MOLECULAR AND CELLULAR BIOLOGY, Oct. 1996, p. 5683–5690 Vol. 16, No. 10
0270-7306/96/$04.00+0

DNA Sequence and Structure Requirements for Cleavage of V(D)J Recombination Signal Sequences

CHRISTINA A. CUOMO, CYNTHIA L. MUNDY, AND MARJORIE A. OETTINGER*

Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, and Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts 02114

Received 6 June 1996/Returned for modification 30 June 1996/Accepted 15 July 1996

Purified RAG1 and RAG2 proteins can cleave DNA at V(D)J recombination signals. In dissecting the DNA sequence and structural requirements for cleavage, we find that the heptamer and nonamer motifs of the recombination signal sequence can independently direct both steps of the cleavage reaction. Proper helical spacing between these two elements greatly enhances the efficiency of cleavage, whereas improper spacing can lead to interference between the two elements. The signal sequences are surprisingly tolerant of structural variation and function efficiently when nicks, gaps, and mismatched bases are introduced or even when the signal sequence is completely single stranded. Sequence alterations that facilitate unpairing of the bases at the signal/coding border activate the cleavage reaction, suggesting that DNA distortion is critical for V(D)J recombination.

Immunoglobulin and T-cell receptor genes are assembled from gene segments by a series of site-specific recombination reactions collectively termed V(D)J recombination. By joining different combinations of V, D, and J coding segments, this process generates much of the enormous diversity in antigen receptor specificity that is required for a functional immune system (for recent reviews, see references 7 and 12). Expression of the *RAG1* and *RAG2* genes is required for this joining reaction to occur; lymphoid cells with targeted disruptions of either gene do not carry out V(D)J joining (15, 25). Conversely, expression of the *RAG* genes in fibroblasts is sufficient to allow recombination to occur in these nonlymphoid cells (16, 23). Recent work has shown that the two RAG proteins together are the only protein factors required to carry out the initial cleavage step of V(D)J recombination in vitro (13).

Recombination signal sequences (RSSs) flank all recombinationally active gene segments and are recognized by one or both RAG proteins in vitro (13). RSSs consist of conserved heptamer and nonamer motifs separated by a nonconserved spacer region of either 12 or 23 bp (called 12-spacer or 23-spacer RSSs). V(D)J recombination requires two RSSs, one of each spacer length. Because the segments at a given locus are flanked by the same arrangement of RSSs (e.g., 12-spacer signals at Vκ and 23-spacer signals at Jκ), this 12/23 rule restricts segment joining to events that could be biologically productive. V(D)J recombination results in the formation of a signal joint by the precise fusion of signal sequences heptamer to heptamer and a more varied coding junction at which base loss and addition have usually occurred.

Mutational analysis of RSSs has shown that the consensus heptamer (5′CACAGTG3′) and nonamer (5′ACAAAAACC 3′) sequences are also the optimal sequences for recombination of a plasmid substrate (9, 18). Indeed, the three bases of the heptamer (5′CAC3′) which are the most highly conserved are also found to have the strongest influence on recombination efficiency. A heptamer is required in each RSS for V(D)J recombination to occur; no recombinants are detected with substrates in which one RSS lacks a heptamer, even if the other RSS has an optimal sequence. In contrast, a substrate containing one normal RSS and an RSS lacking a nonamer can still undergo V(D)J recombination, albeit at a reduced frequency. Alterations of more than one base pair in the spacer length also greatly decrease recombination frequency. Although an RSS is the only *cis*-acting DNA element required for recombination, the first few nucleotides of the adjoining coding segment can influence recombination frequencies (2, 6, 8, 21).

Recent work has shown that the RAG proteins specifically recognize DNA at an RSS and introduce a double-strand break at the border between the heptamer and coding sequence (13). This break is generated in two steps (Fig. 1). First a nick is introduced at the 5′ end of the coding side of the heptamer, leaving a free 3′ hydroxyl on the last nucleotide of coding sequence. In a second step, this 3′ hydroxyl attacks the phosphodiester bond between the coding sequence and the RSS of the opposite strand, leaving a blunt 5′ phosphorylated signal end and a coding end covalently sealed in a hairpin structure. These signal and coding ends are the same as the broken species previously observed in vivo in T and B cells by DNA hybridization and ligation-mediated PCR (19, 20, 24).

