Volume 16 Number 3 1988 Nucleic Acids Research

Joining of nonhomologous DNA double strand breaks *in vitro*

Petra Pfeiffer and Walter Vielmetter

Institut für Genetik der Universität zu Köln, Weyertal 121, D-5000 Köln 41, FRG

Received October 22, 1987; Revised and Accepted December 23, 1987

ABSTRACT

Extracts of *Xenopus laevis* eggs can efficiently join ends of duplex DNA that differ in structure and sequence. This was analysed by recircularisation of linear plasmid DNA molecules with dissimilar termini, generated by successive cuts with two different restriction enzymes within the pSP65 polylinker. Use of various enzymes provided blunt ended or 4 nucleotides long 3' and 5' protruding single strand (PSS) termini which were successfully joined *in vitro* in any tested combination. Sequence analysis of numerous junctions from cloned reaction products of 7 terminus combinations reveal: apart from very rare base exchanges and single nucleotide insertions (10% deletions (1 to 18 nucleotides long) were detected. Blunt/PSS or 3'PSS/5'PSS terminus pairs undergo simple "blunt end" joining which preserves PSS ends by fill-in. In contrast, equally polar 3'PSS/3'PSS or 5'PSS/5'PSS terminus pairs are joined by a complex mode: PSS ends overlap by a defined number of nucleotides, set by matching basepairs. Even one basematch suffices to define the setting. This then determines the final mismatch repair and fill-in pattern. We propose that yet unknown terminal DNA-binding proteins stabilize the energetically highly unfavorable configuration of single matching basepairs and help to support defined overlap structures.

INTRODUCTION

Living organisms have evolved repair mechanisms to eliminate highly lethal double-strand breaks (DSB) from chromosomal DNA (1 for review). Such mechanisms are not only related to homologous recombination pathways (2,3), in which recombinogenic DSB termini interact end-to-strand with duplex DNA (4-7), but also to nonhomologous recombination pathways, in which DSB termini are joined end-to-end (8-11).

Nonhomologous DSB repair may be involved in illegitimate recombination pathways, such as chromosomal rearrangements, transposition of trans- and retroposons or integration of viral DNA into chromosomes (12 and 13 for review). One way to join nonhomologous DSB termini end-to-end would be to shape them to blunt ends (9,14) and ligate them by type I ligase (15). Protruding single-strands (PSS) created by staggered breakage of duplex DNA could undergo fill-in or degradation repair to produce small duplications or deletions at junc-

tional flanks (9,16). However, generation of duplications and junctional insertions demonstrate that more complex reaction schemes also exist (14,16). Fortuitious complementarities between short, unrelated DNA-sequences may be called "pseudohomologies" in contrast to "homology", defined as a relation between sequences of common origin. A role of short pseudohomologies in junction formation during nonhomologous joining has been conjectured (reviewed in 17). Direct comparisons between substrate and product DNA sequences unequivocally demonstrates use of pseudohomologies down to singlets as alignment markers in nonhomologous joining reactions (17).

Elementary recombination processes in eukaryotes may be studied *in vivo* by transfer of plasmid or viral DNA into cell systems to induce a strong transient rise of repair and recombination (18-21). Another fruitful approach analyses nonhomologous joining by transfection of linear SV40-DNA with defined ends under conditions selective for joined products (9,11,14,17). To explore nonhomologous joining *in vitro* we have employed oocytes and eggs of *Xenopus laevis*, known to be a rich source of ligation and recombination activities. (22-26).

We report here that extracts of *Xenopus* eggs can efficiently join nonmatching DSB termini in various combinations. In some of them the repair system creates ligatable blunt ends by fill-in of PSS ends. However, terminus pairs with PSS of equal polarity are joined by a complex mechanism: noncomplementary PSS ends overlap by a defined number of nucleotides, set by even single pseudohomologous basematches and are maintained during junctional mismatch repair until final ligation.

MATERIALS AND METHODS.

Extract-preparation from fertilized *Xenopus laevis eggs*

Artificially fertilized *X. laevis* eggs prepared according to Newport and Kirshner (27) were extracted within 90 minutes and before cleavage onset. Sorted out, intact, dejellied eggs overlayed with an equal volume of extraction medium (28) were centrifuged in a Beckman SW50.1 Ti-rotor at 30 krpm at 4°C for 30 minutes. The interphase (clear plus turbid layer) was removed and thoroughly mixed. Aliquots frozen in liquid N_2 and stored at -70°C retain activity for up to 6 months.

Preparation of substrate DNA

Source of substrate DNA was plasmid pSP65 (29) which carries an amp^R gene and a polylinker (top panel of Fig.6). In pSP65(Kpn) a KpnI-linker (8-mer, Boehringer, Mannheim) was ligated into the SmaI-site of pSP65 (Fig.6). Plas-

mids were propagated in *E.coli* HB101(recA) (30). Substrate DNA was prepared by successive cuts with two different restriction enzymes (Boehringer, Mannheim) in the plasmid's polylinker. After each restriction, bands of linearized DNA were eluted from low melting agarose (Type A-9414; Sigma, Munich). Perfectly doublecut molecules cannot be recircularized by ligation with T4-DNA-ligase (Boehringer, Mannheim) (Fig1:c). This was checked by the absence of colonyformers in transformation assays. Purified substrate DNA was stored in TE (10 mM Tris pH 7.6; 1 mM EDTA) at 25 ng/µl.

Joining of substrates in extract and T4-DNA-ligase controls

Usually 1/5 volume of substrate DNA in TE was joined in 4/5 volume of undiluted extract (25-50 ng DNA /25 µl) at 13°C for 90 minutes. Joining reactions were terminated by freeze shock in liquid N_2 and samples stored at -70°C. T4-DNA-ligase controls (0.01 units/µl) in 50 mM Tris pH 7.6, 10 mM $MgCl_2$, 70 mM ATP, 10 mM DTT and 0.2 mg/ml BSA were performed in the same volume under identical conditions.

