A modified single-strand annealing model best explains the joining of DNA double-strand breaks in mammalian cells and cell extracts

Andrea L. Nicolás, Patricia L. Munz and C. S. H. Young*

Department of Microbiology, Columbia University, New York, NY 10032, USA

Received October 3, 1994; Revised and Accepted January 26, 1995

ABSTRACT

The joining of DNA double-strand breaks *in vivo* is frequently accompanied by the loss of a few nucleotides at the junction between the interacting partners. *In vitro* systems mimic this loss and, on detailed analysis, have suggested two models for the mechanism of end-joining. One invokes the use of extensive homologous side-by-side alignment of the partners prior to joining, while the other proposes the use of small regions of homology located at or near the terminus of the interacting molecules. To discriminate between these two models, assays were conducted both *in vitro* and *in vivo* with specially designed substrates. *In vitro*, molecules with limited terminal homology were capable of joining, but analysis of the junctions suggested that the mechanism employed the limited homology available. *In vivo*, substrates with no extensive homology end-joined with equal efficiency to those with extensive homology in two different topological arrangements. Taken together, these results suggest that extensive homology is not a prerequisite for efficient end-joining, but that small homologies close to the terminus are used preferentially, as predicted by the modified single-strand annealing model.

INTRODUCTION

The mechanisms by which mammalian cells repair double-strand breaks are not clearly understood, but it is likely that most such repair involves the rejoining of the broken ends without the intervention of homologous recombination with an intact chromosome or chromatid [reviewed in (1)]. The rapid rejoining of broken ends by non-homologous means has two consequences. First, the junction frequently will not restore the integrity of the sequence existing before the break occurred. This may be, however, a small price to pay to avoid the disastrous effects of the loss of part of a chromosome or chromatid, and of course the mutated sequence may not be expressed, or it may be masked by expression from the wild type allele. Secondly, DNA presented to cells by transfection is rapidly joined together, and integrates at naturally occurring breaks in the chromosome, rather than by

homologous recombination with the resident sequence. As is well known, the failure to recombine homologously is one of the practical limitations to gene targeting strategies in mammalian cells [reviewed in (2)].

Because of the importance to the cell of being able to repair double-strand breaks, we and others have begun the biochemical analysis of non-homologous end-joining in extracts from mammalian cells (3–7) or from *Xenopus* eggs (8–11). Potent end-joining activities are observed, but the precise nature of the joining is different in the different systems. In our own experiments and those of our collaborators (3,4), one of the most striking results was the observation that the joined products were predominantly arranged to form a set of linear multimers arranged in head-to-head (H:H) and tail-to-tail (T:T) orientations, but rarely in the head-to-tail (H:T) orientation. Random joining of linear molecules would be expected to create twice as many H:T as either H:H or T:T molecules, and intramolecular joining of the linear substrate would create circular molecules, which were absent among the products. A further notable feature was that the sequences at the junctions of the multimers showed the loss of a specific number of base pairs (3). Two models could plausibly account for these observations (Fig. 1). In one, the fragments to be joined are aligned side-by-side, presumably by some form of homologous pairing, and then the juxtaposed ends would necessarily be joined in H:H or T:T configuration. The joining reaction would involve a loss of nucleotides, because of the involvement of an exonuclease (4) to create single strands in each of the partners. This 'homologous alignment' model is a molecular analogue of the breakage and fusion steps of the breakage–fusion–bridge cycle, first observed by McClintock in *Zea mays* (12). An alternative model does not invoke complete homologous alignment, but suggests that small terminal homologies are used to create joint linear molecules (3), as has been suggested earlier for the creation of circular molecules from linear plasmid substrates in a different *in vitro* system (13). The predominance of the observed H:H and T:T multimers in the experiments reported previously (3,4) would occur because the linear plasmid substrates employed were cut within a polylinker sequence by restriction enzymes. Because of the palindromic nature of the polylinker, there are limited regions of complete homology between two substrates arranged in H:H and T:T orientations, but these regions are much less extensive in the H:T

* To whom correspondence should be addressed

Figure 1. Models to explain the preponderance of H:H and T:T products in the *in vitro* end-joining reaction. (A) Linear molecules are paired along the length of their homology, presumably by a pairing protein (not shown). How the ends are joined is not known, but the diagram shows some resection in the 3' → 5' direction. The exposed single-stranded ends are then ligated together, with or without any fortuitous homology in the exposed single strands, to give the H:H molecule shown (orientation is indicated by the arrows). (B) Molecules are resected first (or they could be unwound by a helicase), complementary single strands (dotted lines) anneal, and the joint molecule is then resolved by nuclease digestion of the unpaired sequences, and ligation of the nicks. A more detailed example is given in Figure 4.

arrangement (3). In this alternative model, pairing occurs in the region of homology, and a precise number of base pairs are lost from the junctions created from substrates with differing terminal sequence or structure. In essence, this model is a variant of the non-conservative single-strand annealing model for homologous recombination in both mammalian cells (14,15) *Xenopus* (16,17) and *Saccharomyces cerevisiae* (18–21). A prediction of this modified single-strand annealing model is that end-joining of non-palindromic DNA would be less efficient, and the precise number of base pairs lost from the junction would be dependent on small fortuitous regions of homology. In this paper, we describe end-joining in the *in vitro* extract, using linear molecules whose terminal sequences are non-palindromic. As predicted from the single-strand annealing model, the efficiency of the reaction was lower than with palindromic DNA, but analysis of the sequences of the junctions suggest that the reaction still involves limited sequence homology.

