The EMBO Journal vol.15 no.12 pp.3197–3206, 1996

Distinct DNA sequence and structure requirements for the two steps of V(D)J recombination signal cleavage

Dale A.Ramsden, J.Fraser McBlane[1], Dik C.van Gent and Martin Gellert[2]

Laboratory of Molecular Biology, Building 5, Room 241, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892-0540, USA

[1]Present address: Leukaemia Research Fund Centre at the Institute of Cancer Research, Chester Beatty Laboratories, 237 Fulham Road, London SW3 6JB, UK

[2]Corresponding author

D.A.Ramsden and J.F.McBlane contributed equally to this work

Cleavage of V(D)J recombination signals by purified RAG1 and RAG2 proteins permits the dissection of DNA structure and sequence requirements. The two recognition elements of a signal (nonamer and heptamer) are used differently, and their cooperation depends on correct helical phasing. The nonamer is most important for initial binding, while efficient nicking and hairpin formation require the heptamer sequence. Both nicking and hairpin formation are remarkably tolerant of variations in DNA structure. Certain flanking sequences inhibit hairpin formation, but this can be bypassed by base unpairing, and even a completely single-stranded signal sequence is well utilized. We suggest that DNA unpairing around the signal–coding border is essential for the initiation of V(D)J recombination.

Keywords: DNA double strand breaks/DNA hairpins/DNA unpairing/RAG1/RAG2

Introduction

The V(D)J recombination process that assembles functional immunoglobulin and T cell receptor genes is targeted by sequence motifs flanking the germline gene segments. A V(D)J recombination signal sequence (RSS) is composed of conserved 7 and 9 bp motifs (heptamer and nonamer), separated by a spacer of relatively non-conserved sequence, but conserved length. Spacers are typically 12 or 23 bp long, and recombination occurs between two signals of different length (reviewed in Gellert, 1992; Lewis, 1994). Recombination is initiated by cleavage at the border between the recombination signal and the coding segment (Roth *et al.*, 1992a,b). These double strand breaks leave blunt, 5′-phosphorylated signal ends and hairpin coding ends (Roth *et al.*, 1993; Schlissel *et al.*, 1993). Coding ends are then joined rapidly and with some sequence variability, while signal ends appear to be joined precisely, but more slowly (Ramsden and Gellert, 1995).

Recombination signals were first described by the alignment of the many examples available from the sequencing of germline V, D and J segments (Sakano *et al.*, 1979). In addition to permitting the identification of the separate heptamer and nonamer motifs, this analysis indicated that several positions within both motifs were usually conserved in signals adjacent to gene segments that were known to rearrange. Specifically, the three positions of the heptamer immediately adjacent to the coding region were present at almost all utilized coding segments; in the nonamer, the 5th and 6th positions were also highly conserved (Hesse *et al.*, 1989; Ramsden *et al.*, 1994).

Extra-chromosomal substrates permitted direct testing of the relative abilities of different recombination signals to mediate recombination (Hesse *et al.*, 1989). For the most part, the functional importance of any position in the sequence was reflective of the degree to which it had been found conserved. For example, the highly conserved first three positions of the heptamer were shown to be individually critical for efficient V(D)J recombination in extra-chromosomal substrates. Further investigations of these substrates have suggested that differences between endogenous recombination signals might affect coding segment usage in the antigen receptor loci, and thus skew the repertoire available to the adaptive immune response (also reviewed in Lewis, 1994).

Coding flank sequence may also affect the ability of recombination signals to mediate recombination (Boubnov *et al.*, 1993; Gerstein and Lieber, 1993; Ezekiel *et al.*, 1995). Polythymidylate tracts in the coding flank immediately adjacent to the heptamer have been shown to reduce recombination, and other coding sequence effects have been described. Furthermore, work from our laboratory has shown an interaction between the effect of coding flank sequence and certain mutated versions of the RAG1 protein (Sadofsky *et al.*, 1995). Recombination of extra-chromosomal substrates mediated by these mutants was hypersensitive to coding sequence variations that did not affect recombination with normal RAG1.

Until recently, the mechanism of V(D)J recombination could only be investigated by analysis of the DNA products made in cells. Double strand breakage at recombination signals was reproduced first with a cell-free system using nuclear extracts and purified RAG1 protein (van Gent *et al.*, 1995), then subsequently with only the RAG1 and RAG2 proteins (McBlane *et al.*, 1995). Thus, it is now possible to determine DNA sequence and structure requirements much more accurately, taking into account that cleavage occurs in two identifiable steps. A nick is first introduced immediately 5′ of the heptamer, between the heptamer and the coding flank. In the second step, the nick is converted into a hairpin coding end and a blunt signal end; these same intermediates have been observed in cells.

Here we show that the different elements of the recombination signal have different functions in these steps.

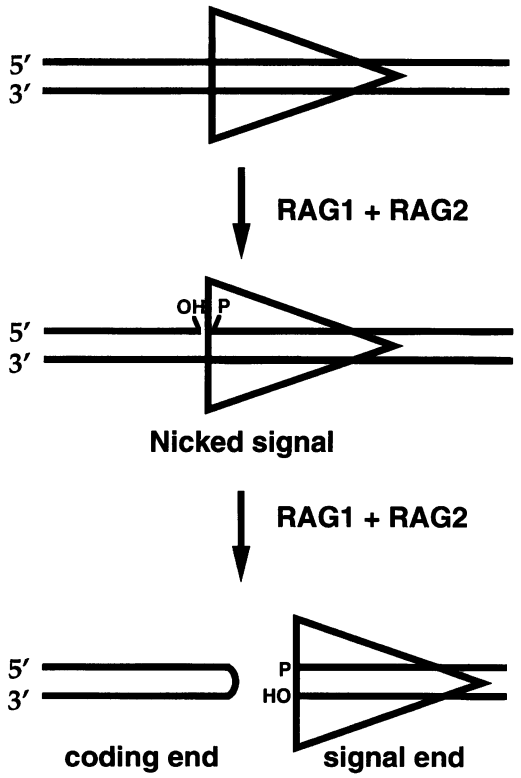

Fig. 1. Cleavage of a V(D)J recombination signal by RAG proteins. Parallel lines denote DNA strands of the recombination substrate. The recombination signal sequence is represented by a superimposed triangle. The strand where cleavage is initiated (nicking step) is defined as the top strand, and its complementary strand as the bottom strand (see text).

We also demonstrate that recognition and cleavage of the recombination signal occurs when DNA near the site of cleavage is unpaired, an observation that is consistent with the distortion of DNA structure required to form the intramolecular hairpin product. DNA unpairing near the site of cleavage also appears to play a related role in the Mu transposition (Lavoie *et al.*, 1991; Savilahti *et al.*, 1995) and human immunodeficiency virus (HIV) integration reactions (P.O.Brown, personal communication), suggesting it may be a common feature of transpositional recombination. This extends the recent observation that V(D)J signal cleavage by RAG1 and RAG2 displays several mechanistic similarities to these transposition reactions (van Gent *et al.*, 1996). Our observations indicate several ways in which mechanistic requirements for cleavage of the V(D)J recombination signal could regulate substrate choice in cells.

