4946–4953 *Nucleic Acids Research, 1996, Vol. 24, No. 24*

The joining of non-complementary DNA double-strand breaks by mammalian extracts

Rebecca M. Mason[+], John Thacker and Micaela P. Fairman*

DNA Repair and Mutagenesis Group, MRC Radiation and Genome Stability Unit, Harwell, Didcot, Oxon OX11 ORD, UK

Received August 15, 1996; Revised and Accepted November 4, 1996

ABSTRACT

We have developed a high efficiency system in which mammalian extracts join DNA double-strand breaks with non-complementary termini. This system has been used to obtain a large number of junction sequences from a range of different break-end combinations, allowing the elucidation of the joining mechanisms. Using an extract of calf thymus it was found that the major mechanism of joining was by blunt-end ligation following removal or fill-in of the single-stranded bases. However, some break-end combinations were joined through an efficient mechanism using short repeat sequences and we have succeeded in separating this mechanism from blunt-end joining by the biochemical fractionation of extracts. Characterization of activities and sequence data in an extensively purified fraction that will join ends by the repeat mechanism led to a model where joining is initiated by 3′ strand invasion followed by pairing to short repeat sequences close to the break site. Thus the joining of double-strand breaks by mammalian extracts is achieved by several mechanisms and this system will allow the purification of the factors involved in each by the judicial choice of the non-complementary ends used in the assay.

INTRODUCTION

The processes of DNA replication and transcription rely on the presence of a continuous DNA template. For this reason double-strand breaks (dsb) in the DNA helix can have catastrophic consequences for the cell if they are not rapidly repaired. Previous work has shown that in lower eukaryotes the presence of a single unrepaired dsb can be lethal (1), even when the break is in a piece of DNA that is not essential for viability (2). Additionally, in higher eukaryotes, agents that cause a high proportion of dsbs, such as ionizing radiation, are effective in causing genetic changes such as complex mutations and chromosomal aberrations (3,4). For such reasons the efficient repair of dsbs is crucial for the maintenance of genomic stability in all organisms.

In lower eukaryotes such as the yeast *Saccharomyces cerevisiae*, the major pathway for the removal of dsbs appears to be through the mechanism of homologous recombination (5). However, in higher eukaryotes many studies have shown that dsbs are commonly joined through end-to-end mechanisms by a variety of processes termed non-homologous or illegitimate recombination. The majority of these studies have utilized site-specific DNA dsbs produced by cutting DNA with restriction enzymes to produce DNA termini with defined end structures. This has been achieved by either transfection of linear DNA into cultured cells (6–9) or direct electroporation of enzymes into cells (10,11). Additionally, the rejoining of linear plasmids after incubation in cell-free extracts from either mammalian sources (12–16) or from *Xenopus laevis* (17–20) has been studied. In all cases, it has been shown that the majority of plasmids cut to produce complementary ends are joined faithfully by cells or cell-free extracts from normal tissue sources.

However, to examine the pathways of joining used when such ends are not available, as in the breaks caused by DNA damaging agents, junctions produced by the joining of non-complementary ends have also been examined. In general, three major mechanisms have been described. The first (fill-in mechanism) requires the filling in of single-stranded overhangs followed by blunt-end ligation. The second (overlap mechanism) appears to proceed through the pairing of one or more bases in the termini, followed by filling in of the missing bases combined with correction of the mismatched bases. The third mechanism (repeat mechanism) brings together short regions of homology with subsequent removal of the displaced single-stranded sequences. However, although these three mechanisms have been found in a variety of different systems, the proportions of each varies widely. In addition, no attempt has been made to date to identify or purify the proteins involved in any of these pathways.

Previous studies of non-complementary end joining have used a strategy where a single DNA molecule is restricted with two separate enzymes and the (circular) products analysed after transfection into bacteria. However, a large fraction of joined products occur as linear dimers and multimers rather than circular monomeric products and it is possible that the analysis of circular products alone may detect only joining mechanisms that occur in conjunction with the process of recircularization. In addition, our extract fractionation studies have shown that many partially purified fractions lost the ability to form circular products, while the formation of linear forms was not affected. Therefore, to avoid both these problems we developed a system based on the formation of linear dimers.

*To whom correspondence should be addressed. Tel: +44 1235 834393; Fax: +44 1235 834776

[+]Present address: Department of Radiation Genetics and Chemical Mutagenesis, Sylvius Laboratory, Wassenaarseweg 72, 2333 AL Leiden, The Netherlands

In this report we present a systematic analysis of the junctions in linear molecules produced from DNA with a variety of non-complementary end termini by mammalian extracts. In addition we present the extensive purification and characterization of an activity that joins non-complementary ends through a mechanism that utilizes short repeat sequences.

MATERIALS AND METHODS

Calf thymus extract preparation

Thymus glands (500 g) from newly slaughtered calves (3–6 months old) were frozen in liquid nitrogen and stored at –70°C. Protease inhibitors (1 µg/ml pepstatin A and 1 µg/ml leupeptin) were added to all buffers directly before use and all steps were carried out at 4°C. The frozen thymus was disrupted in a Waring blender in 750 ml homogenization buffer (100 mM Tris–HCl, pH 7.5, 20 mM phosphate buffer, pH 7.5, 100 mM KCl, 1 mM EDTA, 1 mM DTT, 1 mM PMSF) until homogeneous. This was stirred for 30 min and centrifuged at 10 000 *g* for 30 min. The supernatant was filtered through gauze and added to 1 l phosphocellulose resin (100 g dry weight) which had been prepared according to the manufacturer's instructions (Whatman) and equilibrated in buffer A (25 mM Tris–HCl, pH 7.5, 1 mM EDTA, 0.01% Nonidet P-40, 10% glycerol, 1 mM DTT, 1 mM PMSF) containing 25 mM NaCl. The slurry was stirred for 15 min and filtered through Whatman No. 1 paper. The cake of resin was resuspended in ~600 ml buffer A/150 mM NaCl, restirred and filtered and the protein eluted with 600 ml buffer A/600 mM NaCl. To this wash was added solid ammonium sulphate to 40% saturation and the precipitated protein recovered by centrifugation at 10 000 *g* for 30 min. The supernatant was adjusted to 70% saturation and the proteins precipitated and recovered as before. For activity assays this pellet (Fraction I) was resuspended in buffer A/25 mM NaCl, dialysed overnight against the same buffer and stored as small aliquots at –70°C. For fractionation studies, the ammonium sulphate pellet was stored at –70°C.