Although the characteristics of the joining reactions in the *in vitro* extracts are broadly consistent with the products of end-joining observed in transfection [reviewed in (1)], it is possible that reactions *in vivo* may occur by both single-strand annealing and homologous alignment mechanisms. To date, no attempt has been made to test *in vivo* if extensive homology between the end-joining partners increases the efficiency of the joining reaction, as would be predicted in the homologous alignment model. We have designed a set of plasmids to test this directly, using transfection experiments with adenovirus DNA fragments and plasmids containing adenovirus sequences in two opposing orientations. There is no evidence that extensive homology increases the efficiency of end-joining, suggesting that the homologous alignment model is not a major pathway *in vivo*.

MATERIALS AND METHODS

Plasmids and viruses

The plasmids used in the *in vitro* experiments were either pUC18, or derivatives thereof. Some plasmids contain DNA sequences

derived from the genomes of adenovirus serotypes 2 and/or 5. These sequences are numbered according to type 2 conventions (22) for ease of comparison. Plasmid pSac3del was derived from pSac3 as follows. Plasmid pSac3 contains a 928 bp *SacI* fragment of recombinant P54 adenovirus DNA (23), extending from base pairs 27634 to 28562, cloned into the *SacI* site of pUC18. The leftmost adenovirus sequences of pSac3 were deleted by removal of an *SphI* to *ApaI* fragment, followed by removal of a *PpuMI* to *SacI* fragment, on the right of the adenovirus sequence. The remaining adenovirus DNA is 100 bp long (Fig. 3) and contains a central *BsaAI* site flanked by the *EcoRI* and *HindIII* sites of the pUC18 vector polylinker. These two sites were used for the cloning of junctions, from the *in vitro* reaction products, into pA100 as described previously (3). The plasmids for use in the viral transfection experiments were created from pPF446, originally made by Paul Freimuth (unpublished). This plasmid contains adenovirus type 5 sequences extending from the *BamHI* site at base pair 21606 to the right hand terminus, to which a *SalI* linker was added, before cloning into the respective sites in pBR322. A series of plasmids was derived from pPF446, first by removal of the sequence from the *BamHI* site at base pair 21606 to the *XhoI* site at base pair 29788, and the filling-in of the ends. These manipulations resulted in the restoration of the *XhoI* site. The adenovirus serotype 5 sequence from the *XhoI* site at base pair 29788 to the *XbaI* site at base pair 30461 was replaced with the equivalent sequence from adenovirus serotype 2, to create plasmid p83-29.2. Plasmids pSna83, pSna84 and pSna83rev (Fig. 5) were derived from p83-29.2, using standard cloning procedures. They contain, respectively, a *SnaBI* linker at the *XhoI* site at base pair 29788, a *SnaBI* linker at the *XbaI* site at base pair 30461, and a reversed sequence between the two sites, with a *SnaBI* linker at base pair 29788. All linkers were added to restriction sites with 5' extensions that had been repaired with the Klenow fragment of DNA polymerase. Sequencing across the linkers revealed that the sequences were as expected. Virus Ad2.5 (Sna84) (Fig. 5) was made by overlap recombination (24,25) between DNA protein complex (DNA-PC) from adenovirus serotype 2, cleaved with *AseI* to give a left terminal fragment extending rightward to base pair 31900, and pSna84, cleaved with *XhoI* to give a right hand fragment extending from base pair 29788. Following co-transfection into A549 cells, recombinant virus was isolated and screened for the presence of the new *SnaBI* site. The recombinant virus has Ad2 sequences up to the *SnaBI* site at base pair 30461, and Ad5 sequences beyond that point. Our strain of adenovirus type 5 does not have the *SnaBI* site reported to be immediately to the left of the right hand inverted terminal repeat (26).

Preparation of nuclear extracts

The methods have been described in detail previously (3). Briefly, nuclear extracts were prepared from cells grown either in suspension, or as monolayers in 15 cm tissue culture dishes (Falcon), using a protocol adapted from that of Peterson, Orth and Calame (27). Extracts were routinely prepared from 10 to 12 monolayer plates of human A549 cells, or from 1 to 3l of HeLa cell suspension culture. All procedures were performed at 0–4°C. The nuclear extract was clarified at 100 000 g for 60 min in a Beckman SW50.1 rotor, and the protein concentration of the supernatant was determined by Bradford colorimetric assay (Bio-Rad). The clarified extract was then distributed in 100 µl aliquots, and stored at –80°C.