Gel assay of reaction products (Southern blots)

Reaction products were processed by methods modified from Rusconi and Schaffner (24): 0.3 mg proteinase K (Sigma, Munich) in 275 µl of 20 mM Tris pH 7.6, 300 mM NaCl, 10 mM EDTA, 1% (w/v) SDS was added to frozen 25 µl samples. After 60 min at 65°C samples were phenolized and precipitated with ethanol. Southern blots (31): 0.5 or 1 ng DNA/slot was electrophoresed on 0.7% agarose minigels containing 1 µg/ml EthBr in Loening E buffer (32). Gels, blotted onto nitrocellulose sheets (Schleicher & Schuell, Dassel) according to Wahl et al. (33) were hybridized with ^{32}P nicktranslated pSP65 as probe.

Transformation assay of reaction products

Aliquots of competent HB101 cells (5 x 10^8) were transformed with 1 ng of extract-joined substrate or pSP65 control DNA and directly plated on ampicillin YT agar to yield 3 x 10^3 transformants per 1 ng pSP65 DNA. Transformations with strain BMH71-18 (mutS) were performed according to Kramer and Fritz (34). MutS belongs to a group of genes (mutH, L, S, uvrD) which control the postreplicative, methyl-directed major pathway of mismatch repair in *E.coli* (35). Strain BMH71-18 is virtually unable to repair mismatches in transformed plasmid DNA (34). To monitor mutS maintenance, subcultures of 5 to 10 independently isolated bacterial colonies from each assay were tested for the mutator phenotype, measured as an enhanced rate for strA mutations (5 x 10^{-7}).

Analysis of cloned joined products

Transformant clones were propagated in 5 ml cultures and plasmids prepared according to Crouse et al. (36) with modifications: isopropanol precipitated

DNA pellets, resuspended in 200 µl distilled water were digested with 1 µg/ml RNase A (Sigma, Munich) at 25°C for 30 minutes. Samples, adjusted to 300 mM LiCl, were phenolized and precipitated with ethanol. Washed pellets (about 10 µg plasmid DNA) resuspended in distilled water at 200 ng DNA/µl were rechecked for purity and concentration by electrophoresis.

Restriction assays: Several samples for a given clonal DNA were treated each with one enzyme selected from a set of restriction enzymes including XbaI (see Fig.5) and assayed on EthBr agarose gels for the presence (linearisation) or absense (maintenance of circularity) of restriction sites.

Size determinations of polylinker fragments: EcoRI-cut pSP65-DNA from XbaI resistant clones was filled in at 5'PSS ends by Klenow enzyme (Boehringer, Mannheim) in the presence of ^{32}P-α-dATP and the polylinker fragment released by a HindIII cut. Fragment sizes were analysed electrophoretically in relation to a size marker ladder on 10% denaturing PAGE.

DNA sequence analysis of selected clones from joined products: DNA was sequenced by the dideoxy method of Sanger et al. (37) as modified by Seedorf et al. (38) which permits direct use of pSP65 duplex DNA. Additional modifications were: (i) Incubation of Klenow reactions at 50°C; (ii) 200 µM concentrations of ddATP and ddTTP. The primer (SP6 promoter specific 20-mer, Boehringer, Mannheim) hybridizes to the 3' to 5' strand of pSP65, 23 bp upstream of the EcoRI site. It was end-labeled with ^{32}P-gamma-ATP by T4-polynuleotide-kinase (Boehringer, Mannheim). Samples from 8 clones were sequenced in parallel on one denaturing 8% PAGE.

RESULTS

Experimental system

We report here that extracts from fertilized *Xenopus* eggs join unrelated termini of duplex DNA by reactions which achieve more than mere ligation. Fig.1 outlines, how these reactions are assayed and their joined products are analysed. In essence: Linearization of plasmid pSP65 (29) by restriction cuts at unique polylinker sites (Fig.1:a,b) provides substrate DNA molecules with blunt and 3' or 5' protruding single strand (PSS) ends. Controlled (Fig.1:c), successive cuts with two different restriction enzymes generate linear substrates with termini that differ in structure (blunt versus PSS ends) and lack complementary sequences. They are referred to as "nonhomologous" (in contrast to single cut generated "homologous" terminus pairs which share complementary nucleotide quadruplets of common origin). These terms are also adopted for corresponding joining reactions and their products. By various

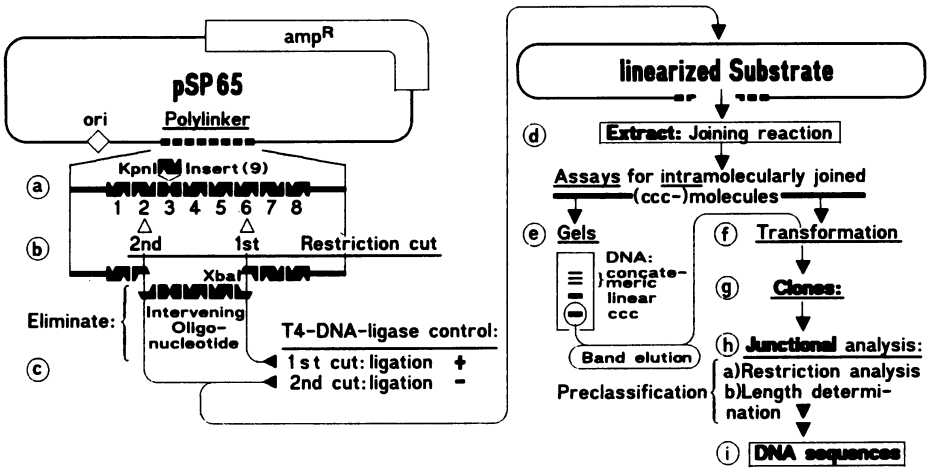


Fig.1 Experimental outline for the analysis of joining of nonhomologous DNA termini. Plasmid pSP65 carries a) a ***polylinker*** with 8 unique restriction sites (and an additional KpnI site (9) in derivative pSP65(Kpn) (see top panel of Fig.5). Cleavage at restriction sites (black blocks with identations to mark cleavage positions) provides blunt (3) and 4 nucleotides long protruding singlestrand (PSS) ends of 3'(2.7.9) or 5'(1.4.5.6.8) polarity. b) ***Restriction cuts*** generate linear DNA molecules with terminus pairs: (i) blunt or "homologous" upon a single cut, (ii) "nonhomolgous" upon two successive cuts (which excise an intervening oligonucleotide). c) ***Control for correct cutting:*** singlecuts can, doublecuts cannot be ligated by T4-DNA-ligase. d) Linearized substrates treated in the ***egg extract*** are assayed for intramolecularly joined products (ccc-molecules; ccc = covalently closed circular) (e) on ***EthBr-gels*** or (f) by ***transformation*** into *E.coli* cells (which selectively propagate ccc-molecules). g) Band-eluted ccc-molecules ***cloned*** via the transforming system are ***preclassified*** (as described in the text) to yield clonal subclasses from which representative clone sets are selected for (i) DNA-sequence analysis of junctions.