Fractionation of the calf thymus extract

All columns were run at 4°C and all dialysis steps were against buffer A/25 mM NaCl for 14–17 h. Fraction I was dissolved in buffer A/25 mM NaCl (~20 mg/ml) and loaded onto a Superose 6 FPLC column (Pharmacia). Pooled active fractions (II) were loaded onto a native DNA cellulose (US Biochemicals) column (4 mg/ml resin) in the same buffer and the protein eluted at 200 mM and 1 M NaCl. After dialysis the active fraction (1 M) was adjusted to buffer A/0.9 M ammonium sulphate/25 mM NaCl and loaded onto a FPLC Phenyl Superose column (Pharmacia) in the same buffer. The column was eluted in steps and the activity (Fraction IV) eluted in buffer A/25 mM NaCl. This was dialysed against the same buffer containing 20% sucrose and loaded onto a FPLC Mono S ion exchange column (Pharmacia). Proteins were eluted using a gradient of buffer A containing 25–400 mM NaCl. Activity from this column could only be detected after the addition of T4 DNA ligase to the joined products before electrophoresis (see below).

DNA substrates

pUC18 plasmid was propagated in *Escherichia coli* HB101, isolated using alkali lysis and purified by double banding on caesium chloride/ethidium bromide gradients. Plasmids were linearized with restriction enzymes as indicated and the digestion monitored by agarose gel electrophoresis. Bacteriophage λ DNA was restricted with *Kpn*I, the unique 1.5 kb fragment isolated and cloned into the multicloning site of pUC18. This plasmid (pBK-1) was then grown in *E.coli* HB101 and isolated as described above. The 1.5 kb λ fragment was released by digestion with *Kpn*I and purified by gel isolation.

Joining reactions

Reactions (25 µl) contained 70 mM Tris–HCl, pH 7.5, 10 mM $MgCl_2$, 10 mM DTT, 4 mM ATP, 40 mM phosphocreatine, 1 µg creatine phosphokinase, 100 µM each dATP, dCTP, dGTP and dTTP, 45 pM both linear pUC18 (1 µg) and λ/*Kpn*I (0.55 µg) and protein extracts as indicated. Reactions were set up on ice and incubated for 15 min at 37°C unless otherwise stated. Reactions were terminated by addition of an equal volume of 100 mM EDTA/1% SDS and the DNA purified as described previously (14). Where indicated, the precipitated DNA products were resuspended in 1× One-Phor-All buffer (Pharmacia) and incubated with 2 U T4 DNA ligase for 10 min at 37°C before addition of gel loading buffer (2.5% Ficoll, 0.025% bromophenol blue). In all cases the joined products were separated by agarose gel electrophoresis (1%) in 1× TBE (89 mM Tris base, 89 mM orthoboric acid, 2 mM EDTA) containing 0.5 µg/ml ethidium bromide for 15–17 h at 2.25 V/cm.

Detection and quantification of joined products

For calculation of specific activities (Table 2), an image of the gel was produced using a charge coupled device (CCD) camera and the image was stored as a TIFF file (Apple Macintosh). The relative amount of pUC/λ DNA forms was calculated as a percentage of the unjoined pUC + pUC/λ forms using the Collage software package (Fotodyne). For gels in which the joined product was not visible by this method, the gels were blotted onto reinforced nitrocellulose (Sartorius) and probed using the Enhanced Chemiluminescence (ECL; Amersham) kit with pUC18 labelled with horseradish peroxidase and hybridized according to the manufacturer's instructions. Blots were washed at a final stringency of 0.4% SDS at 55°C for 10 min and exposed to X-ray film at –70°C.

Purification of the joined products

To obtain sufficient material for further analysis, multiple joining reactions (2–8) were performed and the reactions combined before gel electrophoresis. Sequence analysis was performed on joined products produced by several different extracts in separate experiments. The product of non-complementary joining was excised from the gel and purified using Gene Clean II (Bio 101) according to the manufacturer's instructions.

PCR junction amplification and sequence analysis

As two linear molecules are used in the reaction, there are four different orientations in which the ends could combine (Fig. 1). To be able to study these four orientations separately, PCR primers were designed which would selectively amplify the junction in each orientation in each case. To amplify the junction produced in orientation 1, the forward pUC primer 196 (5′-CTTCGCTATTA-CGCCAGCTGGCGA) was used in combination with the λ