Preparation of substrate DNA for the *in vitro* reaction

The substrates used in the end-joining reactions were made from plasmid DNA, prepared from 500 ml cultures, and processed by column chromatography (Qiagen Inc.). The plasmid was exhaustively digested with the restriction enzyme appropriate for each experiment, until no nicked circular DNA was observable on an ethidium bromide-stained gel. The linear substrate was phenol-chloroform extracted, ethanol precipitated and resuspended to 1 mg/ml concentration.

In vitro reaction conditions

These have been described in detail previously (3). When junctions were to be cloned from the products of the reaction, 20 µg of linearized substrate was incubated with 1–2 µg of nuclear extract protein, for 2 h at 25°C in 50 mM Hepes, pH 7.5, 0.5 mM DTT, 0.2 mM ATP and 10 mM MgCl₂, in 0.2 ml total volume. After incubation, Proteinase K and SDS were added to 100 µg/ml and 0.5% respectively, and incubated at 37°C for 1 h. The end-joined DNA was then extracted with phenol-chloroform, ethanol precipitated, washed with 70% ethanol and resuspended to 1 mg/ml in TE. For routine examination of the *in vitro* reaction, ~1 µg of the linearized plasmid and a correspondingly lower amount of the nuclear extract were incubated in a reaction volume of 50 µl.

Cloning and sequencing of junction fragments

The methods were exactly as described previously (3). H:H and T:T junctions were initially isolated as large fragments, and the smaller junctional fragments were then cloned into pA100 using the flanking *Hind*III and *Eco*RI sites. It was observed that cloning of potentially palindromic DNA was dependent on electroelution of the desired bands from the separating gel at 4°C. The sequence of each junction reported in this paper was derived from a bacterial colony arising from an independent transformation event. DNA was prepared for sequencing using the 'Speedprep' method (28), and one half of the final yield of DNA was denatured, and T3 primer was annealed. Sequencing was performed using the 'Sequenase' kit (US Biochemicals) with dITP termination solutions.

Transfection of human A549 cells

General methods for transfection and cell culture have been described in detail previously (25). The appropriate plasmid DNA was digested with *Sna*BI and *Sal*I to liberate the adenovirus sequences, and DNA-PC from Ad2.5 (*Sna*84) was digested with both *Sna*BI and *Ase*I. The left terminal fragment from the viral DNA always extended to the *Sna*BI site at base pair 30461, while the right terminal plasmid sequences either overlapped with the viral DNA, from base pair 29788, in the case of pSna83 and pSna83rev, or were co-terminous with it, in the case of pSna84. The transfection was performed as a direct plaque assay, and the virus in individual plaques was isolated, and the clone expanded for the subsequent isolation of intracellular viral DNA. Individual DNA samples were examined by restriction digestion.

Figure 2. Efficiencies of end-joining of different blunt-ended substrates in the *in vitro* reaction. The substrates were made by digestion with *Hinc*II or *Xmn*I, and were then incubated in the standard reaction with 1 µg of substrate. The substrates are: lane 1: pUC18.*Hinc*II; lanes 2 and 3: pUC18.*Xmn*I; lane 4: pSac3.*Xmn*I. The substrate in lane 2 was not incubated with extract. M: λ BstEII marker. Densitometric scanning of a photographic negative gave the following values for dimer and multimer formation: *Hinc*II substrate: 48%; *Xmn*I substrate from pUC18: 14%; *Xmn*I substrate from pSac3: 18%.

RESULTS

Joining of non-palindromic DNA in the crude nuclear extract

In a previous paper, we examined the ability of a crude nuclear extract to join linear DNA substrates made by the cleavage of pUC18 with restriction enzymes whose unique sites are in the polylinker region. The characterization of these junctional fragments allowed us to develop a model in which small terminal homologies direct the end-joining reaction [(3), and Fig. 1]. Most importantly, the polylinker substrates used in the previous experiments had a series of internal palindromes at their termini due to the presence of multiple restriction enzyme sites. The presence of these palindromes may influence the end-joining reaction in its nature or efficiency.