combinations of duplicate restriction cuts 7 nonhomologous model substrates were constructed and applied in this work (Fig.2).

Concerning the joining ability of extracts (Fig.1:d), high activities are found in developed eggs but not in oocytes *in vitro* and *in vivo* (39). Fertilization, although favourable to stabilize eggs mechanically during manipulations (see Materials and Methods), is itself not required to induce joining activity. The activity of the joining reaction is fairly thermolabile: it is highest between 12°C and 20°C , decays towards higher temperatures and ceases at 41°C. Therefore, 13°C was chosen as standard temperature. Under these conditions more than 50% of the input DNA is recovered as joined products.

Reaction products are assayed either directly on gels or by transformation

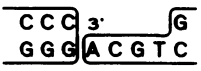
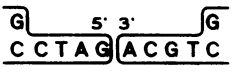
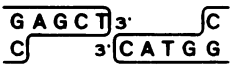
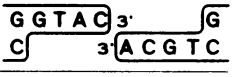
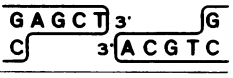
Substrate type	Terminus configuration	Sequence
1) SmaI/SalI	blunt/5'PSS	
2) SmaI/PstI	blunt/3'PSS	
3) BamH1/PstI	5'PSS/3'PSS	
4) BamH1/SalI	5'PSS/5'PSS	
5) SacI/KpnI	3'PSS/3'PSS	
6) KpnI/PstI	3'PSS/3'PSS	
7) SacI/PstI	3'PSS/3'PSS	

Fig.2 Terminus configurations of employed nonhomologous substrate types.

(Fig.1:e,f). For the assays it is essential to consider that only the fraction of intramolecularly recircularized ccc-monomer molecules yields unambiguously nonhomologous reaction products. Besides that concatemer molecules are formed by intermolecular reactions in which nonhomologous substrate molecules may interact head to head (tail to tail) to join equal i.e. homologous termini. In accord with known critical DNA concentrations for our plasmid size (40,32 for review), sufficiently high ratios of ccc-monomers versus concatemers were formed in our reactions as illustrated in Fig.3 and 4. In gel assays ccc-monomers can be readily selected by band position, while in transformation assays a mixture of products is processed. However, transformation selectively filters out linear or head to head formed concatemers (41) and therefore propagates ccc-monomers with much higher probabilities. In fact, most of the transforming activity resides in gel eluted ccc-monomer bands and is almost absent in eluted joint concatemer bands (data not shown). Thus, both assay procedures provide tools to unequivocally prove joining of nonhomologous substrates in the extract and provide cloning conditions for the junctional analysis (Fig.1:g-i).

The extract efficiently joins nonhomologous terminus pairs

Gel assays (Fig.3) show that only extract treatment can convert nonhomolo-

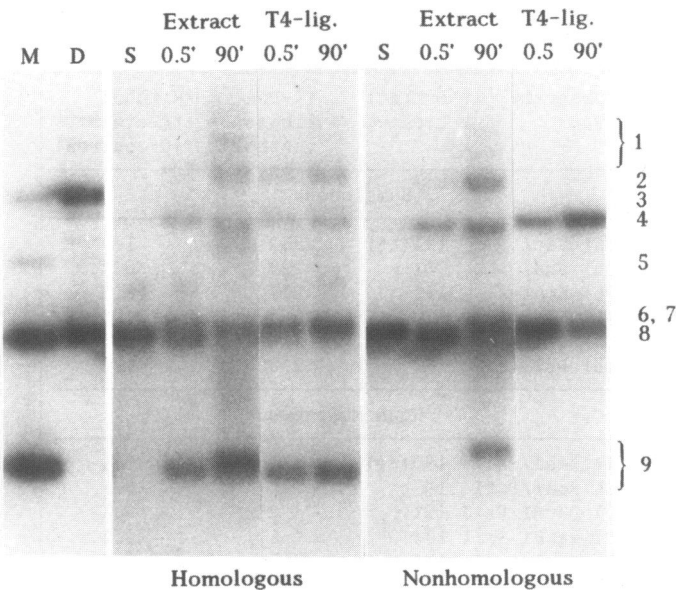


Fig.3 Gel assays (Southern blots) of homologous and nonhomologous joined products and T4-DNA-ligase controls. ***Homologous*** substrate: pSP65 cut by BamH1. ***Nonhomologous*** substrate: type 4, BamH1/SalI (see Fig.2). Reactions were performed at 13°C for 0.5 and 90 minutes respectively at a DNA-input of 2 ng/µl. Reaction products were separated on EthBr agarose minigels. Bands were visualized on Southern blots by use of ^{32}P-nicktranslated pSP65 as probe. This excludes RNA and chromosomal DNA contaminants and greatly increases the sensitivity of the assay. Markers: nicked and circular monomer (M) and dimer (D) of pSP65; nontreated linear substrate (S). Interpretation of bands: 1) higher concatemers; 2) linear trimer; 3) nicked dimer; 4) linear dimer; 5) closed circular trimer; 6) closed circular dimer; 7) nicked monomer; 8) linear monomer; 9)closed circular monomer. Only bands 7, 8 and 9 are relevant for assays.

gous substrates into ccc-monomers. Monomers are absent in the corresponding T4-DNA-ligase controls, although concatemer bands indicate full ligase activity. Similar results were obtained for all nonhomologous substrates listed in Tab.1 (results not shown). In contrast, homologous substrates are readily ligated to ccc-monomers both by the extract and by T4-DNA-ligase.