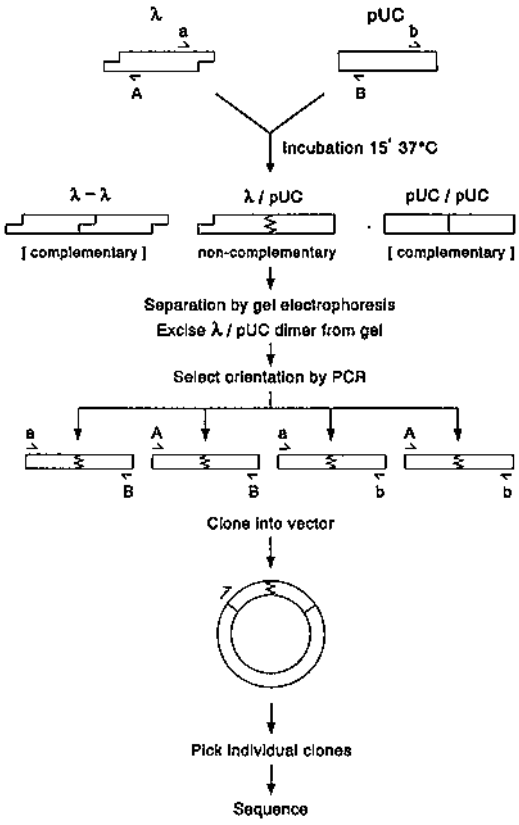

Figure 1. Schematic of the production and analysis of non-complementary end joining. The λ/*Kpn*I fragment and pUC, linearized with an appropriate enzyme, were incubated with cellular extracts to produce a mixture of products joined by complementary or non-complementary termini. These were separated by gel electrophoresis and the λ/pUC band excised. The four possible junction orientations were selected by PCR using different combinations of primers as indicated. PCR products were directly cloned and sequenced.

primer LK3 (5′-GCAATCCTCCGGCCTTTTCCCTGA). For orientation 2, primer 196 was used in combination with the λ primer LK4 (5′-ACGTGCCGTCGTCGTTCTCACGA). For orientation 3, the reverse pUC primer RC (5′-TTGTGAGCGGA-TAACAATTTC) was used with LK3 and orientation 4 was amplified using the primers RC and LK4. PCR reactions (25 or 50 µl) contained 10 mM Tris–HCl, pH 8.3, 1.5 mM $MgCl_2$, 50 mM KCl, 200 µM dATP, dCTP, dGTP and dTTP, 1.5% DMSO, 400 nM each primer, 1–5 ng DNA and 1 U Taq polymerase (Boehringer Mannheim). Routinely, the initial denaturation step was at 94°C for 5 min followed by 30 cycles of 55°C for 30 s, 72°C for 1 min, 94°C for 30 s, followed by a final cycle of 55°C for 30 s, 72°C for 5 min. Aliquots of 1–2 ng of PCR products were cloned into pCRII using a TA Cloning kit (Invitrogen), single transformants picked and the orientation of the cloned fragment checked by PCR analysis. Single-stranded DNA was prepared from suitable clones using 5×10^9 p.f.u. of the helper phage M13K07 (Pharmacia) and sequenced using a USB Sequenase 2 kit (Amersham) according to the manufacturer's instructions.

Other procedures

SDS–PAGE was performed according to Laemmli (21) and gels stained using a Silver Stain Plus kit (BioRad). Western blots were treated and immunodetected using the ECL system (Amersham); antibodies were checked against a positive control (human 293 cell extract) and shown to cross-react with Fraction I of calf thymus extracts. Protein concentrations were measured using the assay of Bradford (22) using BSA as a standard.

Nuclease assays

To detect 3′→5′ exonuclease activity, labelled templates were made by filling in pUC18/*Xba*I with Klenow fragment and [α-^{32}P]CTP. To detect 5′→3′ exonuclease activity, aliquots of the pUC18/*Sac*I or the λ/*Kpn*I fragment used for joining reactions were dephosphorylated with calf intestinal phosphatase and end-labelled with [γ-^{32}P]ATP and T4 polynucleotide kinase using standard protocols (23). Reactions (80 µl) contained 20 000 c.p.m. of appropriately labelled template in 70 mM Tris–HCl, pH 7.5, 10 mM DTT, 10 mM $MgCl_2$, 40 µg/ml BSA and various volumes of column fractions. Aliquots of each reaction were removed at 0 and 60 min and stopped by addition of an equal volume of 20 mM EDTA/1 µg/ml sheared salmon sperm DNA. The DNA was precipitated with 1 ml 8% TCA/1% sodium pyrophosphate on ice for 30 min, filtered onto GF-C filters and counted in a scintillation counter.

RESULTS

Joining assay

To ensure that joining was not occurring by homologous recombination, reactions contained two DNA substrates that were unrelated (Fig. 1). One substrate was pUC18 (pUC), chosen for the availability of unique restriction sites (*Sac*I, *Eco*RI, *Hin*cII, *Pst*I, *Hin*dIII and *Bam*HI) in the multicloning site. The second substrate, a *Kpn*I–*Kpn*I fragment from bacteriophage λ, was selected because of its convenient size (1.5 kb) and absence of cross-reactivity to pUC as determined by Southern analysis. It was possible to monitor the joining of DNA with complementary and non-complementary ends concurrently from the different sizes of products. The major product of non-complementary joining is the λ/pUC dimer (4.1 kb), while dimers of complementary joining were 5.2 and 3 kb for the pUC/pUC and λ/λ dimers respectively. Probing the resulting Southern blots with pUC increased the sensitivity of the assay and simplified the identification of the various joined forms (see Fig. 2).