The first suggestion that the palindromic nature of the terminal sequence might influence the efficiency of the reaction, came from the observation that linear substrates created by cleavage of pUC18 or pSac3 by *Xmn*I, whose unique site lies outside the polylinker, were not as efficiently joined together as those restricted by *Hinc*II (Fig. 2). Both enzymes create blunt-ended substrates, yet the efficiency of end-joining with the *Hinc*II substrate was ~3-fold greater than with either of the *Xmn*I substrates. Similar results were obtained with blunt-ended substrates created by *Sca*I, whose site also lies outside the polylinker (data not shown). In order to examine joining of non-palindromic substrates more carefully, we designed a substrate which could be manipulated in the same manner as the polylinker substrates, but would not contain extensive palindromic sequence. This was accomplished by the insertion of 100 bp of adenovirus DNA into a pUC18 vector at the *Sac*I site. This substrate was linearized by restriction at a *Bsa*AI site, roughly in the middle of the fragment, to give a blunt linear duplex (Fig. 3A). The substrate was incubated with the extract, and the junctional fragments were cloned and sequenced using the methods previously described for the polylinker substrates (3). A parallel incubation with a *Hinc*II substrate revealed that the efficiency of formation of the H:H junctional fragment was ~37% greater with the latter. A total of 12 junctions were sequenced from the H:H orientation. Unlike the situation with the polylinker substrates,

Figure 3. The sequence of pSac3del and of the junctions produced in the *in vitro* reaction with *BsaAI*-linearized pSac3del. (A) The sequence is shown from the *HindIII* to the *EcoRI* sites present in the polylinker of pUC18. Palindromic restriction sites within the sequence are indicated, and the two sequences are given the arbitrary designations 'Head' and 'Tail'. (B) The classes of products are classified by the number of nucleotides lost at the junction, and the numbers of observations in a given class are indicated. The two fragments are aligned to give the number of nucleotides lost at the junction, the number of nucleotides of homology are shown by asterisks, and the region of crossing over is indicated either by a single gap (class 1) where the site is unambiguous, or by two gaps immediately flanking a region of homology (all other classes).

where a unique class of sequence was observed for each substrate (3), five classes were found, with losses of 8, 14, 15, 30 or 32 nucleotides at the junction (Fig. 3B). The non-palindromic nature of the substrate also allowed us to determine more precisely which nucleotides in the junction had been derived from each partner. The linear substrates are shown in duplex form, the stars indicate the positions of nucleotide homology, and the regions in which the crossovers occurred are indicated by gaps in the two duplexes. In all five classes, there is significant homology at the region of crossover. These data can best be explained if complementary single strands are exposed and paired, as in the single-strand annealing model. The manner in which this can be accomplished for one of the classes is shown in Figure 4. Although the diagram is shown as though there is no pairing of single strands flanking the central AGCT, the overlapping sequence encompasses an imperfect palindrome, and the single

Figure 4. An example of the single-strand annealing model as applied to one class of product. The exposed single strands are shown as pairing only in the region of perfect homology, but they could pair throughout some or all of their length if mis-matched base pairs form.

strands may anneal over a distance greater than the central homology. The range of possible junctions, and the absence of specific types, will be considered further in the Discussion.

Does the homologous alignment mechanism operate *in vivo*?

The data above, and several other examples alluded to in our previous publication (3), suggest that the single-strand annealing model is a better explanation for the potent end-joining of linear DNA molecules observed *in vitro*. However, it is still possible that the homologous alignment mechanism is operative *in vivo*. To test this idea, a set of plasmids was created to examine the efficiency of end-joining when the ends could be aligned homologously before joining. The region of adenovirus chosen for this purpose lies within the dispensable early region 3 (29), in which extensive sequence modification can occur without effects on replication kinetics. Figure 5 illustrates the topologies of the three plasmid fragments, when they enter into a recombinational interaction with a large left hand terminal DNA fragment derived from a specially constructed adenovirus, Ad2.5 (Sna84), which contains a unique *SnaBI* site at base pair 30461. The fragment from pSna83 (Fig. 5A) would be expected (30) to recombine in two ways with the viral left terminal fragment, either homologously within the 673 bp of overlapping sequence, or non-homologously by end-joining at the *SnaBI* sites at base pairs 29788 and 30461. The fragment from pSna84 (Fig. 5B) can only join end-to-end with the viral fragment. The key interaction is between pSna83rev and the viral fragment (Fig. 5C). Because the adenovirus sequence between base pairs 29788 and 30461 is reversed in the two partners, homologous recombination between the two viral sequences will not yield full length viral genomes. The only way in which a viable virus genome can be created is by end-joining. If pairing by extensive sequence homology is a rate-limiting step in end-joining, as would be predicted in the homologous alignment model, the number of end-joined molecules, and hence

Figure 5. Arrangements of the interacting DNA fragments. (A) pSna83 can interact with the viral fragment in two ways, either by homologous recombination, as shown on the left, or by end-joining, as shown on the right. In (B and C) only end-joining is possible. The leftward terminal fragment in each case was made by digesting DNA-PC of Ad2.5 (Sna84) with *Sna*BI and *Ase*I. The various plasmids were digested with *Sna*BI and *Sal*I, and, in the case of pSna84, with *Xho*I also. Transfections were performed with unpurified fragment mixtures. The abbreviations are as follows: S: *Sna*BI half site in the fragments and the full site in the products. (S) indicates that the *Sna*BI site may or may not be present, depending on the faithfulness of the joining reaction. Horizontal arrows: direction of E3 coding sequences. Vertical dotted line: end-joining. Curved line: homologous recombination.

viable virus, should be markedly increased in the pSna83rev transfection as compared with that of pSna84.