Substantial amounts of all ligation products are formed within the first 0.5 minutes. At this time no trace of nonhomologously joined ccc-monomers is visible (Fig.3). Fig.4 illustrates the slow time course for the formation of nonhomologously joined reaction products. Approximate reaction time constants (estimated from gel scans) are in the order of 10 seconds for ligation and 30 minutes for nonhomologous joining. Thus, the extract ligates homologous termi-

Tab. 1 Transformation assays of joined products(*)

Substrate type	Extract treated	T4-DNA ligase control	Without treatment (0-control)
	Homologous		
1h) SacI	194(**)	352	1
2h) SmaI	81	229	0.7
3h) KpnI	192	274	0.9
4h) BamH1	170	283	1.6
5h) SalI	154	324	1.6
6h) PstI	167	218	1.2
	Nonhomologous		
1) SmaI/SalI	190(**)	17	<0.1
2) SmaI/PstI	197	20	<0.1
3) BamHI/PstI	191	2	<0.1
4) BamH1/SalI	184	6	<0.1
5) SacI/KpnI	171	5	<0.1
6) KpnI/PstI	146	9	<0.1
7) SacI/PstI	137	1	<0.1

*) Substrates were treated in extracts, T4-DNA-ligase controls and 0-controls at 13°C for 90 minutes.
**)Average of colonies/plate at a DNA input of 1 ng/plate.

ni at a rate comparable to that of T4-DNA-ligase, but in addition it efficiently joins nonhomologous termini at an approximately two orders of magnitude slower rate.

Transformation assays confirm the results from gel assays. Tab.1 shows that both extract and T4-DNA-ligase can ligate various homologous substrates at about equal relative efficiencies except for reduced blunt/blunt end ligation (Tab.1:2h). Furthermore, the extract can join nonhomologous and ligate homologous substrates at comparable final efficiencies. T4-DNA-ligase apparently ligates nonhomologous substrates at very low levels, if at all. The latter result underlines the selective power of the transformation system to preferentially process intramolecularly joined substrates, an important feature to reliably isolate "correct" clones during clonal analysis.

Clonal analysis of junctions

Useful information on the structures of joined products can only be obtained by DNA sequence analysis if a sizeable number of junctions is considered. Sets

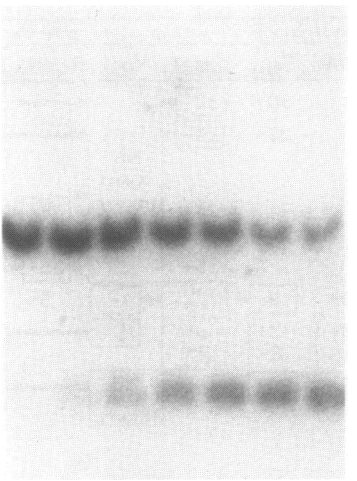


Fig.4 Time course for generation of joined products in extracts at 13°C. Substrate: type 7, SacI/PstI (see Fig.1) at low DNA input (0.5 ng/µl) to suppress concatemer formation. Gel assay (Southern blot) as in Fig.3. Reaction times (from left to right): 5, 15, 30, 60, 90, 180, 360 minutes.

of around 20 transformant clones from the joined products of each substrate type (see Fig.2) were isolated and ccc-DNA from each clone was preclassified by comparatively fast assays. From subclasses so derived, representative subsets of clones were chosen for DNA sequencing.

Preclassification includes the screening of clones for gain or loss of polylinker restriction sites. First, 2% of defective clones were discarded, stemming from incompletely restricted substrate molecules which had retained intervening oligonucleotides including the XbaI site (Fig.1:b; horizontal bars in Fig.5 and 6). XbaI resistant clones were further tested with selected restriction enzymes. Fig.5 summarizes the results for 5 newly derived subclasses. It can be seen that disrupted restriction sites at substrate termini (boxed numbers) are restored in some cases during junction formation (2a, 3a) and these may be abolished in adjoined subclasses (2b, 2c, 3b). On occasion loss of neighbouring restriction sites was observed (2d, 6d, 7e), events which we ascribed to deletions.

In another procedure gain or loss of nucleotides at junctions was assayed by length determinations of single stranded polylinker fragments. Tab.2 shows that while one substrate type produces junctions with preserved PSS termini, subclasses from others exhibit sequences shortened by a small, defined number

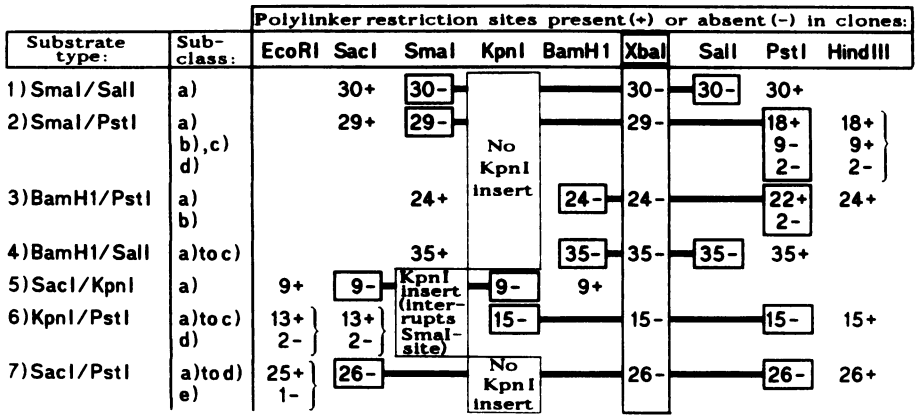


Fig.5 Classification by restriction analysis of polylinker restriction sites of clones derived from nonhomologous joined products. Letters for subclasses correspond to definitive subclass assignments (see Fig.6). ***Numbers:*** amounts of clones. Signs: presence (+) or absence (-) of a restriction site. Restriction sites are drawn according to their order along the polylinker. This includes the KpnI site inserted into derivative pSP65(Kpn) (relevant for substrate types 5 and 6). ***Boxed numbers:*** positions of terminal restriction sites at which molecules were cut during substrate preparation (junctional breakpoints). ***Bars:*** symbolize intervening oligonucleotides (cut out from the polylinker during substrate preparation). Note that except for substrate type 5 the XbaI site is adjoined to intervening oligonucleotides.

of nucleotides. Defined sequence shortening is only seen in substrate types with PSS termini of equal polarity (3'PSS/3'PSS, 5'PSS/5'PSS). This indicates that sequence shortening might arise by defined PSS overlap.

