercaptoethanol, 0.5 mM phenylmethylsulfonyl fluoride, 1 μM pepstatin, 1 μM leupeptin) supplemented with 10 mM KCl was used. Fractions containing the total ligase activity of the extract, referred to as PPL, were eluted in a single step with 150 mM KCl in buffer A and subsequently dialyzed against reaction buffer B′ (50 mM Tris · HCl [pH 8.3], 50 mM KCl, 6 mM magnesium acetate, 5 mM β-mercaptoethanol, 0.5 mM phenylmethylsulfonyl fluoride, 1 μM pepstatin, 1 μM leupeptin). Ligation assays were carried out by the method of Thode et al. (25).

RESULTS

Experimental setup. The experiments reported here were performed in an attempt to acquire new information on DNA end-joining mechanisms in the overlap mode promoted in vitro by extracts derived from *X. laevis* eggs. For this, linear substrates with defined nonhomologous ends were generated by two different successive restriction cuts within the polylinker region of modified pSP65 plasmids. All chosen enzymes produced termini with 4-nucleotide (nt) PSS tails so that 17 substrates with 5′PSS/5′PSS and 9 with 3′PSS/3′PSS terminal configurations were generated (Fig. 2; see also Fig. 5).

Extract treatment converts such linear substrate molecules into various mono- and multimeric joined products. Since multimers usually result from ligation of complementary restriction ends, they are not further considered here. Circular monomers, however, which arise by self joining, contain genuine nonhomologously joined products (16). These may exist as nicked or gapped open circle (OC) intermediates or as covalently closed circles (CCC) which can be separated and visualized in gel assays (see Fig. 4).

Reaction products are cloned by transformation in *E. coli*, which processes monomeric circles with high preference to produce individual clones originating from single nonhomologously joined substrate molecules (16). Resulting clones were subjected to junctional sequence analysis. Sequence data derived from terminal configurations providing at least one base match are shown in Fig. 2; for those derived from nonmatching terminal configurations, see Fig. 5.

General features of junction formation in the overlap mode. Joining processes in the overlap mode are thought to pass through a state in which noncomplementary PSS tails overlap physically. Although overlap intermediates are not yet amenable to physical analysis, their existence may be deduced from characteristic differences between the original termini and the junctional sequences (16). Depending on sequence patterns of both involved PSS tails, these differences include (i) terminal-sequence shortening (base loss) strictly correlated to existing base match positions (ii) template-directed filling in of small gaps, and (iii) base pair substitutions related to traits of mismatch correction. These junctional features are not reconcilable with joining models of random and independent processing of PSS tails but, rather, are consistent with models of active overlap interaction between noncomplementary PSS tails.

Whether these overlap criteria hold true for a larger set of data and whether rules governing preferences for certain overlap settings can be established were tested by the analysis of 298 typical overlap junctions derived from 23 different match-containing terminal configurations (Fig. 2). For this study, the *Xenopus* in vitro system was particularly well suited because formation of deletions affecting the integrity of neighboring duplex regions is minimized under our reaction conditions used. Among a total of 361 analyzed junctions, only 15 deletions were found. The inherent limit of this system has the analytical advantage to greatly reduce the diversity of possible junctional configurations, often encountered in in vivo joining systems (8, 20, 27), to a comprehensible spectrum of DNA sequences which may be directly related to involved PSS sequences.

In theory, each PSS pair can overlap at four different positions. Thus, the 23 match-containing terminal configurations (Fig. 2) can produce 92 possible overlap settings, roughly half of which (45%) contain base matches. The breakpoints of all analyzed junctions exhibiting sequence