Results

We have shown previously that RAG1 and RAG2 proteins alone are sufficient for cleavage of the V(D)J recombination signal (Figure 1) (McBlane *et al.*, 1995). Here we investigate the substrate requirements for each step of the cleavage reaction. Most oligonucleotide substrates used in this study are based on a 50 bp DNA fragment containing a recombination signal with a 12 bp spacer, and 16 bp of flanking DNA that would correspond to the coding segment DNA adjacent to recombination signals. A ^{32}P label at the 5′ end of the coding flank allows identification of both the nicked precursor and final hairpin product after analysis by denaturing gel electrophoresis.

Distinct roles of nonamer and heptamer motifs

To examine the role of different elements of the recombination signal in the cleavage reaction, we first performed cleavage on substrates possessing either a heptamer or nonamer alone. Mutation of the entire heptamer resulted in low levels of nicking distributed across several sites, remarkably enough centered about the position relative to the nonamer that is nicked in a fully consensus recombination signal (Figure 2A, lane 6). Thus, the nonamer alone is capable of targeting some nicking activity to an appropriate site. Nevertheless, nicking was slow (data not shown) and inaccurate, indicating the importance of the heptamer in this cleavage step. The heptamer is even more critical for hairpin formation: no hairpins were formed in the absence of this element.

When the heptamer was retained but the entire nonamer was substituted with random sequence, both nicks and hairpins were still found, but overall cleavage was reduced 7-fold (Figure 2B). Nicks were still introduced rapidly and accurately (data not shown), and were converted to hairpins as efficiently as in substrates with an intact nonamer. Mutation of the nonamer therefore does not appear to inhibit specifically either of the chemical steps in cleavage. A single base mutation at the most conserved position of the nonamer, as well as mutations in all nonamer positions except the two most conserved, resulted in a similar reduction in overall cleavage levels (Figure 2B). Thus, the heptamer alone can specify accurate cutting, and the nonamer at the correct distance improves its efficiency (see also below).

Requirements for nicking and hairpin formation

We have identified the heptamer as the minimal element necessary for completion of the cleavage reaction. As mentioned in the Introduction, only the first three positions of the heptamer are critical for efficient V(D)J recombination in cells (Hesse *et al.*, 1989). Substitution of these three positions also blocked the ability of RAG1 and RAG2 to mediate production of double strand breaks (Figure 2A, lane 8 and Figure 2B). Nicks were slightly more abundant and accurate than observed when the heptamer was completely mutated but, again, no hairpins were observed. In contrast, when the sequence of the last four positions of the heptamer was changed, both nicks and hairpins were reduced <2-fold (Figure 2A, lane 10 and Figure 2B).

The importance of the heptamer in mediating nicking and hairpin formation therefore rests principally on the first three positions. When these positions were altered individually, the substrates were also unable to progress efficiently through all steps in the cleavage reaction. However, different steps were inhibited by the various mutations.

Mutation of the first position to any other nucleotide allowed as much as half of the nicking observed in a normal substrate, but in no instance were hairpins formed (Figure 2A, lane 12 and Figure 2B). Hairpin formation was also blocked when a substrate containing a pre-existing nick at the heptamer border was mutated at this