With the help of established subclasses containing 165 preclassified clones, 85 clones were selected for DNA sequence analysis (see Materials and Methods). Since the polylinker sequence is well known, only one strand was sequenced once per clone. Fig.6 summarizes the 20 sequenced subclasses of junctions.

Survey of junctional DNA sequences

Sequence data listed in Fig.6 show that nonhomologous joining in the extract proceeds without gross rearrangements. Besides one base exchange (4b) one insertion of a single base is observed (7d). 1 to 18 nucleotides long deletions (arrow head strings) were found in 10% of the sequenced clones and these start exclusively from 3'PSS ends at junctional breakpoints. About 70% of the deletions are clustered in substrate type 2. Deletions are possibly induced by 3' to 5' exonucleolytic attack at ends of linear substrate molecules. Due to reduced joining rates termini might be more accessible to attack in certain substrate types.

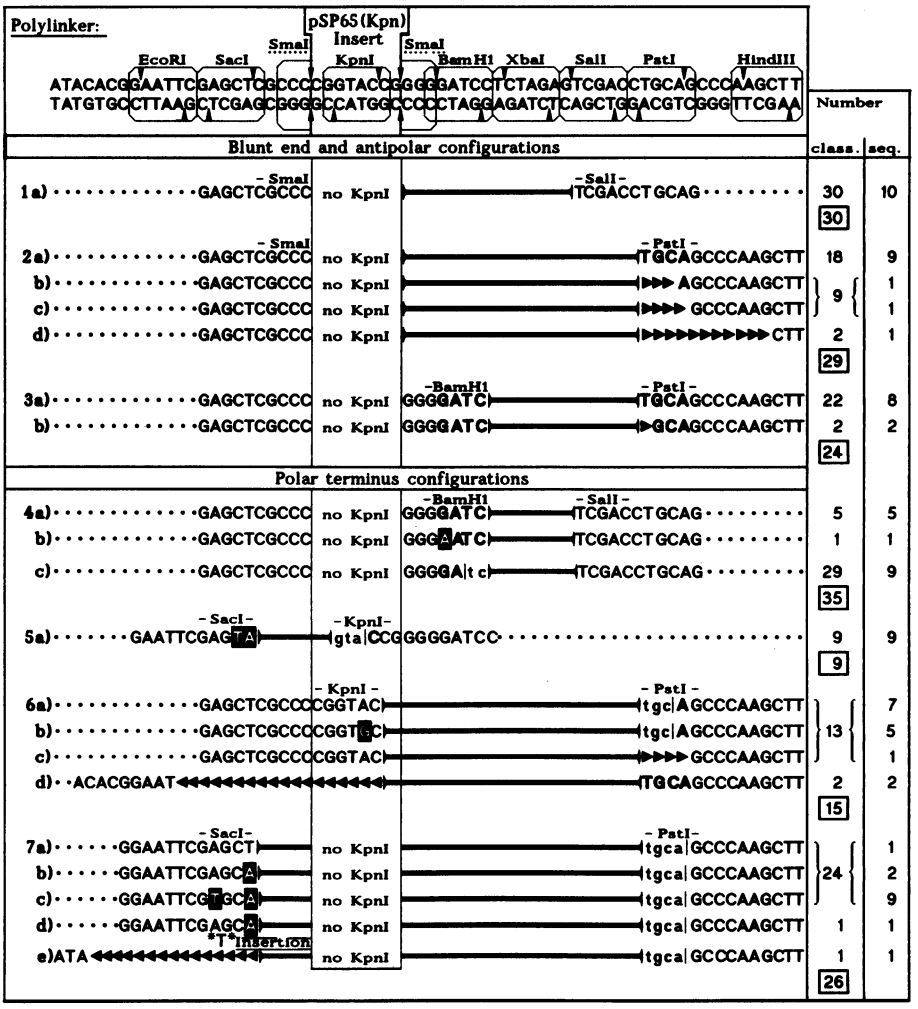


Fig.6 DNA sequences at junctions of the joined products from 7 substrate types. Letters attached to substrate types: definitive subclass assignments. ***Top panel:*** pSP65 polylinker sequence including restriction sites. The KpnI site of derivative pSP65(Kpn) (relevant for substrate types 5 an 6) is additonally inserted in this sequence. Determined sequences are aligned to the polylinker sequence and are therefore interrupted at junctional breakpoints. ***Horizontal bars:*** intervening oligoncleotides (cut out from the polylinker during substrate preparation). ***Capital letters:*** determined sequences. ***Bold letters:*** fill-in positions. ***Lower case letters:*** putative overlaps between PSS of both termini (in complements of the 3' to 5' strand). ***White letters on dark background:*** base substitutions. ***Arrowhead strings:*** deletions. ***Asterisk pairs:*** insertions (in 7d). ***Number class.:*** number of clones/subclass obtained by preclassification procedures (see Fig.5 and Tab.2). ***Number seq.:*** number of actually sequenced clones/subclass.

Tab. 2 Length of polylinker fragments from joined products.

Substrate type	Subclass	Normal length (bp) (*)	Length of joined fragment (bp)	Number per subclass
3) BamH1/PstI	a) (**)	35	35	12
4) BamH1/SalI	a)	41	39	29
	b)		41	5
7) SacI/PstI	a) to c)	23	19	24
	d)		20	1

(*) "Normal length" corresponds to the added lengths of the two excised, residual polylinker fragments on either end of a linear substrate molecule aligned so that ends are just juxtaposed.
(**) Letters indicate subclass assignments (see Fig.5).