The requirement for a pair of signals with 12- and 23-bp spacers seems to be imposed at this initial cleavage stage of the reaction rather than at a later joining stage. Cleavage of plasmid substrates in vivo is coupled; in the presence of both a 12-spacer and a 23-spacer signal, 30-fold-higher coupled cutting than cleavage at a single site is observed (26). This 12/23 dependence can be reproduced in a cell extract (5) and is an intrinsic property of the RAG1 and RAG2 proteins (31). When cleavage is carried out with purified proteins in the presence of Mg^{2+} as the divalent cation, cleavage requires both a 12-spacer and a 23-spacer signal, and the vast majority of cutting occurs at both signals rather than at either one alone (31). In contrast, cleavage in the presence of Mn^{2+} is uncoupled; cuts are introduced independently if two signals are present and occur efficiently even if there is only a single signal.

In this work, we used the cell-free reaction to dissect various aspects of RSS recognition and the role of the RSS in the initial V(D)J cleavage event. Compared with earlier studies in vivo, this system permits separate analysis of the two cleavage steps, characterization of requirements at an individual RSS, and the analysis of nicked, gapped, and heteroduplex sub-

* Corresponding author.

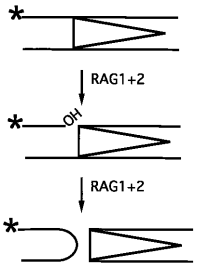

FIG. 1. RAG-mediated cleavage at a V(D)J recombination signal occurs in two steps. The recombination signal sequence is indicated as a triangle, with the vertical face of the triangle marking the site of cleavage. DNA strands of the recombination substrate are indicated as parallel lines. The typical position for 5′ end labeling of substrates is indicated as an asterisk, and this strand is defined as the top strand in the text.

strates. While cleavage in vivo is a coupled reaction, many similarities between cleavage at a single signal and the properties of V(D)J recombination in vivo are evident. Furthermore, an understanding of the interaction between sites and the constraints imposed by coupled cleavage requires a prior understanding of cleavage at an individual site. This work demonstrates the importance of the heptamer in the initiation of V(D)J recombination and reveals a specific and independent role for the nonamer. Further, we show that the coding sequence adjacent to the heptamer can greatly influence the efficiency and precision of cleavage at an isolated RSS and can differentially affect the two steps in the cleavage pathway. Finally, we find that RAG1 and RAG2 can specifically recognize and cleave a single-stranded RSS, and we suggest that the RAG proteins may play a role in unwinding the DNA at the target site for cleavage.

MATERIALS AND METHODS

Oligonucleotide cleavage assay. Truncated mouse RAG1 and mouse RAG2 proteins used in the cleavage assay were expressed and purified as previously described (13, 29). Standard cleavage reaction mixtures contained approximately 100 ng of RAG1 and 50 ng of RAG2 proteins, along with 0.2 pmol of duplex oligonucleotide substrate (an approximate molar ratio of 5 [protein/DNA]). Samples were incubated in 11 μl (total volume) of a reaction mixture with final concentrations of 23 mM *N*-2-hydroxyethylpiperazine-*N*′-ethanesulfonic acid (HEPES)-KOH (pH 7.5), 1 mM Tris-HCl, 1 mM $MnCl_2$, 90 mM KCl, and 2 mM dithiothreitol. Samples were incubated for 4 h at 30°C (except as indicated), and reactions were stopped by the addition of formamide loading buffer containing 95% formamide and 20 mM EDTA. Samples were heated to 95°C and directly analyzed by denaturing gel electrophoresis in a Tris-borate-EDTA-urea gel containing 10% polyacrylamide, 30% formamide, and 1.25 mM HEPES-KOH; gels were visualized by autoradiography and quantified with a PhosphorImager (Molecular Dynamics) and ImageQuaNT software.