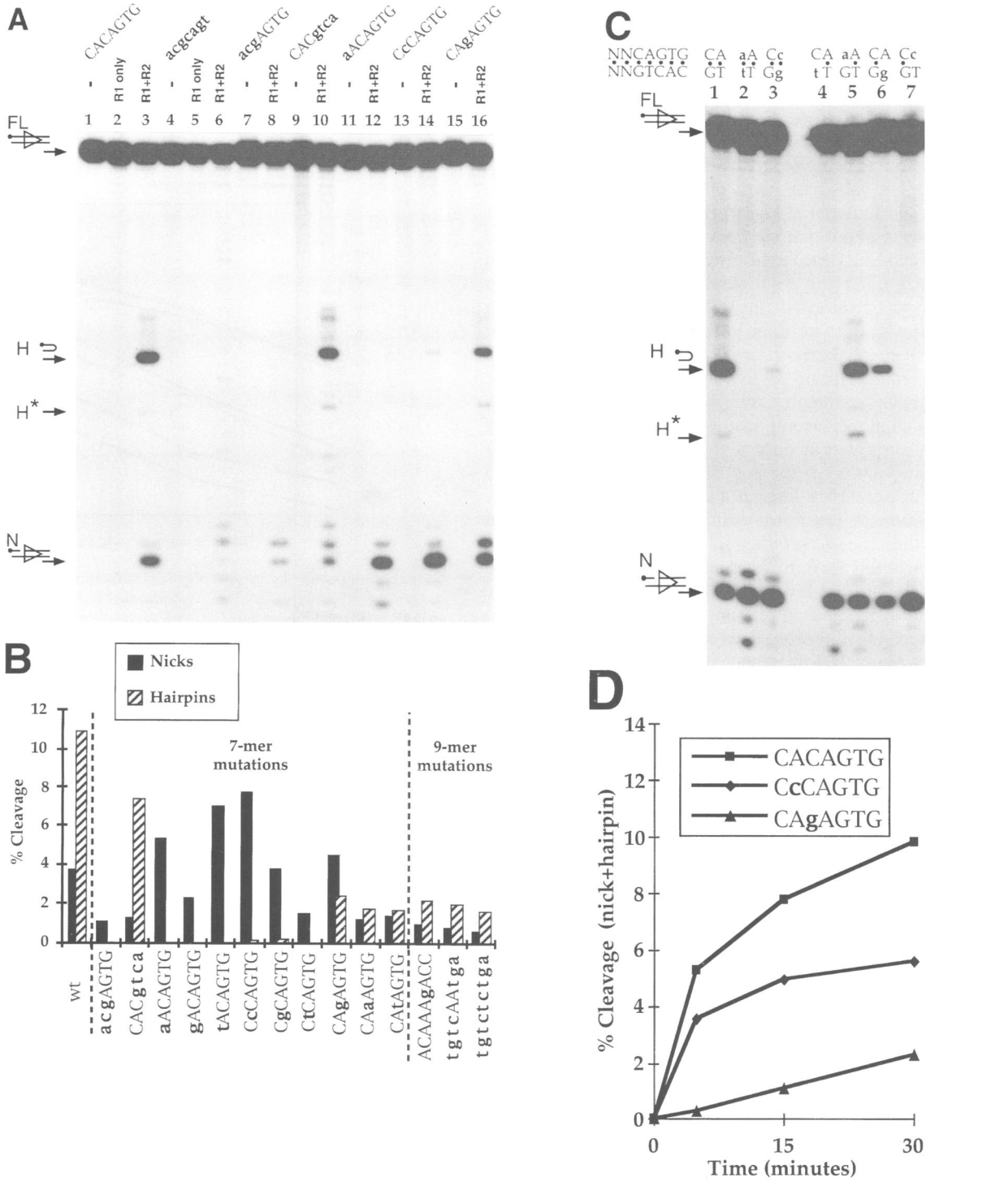

Fig. 2. Effects of mutations on cleavage by RAG proteins. Mutations were made in the context of the 12 bp spacer substrate, DAR39/40. In all figures, mutations are denoted by bold lower case lettering. (**A**) Mutations of the heptamer sequence. R1: RAG1 protein included. R2: RAG2 protein included. Reaction conditions and controls are as described in Materials and methods. Reaction products were separated by denaturing gel electrophoresis, quantified using a phosphorimager and visualized by autoradiography. Each of the following products is similarly labeled in all figures, and their identity was confirmed by co-migration of synthetic oligonucleotides with the predicted sequence. FL, full-length substrate; H, full-length hairpin (including all coding sequence); H*, smaller hairpin species (see Results, and Figure 6A); N, nick at heptamer–coding border. The minor species observed above H in this figure and others is also observed in electrophoresis of the synthetic oligonucleotide hairpin, and therefore represents an alternate form of H. (**B**) Quantitation of nicks and hairpins formed with mutated heptamer and nonamer sequences. wt; wild-type substrate (DAR39/40). Values are averaged from at least three experiments. Only precise nicks and hairpins are represented. (**C**) Cleavage of signals with paired or unpaired bases. Reaction products are labeled as in (A). Shown again are cleavage reactions with duplex forms of substrates with wild-type sequence (lane 1), or mutations at either the first (lane 2) or second (lane 3) position of the heptamer. Positions that are base-paired are noted by filled circles between the nucleotide of each strand; thus lanes 4 and 5 show cleavage reactions with heteroduplex substrates where the first position was unpaired, and lanes 6 and 7 show cleavage reactions with heteroduplex substrates where the second position was unpaired. (**D**) Kinetics of nicking. Aliquots of reactions were taken 5, 15 and 30 min after transfer to 37°C. Since nicks are the precursors of hairpins, we represent cleavage by plotting the sum of both species.

position (data not shown), indicating that the second step of cleavage is specifically inhibited. Substitution of the second position with any other nucleotide led to similar behavior (Figure 2A, lane 14 and Figure 2B).

The CA/TG dinucleotide in the first two positions of the heptamer is evidently essential for forming hairpins. Exactly what is being recognized at these positions? Using heteroduplex substrates with one normal strand and the other strand mutated at one of the first two positions, we were able to show that only one strand needs to possess the wild-type sequence to allow efficient hairpin formation (Figure 2C). At the first position, the ability to form hairpins was retained if the wild-type nucleotide was in the bottom strand (we refer to the strand that is nicked in the first step of the reaction as the 'top' strand, and the complementary strand as the'bottom strand'; see Figure 1) (Figure 2C, compare lanes 4 and 5). A substrate with a mismatch at the second position also retained the ability to form hairpins, but only if the wild-type nucleotide at this position was in the top strand (the nicked strand) (Figure 2C, compare lanes 6 and 7).

Mutations at the third position of the heptamer did support the efficient production of hairpins. For example, a heptamer with the third position changed to a G generated ~1/4 of the normal level of hairpins (Figure 2A, lane 16 and Figure 2B), and more than half the normal level if a pre-nicked substrate was used (data not shown). However, nicking was much slower than with the wild-type substrate, or even with a mutant specifically inhibited in its ability to form hairpins (Figure 2D). It appears, therefore, that mutations at the third position of the heptamer mainly impair the nicking step of the cleavage reaction. Consistent with this argument, cleavage of substrates with mutations at the third position produced a high proportion of imprecise nicks.

In summary, the heptamer is critical for generation of double strand breaks by RAG1 and RAG2. As previously found for recombination of extra-chromosomal substrates in cells (Hesse *et al.*, 1989), the first three positions of the heptamer are most important. Surprisingly, several of the mutations of the first three positions supported high levels of one cleavage step or the other, suggesting that RAG1 and RAG2 were still able to bind to these mutated signals. However, cleavage was impaired either at the nicking step or the hairpin formation step.

Role of the nonamer in sequence-specific binding

We have already shown above that the nonamer augments cleavage at the heptamer border. Since nonamer mutations led to reduced levels of cleavage without specifically inhibiting either cleavage step, it seemed possible that the nonamer was primarily important for sequence-specific binding. We therefore assayed the relative ability of mutant substrates to act as competitors in the cleavage reaction.

In these experiments, radiolabeled consensus recombination signal DNA was mixed with increasing amounts of different unlabeled competitors. As expected, when the consensus recombination signal was used as the competitor, the degree to which cleavage was inhibited was equal to the fraction of substrate present as competitor (a 1:1 ratio of competitor to target resulted in ~50% inhibition of cleavage) (Figure 3). Addition of only 10-fold more of a non-specific DNA species was able to compete to a similar degree, suggesting that RAG1 and RAG2 are not capable of a high level of sequence discrimination at the binding step, at least under the conditions used here.

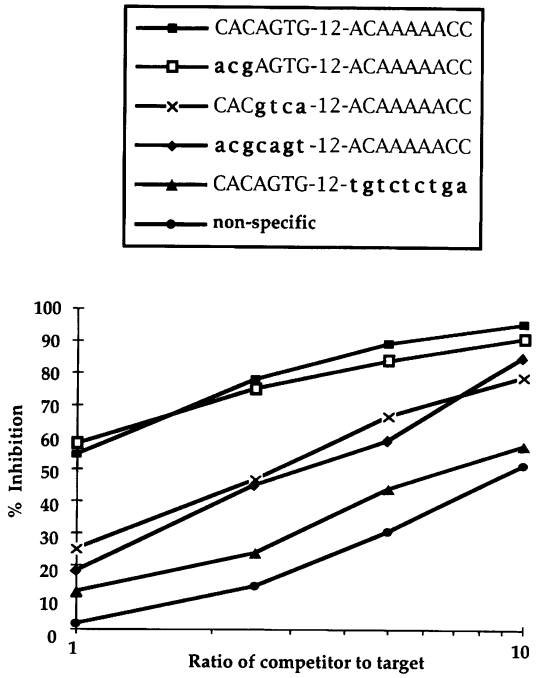

Fig. 3. Competition by various oligonucleotides in cleavage. Standard reactions were assembled on ice, with labeled wild-type substrate (DAR39/40) and the specified amount of unlabeled competitor DNA substrates. Reactions were then started by addition of RAG proteins, and incubated for 60 min at 37°C. Percentage inhibition is the reduction of substrate cleavage relative to a reaction without competitor.

The ability to act as an effective competitor was mainly dependent on the presence of the nonamer; DNA with an intact heptamer and a randomized nonamer was scarcely more effective than a non-specific fragment (Figure 3). Mutations in the heptamer had a modest effect on competition, and only if the last four positions were altered. Mutation of these positions reduced competition almost as much as ablation of the entire heptamer. Changes only in the first three positions of the heptamer did not reduce the ability of the recombination signal to act as a competitor, supporting the argument that these positions are involved specifically in cleavage, rather than binding.

Importance of spacer length

We have shown that either the heptamer or nonamer motif alone can partly support the activity of RAG1 and RAG2, although full activity requires the cooperation of both motifs. How is their interaction transmitted across the intervening spacer?