Excluding deletions, the remaining subclasses reflect the more interesting junctional sequences generated by the major joining pathways (1a; 2a; 3a; 4a,c; 5a; 6a,b; 7a,b,c). They have been grouped into "nonpolar" substrate types 1 to 3 and "polar" substrate types 4 to 7 (see Fig.2 and Fig.8).

Nonpolar substrate types generate junctions in which the original terminal sequences remain strictly preserved (1a; 2a; 3a) presumably by nucleotide fill-in of PSS ends and subsequent "blunt end" ligation. Polar substrate types generate junctions by a more complex route which includes (i) sequence shortening and (ii) in most cases exchange of bases (4c; 5a; 6a,b; 7a,b,c). The degree of sequence shortening is invariably determined for a given substrate type (except type 4 which produces a minor subclass of junctions with preserved termini). This invariance distinguishes the phenomenon of sequence shortening from the random appearence of deletions. The observed base exchanges appear in a defined and narrow pattern which indicates that they stem from mismatch repair events directly related to sequence shortening. Implications of this feature will be discussed below.

Mismatch correction is a property of the egg extract

DNA from mismatch corrected junctional sequences must pass during cloning procedures through a bacterial system. This might raise doubts as to whether the observed mismatch correction stems from bacteria rather than from the egg extract. Possible interference of bacterial systems was tested by clonal passages through an *E.coli* mutS defecive strain which is virtually unable to correct mismatches in plasmid DNA (34).

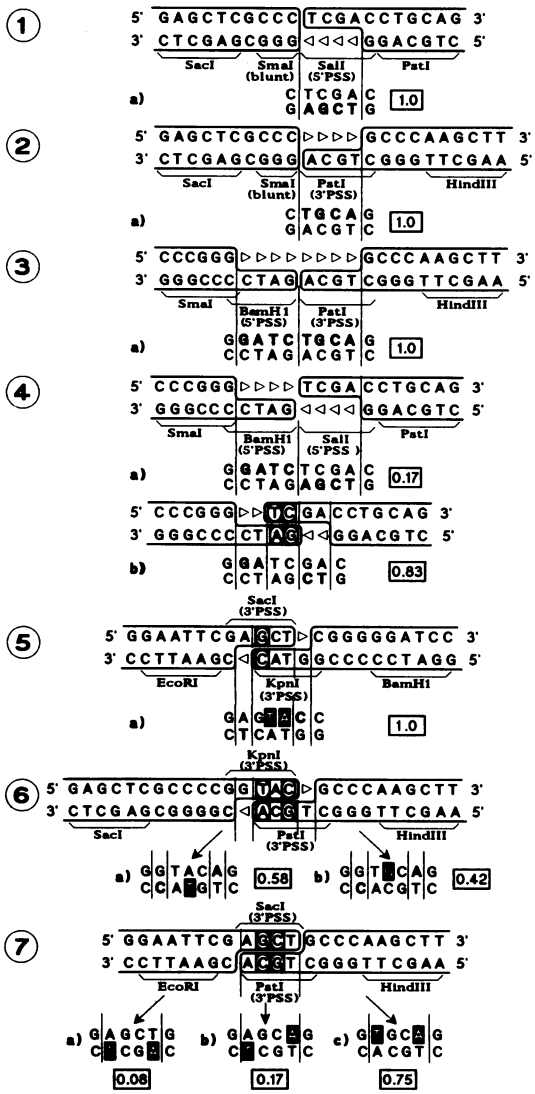


Fig.7 Interpretative models for the generation of junctions summarized for 7 nonhomologous substrate types and their affiliated subclasses (excluding insertions and deletions). ***Boxed numbers:*** Proportions of clones found in given (sub)classes. ***Framed sequences:*** terminus pairs including neighbouring restriction sites. Termini are aligned into a register which corresponds to the shortening of determined junctional sequences. ***Medallion-framed letters:*** matching basepairs between overlapping PSS. ***White arrowhead strings:*** putative fill-in stretches in synthesis direction. ***Unframed sequences:*** determined sequences of junctions and their complements. ***Bold letters:*** fill-in stretches. ***White letters on dark background:*** bases exchanged by mismatch correction.

Uncorrected base mismatches transmitted during transformation of joined DNA-molecules into mismatch repair deficient bacterial hosts segregate upon replication to generate mixed clones. Sequence gels for the DNA of such clones should exhibit two equally strong bands for different bases at adjoined positions, provided, premature division of colonyformers has been avoided prior to plating. Continuous growth arrest monitored during transformation procedures by serial Coulter counts ruled out premature segregation.

On this premise joined DNA of two suitable substrate types was cloned via mutS cells to examine base splits in the DNA sequences. Six clones from type 6 (Fig.7) and seven clones from type 7 (Fig.7) were sequenced to yield subclass configurations in roughly the expected proportion. The main result is that none of the samples exhibited on gels even the faintest indication of bands for two different bases at one sequence position. This precludes measurable transmission of retained base mismatches from joined DNA into the bacterial system. Therefore, the observed mismatch correction is concluded to be a property of the extract itself.

DISCUSSION

Extracts from *X.laevis* eggs provide a selfcontained *in vitro* system which can join nonhomologous termini of duplex DNA. Joining was probed by use of linear plasmid DNAs with dissimilar ends, generated by cuts with different pairs of restriction enzymes. Alltogether 7 model substrates (Fig.2) were offered to the system and these were without exception efficiently joined. DNA sequence analyses of numerous junctions from the available substrate types yielded 20 subclasses of junctional sequences (Fig.6). This information may be used to elucidate features of the joining mechanism itself.

Nonhomologous joining has been studied before *in vivo* by use of linear SV40-DNA (9-11), into which various mismatching terminus pairs had been introduced by appropriate restriction cuts (14,17). One class of junctions in this system creates rare rearrangements and deletions running from junctional breakpoints in both directions (14). In contrast, in our *in vitro* system occasional short deletions (preferentially clustered in substrate type 2) are induced exclusively at 3'PSS ends. Their paucity clearly distinguishes these events from the major junctional sequence subclasses which are considered here and which are summarized in Fig.7.