A) 5'PSS/5'PSS terminus configurations

Eco/Sal Σ 10
- G TCGAC / CTTAA G
- GAATTTCGAC 1
- GAATTCGAC 8
- GAAACGAC [T] 1

Sal/Hind Σ 15
- G AGCTT / CAGCT A
- GTCGAAGCTT 4
- GTCGAGCTT 11

Bam/Xba Σ 6
- G CTAGA / CCTAG T
- GGATCTAGA 6

Bam/Nco Σ 10
- G CATGG / CCTAG C
- GGATCATGG 9
- GGATCG [B] 1

Ia | Ib

Ava/Hind Σ 10
- C AGCTT / GGGCC A
- CCCGGAGCTT 8
- CCCAGCTT [T] 1
- CCAGCTT [T] 1

Nco/Sal Σ33
- C TCGAC / GGTAC G
- CCATGTCGAC 21
- CCATCGAC [T] 6
- CCAT(C/G)GAC [MM] 2
- CCATGAC [B] 4

Bam/Sal Σ35
- G TCGAC / CCTAG G
- GGATCTCGAC 5
- GGATCGAC 30

Bam/Hind Σ 12
- G AGCTT / CCTAG A
- GGATCAGCTT 1
- GGAGCTT [T] 6
- GGATCTT [B] 5

II | IIIa | IIIb

Xba/Sal Σ12
- T TCGAC / AGATC G
- TCTAGTCGAC 2
- TCT(C/A)GAC [MM] 10

Eco/Hind Σ 9
- G AGCTT / CTTAA A
- GAGCTT [T] 3
- GAATTT [B] 6

Ava/Xba Σ12
- C CTAGA / GGGCC T
- CCCGGCTAGA 2
- CCCTAGA [T] 1
- CCTAGA [T] 6
- CCCGGA [B] 3

Nco/Xba Σ 7
- C CTAGA / GGTAC T
- CCATGCTAGA 1
- CCATGA [B] 6

IIIb | IVa

Bam/Asp Σ20
- G GTACC / CCTAG G
- GGTACC [T] 3
- GG(A/T)ACC [MM] 14
- GGATCC [B] 3

Eco/Bam Σ 14
- G GATCC / CTTAA G
- GAATTGATCC 6
- GGATCC [T] 4
- GAATCC [MR] 1
- GAATTC [B] 3

Eco/Nco Σ 9
- G CATGG / CTTAA C
- GAATTCATGG 3
- GAATGG [MR] 6

Ava/Sal Σ 14
- C TCGAC / GGGCC G
- CCCGGAC [B] 1
- CTCGAC [T] 3
- C(C/T)CGAC [MM] 4
- CCCGAC [MR] 3
- CCCGGC [B] 3

IVa | IVb

B) 3'PSS/3'PSS terminus configurations

BstX/Pst Σ 7
- TACGT G / A ACGTC
- TACGTGCAG 6
- TACGCAG [MR] 1

Sac/Pst Σ 10
- GAGCT G / C ACGTC
- GAGCTG [T] 1
- GAGC(A/T)G [MM] 1
- GAGCAG [MR] 6
- GTGCAG [B] 2

Sac/Kpn Σ 15
- GAGCT C / C CATGG
- GAGTAGC [MR] 15

Kpn/Pst Σ 12
- GGTAC G / C ACGTC
- GGTACAG [T] 5
- GGTGCAG [B] 7

Ia | Ib | II | IIIb

Sac/BstX Σ 9
- GAGCT A / C TGCAT
- GAGCTA [T] 2
- GA(G/C)CTA [MM] 2
- GACGTA [B] 5

Kpn/Apa Σ 7
- GGTAC C / C CCGGG
- GGTACC [T] 3
- GGGCCC [B] 4

Sac/Apa Σ 10
- GAGCT C / C CCGGG
- GAGCTC [T] 2
- GGGCCC [B] 8

IVa | IVb

JUNCTION TYPES	5'PSS/5'PSS							3'PSS/3'PSS					
	Ia	Ib	II	IIIa	IIIb	IVa	IVb	Ia	Ib	II	IIIb	IVa	IVb
A	16%	0%	67%	14%	13%	6%	24%	0%					
B	81%	90%						85%	0%				
C	3%		33%			2%		15%		100%		0%	
D				86%							100%		
E					87%								
F		10%				92%						100%	
G							74%		100%				100%

FIG. 3. Diagram of the frequency distributions of junction types. These types are drawn in the form of overlap intermediates (A to G) along the ordinate and represent all types found by sequence analysis of cloned junctions derived from the 23 match-containing terminal configurations in Fig. 2. Junction types: A, fill-in type (zero overlap); B and C, types with base match singlets; D to G, types with base match doublets. Solid boxes indicate base matches, and hatched boxes indicate external mismatches in junction types C and G and internal mismatches in types E and F. The abscissa is separated into two blocks representing terminal configurations of the 5′ and 3′ classes. Blocks are further subdivided into subsets I to IV (Fig. 2). To give a more comprehensible survey of the competitive interrelations between different overlap types in each subset of substrates, frequency distributions are given in percentages, although the numbers of clones analyzed are sometimes only small. Percentages are based on the total numbers of clones found for the junction types formed by all substrates of each subset (Fig. 2).

shortening of 1 to 4 nt remarkably always coincided with base match positions in postulated overlaps, whereas not a single base-match-free "overlap" junction was detected (Fig. 2). This result rigorously confirms our previous assertion (16) that the extent of overlap during junction formation is set without exception by base matches and in some junction types even by a single base match.

Most of the terminal configurations produced more than one junction type. Configurations yielding similar patterns of frequency distributions were collected in subsets (I to IV). It is seen that one junction type strongly dominates within each subset over often rarely represented alternative types, which in turn may prevail in other subsets (Fig. 2). A clearer survey over the relative frequency distributions of all junction types found is presented in Fig. 3. Junction types are illustrated in the graphic insert as overlap intermediates exhibiting characteristic base match/mismatch patterns and are ordered by increasing lengths of overlaps (0 to 4 nt): 1- to 3-nt partial overlaps, which contain gaps (types B to E), and 4-nt complete overlaps, which contain nicks (types F and G). Owing to the palindromy of restriction sequences, only 8 of 13 theoretically possible base match-mismatch patterns are actually realized. The competitive interrelations between alternative overlap settings within each subset of substrates provide hints about which structural features of overlap patterns are relevant for the formation of a given junction.