The spacing between the heptamer and nonamer in 12 and 23 signals differs by approximately one turn of a B form DNA helix. The relevance of this spacing to the ability of the heptamer and nonamer to cooperate in cleavage was investigated by comparing cleavage with substrates possessing 12 and 23 bp spacers with substrates with a 18 bp spacer (1/2 turn longer than 12 bp spacer), a 29 bp spacer (1/2 turn longer than 23 bp spacer) or a 34 bp spacer (one turn longer than 23 bp spacer).

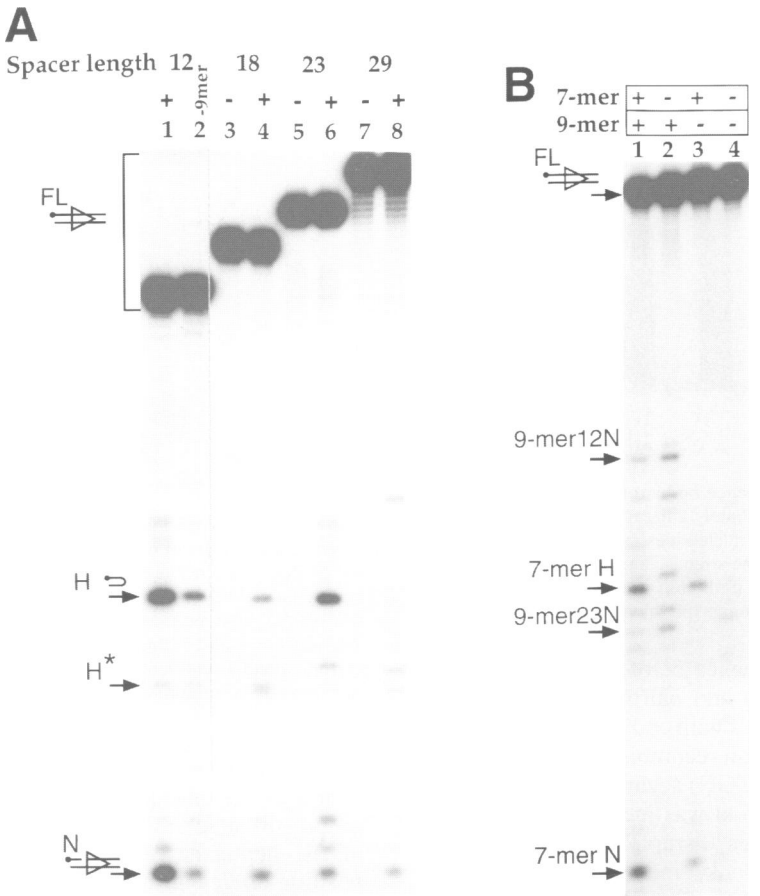

Fig. 4. Effects of spacer length on cleavage. (**A**) Reactions of substrates with the indicated spacer lengths are shown. –9mer; the 12 signal substrate with a mutated nonamer, as described in Figure 2B. (+): RAG proteins included. Reaction products are labeled as in Figure 2A. (**B**) Substrate with 34 bp spacer. The products of cleavage at the heptamer–coding flank border (7-merN and 7-merH) were identified in the same manner as in previous experiments. 9-mer12N and 9-mer23N are located 19 nucleotides and 30 nucleotides 5′ of the nonamer, respectively, as determined by comparison with a ladder of DNase I digestion products of the full-length substrate. Cleavage products were characterized as nicks (N) and hairpins (H) by 2-D gel electrophoresis (data not shown).

The presence of a nonamer 12 or 23 bp away from a heptamer results in increased cleavage at the heptamer (Figure 4A, compare lanes 1 and lane 6 with cleavage of a substrate without a nonamer, lane 2), as expected. In contrast, presence of the nonamer 18 or 29 bp from the heptamer even interferes with cleavage at the heptamer; nicks are observed at levels similar to those in a substrate without a nonamer, but hairpin formation is somewhat reduced (Figure 4A, compare lanes 4 and 8 with lane 2).

The action of RAG1 and RAG2 on a 34 bp spacer substrate generated nicks immediately 5′ of the heptamer, and efficiently converted these nicks to hairpins (Figure 4B, lane 1). To see more clearly whether the heptamer and nonamer are cooperating, functioning independently or interfering with each other in this substrate, either the heptamer, nonamer or both motifs were mutated in the same context. Both nicks and hairpins at the heptamer border were observed at significantly higher levels in the substrate with an intact heptamer and nonamer than in a substrate with a mutated nonamer (Figure 4B, compare lanes 1 and 3), arguing that the heptamer and nonamer cooperate even when separated by 34 bp.

In addition to cleavage occurring at the heptamer border, nicks were observed both at and between the appropriate sites of cleavage for recombination signals with 12 and 23 bp spacers, relative to the nonamer, in both the 29 bp spacer substrate (Figure 4A, lane 8) and the 34 bp spacer substrate (marked 9-mer12N, and 9-mer23N in Figure 4B, lanes 1 and 2). That the nonamer alone directs this nicking activity has been confirmed in substrates that do not possess an intact heptamer (Figure 4B, lane 2; see also Figure 2A, lane 6).

Coding flank sequence influences the ability to form hairpins

As mentioned in the Introduction, certain coding flank sequences reduced recombination as much as 100-fold with a mutant version of RAG1 ('D32'), while these same coding flank sequences did not significantly affect recombination with wild-type RAG1 (Sadofsky *et al.*, 1995). The two base pairs immediately adjacent to the site of cleavage were mainly responsible for this effect.

Surprisingly, cleavage mediated by the RAG proteins showed a similar pattern of coding flank preference even though wild-type RAG1, rather than RAG1-D32, was used. This preference for coding flank sequence was observed only at the hairpin formation step, however. When substrates with coding flank sequences that were shown to be relatively insensitive to the RAG1-D32 mutation in cells were used in the cell-free cleavage

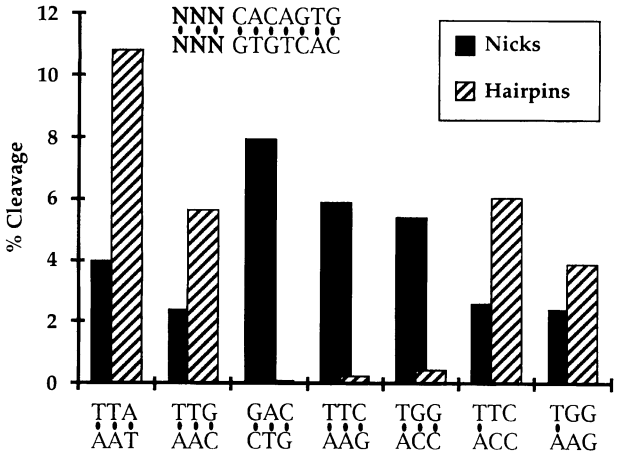

Fig. 5. Effect of coding flank sequence on hairpin formation. The three coding flank positions immediately adjacent to the heptamer were varied. Positions with paired nucleotides are noted with a filled circle between the sequences for the strands. The sequence of the substrate with a TTA coding flank in the top strand is that of the 'wild-type' substrate used in previous experiments.