Substrate construction. Oligonucleotides were synthesized on a Millipore Expedite synthesizer and purified by polyacrylamide gel electrophoresis under denaturing conditions. Except as noted, the top-strand coding-flank oligonucleotide was 5′ end labeled with ^{32}P under standard conditions (1). This labeled oligonucleotide was annealed to a complementary oligonucleotide in 100 mM KCl to make duplex DNA and separated from unincorporated ^{32}P on a Sephadex spin column equilibrated in 10 mM Tris (pH 8) and 100 mM KCl. In substrates which contain nicks and gaps, one oligonucleotide, usually the top-strand coding flank, was kinase treated and then annealed to the other two oligonucleotides.

The duplex substrates contain a 5′ overhang of one nucleotide at each end boldface in the VDJ100 and VDJ101 sequences presented below. The 12-spacer RSS was made by annealing VDJ100 (5′**G**CTGCAGGTCGACCTGCACAGTG CTACAGACTGGAACAAAAACCCGGTCTC) to VDJ101 (5′**T**GAGACCTG GGTTTTTGTTCCAGTCTGTAGCACTGTGCAGGTCGACCTGCAG).Heptamer and nonamer replacements and coding flank mutations were all made in the context of this 12-bp signal. The nonamer was replaced by 5′CTGGCTCA G3′ and the heptamer was replaced by 5′ACGCCGT3′, while the remainder of the RSS was left intact. The sequences of the top strand of the spacer sequence for the RSSs with altered spacer lengths are as follows: 10-spacer, 5′CTACAG ACTG; 14-spacer, 5′ CTACAGACTGGAAG; 18-spacer, 5′ACATAGCTACA GACTGGA; and 23-spacer, 5′GTAGTACTCCACTGTCTGGCTGT. The sequences used for markers are as follows: standard nick (M, lower band), 5′ GCTGCAGGTCGACCTG; and standard hairpin (M, upper band), 5′GCTGC AGGTCGACCTGCAGGTCGACCTGCAG.

Oligonucleotides corresponding to potential products from the hairpinning of various substrates were synthesized to serve as markers. For example, if a prenicked substrate with the nick two nucleotides into the heptamer were to hairpin directly across to a position two nucleotides into the heptamer on the bottom strand, a hairpin four nucleotides longer than standard, with the specific sequence shown in Fig. 5A, lane M1, +4, would arise.

In Fig. 5A, lane M1 contains two hairpin markers, +4 (5′GCTGCAGGTCG ACCTGCATGCAGGTCGACCTGCAG) and −4 (5′GCTGCAGGTCGACA GGTCGACCTGCAG). Lane M2 contains two hairpin markers, +2 (5′GCTG CAGGTCGACCTGCACAGGTCGACCTGCAG) and −2 (5′ GCTGCAGGT CGACCCAGGTCGACCTGCAG). In Fig. 5C, markers for hairpins from gapped substrates are M1-hairpin (−4) (5′GCTGCAGGTCGACAGGTCGAC CTGCAG) and M2-hairpin (−8) (5′GCTGCAGGTCGATCGACCTGCAG). In Fig. 6B, the sequence of M1, the marker for HP′, is 5′GCTGCAGGTCGA CCTGCACTGTGCAGGTCGACCTGCAG.

RESULTS

It was previously demonstrated that purified RAG1 and RAG2 proteins cleave recombination signal sequences in two steps and in an RSS-dependent fashion. Here we use this system to study variations in the RSS. For all experiments, we use an N-terminally truncated RAG1 protein (retaining amino acids 384 to 1008) purified from a baculovirus expression system and a C-terminally truncated RAG2 protein (amino acids 1 to 383 out of 527) purified from mammalian cells infected with a vaccinia virus expressing RAG2. These two truncated proteins together are fully functional in mediating V(D)J recombination of both plasmid and integrated substrates in vivo (10). The various oligonucleotide substrates used were radioactively labeled at the 5′ end of the top strand of coding sequence (Fig. 1), except as noted, and the reaction products were resolved on denaturing polyacrylamide gels that allowed us to distinguish the nicked and hairpin species by size. Under standard conditions, efficient cleavage (up to 40%) of an oligonucleotide substrate containing a single signal sequence was achieved.