Prevalence of overlap junctions over fill-in junctions. A particularly instructive case of competitive junction formation is observed between junctions of the overlap (Fig. 3, rows B to G) and zero-overlap type (Fig. 3, row A). We have designated the latter as the fill-in type because PSS tails are completely preserved by DNA fill-in synthesis and are thus converted into ligatable blunt ends. Among terminal configurations of the 5′ class, fill-in junctions have only a low (10 to 20%) chance of being formed in competition with overlap junctions. Solely in subset II do they occur more frequently than a special type of overlap junctions (Fig. 3, row C). Thus, the joining system strongly prefers the formation of comparatively complex overlap structures over seemingly simpler fill-in structures. This situation is drastically more pronounced for terminal configurations of the 3′ class, in which among 70 cloned junctions not a single example of the fill-in type was detected. Although free 3′PSS tails cannot be directly filled in for topological reasons, they are known to readily undergo filling in if they abut blunt or 5′PSS termini within junctional complexes (25). Apparently, this particular fill-in mechanism can rarely, if at all, prevail over the strong tendency to form overlaps between two antiparallel 3′PSS tails.

Competitive interrelations among overlap junctions. Frequency distributions of junctional sequences within a given subset of substrates and among different subsets reflect which overlap types are preferentially formed (Fig. 3). Such preferences may be related to the differing structural features of the corresponding overlap intermediates defined by the polarity of PSS tails as well as the number, type (G · C versus A · T), and position of base matches and mismatches. In particular, internal mismatches flanked by base matches on either side (types E and F) are expected to counteract the stabilizing forces of base matches whereas external mismatches located at PSS tips (types C and G) may in addition interfere with filling in and ligation during the processing of junctional intermediates.

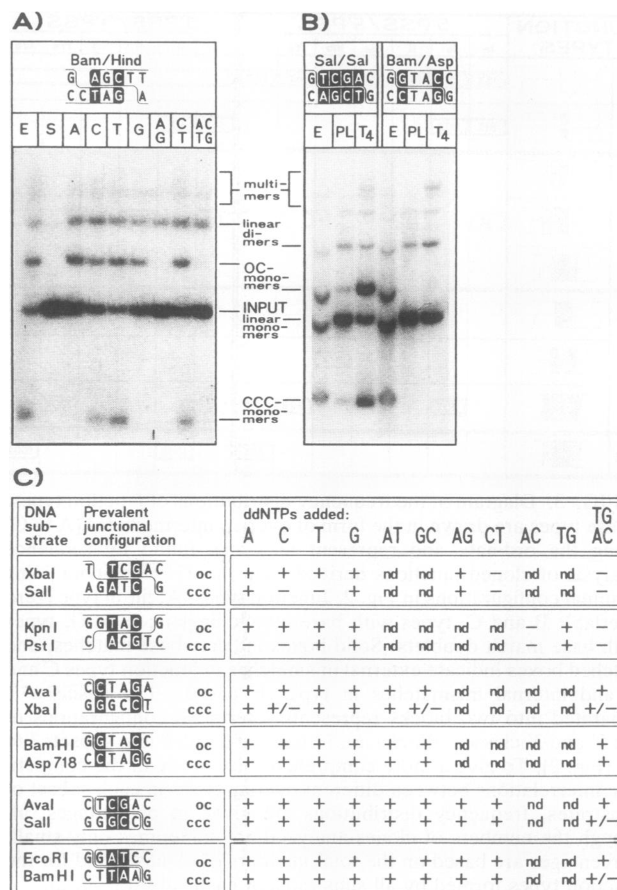
Inspection of the frequency distributions in Fig. 3 reveals that in fact the most distinct competition asymmetry is found

FIG. 2. List of the 16 5′PSS/5′PSS (A) and 7 3′PSS/3′PSS (B) match-containing terminal configurations. Restriction enzyme designations refer to the enzymes used to generate corresponding terminal configurations. Possible base matches arising by overlap interaction of PSS tails are emphasized as white-on-black letters (predominantly found), medallion-framed letters (rarely found), or open letters (not found). When two bases in the bottom strand share the same match partner in the top strand, only the two bases of the bottom strand are marked differentially. Terminal configurations with similar match properties are collected in subsets (I to VI; compare Fig. 3) as indicated. Junctional sequences derived from these terminal configurations are presented as the 5′→3′ (top) strand of cloned junctions between the dotted lines. Total numbers of clones analyzed per terminal configuration are indicated by the Σ value. Numbers to the right of the sequences represent the numbers of clones with the appropriate sequence pattern. Sequences in which no base is specifically marked represent fill-in junctions. White-on-black letters in sequences label the base matches thought to be used to set the overlap register, and medallion-framed letters mark less frequently used base match positions in alternative overlaps. Original mismatch positions are framed; double bands detected in these positions in sequencing patterns are indicated by two bases in one place. Coordinate base pair substitutions in either of the two strands, probably caused by cosegregation of adjoined mismatches, lead to the preservation of either the top (T) or the bottom (B) strand sequence; double bands in sequence patterns indicate maintenance of the junctional mismatches (MM); independent base pair substitutions in both strands indicate mismatch resolution in the joining process (MR).

between junction types carrying external mismatches (type C) with their strongly prevalent competitors lacking mismatches (type B). Junction type C is always the least frequently used and outstripped by fill-in and alternative overlap reactions, although this overlap pattern is provided by the PSS sequences of most terminal configurations (Fig. 2, subsets Ia, II, and IVa and b). The existence of internal mismatches, however, appears to have a much less pronounced influence. For instance, junction types D (no mismatches) and E (one internal mismatch) prevail over the fill-in junction type with almost equal probability (Fig. 3, IIIa and b). These relations imply that the existence of external mismatches may constitute the most difficult handicap during junction processing while internal mismatches, hidden between base matches, are better accepted or possibly even neglected by the system.