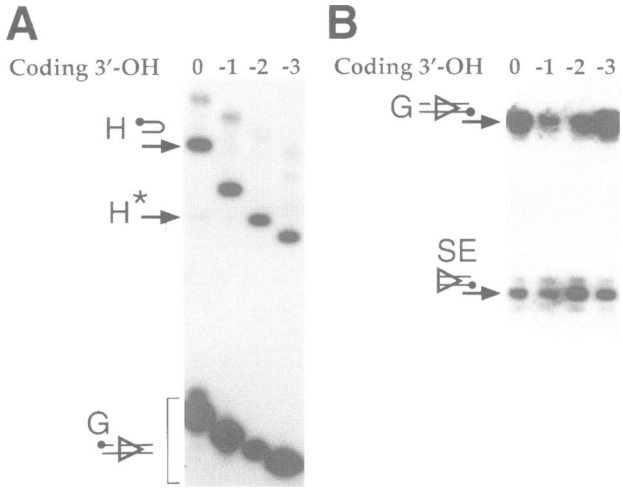

Fig. 6. Hairpin formation from misplaced 3′-OH ends. Substrates with a gap in the top strand were generated by removing nucleotides from the 5′ side of the heptamer–coding flank border. The location of the coding flank 3′-OH is noted in terms of the number of coding flank nucleotides removed from the top strand. (**A**) Gapped substrates (G) with the top strand 5′ end-labeled. (**B**) Gapped substrates (G) with the bottom strand 5′ end-labeled. SE; the signal end cleavage product, identified by co-migration of an oligonucleotide with the predicted sequence.

reaction (i.e. when the sequence 5′ of the heptamer was TTA or TTG) both nicking and hairpin formation were efficient (Figure 5). Substrates with coding flank sequences that have been shown to inhibit recombination with RAG1-D32 in cells (GAC, TTC or TGG) accumulated high levels of nicks, but hairpin formation was inefficient.

Hairpin formation requires disruption of base pairing at the end of the coding flank, and coding flank DNA strands must be severely bent. The inability of RAG1/RAG2 to form hairpins with certain coding flank sequences may reflect limitations imposed by differences in flexibility. Flexibility can, in principle, be increased by introducing unpaired bases in the flanking DNA; one might then expect that hairpin formation would be restored in such substrates. Indeed, substrates with unpaired bases in the coding flank readily made hairpins, even if both strands were composed of flanking sequences that in homoduplex form did not permit efficient hairpin formation (Figure 5).

Imprecise nicks can be converted to hairpins

In the previous experiment, when heteroduplex DNA was present in the coding flank, several smaller hairpins were observed in addition to the normal hairpin product (data not shown); a smaller hairpin species has also been observed in experiments when the coding flank was in homoduplex form (H* in Figures 2A and C and 4A). These smaller hairpin species must be derived from cleavage within the coding flank on one or both strands. To determine if nicks located in the coding flank are precursors to hairpins, we assembled several pre-nicked substrates where the nicks were displaced one, two or three nucleotides 5′ of the normal site of nicking.

Substrates with a nick located in the coding flank could readily form hairpins, which became progressively smaller as the nick was moved further into the coding flank (data not shown). Similar amounts of the identical products were also obtained if the nucleotides between the coding strand terminal 3′-OH and the normal site of nicking were removed ('gapped' substrates) (Figure 6A). Thus 3′-OH groups displaced into the coding strand are incorporated efficiently into hairpins.

The misplaced 3′-OH might attack the complementary strand at the phosphodiester bond directly opposite, or might still attack the phosphodiester bond immediately 3′ of the heptamer. We investigated this by locating the position of cleavage in the bottom strand of the reciprocal product (the signal end). Using substrates that were 5′ end-labeled in the bottom strand, we determined that cleavage of this strand occurred exactly at the border between the heptamer and coding flank, independently of the position of the coding flank 3′-OH (Figure 6B).

In summary, smaller hairpins appear to be the product of joining the misplaced 3′-OH diagonally to the coding flank nucleotide on the complementary strand immediately 3′ of the heptamer. We have confirmed this for the hairpin species produced when the 3′-OH is displaced two nucleotides into the coding flank, by comparison with a synthetic hairpin oligonucleotide with the predicted sequence. This marker oligonucleotide also co-migrated with the minority hairpin species observed in previous experiments (species 'H*' in Figures 2B and C and 4A), indicating that this species was also the product of the attack of a 3′-OH located in the coding flank 2 bp 5′ of the coding flank–heptamer border. Aside from the normal hairpin, this is the only other detectable hairpin species in reactions with the wild-type substrate, suggesting that this reaction is a favored alternative to the normal reaction.

Pre-nicked substrates where the nick was displaced 3′ of the normal site (into the heptamer) were also efficient substrates for the formation of hairpins. The hairpin products of these substrates, however, were identical to those produced when a standard pre-nicked substrate was used. This reaction involves the introduction of a second nick at the normal site prior to hairpin formation (data not shown).

Single-stranded recombination signals are substrates for cleavage

Experiments with heteroduplex substrates have shown that cleavage by RAG1 and RAG2 tolerates the presence of

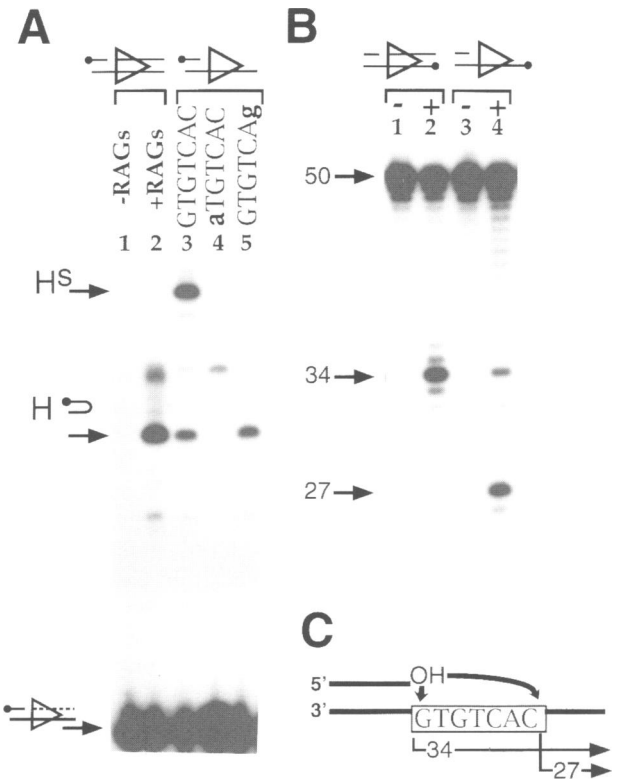

Fig. 7. Cleavage of a single-stranded DNA signal. In (A) and (B), lanes 1 and 2 for both panels are reactions with a substrate containing a pre-existing nick at the heptamer–coding flank border; all other lanes are reactions with a substrate whose top strand contains only coding flank sequence. The substrates are diagrammed at the top of the figure. (**A**) Cleavage products of substrates 5′ end-labeled in the top strand. Heptamer sequences are written for the bottom strand, 3′ to 5′, as in (C). H^{s}; hairpins specific to reactions with a ssDNA signal. (**B**) Cleavage products of substrates 5′ end-labeled in the bottom strand. Sizes are determined by comparison with a ladder of DNase I digestion products of the full-length substrate. (**C**) Diagram of the products.