Human A549 cells were co-transfected with the appropriately digested DNAs and the numbers of plaques counted. As can be seen from Table 1, the arrangement with pSna83, in which either homologous recombination or end-joining can occur, yields the greatest number of plaques. This is to be expected if the homologous recombination pathway is an alternative to the end-joining pathway. On the other hand, the DNA fragments from pSna83rev joined together with the viral DNA fragments to give somewhat fewer plaques than the standard end-joining with pSna84. This suggests that the presence of homologous DNA does not promote the end-joining of molecules, as would have been predicted by the homologous alignment model. The structures of the genomes of representative plaques arising from all three transfections were examined by restriction digestion of intracellular viral DNA, to gain some insight into the nature of the joining reaction (Table 2). With pSna83, 15 genomes from experiment no. 2 were examined. Fourteen of them were of standard size and lacked the *Sna*BI site, suggesting that they had arisen by homologous recombination within the overlapping sequence. The sole exception had arisen by faithful end-joining, as deduced from the fact that the genome was appropriately bigger, and an extra *Sna*BI site was present. This 14:1 ratio of homologous recombination to end-joining is rather higher than we have observed previously in other regions of the genome (30,31), but is consistent

with the higher number of plaques observed when comparing the results with pSna83 to those with pSna84 in experiment no. 2. The junctions formed with pSna84, fell into two classes. Eleven of the 28 junctions examined (39%) retained all of the terminal sequences, because the *Sna*BI site was retained. The remaining isolates had genomes with small alterations at the junction, as deduced from the absence of the *Sna*BI site. The results with pSna83 and pSna84 are the first examples of bi-molecular joining of blunt-ended DNA molecules in the adenovirus co-transfection system, but they are entirely consistent with previous results in this and in other systems, which showed that the cell frequently can join ends of several different terminal topologies without loss of terminal nucleotides (30,32). The structures of the genomes created with pSna83rev show a somewhat different pattern of end-joining. None of them had retained the *Sna*BI site, although some were of approximately the correct size and genomic structure for end-joining, with the loss of very few nucleotides. Note that these genomes would contain an almost perfect palindrome of some 1.3 kb. Other genomes arising from the pSna83rev transfection had sizable deletions extending towards the left end of the genome, and/or additional sequences at the junction. The additional sequences probably derive from the salmon sperm DNA used as carrier in the transfection protocol; similar additions have been observed before (30). However, although the precise range of junctions arising from the transfections with pSna84 and pSna83rev is somewhat different, both sets of observations fall within the spectrum observed from other experiments with a variety of substrates (31). We also observed that those genomes, produced by end-joining with pSna83rev, which were closest in size and structure to the expected palindrome, gave rise to virus with slower growth rates, and thus smaller plaque size. The magnitude of this effect is probably not enough to negate an erstwhile increase in the number of plaques, but to compensate for any difficulty in observing smaller plaques in the pSna83rev transfection, the assays were extended for a week beyond the normal termination date. The causes for the slower replication rate of the viruses containing a palindrome are unknown, but they are not related to the size of the genome, because the virus with a direct repeat of the region between base pairs 29788 and 30461, produced by end-joining of the DNA fragments from pSna83 and Ad2.5 (Sna84), replicates as efficiently as the wild type (data not shown). Furthermore, the size of genomes containing a single repeat in either orientation is well within the packaging limit for adenovirus (33). DNA replication rates of the palindrome-containing viruses were also normal, and the genomic structures were stable (data not shown).

DISCUSSION

The study of DNA end-joining *in vitro* has led to the development of two molecular models that are distinguished by the degree of homology involved in the pairing mechanism prior to joining (Fig. 1). The homologous alignment model (4) proposes that the interacting duplexes pair throughout an extensive region of homology, and then join at the juxtaposed ends. The modified single-strand annealing model on the other hand (3,13) requires only limited homology, located at or towards the molecular termini, which is used to hold the ends together prior to the joining reaction. We have used complementary *in vivo* and *in vitro* assays to determine which of these two models is of greater biological significance.