The strong effects on competitive relations imposed by the presence of mismatches overshadow effects due to other influences of interest, such as the number or type of base matches which can be unequivocally observed, only in the few cases providing comparable competitor pairs. Among junction types C and G (subset IVb), for instance, which are both characterized by external mismatches, the complete overlap (type G) containing a base match doublet strongly suppresses the partial overlap (type C) containing only a base match singlet (Fig. 3). Nevertheless, the proportion of type G is significantly reduced compared with types B, D, E, and F. A striking example for the influence of base match types on junction formation is seen in subset Ib, in which the competition between the mismatch-free terminal overlap (type B) and the externally mismatched complete overlap (type G) is adjoined to the 5′ class as well as to the 3′ class but with different pairs of base matches (A · T G · C [Fig. 2, subset Ib]). This causes a complete inversion of the preferences of overlap formation in favor of the competitor carrying the G · C base match(es). A further determinant of overlap setting is represented by the polarity of the involved PSS pairs. This is observed in subset II of the 5′ and 3′ class between competitors of the externally mismatched partial overlap junction type C. With *Nco-Sal* (Fig. 2A) and *Sac-Kpn* (Fig. 2B), PSS sequences are exactly inverted relative to each other, permitting equal patterns of alternative-overlap settings. Although the 3′ terminal configuration preferentially uses the G · C match for overlap formation, the corresponding 5′-PSS pair uses both available base matches with a slight bias for A · T. The understanding of these relationships requires a closer investigation of the fate of mismatches in the joining process.

Removal and maintenance of mismatches in overlap junctions. (i) Evidence from sequence patterns. The junctional sequence data provided several pieces of evidence that internal mismatches were conserved within partial and complete overlap junctions (Fig. 2): (i) with respect to the mismatch position, nearly all sequences displayed an approximate 1:1 clonal distribution of their expected perfectly matched products (subset IIIb, T and B), (ii) adjoined mismatches in complete overlaps cosegregated (subset IVa, T and B), and (iii) some sequence patterns contained double bands in the positions of internal transversion mismatches (subset IVa, MM). These results indicate that base pair substitutions at mismatch positions in junctional sequences are due to either bacterial methyl-independent patchwise mismatch repair (5) or postreplicative segregation of mismatched junctions during amplification in *E. coli*. This suggests that internal mismatches are conserved within overlap junctions.



DNA substrate	Prevalent junctional configuration		A	C	T	G	AT	GC	AG	CT	AC	TG	TGAC
XbaI	T TCGAC	oc	+	+	+	+	nd	nd	+	−	nd	nd	−
SalI	AGATC G	ccc	+	−	−	+	nd	nd	+	−	nd	nd	−
Kpn I	GGTAC G	oc	+	+	+	+	nd	nd	nd	nd	−	+	−
Pst I	C ACGTC	ccc	−	−	+	+	nd	nd	nd	nd	−	+	−
Ava I	CCTAGA	oc	+	+	+	+	+	+	nd	nd	nd	nd	+
Xba I	GGGCCT	ccc	+	+/−	+	+	+	+/−	nd	nd	nd	nd	+/−
Bam H I	GGTACC	oc	+	+	+	+	+	+	nd	nd	nd	nd	+
Asp 718	CCTAGG	ccc	+	+	+	+	+	+	nd	nd	nd	nd	+
AvaI	CTCGAC	oc	+	+	+	+	+	+	+	+	nd	nd	+
SalI	GGGCCG	ccc	+	+	+	+	+	+	+	+	nd	nd	+/−
Eco R I	GGATCC	oc	+	+	+	+	+	+	+	+	nd	nd	+
Bam H I	CTTAAG	ccc	+	+	+	+	+	+	+	+	nd	nd	+/−

FIG. 4. Gel assays of joining reactions. The presence of ethidium bromide in the agarose gels allows separation of the different topological forms of the various circular and linear products generated by nonhomologous DNA end joining and ligation. Genuine joined products arise from linear input monomers as intermediate OC and final CCC monomer products. Linear dimers and multimers are indicative of head-to-head and tail-to-tail ligations, respectively, of linear input monomers. (A) Dideoxy inhibition of the joining reaction of the terminal configuration *Bam-Hind*. Lanes: E, control joining reaction in complete uninhibited extract; S, linear substrate input; A, C, T, G, inhibition of the joining reaction with the corresponding ddNTPs; AG, CT, ACTG, inhibition with combinations of the corresponding ddNTPs. (B) Assay for the ability of DNA ligases to ligate the indicated homologous *Sal-Sal* and nonhomologous *Bam-Asp* terminal configurations. Lanes: E, control reaction with complete unfractionated extract; PL, PPL fraction of the egg extract; T_4, control ligation reaction with T4 DNA ligase. (C) ddNTP inhibition patterns (cf panel A). Inhibition experiments were performed as indicated with all terminal configurations of subsets III and IVa/b of the 5′ class. Since we obtained the same or very similar results in all experiments, only some typical examples are listed here. Symbols: +, uninhibited formation (presence) of OC or CCC monomer product; −, complete inhibition of formation (absence) of OC or CCC products; +/−, reduced formation (less intense band) of OC or CCC products; nd, not determined.