Optimization of Fraction I

Unfractionated extracts from calf thymus would not join substrates with non-complementary termini and the extract preparation was modified to include a batch elution step from phosphocellulose resin and a 40–70% precipitation using ammonium sulphate (Materials and Methods) to give Fraction I. This fraction was then incubated with λ/*Kpn*I and pUC restricted with various enzymes to give either 3′-3′, 3′-5′ or 3′-blunt combinations to estimate the amount of λ/pUC dimer produced. The greatest amount of λ/pUC dimer was produced by the combination of λ/*Kpn*I fragment with pUC/*Sac*I and this combination was used to determine optimal joining conditions for Fraction I. Production of λ/pUC dimer increased with protein concentration to 10 µg and a time course study showed an optimal incubation time of 15 min at 37°C (data not shown). Although longer incubations and higher extract concentrations resulted in the formation of more product, nuclease activity also increased,

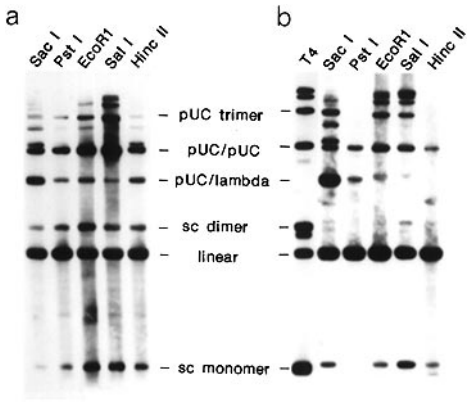

Figure 2. Southern blots hybridized to pUC18, showing the joining activity of calf thymus fractions on non-complementary termini. (**a**) Fraction I (10 µg/reaction); (**b**) Fraction V (400 ng/reaction). Reactions were as in (a) but the reacted products were treated with 2 U T4 DNA ligase before electrophoresis (see Materials and Methods). Lane 1 shows substrates treated with 2 U T4 DNA ligase alone, showing the absence of non-complementary product. The identity of the different forms is indicated. sc, supercoiled DNA.

making dimer production difficult to quantitate. Using these optimal conditions, Fraction I was found to join all the end combinations with similar efficiency; however, the *Kpn*I/*Sac*I combination routinely gave ~3-fold more non-complementary joined product than any of the other combinations tested (Fig. 2a).

Sequence analysis of products joined by Fraction I

Dimers were gel purified, amplified by PCR and cloned to determine by sequencing the mechanisms of joining of the different end combinations by Fraction I (Table 1a). The *Kpn*I/*Sac*I (3′-3′), *Kpn*I/*Eco*RI (3′-5′) and *Kpn*I/*Hin*cII (3′-blunt) combinations were assayed in each of the four orientations (see Materials and Methods and Fig. 1), while *Kpn*I/*Pst*I (3′-3′), *Kpn*I/*Hin*dIII (3′-5′) and *Kpn*I/*Bam*HI (3′-5′) termini were analysed in a single orientation. It is clear that the predominant mechanism used for the joining of non-complementary termini proceeded by a final step of blunt-end ligation. This was achieved by either removal of the single-stranded 3′ regions and/or by the fill-in of the 5′ overhangs respectively (shown schematically in Fig. 3a). Moreover, this mechanism was observed in all four orientations for the 3′-5′ (*Kpn*I/*Eco*RI) and 3′-blunt (*Kpn*I/*Hin*cII) combinations, indicating that the DNA sequence adjacent to the termini had little effect on the mechanism used. However, the combination that was joined with the greatest efficiency (*Kpn*I/*Sac*I), showed a more varied pattern, with only one of the four orientations (no. 3) joined primarily by this mechanism. For the remaining three orientations (nos 1, 2 and 4) the sequences obtained indicated that the majority of junctions (55–100%) had been joined through a homology-dependent mechanism involving the pairing of short (2–6 bp) repeat sequences. A model proposed for this repeat-driven mechanism (13) is shown in Figure 3b. This variability in mechanism observed between the orientations indicates that for this combination, the sequence around the termini has a marked affect on the mechanism used. The sequences at each junction obtained for *Kpn*I/*Sac*I and the positions and lengths of the repeat sequences used in each orientation are shown schematically in Figure 4. For all

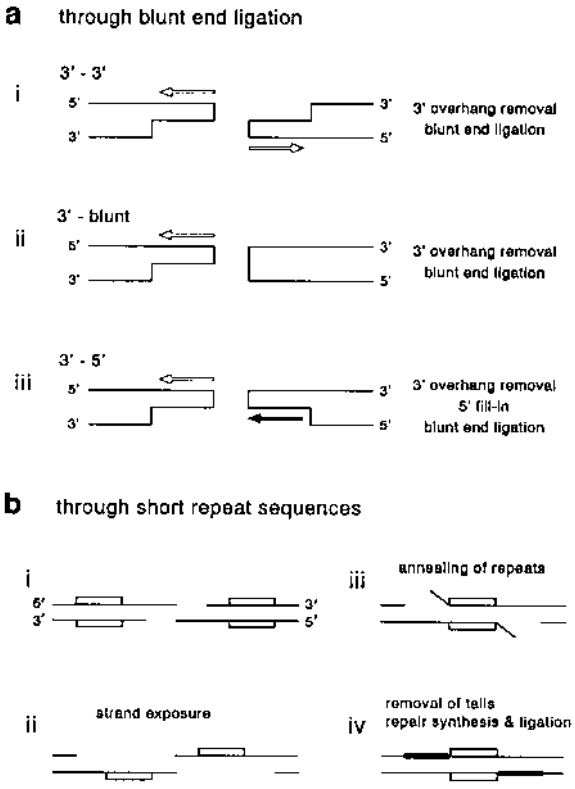

Figure 3. Schematic representation of the major mechanisms found by sequence analysis. (**a**) Joining using a final step of blunt-end ligation. (i) 3′-3′ combinations showing removal of both 3′ termini (white arrows) followed by ligation. (ii) 3′-blunt combinations showing removal of the 3′ termini as in (i) above. (iii) 3′-5′ combinations showing removal of the 3′ termini and fill-in of the 5′ termini (black arrow). (**b**) Joining using short repeat sequences. Model of joining through short regions of homology showing exposure of the repeat sequences followed by pairing, removal of the single-strand tails and repair synthesis.

combinations of ends used, a small percentage of junctions was found that did not fall into either of these two categories (termed ‘other’ in Table 1). Such sequences showed small deletions that had no homology at the junction site. Interestingly, using this system, no sequences were found that had been joined using the overlap mechanism.