one or two unpaired bases in the heptamer. This suggested the possibility that the RAG proteins might recognize the signal in single strand form. Indeed, hairpins were also made from a substrate where the entire recombination signal was single-stranded, while only the coding flank was present in duplex DNA form (Figure 7A). This substrate is comparable with previously described pre-nicked substrates, in the sense that a 3′-OH is present at the terminus of the coding flank, and thus can act as a nucleophile to form hairpins. Two hairpin species were formed in reactions with this substrate (Figure 7A, lane 3); one was identical to the hairpins produced from pre-nicked substrates with duplex DNA signals (marked 'H' in Figure 7A), but reduced in amount. The other hairpin species (marked 'H^{s}' in Figure 7A) was larger than the normal one, and almost as abundant as the normal hairpin product from fully duplex pre-nicked substrates.

Mutation of the first position of the heptamer blocks formation of both hairpin products (Figure 7A, lane 4); the same mutation in the context of a duplex signal also blocked formation of the normal hairpin product (Figure 2A, lane 12). Thus, recognition of the single-stranded signal displays some similarity to recognition of normal duplex DNA substrates. However, presentation of the recombination signal in single-stranded form apparently relieves the requirement for the correct nucleotide at the second position, as the top strand is no longer present in this substrate. Also, mutation of the nonamer had little effect.

The length of the cleaved bottom strand confirmed that the normal sized hairpin was made by the coding flank 3′-OH attacking the normally targeted phosphodiester bond (immediately 3′ of the heptamer) (Figure 7B, lane 4 and Figure 7C). The larger hairpin product ('H^{s}') apparently resulted from attack by the coding flank 3′-OH on the bond immediately 5′ of the opposite end of the heptamer (Figure 7B, lane 4 and Figure 7C).

The sequence of the site of nucleophilic attack for the larger hairpin species is similar to the site of nicking in a normal double-stranded signal (5′ of CAC), due to the palindromic nature of the heptamer. This raises the possibility that the heptamer is recognized in the reversed orientation in this reaction. The observation that mutation of the last nucleotide of the heptamer inhibits formation of the larger hairpin product (H^{s}), but not the normal hairpin (Figure 7A, lane 5), supports this hypothesis. Thus, recognition and cleavage of a single-stranded signal by the RAG proteins can occur at the normal site, or cleavage can be directed to the opposite end of the heptamer in a reaction involving recognition in the reversed orientation (Figure 7C).

Discussion

Substrate requirements change during cleavage

It was shown previously that RAG1 and RAG2 proteins alone are capable of cleaving DNA at recombination signals (McBlane *et al.*, 1995). This reaction occurs in three separable stages. RAG1 and/or RAG2 must first bind to the recombination signal in a sequence-specific manner. The RAG proteins then introduce a nick immediately 5′ of the heptamer, after which the nick is converted to a hairpin coding end and a blunt signal end. We have now determined that different elements of the recombination signal are required for each step.

The nonamer is probably the most important element in initial binding. This was determined through the use of mutated substrates as competitors in cleavage, as specific binding of the active complex to a recombination signal has not yet been demonstrated by direct physical means.

In contrast, nicking and hairpin formation are mainly dependent on the heptamer. Nicking is more tolerant of substrate variation than is hairpin formation. Accurate and efficient nicking is most dependent on the first three positions of the heptamer, with other conserved sequences playing a secondary role.

The substrate requirement for hairpin formation is the most stringent. The first two positions of the heptamer are critical. However, a substrate possessing only the first three positions of the heptamer supported barely detectable levels of nicking and no hairpin formation (data not shown), indicating that this later step also requires either the remainder of the heptamer or the nonamer. Hairpin formation also requires 'permissive' coding flank sequence, a point to which we return below.

Comparison with the substrate requirements for recombination in cells

The first three nucleotides of the heptamer have been shown previously to be individually critical for V(D)J recombination in cells (Hesse *et al.*, 1989). In RAG-mediated cleavage, the first two nucleotides are critical only for completing the final step in signal cleavage which converts nicks to hairpins. Mutations in these positions allow efficient nicking, and do not significantly affect the ability of a signal to act as a competitor. This portion of the recombination signal is thus likely to be comparatively insignificant in any steps prior to hairpin formation. This last observation suggests that these same mutations may also allow nicking of signals in cells, a possibility that could be tested readily at similarly mutated endogenous recombination signals or in artificial substrates. In actively rearranging cells, such nicks in the many 'pseudo-signals' in the genome might lead to DNA damage, with ensuing effects on cell fate. Although a single mutation at the third position severely impairs V(D)J recombination on plasmid substrates in lymphocytes, the same mutation allows both nicking and hairpin formation, but at a much slower rate. This suggests that the time required for cleavage could be a much more limiting factor when complete recombination is demanded.

A heptamer alone is capable of supporting RAG-mediated cleavage. Nonamer-independent recombination also occurs in cells both at endogenous loci [e.g. V gene replacement (Kleinfield *et al.*, 1986; Reth *et al.*, 1986)] and in artificial substrates (Hesse *et al.*, 1989), although at a very low level. In contrast, nonamer-independent cleavage in the cell-free assay is only 7-fold below that of an intact 12 bp spacer signal. Such limited sequence discrimination would probably not be tolerated in recombinationally active cells. However, we have shown that cell-free cleavage is very sensitive to the presence of competitor DNA, even when the competitor sequence is completely non-specific; moreover, the nonamer is probably the most important feature of the signal in determining competitor sensitivity. Therefore, the nonamer is likely to play a much more important role in recombination in cells, where stability of the cleavage complex in the presence of the vast amount of genomic DNA will be critical.

Although RAG-mediated cleavage can occur with a heptamer alone, cleavage nevertheless is most efficient when both heptamer and nonamer are present and separated by either 12 or 23 bp. We find that a heptamer and nonamer separated by a 34 bp spacer can also cooperate in cleavage at the signal–coding border, indicating that recognition by the active cleavage complex must align proteins bound to the heptamer and nonamer similarly with respect to the helical phase.

Some coding flank sequences that have only modest effects on recombination in cells inhibit hairpin formation by RAG1 and RAG2 as much as 100-fold (Sadofsky *et al.*, 1995). These inhibitory ('non-permissive') coding flank sequences are the same as those that showed reduced recombination with RAG1-D32; this is a surprising observation, as wild-type RAG1 was used in the cell-free assay. Why then is recombination with wild-type RAG1 in cells not sensitive to these coding flank variations? Perhaps hairpin formation with non-permissive coding flanks is rescued in cellular conditions by some other protein factor; the D32 mutation might render RAG1-D32 unable to interact with this other factor. Alternatively, conditions in cells may increase the reactivity of substrates sufficiently for a critical threshold to be surpassed, and wild-type RAG1 no longer discriminates between different coding flank sequences. RAG1-D32 may be an attenuated version of RAG1 that remains unable to surpass this threshold under cellular conditions.