All sequences of externally mismatched partial overlap junctions of the 3′ class (Fig. 2B, subset II, *Sac-Kpn*) exhibited base pair substitutions exclusively in the junctional top strand, suggesting that the two bases at the tip of the *Sac*-PSS have been selectively replaced by the complement of the *Kpn*-PSS in the joining reaction. By contrast, the sequences derived from subset IVb (Fig. 2) suggest that

<table>
<tr><th colspan="2">5'/5'</th><th colspan="4">3'/3'</th><th></th></tr>
<tr><th>Eco/Ava</th><th>Σ10</th><th>BstX1/Apa</th><th>Σ20</th><th>BstX2/Apa</th><th>Σ18</th><th></th></tr>
<tr><td>G CCGGG
CTTAA C</td><td></td><td>TAATT C
A CCGGG</td><td></td><td>TTTAA C
A CCGGG</td><td></td><td>terminal nt-loss</td></tr>
<tr><td>GAATTCCGGG</td><td>7</td><td></td><td></td><td></td><td></td><td>0nt</td></tr>
<tr><td>GAAT|CCGGG</td><td>1</td><td>TAA[+A]GGCCC</td><td>1</td><td>TTT[+T]GGCCC</td><td>1</td><td>-1nt</td></tr>
<tr><td>GAA|CCGGG</td><td>2</td><td>TAA|GGCCC</td><td>2</td><td>TTT|GGCCC
TTTAA|CCC</td><td>3
1</td><td>-2nt</td></tr>
<tr><td></td><td></td><td>TA|GGCCC
TAA|GCCC
TAATT|CC</td><td>10
2
1</td><td>TT|GGCCC
TTT|GCCC</td><td>9
1</td><td>-3nt</td></tr>
<tr><td></td><td></td><td>T|GGCCC
TA|GCCC
TAAT|CC</td><td>1
2
1</td><td>TT|GCCC
TTT|CCC</td><td>2
1</td><td>-4nt</td></tr>
</table>

FIG. 5. Survey over the sequences (5'-3' strand) obtained from the three nonmatching terminal configurations. Vertical bars in sequences indicate junctional breakpoints. Framed base pairs marked with a + in the two terminal configurations of the 3' class indicate base pair insertions. Otherwise, symbols are as in Fig. 2.

mismatch removal is not a prerequisite for the closure of the nicks in externally mismatched complete overlaps (type G). Coordinate base pair substitutions in both mismatch positions of either strand (subset IVb, T and B), as well as sequence ambiguities (subset IVb, MM), indicate bacterial postreplicative segregation of mismatches originally conserved within the overlap junction. In several cases, however, independent base exchanges in both strands indicate that they were produced by mismatch resolution during junction formation (subset IVb, MR). This suggests that mismatched bases adjacent to nicks may be removed by exonucleolytic excision and replaced by the complements of the counter strand in the course of the joining reaction itself. From these data we conclude that removal of external mismatches probably precedes gap filling in partial overlaps of the 3' class but occurs only occasionally in the corresponding partial overlaps of the 5' class and in externally mismatched complete overlaps of both classes. In this context, it is noteworthy that nick ligation in these latter junction types can proceed without conversion of the adjacent mismatch into a base match.

(ii) Investigation of mismatch maintenance in the joining process by ddNTP inhibition. To analyze in more detail the problem of mismatch maintenance or removal in the joining process, we used inhibition of the joining reaction with ddNTPs which block elongation of fill-in DNA synthesis presumably mediated by DNA polymerase β (25). In this way, joining reactions involving gap filling are selectively blocked at the site of ddNTP incorporation. The affected strand retains a gap or a nick which leads to the suppression of final CCC product formation and accumulation of intermediate OC products.

The gel in Fig. 4A shows the differential ddNTP inhibition pattern of the joining reaction for the *Bam-Hind* terminal configuration which produces a partial overlap containing an internal G-A transversion mismatch (Fig. 2A). As expected, the addition of ddGTP or ddATP inhibits CCC formation and the combination of both ddNTPs even abolishes OC formation because these nucleotides are incorporated by gap filling. Nucleotides required for the conversion of the mismatch into one of two possible base matches, however, are not incorporated, as seen by the addition of ddCTP and ddTTP: neither one alone nor the combination of both blocks CCC formation. Similar results were obtained for the other two substrates of subset IIIb of the 5' and 3' class shown in Fig. 4C. This strongly suggests that junction formation can be completed without conversion of internal mismatches into basematches.