Independent functions of heptamer and nonamer. We first considered what effect the nonamer sequence of the RSS has on nicking and hairpin formation. Starting with an oligonucleotide containing an optimal RSS with a 12-bp spacer, we replaced the canonical nonamer sequence with a sequence that replaces each nucleotide of the nonamer with one that has been shown to be least favorable for V(D)J recombination in vivo (9, 18). In comparison with the reaction products formed with a standard 12-spacer RSS (Fig. 2A, lane 1), the nonamer-less RSS (Fig. 2A, lane 5) showed a 5- to 10-fold reduction in both nicking and hairpin formation. However, the sizes of both the nicked and hairpinned species were the same as with a full RSS, indicating that the heptamer was sufficient to specifically direct nick and hairpin formation to the heptamer/coding border.

In the converse experiment, we replaced the heptamer sequence with nucleotides at each position that are unfavorable to the complete recombination reaction. Surprisingly, this substrate can still direct specific nicking, but the efficiency is much reduced and the specificity is altered. Rather than introducing a nick precisely where the heptamer border would have been (i.e., 19 nucleotides from the near side of the nonamer), a conspicuous band is present 1 nucleotide closer to the nonamer sequence (Fig. 2A, lane 3; Fig. 2B, lane 4; Fig. 2C, lane 1). Furthermore, the presence of additional nicked species indicates that nicking is now less precisely targeted (Fig. 2B, lane 4; Fig. 2C, lanes 1 and 3). A comparison with reactions containing only RAG1 or RAG2 or with a substrate lacking both a heptamer and a nonamer (Fig. 2B, lanes 2, 3, and 7) indicates that this low level of nicking depends on the presence