Substrate types will be discussed according to the structurally different groups, referred to as nonpolar and polar (Fig.8):

(i) Nonpolar substrates (Fig.7:1-3) are joined by simple blunt end ligation

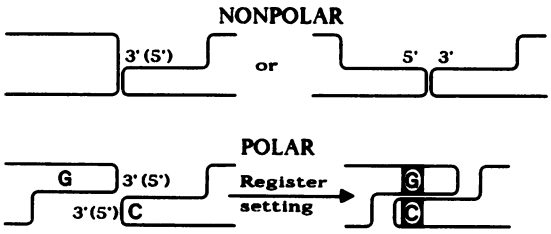


Fig.8 Scheme for substrate type groups: 1) "nonpolar" substrate types with blunt/PSS or 3'PSS/5'PSS terminus configurations and 2) "polar substrate types with 3'PSS/3'PSS or 5'PSS/5'PSS terminus configurations which can overlap by a number of nucleotides determined by matching basepair(s).

which preserves PSS ends by nucleotide fill-in. With interesting deviations (see below), the same was observed for corresponding substrate types *in vivo* in the SV40-system (17). This type of joining requires, in principle, only DNA-ligase and presumably DNA-polymerase β which mainly fills in small DNA stretches (42). In fact, nonhomologous joining in the extract is aphidicolin resistant and highly sensitive towards dideoxy-nucleotides (S.Thode, unpublished).

In contrast to 5'PSS termini (Fig.7:1), 3'PSS termini (Fig.7:2,3) cannot prime nucleotide fill-in. However, ligation of PSS ends in advance of fill-in would produce gapped molecules providing 3'OH primers for fill-in DNA synthesis. Some pro- and eukaryotic DNA-ligases can in fact ligate short PSS ends to blunt ends although at very low rate (43). Thus, it remaines open, how comparably high joining efficiencies observed for substrate types 1 to 3 are achieved in the system. Results in the SV40-System pose the same problem which led to the suggestion of the existence of a novel DNA-ligase activity (17).

(ii) Polar substrates (Fig.7:4-7) reveal a more complex reaction mechanism. During joining, junctional sequences are shortened by a fixed number of nucleotides characteristic for a given substrate type. This suggests a defined overlap interaction between noncomplementary PSS ends of equal polarity rather than mere nucleotide loss. The degree of overlap consistently coincides with pseudohomologous base matches (Fig.7) unique for each substrate type. (Substrate type 5 is an exception that permits besides a GC match a weaker AT match. Use of the AT match was not observed). In no case does the degree of overlap exceed the PSS length. However, in the SV40-system PSS termini frequently invade double-strand regions of terminus partners to generate junctions with up to 17 nucleotides long overlaps associated with pseudohomologous

basematches ranging from singlets to quadruplets (17). Possibly in this *in vivo* system terminal DNA-doublestrand regions are opened up by helicase activities which might be lacking in our *in vitro* system.

In polar 5'PSS substrates blunt end ligation (Fig.7:4a) competes with the strongly preferred overlap reaction (Fig.7:4b), a preference also seen in the SV40-system (14,17). This demonstrates that setting of basematch directed overlaps is a strong and fast reaction which, despite its higher complexity, can compete with and override simple fill-in reactions. Most surprising is the fact that even one single matching basepair, energetically highly unfavorable within a noncomplementary environment (Fig.7:5a), suffices to set the degree of overlap and maintain it during subsequent repair reactions up to final ligation.

Partial noncomplementarity between aligned PSS ends demands restoration of complementarity by mismatch correction. Exclusion experiments demonstrate that mismatch repair is not just simulated by bacterial hosts during cloning procedures but is a property inherent to the extract system itself. *Xenopus* egg extracts apparently contain a complete mismatch repair system (44). Base exchange patterns derived by mismatch correction (Fig.6:5-7) are without exception consistent with sequence patterns of aligned PSS ends. This provides additional strong evidence for the existence of aligned overlapping PSS structures as joining intermediates. Marked, but as yet unexplained preferences in the direction of mismatch correction are observed. It must be assumed that mismatch repair preceeds the mismatch sensitive reactions of fill-in and ligation. Therefore, this type of repair could dispense with incision functions and only execute excision and nucleotide fill-in. This poses the problem of how aligning base matches in overlap PSS configurations can survive excision events. Some junctional repair patterns (Fig.7) indicate irrevocable erasure of base matches (Fig.6:7a,c) which could be prevented in others, provided, excision operates by a rather delicate mechanism (Fig.6:5a; 6a,b; 7b).

Attempts to explain junction formation in egg extracts leaves two main problems unsolved: (i) how can weak, unusual ligation reactions be enhanced and more important, (ii) how is defined overlapping of PSS ends set and maintained by single base matches. We believe that a solution cannot be provided on the basis of pure DNA enzyme interactions alone but rather needs the existence of appropriate terminal DNA binding protein(s) with exceptional structural properties. The mere ability to increase chances for terminal contacts, though useful for ligation enhancement, would not suffice to explain intrajunctional processes like defined overlap setting and maintenance. It is assumed that the

hypothetical protein (i) stabilizes single base matches by hydrophobic protection from water molecules and (ii) bridges terminus structures by DNA contacts to transiently take over maintenance of the alignment. Whether these assumptions hold must be shown by the isolation of the protein(s) involved. Previously, a DNA binding protein has been purified from Xenopus oocytes which drives the ligation of homologous linear DNA in direction of linear concatemer formation (25). Although terminal binding has been considered, it is as yet unknown whether this protein is to any extent involved in nonhomologous joining reactions.

We are convinced that the system described here for nonhomologous joining is part of a repair network related to illegitimate recombination from which we visualised only these aspects amenable to analysis by our test system. It is obvious that the biological task of the joining mechanism is not to primarily handle restriction cuts, but to accept other and perhaps more complex terminus configurations. Thus, the actual capacities of this intriguing system are still open for further analysis.