Characterization of Fraction I

Using the *Kpn*I/*Sac*I combination, the reaction produced only a small amount of the λ/pUC dimer in the absence of added ATP and was optimal at a concentration of 4–6 mM. Addition of NaCl to the reaction caused a 50% reduction in all joined products at a concentration of 50 mM and complete inhibition at 150 mM NaCl. All joining activity was abolished by heating the extract for 10 min at 100°C.

To assess the involvement of DNA polymerases, reactions containing Fraction I and the end combinations *Kpn*I/*Sac*I (3′-3′), *Kpn*I/*Eco*RI (3′-5′) or *Kpn*I/*Hin*cII (3′-blunt) were assayed in the absence of added deoxynucleotide triphosphates (dNTPs). Only the 3′-5′ *Kpn*I/*Eco*RI combination was dependent on the addition of dNTPs, indicating the involvement of a DNA polymerase, in agreement with the mechanism suggested from the sequence analysis [shown schematically in Fig. 3a(iii)]. Further, the *Kpn*I/*Eco*-

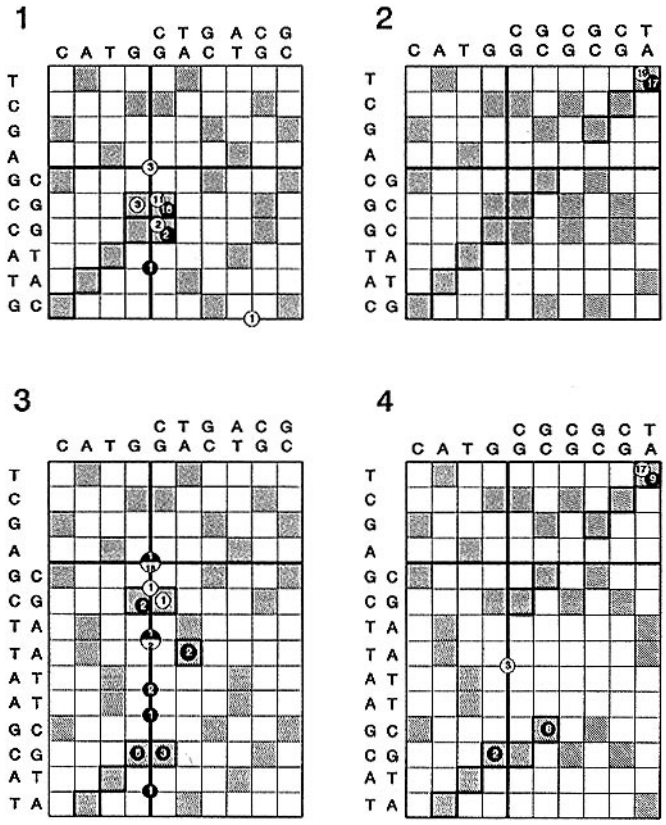

Figure 4. Sequences of the junctions produced using the combination of λ/*Kpn*I and pUC/*Sac*I with Fractions I and V in all orientations. The nucleotide sequences of the pUC and λ fragments are shown along the *x* and *y* axes respectively. Strands drawn at the outside edge are 5′→3′ from bottom to top and left to right. 1–4 indicate the four orientations as defined in the text. Heavy lines indicate the transition from single- to double-stranded sequence. The possible repeat sequences in each case are indicated by shaded squares. The number of junctions formed by each mechanism is indicated in either white or black circles for Fraction I and Fraction V respectively. Junctions that are formed using repeats are indicated by a circle inside a shaded square and the length of the repeat represented by a black border. Junctions formed without repeats are indicated by a circle at the intersection of the gridlines.

RI combination was assayed in the presence of a concentration of either aphidicolin (up to 200 µg/ml) or a dTTP:ddTTP ratio (up to 1:8) which would cause inhibition of DNA polymerases α, γ and δ or of polymerase β respectively. Such reactions were unaffected by the presence of aphidicolin but were 50% inhibited at dTTP:ddTTP ratios of 1:1 and >80% inhibited by ratios >1:4. These results indicate that, as expected, the only end combination that is joined using a fill-in mechanism is affected by the presence of DNA polymerase inhibitors and that the pattern of inhibition observed is consistent with the involvement of DNA polymerase β (24).

In an attempt to identify the DNA ligase(s) involved, adenylation assays were performed across the Superose 6 fractions of Fraction I which separated the majority of the complementary from the non-complementary end joining activities (see below). Although adenylation products were observed, these did not correspond to either peak of joining activity.

Production of Fraction V

The *Kpn*I/*Sac*I combination was used as an assay to purify the proteins involved (Materials and Methods), as this combination was joined predominantly by the repeat mechanism. Using a sizing column as a first step, it was possible to separate the majority of the complementary end joining activity, which eluted as an early peak, from the majority of the non-complementary activity (data not shown). These fractions, however, were highly labile and had to be directly loaded onto the next column (DNA–cellulose) to retain activity. After separation on Phenyl Superose, there was usually a large drop in joining activity which could be restored by treatment of the products with T4 DNA ligase before electrophoresis (see Materials and Methods). As T4 DNA ligase cannot produce the λ/pUC dimer product itself (see Fig. 2b, lane 1), this result strongly indicates that all the enzymes required to produce a ligatable product are present but that the amount of DNA ligase is limiting. The fraction from the Mono S column that gave the greatest amount of activity (Fraction V) contained 15–20 µg protein and had a sp. act. of 80 000 U/mg protein (Table 2). Fractions were stable for 1 week at –70°C, but were inactivated by more than one freeze/thaw cycle.