Corresponding inhibition experiments were performed with all substrates of subset IVa/b of the 5' class. Some typical examples are shown in Fig. 4C. In no case did addition of ddNTPs abolish CCC formation. This indicates that a considerable fraction of the internally as well as externally mismatched complete overlaps can be directly sealed by nick ligation. The sequences of the overlap junctions formed under conditions of complete ddNTP inhibition were verified to be the same as under uninhibited conditions (data not shown). In most cases, the addition of all four ddNTPs led to a significant reduction of CCC formation which is probably due to the suppression of the formation of all alternative junction types requiring DNA fill-in synthesis.

In summary, differential ddNTP inhibition confirms the conclusions drawn from junctional sequences that internal mismatches are conserved in the joining reaction and that the nicks in complete overlaps containing external mismatches may be closed without preceding mismatch removal.

DNA ligases cannot seal complete overlap heteroduplexes. According to the above results, the joining process is able to convert complete-overlap heteroduplexes into covalently sealed junctions by ligation of the nicks between two base pairs or a mismatch and a base pair. We therefore asked whether the DNA ligase fraction partially purified from joining active *Xenopus* egg extract (PPL) (25) or T4 DNA ligase would be able to ligate such structures. In control reactions with a homologous substrate carrying 5' sticky ends (*Sal*I), PPL and T4 DNA ligase readily generated circular monomers (Fig. 4B). Corresponding ligation experiments were performed at equally low DNA concentrations (1 nM ends) with all substrates of subset IVa/b of the 5' class. This is shown in Fig. 4B for the *Bam-Asp* configuration. In contrast to the complete-extract reaction which readily joins the *Bam-Asp* overlap heteroduplex, neither PPL nor T4 DNA ligase is able to ligate this structure, as seen by the absence of OC and CCC monomers. Linear multimers, however, positively indicate the presence of ligation activity in these samples. Comparable results were obtained with all other substrates of subsets IVa/b of the 5' class (Fig. 2A), suggesting that closure of the nicks cannot be achieved in externally or internally mismatched complete overlap heteroduplexes by a ligase alone. Instead, the extract system must contain additional factors to carry out this reaction, which are separable from the overall ligase activity present in the PPL fraction.

Joining of nonmatching PSS pairs. The joining of a completely mismatched terminal configuration of the 5' class was investigated by using the *Eco-Ava* configuration (Fig. 5). As expected, the majority of junctions (70%) were of the fill-in type. The remaining junctions, however, had lost 1 or 2 nt from their *Eco* PSS, which could be due to 5'-3' exonucleolytic trimming. The joining patterns of two totally mismatched terminal configurations of the 3' class containing essentially the same PSS sequences as the corresponding 5' substrate were investigated with the help of two *BstX-Apa* configurations differing in the sequence of their *BstX* PSS (5'-AATT-3' versus 5'-TTAA-3'). Even in this situation,

neither a single fill-in junction nor deletions reaching into adjacent double-stranded regions were generated; all junctional sequences displayed sequence shortening of 1 to 4 nt. The majority of junctions (65% for 5'-AATT-3'; 55% for 5'-TTAA-3') had lost exactly 3 nt from the *Bst*X PSS, whereas the *Apa* PSS remained intact. This preferential preservation of the G+C-rich strand was also observed in junctions from other subgroups and could be accounted for by a preference of the PSS-trimming exonuclease activity for A+T-rich sequences. Template slippage during DNA fill-in synthesis could explain the two insertion junctions which both lost the terminal 2 nt from the *Bst*X PSS and showed an addition of a nucleotide of the same type as the last 2 nt in the remaining PSS (Fig. 5). A · T base pairs were also reported to be the most common 1-bp insertions among all junctions with inserts (18).

DISCUSSION

Several joining systems derived from a variety of widely differing organisms have been shown to join unrelated DNA segments carrying noncomplementary restriction termini end to end (4, 13, 20, 27). The features of junctional sequences produced in such joining events provide evidence that all these different joining systems operate by very similar mechanisms, which manage to process nonhomologous DNA termini in reactions necessarily more complex than mere ligation. In the *Xenopus* system, which presently is the only available in vitro system able to efficiently join nonhomologous terminal configurations, two different pathways of end joining, the overlap and the fill-in modes, were found to be responsible for the joining of structurally differing terminal pairs (16, 25).

In this study, we focused on the mechanisms of the overlap mode, which are investigated in detail with the help of a large set of 26 different nonhomologous terminal configurations. In the overlap reaction scheme, short overlap intermediates are assumed to be formed between 4-nt PSS tails. Without exception, base match doublets or even singlets present within PSS sequences were reliably used to set the extent of overlap. Similar overlap junctions have also been shown to constitute the majority of joining products in mammalian in vivo systems (13, 20, 27). Apart from these, a significant fraction was represented by events involving microhomology regions within parental terminal duplex regions to create larger deletions (19, 20, 23, 27). Recent results from experiments involving injection in *Xenopus* eggs (8) showed that these cells also generate significant fractions of similar deletion junctions in vivo. The strong reduction of the frequency of such deletions in the *Xenopus* in vitro system (15 of 361) is probably due to the reaction conditions used, which have been optimized for overlap interactions between terminal PSS tails to generate a limited comprehensible spectrum of junctions facilitating the analysis of specific mechanisms.

The prevalent overlap pathway is inhibited by a second reaction related to the previously described fill-in mode (25), which leads to the generation of fill-in junctions (zero overlaps) characterized by the complete preservation of both PSS tails. The result that fill-in junctions constitute only minor fractions of all joining events suggests that the fill-in reaction is usually overridden by overlap formation.