Table 1. Numbers of plaques arising from co-transfection of overlapping fragments of adenoviral DNA

Experiment no.	Precipitate no.	DNAs used in transfection			
		cut DNA-PC alone	cut DNA-PC + pSna83	cut DNA-PC + pSna84	cut DNA-PC + pSna83rev
1 ^a	1	0, 0	7, 5	4, 1	2, 4
	2	0, 0	4, 3	0, 0	0, 1
	3	0, 0	6, 6	6, 2	0, 2
	Totals	0	31	13	9
2 ^b	1	0, 0	73, 75	11, 22	5, 2
	2	0, 0	ND ^c	1, 5	0, 0
	Totals	0	148 ^d	39	7

^aIn experiment no. 1, the DNA-PC, digested with *Sna*BI and *Ase*I, was added at 200 ng per 60 mm plate, and the respective plasmids, digested with *Sna*BI, at 600 ng per plate. Undigested DNA-PC gave 31 plaques on one plate with addition of 20 ng. Each transfection was performed as triplicate precipitates.

^bIn experiment no. 2, the DNA-PC was added at 400 ng per plate, and the respective plasmids at either 600 ng (ppt. no. 1) or 1000 ng (ppt. no. 2). Undigested DNA-PC gave a total of 139 plaques on two plates with addition of 15 ng per plate.

^cND; not done.

^dFrom one precipitate only.

Table 2. A summary of the virus genomic structures arising from co-transfection of Ad2.5 (Sna84) and plasmids

Plasmid used in recombination	Experiment ^a	Standard genome ^b	End-joined (no deletion) ^c	End-joined (alterations) ^d	Totals
pSna83	2	14	1	0	15
pSna84	1	NA ^e	4	8	12
	2	NA	7	9	16
	Totals		11	17	28
pSna83rev	1	NA	0	7	7
	2	NA	0	9	9
	Totals		0	16	16

^aThis refers to the experimental number in Table 1.

^bA 'standard' genome arises by homologous recombination, and does not have a *Sna*BI site (Fig. 5A).

^cFaithful end-joining creates a genome with a *Sna*BI site (Fig. 5B and C).

^dLoss or addition of nucleotides creates a genome without a *Sna*BI site (Fig. 5B and C).

^eNot applicable. This applies to genomes arising by homologous recombination only.

In our previous paper (3), we examined the end-joining process *in vitro* using substrates linearized at various positions within a polylinker region. Because of the presence of multiple restriction enzyme sites, and flanking linker-derived base pairs, the sequence of this region contains several 6 and 8 bp palindromes. Using these substrates, end-joining was observed to yield products that were predominantly in the H:H and T:T orientations, and the junctions showed a pattern of nucleotide loss that was uniform and specific for any given substrate. This joining pattern is most easily explained by the single-strand annealing model. We had also noticed that *in vitro* joining of substrates whose ends were not in polylinker sequence, such as those made by digesting pUC18 and its derivatives with *Sca*I, *Xmn*I or *Ssp*I, was much less efficient than those with ends in the polylinker (Fig. 2 and not shown). This strongly suggested that the nature of the polylinker sequence increased the frequency of joining. Further support for

this idea came from the observation with a version of pUC18 that had an *Ssp*I linker inserted in the polylinker *Hinc*II site. This substrate end-joined at a similar efficiency to those of the other polylinker substrates (3) in contrast to that with a substrate created by *Ssp*I cleavage of pUC18 outside the polylinker.

To test the importance of palindromic sequences, we studied the joining of a substrate (pSac3del) which has a small region of 'random' DNA sequence inserted in the polylinker region. This substrate presented the *in vitro* extract with a blunt linear fragment containing only limited palindromic stretches, but still allowed us to clone and sequence the junctions as before (3). The joining reaction was somewhat less efficient than that with another blunt-ended substrate created by cleavage in the polylinker. The results were in marked contrast to those obtained with the polylinker substrates, in which nucleotide loss was specific and uniform, because five classes of junctions were observed (Fig. 3),

despite the small sample size. Nevertheless the sequence analysis suggested that limited homology still played a role in junction formation. All of the observed junctional sequences display blocks of homology, when the two interacting duplexes are shown in overlapping arrangement (Fig. 3). Upon inspection of the sequence in all possible overlapping H:H arrangements up to, and including, 44 bp of overlap, those junctions observed are those with the most homology in blocks of adjacent sequence. Importantly, sequences which were not observed included those with no, or very little, homology, and those with scattered homology. An informative example of the latter is seen if the duplexes are arranged to overlap by 13, 14 or 15 bp. Although the overlap of 13 bp has the highest ratio of homologous bases (6 out of 13), there are never more than two adjacent ones, and this junction was not observed. In contrast, in the overlap with 14 and 15 bp, each with six homologous bases, there are blocks of four or three adjacent homologous bases respectively. Junctions with 14 and 15 base deletions were indeed observed (Fig. 3). Similarly, the amount of H:T products were lower than either the H:H or T:T products (data not shown), and overlaps in this orientation would have fewer instances of blocks of continuous sequence homology. Although the sample of junctions is too small to define the rules governing the precise sequence requirements for junction formation, the observations are most consistent with the predictions of the single-strand annealing model, namely that joining occurs at the intersection of maximum homology per unit of overlap. The alternative homologous alignment model has no ready explanation for the differing efficiencies of joining of the non-palindromic substrates compared with the palindromic ones, nor for the specific sequences observed at the junctions. It is also worth pointing out that a large body of evidence suggests that limited homology close to the terminus is frequently employed during the end-joining of DNA molecules *in vivo* [reviewed in (1)]. This has been confirmed in two recent reports where double-strand breaks were induced *in vivo* by the introduction of restriction enzymes, and repaired junctions were isolated and sequenced (34,35).