ACKNOWLEDGEMENTS

This work has been initiated during a stay of P.Pfeiffer in the laboratory of W.Schaffner, Zurich. We are grateful to him and S.Rusconi for grand hospitality, advise, many stimulating discussions and encouragement to enter the problem. We thank S.Thode and R.Kueppers for valuable experimental contributions, U.Zylajew for excellent technical assistance and B.Kemper and W.Wille for critical comments. We would also like to thank R.Jack for critically reading the manuscript and H.-J.Fritz (MPI, Munich) for the gift of the mutS strain including helpful advise. The "Deutsche Akademische Austauschdienst" (DAAD) provided a travel grant to P.Pfeiffer under the program 312-gen-dk.

REFERENCES

1. Bryant, P.E. (1986) In Simic, M.G., Grossman, L. and Upton, A.C. (eds), Mechanisms of DNA Damage and Repair, Plenum Press, New York, 171-180.
2. Krasin, F. and Hutchinson, F. (1977) J. Mol. Biol. **116**, 81-98.
3. Game, J.C. and Mortimer, R.K. (1974) Mutat. Res. **24**, 281-292.
4. Szostak, J.W., Orr-Weaver, T.L., Rothstein, R.J. and Stahl, F.W. (1983) Cell 33, 25-35.
5. Stahl, F.W., Kobayashi, I. and Stahl, M.M. (1985) J. Mol. Biol. **181**, 199-209.
6. Symington, L.S., Morrison, P. and Kolodner, R. (1985) Mol. Cell. Biol. 5, 2361-2368.
7. Kolodkin, A.L., Klar, A.J.S. and Stahl, F.W. (1986) Cell **46**, 733-740.
8. Johnson, A.D., Barkan, A. and Mertz, J.E. (1982) Virology **12**, 464-469.
9. Wilson, J.H., Berget, P. and Pipas, J. (1982) Mol. Cell. Biol. 2, 1258-1269.
10. Wake, C.T., Gudewicz, T., Porter, T., White, A. and Wilson, J.H. (1984) Mol. Cell. Biol. **4**, 387-398.
11. Roth, D.B. and Wilson, J.H. (1985) Proc. Natl. Acad. Sci. USA **82**, 3355-3359.

12. Shapiro, J.A. (1983) Mobile Genetic Elements, Acad. Press, New York.
13. Weinberg, R.A. (1980) Ann. Rev. Biochem. **49**, 197-226.
14. Roth, D.B., Porter, T.N. and Wilson, J.H. (1985) Mol. Cell. Biol. **5**, 2599-2607.
15. Arrand, J.E., Willis, A.E., Goldsmith, I. and Lindahl, T. (1986) J. Biol. Chem. **261**, 9079-9082.
16. Gerondakis, S., Cory, S. and Adams, J.M. (1984) Cell **36**, 973-982.
17. Roth, D.B. and Wilson, J.H. (1986) Mol. Cell. Biol. **6**, 4295-4304.
18. Subramani, S. and Berg, P. (1983) Mol. Cell. Biol. **3**, 1040-1052.
19. Kopchick, J. and Stacey, D.W. (1984) Mol. Cell. Biol. **4**, 240-246.
20. Ashman, C.R. and Davidson, R.L. (1984) Mol. Cell. Biol. **4**, 2266-2272.
21. Folger, K.R., Thomas, K. and Capecchi, M.R. (1985) Mol. Cell. Biol. **5**, 59-69.
22. Benbow, R.M. and Krauss, M.R. (1977) Cell **12**, 191-204.
23. Bendig, M.M. (1981) Nature **292**, 65-67.
24. Rusconi, S. and Schaffner, W. (1981) Proc. Natl. Acad. Sci. USA **78**, 5051-5055.
25. Bayne, M.L., Alexander, R.F. and Benbow, R.M. (1984) J. Mol. Biol. **172**, 87-108.
26. Carroll, D., Wright, S.H., Wolff, R.K., Grzesiuk, E. and Maryon, E.B. (1986) Mol. Cell. Biol. **6**, 2053-2061.
27. Newport, J. and Kirshner, M. (1982a) Cell **32**, 675-686.
28. Glikin, G.C., Ruperti, I. and Worcel, A. (1984) Cell **37**, 33-41.
29. Melton, D.A., Krieg, P.A., Rebagliati, M.R., Maniatis, T., Zinn, K. and Greene, M.R. (1984) Nucl. Acid Res. **12**, 7035-7056.
30. Boyer, H.W. and Roulland-Dussoix, D. (1969) J. Mol. Biol. **41**, 459-472.
31. Southern, E. (1975) J. Mol. Biol. **98**, 503-517.
32. Maniatis, T. Fritsch, E.F. and Sambrock, J. (1982) Molecular Cloning - a Laboratory Manual, Cold Spring Harbor Laboratory, New York.
33. Wahl, G.M., Stern, M. and Stark, G.R. (1979) Proc. Natl. Acad. Sci. USA **76**, 3683-3687.
34. Kramer, W. and Fritz, H.-J. (1987) Meth. in Enzymol. in press.
35. Modrich, P. (1987) Ann. Rev. Biochem. **56**, 435-466.
36. Crouse, G.F., Frischauf, A. and Lehrach, H. (1983) Meth. in Enzymol. **101**, 78-89.
37. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA **74**, 5463-5467.
38. Seedorf, K., Krämmer, G., Durst, M., Suhai, S. and Röwekamp, W. (1985) Virology **145**, 181-185.
39. Grzesiuk, E. and Carroll, D. (1987) Nucl. Acid. Res. **15**, 971-985.
40. Dugaiczyk, A. Boyer, H.W. and Goodman, H.M. (1975) J. Mol. Biol. **96**, 171-184.
41. Leach, D. and Lindsey, J. (1986) Mol. Gen. Genet. **204**, 322-327.
42. Mosbaugh, D.W. and Linn, S. (1984) J. Biol. Chem. **259**, 10247-10251.
43. Wiaderkiewicz, R. and Ruiz-Carrillo, A. (1987) Nucl. Acid Res. **15**, 7831-7848.
44. Brooks, P. and Radman, M. (1987): Mismatch repair by *Xenopus* eggs; abstract from the July 1987 Rome meeting.