Although terminal configurations of the 5' class readily form mostly small fractions of fill-in junctions, these were never detected in the 3' class. This suggests that two 3'PSS termini cannot be joined in the fill-in mode, although we have shown previously that the 3' end of an abutting unlinked blunt or 5'PSS partner terminus can serve as a primer to promote filling in of a 3'PSS tail (25). Overlap formation between terminal PSS tails may therefore be regarded as a rather strong reaction, which can easily override the fill-in reaction. In this context, the junctions derived from nonmatching terminal configurations of the 3' class must be considered. Two different mechanisms may be envisaged to create the observed sequences: (i) unprotected naked PSS tails could be partially or entirely resected by an exonuclease activity, and remaining PSS sequences would be preserved by the fill-in mode; and (ii) overlaps could be formed between PSS tails despite the absence of sequence homology. PSS tails would be held in alignment during exonucleolytic trimming of mismatched 3'PSS to prepare proper primer conditions for fill-in DNA synthesis. Once properly aligned, filling in could be primed at the unlinked juxtaposed 3'PSS partner. In this way, accidental degradation of adjacent duplex sequences would be prevented. In fact, among 38 nonmatching junctions of the 3' class, not a single larger deletion was found. Instead, all junctions displayed sequence shortening of 1 to 4 nt, characteristic of overlap formation.

Alternative possibilities to form base matches within terminal configurations provided insight in the rules of overlap junction formation, which were found to be determined by (i) the polarity of the PSS pair to be joined; (ii) the positions, number, and types of base matches; and (iii) the locations of mismatches within the overlap. External mismatches were found to be the strongest determinants influencing the overlap reaction.

In contrast to internal mismatches, whose presence appeared to have little influence on junction formation, external mismatches exposed at PSS tips dramatically reduced the efficiency of overlap formation. Sequence patterns derived from partial overlap junctions indicated that external mismatches located at 3'PSS tips (primer position for fill-in DNA synthesis) had been converted into base matches by the joining process. Such mismatches could be eliminated by exonucleolytic proofreading (7). ddNTP-sensitive (but aphidicolin-resistant [24]) filling in of junctional microgaps in the *Xenopus* system is assumed to be performed by DNA polymerase β (25). Although purified polymerase β has been shown to be devoid of proofreading activity (10), it is often associated with DNase V, a nuclease with 3' and 5' exonuclease activity (12, 17) which could fulfill the task of proofreading. External mismatches flanking nicks on the 5' or 3' side appeared to be removed only occasionally (direct evidence for this will be presented elsewhere [15]). This suggests that the complete joining system manages to ligate a mismatch to a base match, a reaction which, however, cannot be achieved by the partially purified ligase activity of the extract alone.

The conservation of internal mismatches as demonstrated by ddNTP inhibition suggests that they are masked by single flanking base pairs and are possibly embedded within a double-helical overlap heteroduplex. Thus, they may be protected against exonucleolytic excision and may allow filling in and ligation to proceed with unreduced efficiency. Nuclear magnetic resonance studies performed with double-stranded dodecamer oligonucleotides containing single-base mismatches indicated that these are easy to integrate in the normal B-DNA helix conformation, where they cause only slight distortion and exert little effect on base-stacking patterns (6). However, the spatial conformation of very short (1- to 4-nt) overlaps set by only one or two base

matches is unknown, and it is doubtful whether such metastable heteroduplexes can persist long enough in naked form to accomplish the subsequent enzymatic reactions needed for completion of junction formation. This notion is substantiated by the result that even the sealing of the nicks in 4-nt complete internally mismatched overlaps, a reaction that in principle resembles a sticky-end ligation, cannot be achieved by a ligase alone. The presence of additional stabilizing factors contained only in the complete extract system is clearly required.

Such stabilizing factors could be represented by highly specialized terminal DNA-binding proteins, so-called alignment factors, which have been assumed to align antiparallel PSS tails in short overlaps as well as to achieve strand juxtaposition between abutting termini (16, 25). Structural maintenance of overlap heteroduplexes set by thermodynamically weak single basematches could be achieved by protecting hydrogen bonds from surrounding water molecules, facilitating stacking interactions, and minimizing the helix-distorting effect of mismatches within a hydrophobic protein cavity. In addition, alignment factors could facilitate the subsequent binding and action of repair enzymes and thus guarantee the high efficiencies of joining reactions. It is likely that mechanisms of nonhomologous DNA end joining are responsible not only for the elimination of simple breaks as generated by restriction enzymes but also for the sealing of other, more complex DNA breaks caused by ionizing radiation or carcinogens. The understanding of the pathways of DNA end joining may provide information about the mechanisms of DNA double-strand break-induced mutagenesis.

ACKNOWLEDGMENTS

This work was funded by grant Vi 23/1-2 from the Deutsche Forschungsgemeinschaft to W.V. Since August 1992, P.P. has been the recipient of a fellowship supplied by the Lise Meitner Program of the Ministerium für Wissenschaft und Forschung des Landes Nordrhein-Westfalen.