The *in vitro* data from the non-polylinker substrates support the modified single-strand annealing model of end-joining, but do not address the issue of whether or not homologous alignment might contribute to end-joining *in vivo*. To approach this mechanistic question, we took advantage of adenovirus DNA-mediated transfection techniques. It is well-established that fragments of adenovirus (from virus or cloned in plasmids) can join end-to-end to give viable virus (31), and that the E3 region of the genome is dispensable for replication, and can tolerate major alterations (29). The design of the transfection assay, with the plasmid constructs shown in Figure 5, allowed us to address directly the importance of homologous alignment to cellular end-joining mechanisms. In the co-transfection with the pSna84 substrate and the viral left hand end fragment, viable virus can only be formed by end-joining. The numbers of plaques that arise is therefore a direct measure of the rate of end-joining in the absence of extensive homology. The frequency of this joining event was compared to that with pSna83rev, which is also limited to end-joining in order to form virus. This substrate however contains a 673 bp region of homology, in inverted orientation, which in theory would allow homologous alignment. If the alignment model is correct, the pSna83rev substrate should join with the left-hand viral fragment to form virus with greater frequency than does the pSna84 substrate. This would be revealed by a significant increase in the number of plaques arising in this

co-transfection compared with that involving pSna84. This is not the case (Table 1); indeed the frequency of end-joining, as measured by plaque formation, is somewhat lower than that with pSna84. It could be argued that the lack of an increase in plaque number in the pSna83rev transfection is caused by the failure to capture those genomes in which palindromes had been created by faithful end-joining. It is known that palindromes can lead to inviability in bacterial viruses, and that special efforts must be made to conserve genomes containing them (36). However, we think this is unlikely to affect the numerical outcome significantly. First, special efforts were made to collect viruses yielding small plaques, and although there is some growth deficiency in those genomes whose structures approached the perfect palindrome, the viruses were relatively easy to propagate, and to examine. Secondly, the palindromic genomes were stable, and the rates of DNA replication were indistinguishable from that of a standard genome. Thirdly, if the perfect palindrome, containing a *SnaBI* site, is inviable, it is likely to comprise less than half of the total number of potential genomes, on the evidence of the frequency with which the *SnaBI* site is retained in the pSna84 transfection (39%). We conclude that the plaque numbers are close to a true reflection of the initial frequencies of end-joining in the three transfections, and that the results suggest that the fate of the ends of broken DNA is not affected by the presence of extensive regions of homology distal to the site of joining. Although these results do not rule out a contribution of the homologous alignment mechanism to end-joining *in vivo*, its magnitude cannot be great.

The data presented here are entirely consistent with the modified single-strand annealing model (Figs 1 and 4) as the most likely mechanism for end-joining both *in vitro* and *in vivo*. The challenge now is to identify the proteins that recognize broken DNA ends, catalyze the exposure of the free single strands, and perhaps stabilize the non-covalently joined intermediate, until resection and ligation can occur. Very recent genetic and biochemical evidence suggest that the end-recognition complex may contain Ku antigen (37–41). A defined *in vitro* end-joining reaction, which mimics that suspected to occur *in vivo*, will be of great value in identifying further proteins involved in this vital repair activity.

ACKNOWLEDGEMENTS

Supported by grant R01 GM31452 from the NIH, with a minority supplement awarded to ALN. We are grateful to Tony Wild for preparation of HeLa cells, to Paul Freimuth for the gift of plasmid pPF446 and to Ulla Lundholm for densitometry.

REFERENCES

- 1 Roth, D., Wilson, J. (1988) In: Kucherlapati, R. and Smith, G.R. (eds) *Genetic Recombination*. American Society for Microbiology, Washington, D.C., pp. 621–653.
- 2 Capecchi, M.R. (1989) *Science*, **244**, 1288–1292.
- 3 Nicolás, A.L. and Young, C.S.H. (1994) *Mol. Cell. Biol.*, **14**, 170–180.
- 4 Derbyshire, M.K., Epstein, L.H., Young, C.S.H., Munz, P.L. and Fishel, R. (1994) *Mol. Cell. Biol.*, **14**, 156–169.
- 5 North, P., Ganesh, A. and Thacker, J. (1990) *Nucleic Acids Res.*, **18**, 6205–6210.
- 6 Fairman, M.P., Johnson, A.P. and Thacker, J. (1992) *Nucleic Acids Res.*, **20**, 4145–4152.
- 7 Tatzelt, J., Fechteler, K., Langenbach, P. and Doerfler, W. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 7356–7360.
- 8 Pfeiffer, P. and Vielmetter, W. (1988) *Nucleic Acids Res.*, **16**, 907–924.