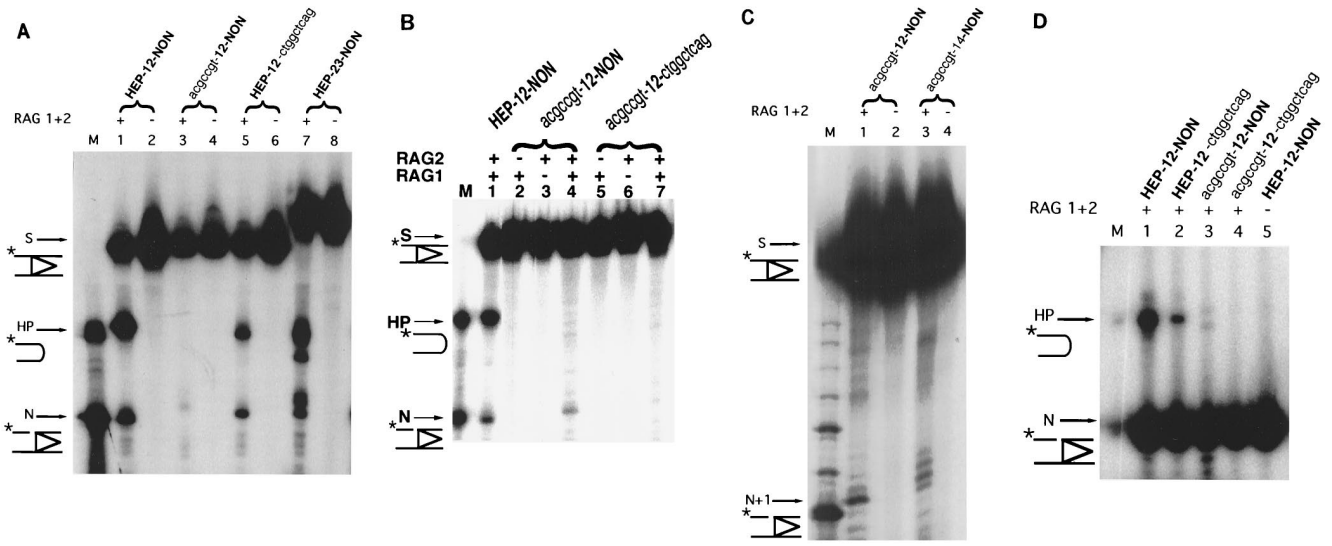

FIG. 2. The heptamer and nonamer motifs of the RSS can function independently. (A) Cleavage is shown for a consensus 12-spacer RSS (HEP-12-NON) along with derivatives in which the heptamer (lanes 3 and 4) or nonamer (lanes 5 and 6) are replaced by the sequences shown in lower case. A 23-spacer signal with consensus heptamer and nonamer sequences is also shown (lanes 7 and 8). (B) Nicking at the nonamer requires both RAG1 and RAG2 and the nonamer sequence. Plus and minus signs indicate whether RAG1 or RAG2 or both were added to a given reaction. Lanes 5 to 7 are reactions in which both the heptamer and nonamer sequences are replaced by the sequences shown in lowercase. (C) Cleavage at a preferred distance from the nonamer. Substrates in which the heptamer is replaced and the spacer length is altered from 12 bp (lanes 1 and 2) to 14 bp (lanes 3 and 4) are shown. (D) The nonamer can direct hairpinning of a prenicked substrate. The substrates shown contain a preformed nick 5′ of the heptamer border. Lanes 2 to 4 show substrates in which the heptamer, the nonamer, and both motifs, respectively, were replaced. Reactions in panels A, B, and D are as described in Materials and Methods; those in panel C were performed for 60 min at 37°C. S, full-length substrate; N, nick at coding/heptamer border; HP, hairpin. The marker (lane M) contains two oligonucleotides with sequences corresponding to a nicked or hairpinned species. For panel A, C, and D, the presence (+) or absence (−) of the purified RAG1 and RAG2 proteins is indicated.

of the nonamer sequence and on both of the RAG proteins. These results suggest that the site of the nick directed by the nonamer alone is determined by distance from the nonamer, not a specific sequence (though local sequence may have some influence). Indeed, when two additional base pairs were introduced in the spacer region of the substrate, nicking still occurred predominantly at the same distance from the nonamer, yielding correspondingly longer products (Fig. 2C).

While little if any hairpin product was seen with a substrate containing the nonamer alone, it was possible to increase the sensitivity of this test by using a prenicked oligonucleotide (with the nick at the heptamer border) as the substrate. When this substrate was analyzed, two species migrating at positions consistent with their being hairpins were apparent, again at a much decreased intensity compared with a consensus RSS (Fig. 2D, lane 3). Two-dimensional gel electrophoresis revealed an off-diagonal spot (data not shown) confirming that at least one of these species was indeed a hairpin. No such off-diagonal spot was revealed when the prenicked substrate lacked both a heptamer and a nonamer or when only one of the two RAG proteins was present (data not shown). Thus, the nonamer and heptamer, completely distinct motifs, can independently direct both the nicking and the hairpin steps. It is particularly surprising that the nonamer has this ability because no V(D)J joining has been detected in vivo with substrates containing only a nonamer sequence (9). However, the extremely low hairpinning activity detected here would put recombination in vivo below the level of detection.

Spacer length affects heptamer-nonamer coupling. The heptamer and nonamer motifs of an RSS are separated by a spacer region of nonconserved sequence with a standard length of 12 or 23 bp. To examine what effect altered spacer length would have on each step of double-strand break formation, reaction products from oligonucleotides with standard spacer lengths of 12 and 23 bp were compared with oligonucleotides containing spacers of 10, 14, and 18 bp or with an oligonucleotide containing just a heptamer alone (Fig. 2A and 3). As expected, the 12-spacer RSS was the most efficient substrate, converting 30 to 40% of substrate into the sum of the two products. Moreover, essentially only one nicked and one hairpinned species were present. In general, the 23-spacer RSS was about three- to fivefold less effective in directing nick and hairpin formation, and additional products, presumably resulting from nicks in-