REFERENCES

1. **Berg, D. E., and M. M. Howe (ed).** 1989. Mobile DNA. American Society for Microbiology, Washington, D.C.
2. **Gahlmann, R., and W. Doerfler.** 1983. Integration of viral DNA into the genome of the adenovirus type 2-transformed hamster cell line HE5 without loss or alteration of cellular nucleotides. Nucleic Acids Res. **11:**7347–7361.
3. **Gellert, M.** 1992. V(D)J recombination gets a break. Trends Genet. **8:**408–412.
4. **Goedecke, W., P. Pfeiffer, and W. Vielmetter.** Mechanisms of nonhomologous DNA end joining in the fission yeast *Schizosaccharomyces pombe*. Submitted for publication.
5. **Grilley, M., J. Holmes, B. Yashar, and P. Modrich.** 1990. Mechanisms of DNA mismatch correction. Mutat. Res. **236:** 253–267.
6. **Kennard, O.** 1987. The molecular structure of basepair mismatches, p. 25–41. *In* F. Eckstein and D. M. J. Lilley (ed.), Nucleic acids and molecular biology, vol. 4. Springer-Verlag, KG, Berlin.
7. **Kunkel, T. A.** 1988. Exonucleolytic proofreading. Cell **53:**837–840.
8. **Lehman, C. W., J. K. Trautman, and D. Carroll.** Illegitimate recombination in *Xenopus*: characterization of end-joined junctions. Submitted for publication.
9. **Lieber, M. R.** 1992. The mechanism of V(D)J recombination: a balance of diversity, specificity, and stability. Cell **70:**873–876.
10. **Linn, S.** 1991. How many pols does it take to replicate nuclear DNA? Cell **66:**185–187.
11. **Meuth, M.** 1989. Illegitimate recombination in mammalian cells, p. 833–860. *In* D. E. Berg and M. M. Howe (ed.), Mobile DNA. American Society for Microbiology, Washington, D.C.
12. **Mosbough, D. W., and S. Linn.** 1983. Excision repair and DNA synthesis with a combination of HeLa DNA polymerase β and DNase V. J. Biol. Chem. **258:**108–118.
13. **Munz, P. L., and C. S. H. Young.** 1991. End-joining of DNA fragments in adenovirus transfection of human cells. Virology **183:**160–169.
14. **North, P., A. Ganesh, and J. Thacker.** 1990. The rejoining of double-strand breaks in vitro. Nucleic Acids Res. **18:**6205–6210.
15. **Pfeiffer, P., S. Thode, J. Hancke, P. Keohavong, and W. G. Thilly.** Resolution and conservation of mismatches in nonhomologous DNA end joining. Submitted for publication.
16. **Pfeiffer, P., and W. Vielmetter.** 1988. Joining of nonhomologous DNA double strand breaks in vitro. Nucleic Acids Res. **16:**907–924.
17. **Randahl, H., G. C. Elliot, and S. Linn.** 1988. DNA repair reactions by purified HeLa DNA polymerases and exonucleases. J. Biol. Chem. **263:**12228–12234.
18. **Roth, D. B., X.-B. Chang, and J. H. Wilson.** 1989. Comparison of filler DNA at immune and oncogenic rearrangements suggests multiple mechanisms of formation. Mol. Cell. Biol. **9:**3049–3057.
19. **Roth, D. B., T. M. Porter, and J. H. Wilson.** 1985. Mechanisms of nonhomologous recombination in mammalian cells. Mol. Cell. Biol. **5:**2599–2607.
20. **Roth, D. B., and J. H. Wilson.** 1986. Nonhomologous recombination in mammalian cells: role for short sequence homologies in the joining reaction. Mol. Cell. Biol. **6:**4295–4304.
21. **Roth, D. B., and J. H. Wilson.** 1988. Illegitimate recombination in mammalian cells, p. 621–653. *In* R. Kucherlapati and G. R. Smith (ed.), Genetic recombination. American Society for Microbiology, Washington, D.C.
22. **Schaal, H., P. Pfeiffer, M. Klein, P. Gehrmann, and A. Scheid.** 1993. Use of DNA end joining activity of a *Xenopus laevis* egg extract for construction of deletions and expression vectors for HIV-1 Tat and Rev proteins. Gene **124:**275–280.
23. **Thacker, J., J. Chalk, A. Ganesh, and P. North.** 1992. A mechanism for deletion formation in DNA by human cell extracts: the involvement of short sequence repeats. Nucleic Acids Res. **20:**6183–6188.
24. **Thode, S.** Unpublished data.
25. **Thode, S., A. Schäfer, P. Pfeiffer, and W. Vielmetter.** 1990. A novel pathway of DNA end-to-end joining. Cell **60:**921–928.
26. **Wake, C. T., F. Vernaleone, and J. H. Wilson.** 1985. Topological requirements for homologous recombination among DNA molecules transfected into mammalian cells. Mol. Cell. Biol. **5:**2080–2089.
27. **Winegar, R. A., L. H. Lutze, J. T. Rufer, and W. F. Morgan.** 1992. Spectrum of mutations produced by specific types of restriction enzyme-induced double strand breaks. Mutagenesis **7:**439–445.