Table 1a. Mechanisms used by Fraction I in joining λ/*Kpn*I to pUC18 restricted with various endonucleases

Termini	Orientation	Mechanism (%)			Total
		Blunt	Repeats[a]	Other	
*Kpn*I/*Sac*I	1	15	55 + 25	5	20
	2	0	100	0	19
	3	79	0 + 5	16	19
	4	0	85	15	20
*Kpn*I/*Eco*RI	1	100	0	0	3
	2	78	0 + 11	11	9
	3	100	0	0	3
	4	100	0	0	1
*Kpn*I/*Hinc*II	1	75	0	25	4
	2	75	0 + 25	0	4
	3	67	0 + 33	0	6
	4	100	0	0	7
*Kpn*I/*Pst*I	3	80	0 + 10	10	10
*Kpn*I/*Hin*dIII	3	100	0	0	7
*Kpn*I/*Bam*HI	3	100	0	0	7

[a]Divided into repeats >1 bp + 1 bp repeats.

Table 1b. Mechanisms used by Fraction V to join the *Kpn*I/*Sac*I substrates

Fraction	Orientation	Mechanism (%)			Total
		Blunt	Repeats[a]	Other	
V	1	0	84 + 10	6	19
V	2	0	100	0	17
V	3	5	28 + 39	28	18
V	4	0	100	0	19

[a]Divided into repeats >1 bp + 1 bp repeats.

Table 2. Purification table showing joining activity using the *Kpn*I/*Sac*I substrates

Fraction	Column/step	Protein (mg)	Total activity (U[a])	Sp. act. (U/mg)
I	Ammonium sulphate (40–70%)	100	5000	50
II	Superose 6	7.5	1071	143
III	Native DNA cellulose	0.65	195[b]	300
IV	Phenyl Superose	0.115	115[c]	1 000
V	Mono S	0.017	1360[d]	80 000

[a]1 unit is defined as 10% λ/pUC dimer product (see Materials and Methods for quantitation).
[b]Fractions assayed at saturation of assay, therefore total units underestimated.
[c]Fractions assayed in the absence of added DNA ligase, therefore total units underestimated.
[d]Joined products treated with T4 DNA ligase (see text).

Activity of Fraction V and sequence analysis of joined products

As seen in Figure 2b, an optimal concentration of Fraction V produced a strong rejoin reaction for the *Kpn*I/*Sac*I combination, while only a small amount of product was observed for the other 3′-3′ combination used (*Kpn*I/*Pst*I). No λ/pUC dimer product was observed for either the 3′-5′ (*Kpn*I/*Eco*RI, *Kpn*I/*Sal*I) or the 3′-blunt (*Kpn*I/*Hin*cII) combinations.

Sequence analysis was limited to the *Kpn*I/*Sac*I combination, since insufficient joined product was produced by the other combinations used. Compared with Fraction I, a larger proportion of orientations 1, 2 and 4 were joined through short repeats and orientation 3 now showed an increased use of the repeat mechanism (use of a 3 bp repeat and increased use of 1 bp 'repeats') as well as random deletions (Fig. 4). A comparison of the repeat sequences used for all four orientations by Fractions I and V is summarized in Figure 6a.

Characterization of Fraction V

Using the *Kpn*I/*Sac*I combination, a comparison of Fractions I and V showed that both fractions were equally sensitive to NaCl and both showed maximum joining at 37°C, with no product at 16°C. Assay of Fraction V in the absence of ATP resulted in the formation of a small amount of λ/pUC dimer, which could be increased to an optimum at 1 mM. Addition of the polymerase inhibitors aphidicolin or ddTTP had no effect. ATPase activity was not observed in Fraction V in the presence or absence of added DNA. DNA binding assays performed across the final fractions showed activity in a wider peak than that of non-complementary joining activity on both single- and double-stranded templates (data not shown). No DNA topoisomerase I and II activity was observed in Fraction V, as assayed by the relaxation of supercoiled plasmid or the decatenation of kDNA respectively (25). Immunoblots of fractions across the Mono S column showed no reaction with antibodies raised against the end-binding protein Ku70 or the structure-specific nucleases FEN-1, XP-G and ERCC1, although these proteins were detected in Fraction I and some were detected with partial purification (data not shown). Monitoring the conversion of closed circle to open circle pUC indicated the absence of any gross endonuclease activity. However, as the mechanism of joining suggests that a nuclease activity would be required to remove the single-stranded overhangs, fractions across the final Mono S column were tested on appropriately labelled templates to detect the presence of 3′→5′ or 5′→3′ exonucleases respectively (see Materials and Methods). In all preparations, weak 3′→5′ and 5′→3′ exonuclease activities which peaked with the joining activity could be detected (data not shown). Aliquots of the same fractions assayed for joining activity were separated by SDS–PAGE and detected by silver staining. Although several protein bands are contained in the most active fractions, the band that most closely follows the distribution of activity in separate preparations from calf thymus indicates a protein of ~73 kDa (Fig. 5).

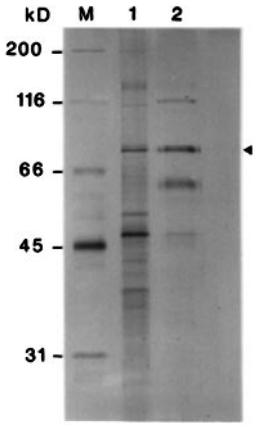

Figure 5. Protein profile of active fractions from two separate calf thymus preparations (lanes 1 and 2) separated by SDS–PAGE after the final step of chromatography (Fraction V). The arrow indicates the 73 kDa protein discussed in the text. Molecular weight markers (M) in kDa are shown to the left.