Cleavage of a favorable substrate usually introduces a nick at the coding–heptamer border and converts this nick to a hairpin that retains all of the coding flank sequence. However, a minor hairpin species is derived from a displaced nick, located 2 bp into the coding flank. The displaced 3′ end of the coding strand is still joined to the usual position on the opposite strand (the border between the coding flank and the heptamer). Such a reaction may have a parallel during recombination in cells; a small fraction of broken signal ends produced at the TCR δ locus has a 2 bp 5′ overhang derived from coding flank DNA (Roth *et al.*, 1993). This end would be the reciprocal product of a 2 bp staggered hairpin. These observations suggest the interesting possibility that some proportion of the deletions and apparent N insertions in coding junctions could be derived from imprecision in cleavage, rather than imprecision in joining.

A role for DNA unpairing at the site of cleavage?

The inhibition of hairpin formation by certain coding flank sequences can be relieved if base pairing of coding DNA next to the signal is disrupted, either by introducing a heteroduplex (Figure 5), or by creating a gap between the 3′ end of the top strand and the heptamer (data not shown). Attack by the 3′-OH on the opposite strand must involve severe bending of the strands, and we argue the enhanced flexibility required for hairpin formation may be lacking in unfavorable coding flank sequences.

Base unpairing has been shown to enter the reaction in another way. A remarkable feature of signal cleavage is the acceptance of unpaired or single-stranded DNA in the heptamer. The use of base mismatches in the signal is one example. Even more striking is the efficient use of an entirely single-stranded signal to generate hairpins. If a double-stranded coding flank is present to supply the attacking 3′-OH group, the heptamer is still recognized when only the bottom strand is present, and hairpins are made almost as well as with a fully native signal. The ability to utilize the heptamer in single-stranded form suggests a role for such a mode of recognition in cleavage of normal duplex DNA substrates.

For example, the RAG proteins might unwind the heptamer in the course of the cleavage reaction. In addition, we note that the CACA sequence that forms part of the heptamer has been reported to have an intrinsically abnormal structure (Cheung *et al.*, 1984; Patel *et al.*, 1987; Timsit *et al.*, 1991), including a disruption of normal base pairing. This distorted structure may even help define the RAG1/RAG2 cleavage site, a possibility supported by the relaxation of the requirement for an A at the second position in the top strand when a single-stranded signal is used. Moreover, it can be noted that coding flank sequences that permit hairpin formation (TA or TG immediately 5′ of the heptamer) continue the pyrimidine–purine alterna-

tion of the first four bases of the heptamer. A reasonable hypothesis is that these sequences promote extension of the heptamer-induced disruption of base pairing into the coding flank. We suggest that unpairing is likely to be important for cleavage by the RAG proteins, because of the intramolecular nature of hairpin formation, with its required distortion of DNA structure.

Previous experiments from this laboratory have pointed to mechanistic similarities of cleavage by the RAG proteins to HIV integration and the transposition reaction of bacteriophage Mu (van Gent *et al.*, 1996). Unpairing near the site of cleavage may be another shared feature of these reactions. Formation of the stable Mu transpososome (Savilahti *et al.*, 1995) and processing of the 3′ end of HIV DNA (P.O.Brown, personal communication) are both aided by disruption of base pairing in flanking DNA.

As pointed out before, alternation of the CA/TG dinucleotide in the heptamer may result in disruption of normal base pairing at the cleavage site, and this unpairing may contribute to the ability of a substrate to support hairpin formation by the RAG proteins. HIV integration (La Femina *et al.*, 1991) and Mu transposition (Burlingame *et al.*, 1986; Surette *et al.*, 1991) also require a terminal CA/TG dinucleotide for cleavage steps, but not sequence-specific binding. Although it is unclear if a single CA/TG dinucleotide has a significantly altered DNA structure, it is of interest that HIV integrase retains activity on substrates with this dinucleotide present in unpaired form (van den Ent *et al.*, 1994). We show here that the RAG proteins are also capable of recognizing this dinucleotide in unpaired form, and recognition of the unpaired CA/TG dinucleotide is similar in that heteroduplex substrates supported both reactions when a G at the first position on the bottom strand and an A at the second position on the top strand were present (van den Ent *et al.*, 1994).

If unpairing of DNA is a normal part of RAG-mediated cleavage, there is an intriguing possibility that distortion of DNA structure by external factors could contribute to regulating recombination. Signal sequences could be more or less available, depending on a state of helix opening modulated by local supercoiling or the action of helicases or other DNA binding proteins.

Materials and methods

Oligonucleotide cleavage assay

Mouse RAG1 and RAG2 fusion proteins [MR1 and MR2 (McBlane *et al.*, 1995)] were co-expressed and purified from baculovirus-infected cells, as previously described. Similar results can be obtained using non-fusion versions of RAG1 and RAG2 proteins [R1 and R2 (McBlane *et al.*, 1995)]. All RAG proteins are truncated derivatives that have been shown to mediate recombination in tissue culture cells (Sadofsky *et al.*, 1993, 1994; Cuomo and Oettinger, 1994).

Standard reaction conditions involve incubation of 0.2 pmol of a ^{32}P 5′ end-labeled duplex oligonucleotide substrate with 2 μl of co-purified RAG1 and RAG2 (~100 ng of each protein) in a volume of 10 μl, for 60 min at 37°C. Reactions were supplemented with 25 mM MOPS pH 7.0, 1 mM dithiothreitol (DTT), 60 mM potassium glutamate and 1 mM $MnCl_2$. A 2-fold increase in RAG protein added (4 μl) resulted in an ~2-fold increase in the amount of cleavage, indicating that the assay is in a proportional range.

To determine the level of contaminating nuclease activity in comparison with cleavage activity derived from the presence of both RAG proteins, we used a similarly prepared sample of RAG1 protein only; compare lane 2 with lane 3, or lane 5 with lane 6, in Figure 2A.

Control lanes (marked '–' in figures) contain an aliquot either of the standard reaction prior to incubation or incubations of the substrate with buffer alone. Reactions were stopped by addition of 2 vols of formamide loading dye, and 1/10 of the sample was analyzed by polyacrylamide gel electrophoresis under denaturing conditions. A Molecular Dynamics phosphorimager and ImageQuaNT software (v4.1) were used for quantification of reaction products; images of gels shown in the figures are from autoradiograms.

Using a wild-type 12 spacer substrate and standard conditions, an average of 3.8% (SD; 1.4%) of substrate was present as accurately nicked DNA, and an average of 10.9% (SD; 3.2%) was fully converted to hairpins.

Two-dimensional gel electrophoresis was used to characterize further the cleavage products in Figure 4B, and was performed as previously described (McBlane *et al.*, 1995).