- 9 Goedecke,W., Vielmetter,W. and Pfeiffer,P. (1992) *Mol. Cell. Biol.*, **12**, 811–816.
- 10 Thode,S., Schäfer,A., Pfeiffer,P. and Vielmetter,W. (1990) *Cell*, **60**, 921–928.
- 11 Pfeiffer,P., Thode,S., Hancke,J. and Vielmetter,W. (1994) *Mol. Cell. Biol.*, **14**, 888–895.
- 12 McClintock,B. (1939) *Proc. Natl. Acad. Sci. USA*, **25**, 405–416.
- 13 Thacker,J., Chalk,J., Ganesh,A. and North,P. (1992) *Nucleic Acids Res.*, **20**, 6183–6188.
- 14 Lin,F.-L., Sperle,K. and Sternberg,N. (1984) *Mol. Cell. Biol.*, **4**, 1020–1034.
- 15 Wake,C.T., Vernaleone,F. and Wilson,J.H. (1985) *Mol. Cell. Biol.*, **5**, 2080–2089.
- 16 Maryon,E. and Carroll,D. (1991) *Mol. Cell. Biol.*, **11**, 3268–3277.
- 17 Maryon,E. and Carroll,D. (1991) *Mol. Cell. Biol.*, **11**, 3278–3287.
- 18 Rudin,N., Sugarman,E. and Haber,J.E. (1989) *Genetics*, **122**, 519–534.
- 19 Ozenberger,B.A. and Roeder,G.S. (1991) *Mol. Cell. Biol.*, **11**, 1222–1231.
- 20 Sugawara,N. and Haber,J.E. (1992) *Mol. Cell. Biol.*, **12**, 563–575.
- 21 Fishman-Lobell,J. and Haber,J.E. (1992) *Science*, **258**, 480–484.
- 22 Roberts,R.J., Akusjärvi,G., Aleström,P., Gelinas,R.E., Gingeras,T.R., Sciaky,D. and Pettersson,U. (1986) In: Doerfler,W. (ed.) *Adenovirus DNA*. Martinus Nijhoff, Boston, pp. 1–51.
- 23 Munz,P.L., Young,C. and Young,C.S.H. (1983) *Virology*, **126**, 576–586.
- 24 Chinnadurai,G., Chinnadurai,S. and Brusca,J. (1979) *J. Virol.*, **32**, 623–628.
- 25 Volkert,F.C. and Young,C.S.H. (1983) *Virology*, **125**, 175–193.
- 26 Chroboczek,J., Bieber,F. and Jacrot,B. (1992) *Virology*, **186**, 280–285.
- 27 Peterson,C.L., Orth,K. and Calame,K.L. (1986) *Mol. Cell. Biol.*, **6**, 4168–4178.
- 28 Goode,B.L. and Feinstein,S.C. (1992) *BioTechniques*, **12**, 148–149.
- 29 Kelly,T.J. and Lewis,A.M. (1973) *J. Virol.*, **12**, 643–652.
- 30 Munz,P.L. and Young,C.S.H. (1987) *Virology*, **158**, 52–60.
- 31 Munz,P.L. and Young,C.S.H. (1991) *Virology*, **183**, 160–169.
- 32 Roth,D.B. and Wilson,J.H. (1986) *Mol. Cell. Biol.*, **6**, 4295–4304.
- 33 Bett,A.J., Prevec,L. and Graham,F.L. (1993) *J. Virol.*, **67**, 5911–5921.
- 34 Phillips,J.W. and Morgan,W.F. (1994) *Mol. Cell. Biol.*, **14**, 5794–5803.
- 35 Rouet,P., Smih,F. and Jasin,M. (1994) *Mol. Cell. Biol.*, **14**, 8096–8106.
- 36 Leach,D.R.F. and Stahl,F.W. (1983) *Nature*, **305**, 448–451.
- 37 Getts,R.C. and Stamato,T.D. (1994) *J. Biol. Chem.*, **269**, 15981–15984.
- 38 Rathmell,W.K. and Chu,G. (1994) *Mol. Cell. Biol.*, **14**, 4741–4748.
- 39 Suwa,A., Hirakata,M., Takeda,Y., Jesch,S.A., Mimori,T. and Hardin,J.A. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 6904–6908.
- 40 Rathmell,W.K. and Chu,G. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 7623–7627.
- 41 Taccioli,G.E., Gottlieb,T.M., Blunt,T., Priestley,A., Demengeot,J., Mizuta,R., Lehmann,A.R., Alt,F.W., Jackson,S.P. and Jeggo,P.A. (1994) *Science*, **265**, 1442–1445.