DISCUSSION

We have developed an efficient cell-free system to examine the joining of non-complementary termini by mammalian enzymes. Two of the three major mechanisms of non-complementary end joining identified previously, the 'fill-in' and 'repeat' mechanisms (see Introduction), were found following examination of a total of 139 junctions joined by Fraction I. We did not observe a third ('overlap') mechanism, where one or more terminal bases can pair, followed by filling in and correction of mismatched bases. The latter mechanism has previously been found in systems relying on the formation of circular products in *Xenopus* extracts (17,18,20) or after mammalian cell transfection (6–8). In these situations there is a high probability that the naked input DNA will be converted to a chromatin-like structure (26,27) and this may stabilize the required alignment of one or

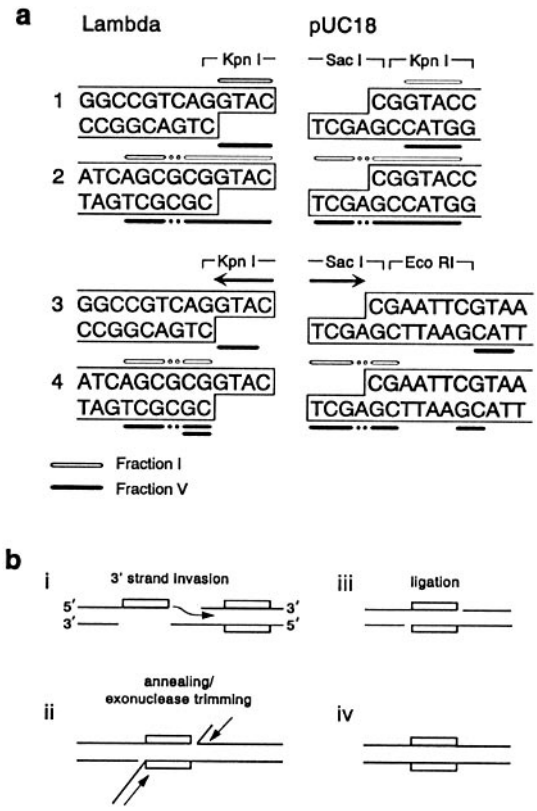

Figure 6. (**a**) Schematic representation of the surrounding sequence for the λ/*Kpn*I and the pUC/*Sac*I termini in orientations 1–4. The major mechanism for joining by Fraction I is indicated above each orientation. Black arrows indicate removal of bases and black bars indicate the major repeat sequence used. A break in the bar indicates a mismatched base. White bars below the sequences indicate the major repeat sequence used by Fraction V. Where two bars are used, both are used at similar frequency. (**b**) Modified model of joining through short regions of homology.

two bases (28). In our system, where there is the possibility of 1 or 2 bp overlaps in the *Kpn*I/*Sac*I and *Kpn*I/*Pst*I combinations respectively, chromatinization may not occur during the short incubation times used. Additionally, we cannot completely rule out some bias in the recovery of different joined products with PCR amplification, but the fact that the proportions of different joined products change with fractionation (Table 1) suggests that our methods do not strongly favour the amplification of one product relative to others.

In this system the major mechanism used for the joining of linear DNA substrates proceeds through a final step of blunt-end ligation. However, if the terminal bases on the 3′ overhang can pair with a complementary sequence <10 bp from the break site, then joining occurs through a pathway utilizing short regions of homology, as described in previous studies with both mammalian cell extracts (13,16) and whole cells (7,10,11,29). A possible explanation why the *Kpn*I/*Sac*I combination showed the greatest proportion of joining using short repeats is that in the pUC polylinker the *Kpn*I and *Sac*I sites are adjacent (Fig. 6a). Therefore, in orientations 1 and 2 (which differ in the orientation of the λ sequence) the 3′ λ overhang has only to find a repeat 6 bp into the pUC sequence to form a 4 bp homology. In orientation 2 this same pairing distance fortuitously forms a potentially more stable 6 + 3 bp region of homology, which probably accounts for the 100% use of this mechanism as compared with the 55% usage for orientation 1. In orientations 3 and 4 the adjacent *Kpn*I site is not present; however, in orientation 4 the 3′ termini from pUC can form a 2 + 3 bp repeat, again situated 6 bp into the λ sequence. This repeat is used exclusively by Fraction I, but a greater range of repeat sequences is used by the purified Fraction V, indicating that factors in the crude extract may be influencing which repeat sequences are used.

Surprisingly, with the other 3′-3′ combination (*Kpn*I/*Pst*I) examined with Fraction I repeats of >1 bp were not used, despite the presence of a 5 bp homology present 2 bp from the ends of both substrates. This shows that for the repeat mechanism to occur efficiently under our conditions, the homology must be complete to the end of the 3′ overhang, suggesting that this process is initiated by strand invasion of the free 3′-end into the helix of the DNA strand to be joined. In addition, the extensively purified Fraction V did not appear to be affected by the inclusion of polymerase inhibitors, indicating that DNA synthesis is not required and unpaired bases are removed at the point where the DNA is again correctly base paired. Taking these factors into account, we propose a new model for repeat-driven joining of non-complementary termini (Fig. 6b) in which, following strand invasion, the single-stranded ends are removed by precise nuclease action, removing the need for a DNA polymerase.

We looked in Fraction V for the presence of several endonucleases that may cleave at single- to double-strand transitions, as predicted by the model for repeat-driven joining. Surprisingly, neither the flap endonuclease FEN-1 (also identified as DNase IV and MF-1; 30) nor the two structure-specific endonucleases involved in excision repair, ERCC1 and XP-G, were found to be present by Western analysis. This result indicates that the weak 3′→5′ and 5′→3′ exonuclease activities detected in the active fractions are sufficient for the removal of the single-stranded tails needed for this mechanism to occur.