Construction of substrates

Oligonucleotides for cleavage substrates were produced using a Millipore 8909 synthesizer, and purified by polyacrylamide gel electrophoresis under denaturing conditions. Substrates typically possess 16 bp of DNA next to the heptamer (the 'coding' flank) and 6 bp of DNA next to the nonamer. Unless otherwise stated, the substrates were 5′ end-labeled with ^{32}P on the coding flank side, by standard techniques (Sambrook *et al.*, 1989). The 12 bp spacer recombination signal is derived from a signal adjacent to a mouse Vκ gene segment, and the 23 bp spacer recombination signal from the recombination signal that mediates rearrangement of the mouse Jκ1 gene segment.

The 12 bp spacer recombination substrate was made by annealing DAR39, 5′GATCTGGCCTGTCTTACACAGTGCTACAGACTGGAACAAAAACCCTGCAG, to an oligonucleotide with complementary sequence (DAR40). All heptamer, nonamer and coding flank mutations were made in the context of this substrate; the sequences of these mutations are provided in each figure. Non-specific substrate (Figure 3) was assembled by annealing DAR81, 5′GATCTCGCCTCTCTTAGGTTAATCCTATAGAACTCGTCCCCGTACCTCGAG, to an oligonucleotide with complementary sequence (DAR82).

The 18, 23, 29 and the non-mutated version of the 34 spacer substrates had the identical coding flank, heptamer and nonamer as the DAR39/40 substrate. Sequence flanking the nonamer for the 18 spacer substrate was identical to DAR39/40; for the 23, 29 and 34 spacer substrates, the sequence flanking the nonamer was 5′CTCGGG. The sequences of the top strand of the spacers are as follows: 18 spacer, 5′CTACAGACTGGATCTGGC; 23 spacer, 5′GTAGTACTCCACTGTCTGGCTGT; 29 spacer, 5′CTACAGGTAGTACTCGGCTGTCTGGCTGT; and 34 spacer, 5′CTACAGACTAGGTAGTACTCGGCTGTCTGGCTGT. Each 72 nucleotide long strand of the 34 spacer substrate was made by ligating together two oligonucleotides. Mutation of the heptamer in the 34 bp spacer substrate was accomplished by substitution of 5′CACAGTG with the sequence 5′ACGCTGA on the top strand; mutation of the nonamer in the 34 bp spacer substrate was accomplished by substitution of 5′ACAAAAACC with the sequence 5′AGTCTCTGT.

In assembly of all substrates, we included an excess of the unlabeled oligonucleotide(s) prior to annealing, to ensure that all labeled DNA was present in duplex form.

Acknowledgements

We are grateful to Joanne Hesse and Moshe Sadofsky for many helpful comments along the way, and to Harri Savilahti for critical reading of the manuscript. We also wish to acknowledge continuing valuable discussions with Marjorie Oettinger, David Roth and our colleagues in the Laboratory of Molecular Biology. D.A.R. is supported by the Medical Research Council of Canada, and D.C.v.G. by the European Molecular Biology Organisation.

References

Boubnov,N.V., Wills,Z.P. and Weaver,D.T. (1993) *Nucleic Acids Res.*, **23**, 1060–1067.

Burlingame,R.P., Obukowicz,M.G., Lynn,D.L. and Howe,M.M. (1986) *Proc. Natl Acad. Sci. USA*, **83**, 6012–6016.

Cheung,S., Arndt,K. and Lu,P. (1984) *Proc. Natl Acad. Sci. USA*, **81**, 3665–3669.

Cuomo,C.A. and Oettinger,M.A. (1994) *Nucleic Acids Res.*, **22**, 1810–1814.

Ezekiel,U.R., Engler,P., Stern,D. and Storb,U. (1995) *Immunity*, **2**, 381–389.
Gellert,M. (1992) *Annu. Rev. Genet.*, **22**, 425–446.
Gerstein,R.M. and Lieber,M.R. (1993) *Genes Dev.*, **7**, 1459–1469.
Hesse,J.E., Lieber,M.R., Mizuuchi,K. and Gellert,M. (1989) *Genes Dev.*, **3**, 1053–1061.
Kleinfield,R., Hardy,R.R., Tarlinton,D., Dangl,J., Herzenberg,L.A. and Weigert,M. (1986) *Nature*, **322**, 843–846.
La Femina,R.L., Callahan,P.L. and Cordingley,M.G. (1991) *J. Virol.*, **65**, 5624–5630.
Lavoie,B.D., Chan,B.S., Allison,R.G. and Chaconas,G. (1991) *EMBO J.*, **10**, 3051–3059.
Lewis,S.M. (1994) *Adv. Immunol.*, **56**, 27–150.
McBlane,J.F., van Gent,D.C., Ramsden,D.A., Romeo,C., Cuomo,C.A., Gellert,M. and Oettinger,M.A. (1995) *Cell*, **83**, 387–395.
Patel,D.J., Shapiro,L. and Hare,D. (1987) In Wells,R.D. and Harvey,S.C. (eds), *Unusual DNA Structures*. Springer, New York, pp. 115–161.
Ramsden,D.A. and Gellert,M. (1995) *Genes Dev.*, **9**, 2409–2420.
Ramsden,D.A., Baetz,K. and Wu,G.E. (1994) *Nucleic Acids Res.*, **22**, 1785–1796.
Reth,M., Gehrmann,P., Petrac,E. and Wiese,P. (1986) *Nature*, **322**, 840–842.
Roth,D.B., Nakajima,P.B., Menetski,J.P., Bosma,M.J. and Gellert,M. (1992a) *Cell*, **69**, 41–53.
Roth,D.B., Menetski,J.P., Nakajima,P.B., Bosma,M.J. and Gellert,M. (1992b) *Cell*, **70**, 983–991.
Roth,D.B., Zhu,C. and Gellert,M. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 10788–10792.
Sadofsky,M.J., Hesse,J.E., McBlane,J.F. and Gellert,M. (1993) *Nucleic Acids Res.*, **21**, 5644–5650.
Sadofsky,M.J., Hesse,J.E. and Gellert,M. (1994) *Nucleic Acids Res.*, **22**, 1805–1809.
Sadofsky,M., Hesse,J.E., van Gent,D.C. and Gellert,M. (1995) *Genes Dev.*, **9**, 2193–2199.
Sakano,H., Huppi,K., Heinrich,G. and Tonegawa,S. (1979) *Nature*, **280**, 288–294.
Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Savilahti,H., Rice,P.A. and Mizuuchi,K. (1995) *EMBO J.*, **14**, 4893–4903.
Schlissel,M., Constantinescu,A., Morrow,T., Baxter,M. and Peng,A. (1993) *Genes Dev.*, **7**, 2520–2532.
Surette,M.G., Harkness,T. and Chaconas,G. (1991) *J. Biol. Chem.*, **266**, 3118–3124.
Timsit,Y., Vilbois,E. and Moras,D. (1991) *Nature*, **354**, 167–170.
van den Ent,F.M.I., Vink,C. and Plasterk,R.H.A. (1994) *J. Virol.*, **68**, 7825–7832.
van Gent,D.C., McBlane,J.F., Ramsden,D.A., Sadofsky,M.J., Hesse,J.E. and Gellert,M. (1995) *Cell*, **81**, 925–934.
van Gent,D.C., Mizuuchi,K. and Gellert,M. (1996) *Science*, **271**, 1592–1594.

Received on February 5, 1996; revised on March 12, 1996