A protein that has been shown to have a major role in the ability of mammalian cells to recover from DNA double-strand breaks is the DNA-dependent protein kinase complex (DNA-PK). This complex consists of a heterodimer of two subunits (Ku70 and Ku86) which can bind to the ends of broken DNA where it then recruits a further subunit (p450) (for a review see 31). It is of interest that we could not detect the Ku70 protein in Fraction V, indicating either that the repeat-driven joining mechanism does not require this protein or that Ku is not directly involved in the joining process but acts as part of a DNA damage signalling pathway.

Fractionation of extracts to purify the proteins involved in the repeat mechanism resulted in the appearance a 73 kDa protein after silver staining as the only common band observed in all active fractions. Recently, Jessberger *et al.* (32) have identified a 72 kDa protein (SRSP) which stimulates homologous recombination in mammalian extracts. It will be important to see if these proteins are the same, suggesting an overlap in homologous and illegitimate recombination mechanisms.

ACKNOWLEDGEMENTS

We thank Profs T.Lindahl and S.Jackson and Drs R.Wood, S.Clarkson and M.Lieber for kindly providing antibodies. We thank Dr S.Bouffler for helpful comments on the manuscript.

This study was supported in part by the Commission of the European Communities contract no. F13PCT920007.

REFERENCES

1 Frankenburg,D., Frankenberg-Schwager,M., Blocher,D. and Harbich,R. (1981) *Radiat. Res.*, **88**, 524–532.
2 Bennett,C.B., Westmoreland,T.J., Snipe,J.R. and Resnick,M.A. (1996) *Mol. Cell. Biol.*, **16**, 4414–4425.
3 Simpson,P. and Savage,J.R.K. (1995) *Int. J. Radiat. Biol.*, **67**, 37–45.
4 Miles,C., Sargent,G., Phear,G. and Meuth,M. (1990) *Mol. Carcinogen.*, **3**, 233–242.
5 Szostak,J.W., Orr-Weaver,T.L., Rothstein,R.J. and Stahl,F.S. (1983) *Cell*, **33.** 25–35.
6 Roth,D.B., Porter,T.N. and Wilson,J.H. (1985) *Mol. Cell. Biol.*, **5**, 2599–2607.
7 Roth,D.B. and Wilson,J.H. (1986) *Mol. Cell. Biol.*, **6**, 4295–4304.
8 Munz,P.L. and Young,C.S.H. (1987) *Virology*, **158**, 52–60.
9 Stary,A. and Sarasin,A. (1992) *Nucleic Acids Res.*, **20**, 4269–4274.
10 Winegar,R.A., Lutze,L.H., Rufer,J.T. and Morgan,W.F. (1992) *Mutagenesis*, **7**, 439–445.
11 King,J.S., Valcarcel,E.R., Rufer,J.T., Phillips,J.W. and Morgan,W.F. (1993) *Nucleic Acids Res.*, **21**, 1055–1059.
12 North,P., Ganesh,A. and Thacker,J. (1990) *Nucleic Acids Res.*, **18**, 6205–6210.
13 Thacker,J., Chalk,J., Ganesh,A. and North,P. (1992) *Nucleic Acids Res.*, **20**, 6183–6188.
14 Fairman,M.P., Johnson,A.P. and Thacker,J. (1992) *Nucleic Acids Res.*, **20**, 4145–4152.
15 Nicolas,A.L. and Young,C.S.H. (1994) *Mol. Cell. Biol.*, **14**, 170–180.
16 Nicolas,A.L., Munz,P.L. and Young,C.S.H. (1995) *Nucleic Acids Res.*, **23**, 1036–1043.
17 Pfeiffer,P. and Vielmetter,W. (1988) *Nucleic Acids Res.*, **16**, 907–924.
18 Thode,S., Scafer,A., Pfeiffer,P. and Vielmetter,W. (1990) *Cell*, **60**, 921–928.
19 Pfeiffer,P., Thode,S., Hancke,J. and Vielmetter,W. (1994) *Mol. Cell. Biol.*, **14**, 888–895.
20 Lehman,C.W., Trautman,J.K. and Carroll,D. (1994) *Nucleic Acids Res.*, **22**, 434–442.
21 Laemmli,U.K. (1970) *Nature*, **227**, 680–685.
22 Bradford,M.M. (1976) *Anal. Biochem.*, **72**, 248–254.
23 Sambrook,K., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
24 Wist,E. (1979) *Biochim. Biophys. Acta*, **562**, 62–69.
25 Jessberger,R., Podust,V., Hubscher,U. and Berg,P. (1993) *J. Biol. Chem.*, **268**, 15070–15079.
26 Laskey,R.A., Mills,A.D. and Morris,N.R. (1977) *Cell*, **10**, 237–243.
27 Reeves,R., Gorman,C.M. and Howard,B. (1985) *Nucleic Acids Res.*, **13**, 3599–3615.
28 Pfeiffer,P., Thode,S., Hancke,J., Keohavong,P. and Thilly,W.G. (1994) *Mutagenesis*, **9**, 527–535.
29 Dar,M.E. and Jorgensen,T.J. (1995) *Nucleic Acids Res.*, **23**, 3224–3230.
30 Harrington,J.J. and Lieber,M. (1994) *Genes Dev.*, **8**, 1344–1355.
31 Jeggo,P.A., Tacciolli,G, and Jackson,S.P. (1995) *BioEssays*, **17**, 949–957.
32 Jessberger,R., Riwar,B., Rolink,A. and Rodewald,H.R. (1995) *J. Biol. Chem.*, **270**, 6788–6797.