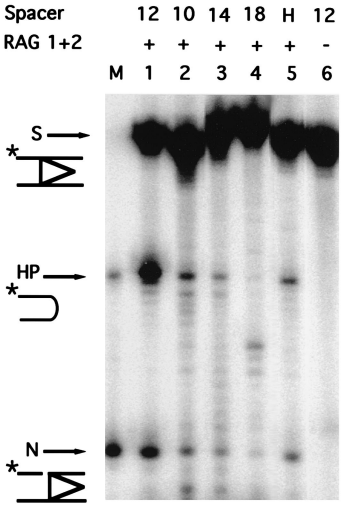

FIG. 3. Effect of spacer length on cleavage. Spacer lengths are as indicated. H indicates a substrate in which the nonamer has been replaced as for Fig. 2. Other symbols and markers are as for Fig. 2.

troduced at positions other than the heptamer border, were observed (Fig. 2A and data not shown). With either the 23-spacer or the 12-spacer RSS, the overall amount of correct nick and hairpin was higher than in a substrate with a heptamer alone, confirming that the presence of the nonamer leads to an increased amount of cleavage. A substrate with a 10-bp spacer, which can serve as a functional RSS in vivo, was roughly equivalent to the 23-spacer RSS; it showed significantly less nick and hairpin product than the 12-spacer RSS and an increase in imprecise nicking. Interestingly, a substrate with an 18-bp spacer supported even less nick and hairpin formation than the heptamer alone, with the hairpin more severely affected (Fig. 3). These observations suggest that an incorrectly positioned nonamer interferes with activity at the heptamer. Furthermore, variation of spacer length results in the reduction of both nick and hairpin formation and alters the sites of nicking.

Multiple effects of coding-flank sequence. While the RSSs are the primary determinants of the efficiency of V(D)J recombination, the sequence of the coding flank has been shown to influence recombination in vivo. Recombination mediated by wild-type RAG1 and RAG2 proteins exhibits sequence preferences for the first few nucleotides immediately adjacent to the coding site. In particular, polythymidylate tracts next to both RSSs have been shown to reduce recombination frequencies by up to 1,000-fold (2, 6, 8). With some mutations in RAG1, other effects of coding-flank sequence on the recombination reaction are also observed. A RAG1 mutant protein (called D32) shows heightened sensitivity to certain dinucleotide sequences adjacent to the heptamer border (21). "Good" coding flanks are readily recombined by the mutant protein, whereas substrates with "bad" coding flanks at both RSSs are recombined at frequencies reduced as much as 1,000-fold.

To learn more about the effect of coding flanks on the initiation of V(D)J recombination, we compared oligonucleotide substrates containing a good flank (top-strand 5′TG) with identical substrates with a bad flank substituted (top-strand 5′AC). Strikingly, flanks defined as good or bad on the basis of their interaction in vivo with D32 also showed differences in this cleavage assay even though wild-type RAG1 protein was used (Fig. 4A, lanes 2 and 3). This effect was only on hairpin formation; nicking was largely unaffected by the presence of a bad flank. The fraction of hairpin product decreased from 31% for a good flank to 2 and 0.7% for two bad flank substitutions. Coding-flank sequences which showed an intermediate phenotype with D32 in vivo showed similar intermediate phenotypes in the oligonucleotide cleavage assay with wild-type protein (data not shown). Thus, cleavage with wild-type RAG proteins shows a sensitivity to the sequence of the coding flank that appears to be the same as that revealed by the D32 mutation in vivo.

To determine if the effect of bad flanks was a result of DNA sequence or structure, we examined the effect of mismatched bases at the signal/coding border. Hairpin formation was fully restored in substrates that were heteroduplex at the coding heptamer border with a bad flank sequence on either the top or bottom strand, 5′AC3′/3′AC5′ or 5′TG3′/3′TG5′. Results with a bad flank sequence in the bottom strand are shown (Fig. 4B). Thus, a bad flank sequence on the bottom strand, when it is unable to pair with the top flank sequence, does not itself prevent the attack from the free 3′ hydroxyl; rather, the structure of the duplex bad flank DNA is responsible for the decreased hairpinning.

We also studied the effect of polythymidylate tracts on RAG-mediated cleavage. In contrast to the bad flank sequences discussed above, the di- and hexanucleotide T flanks reduced both nicking and hairpinning by three- to sixfold, but

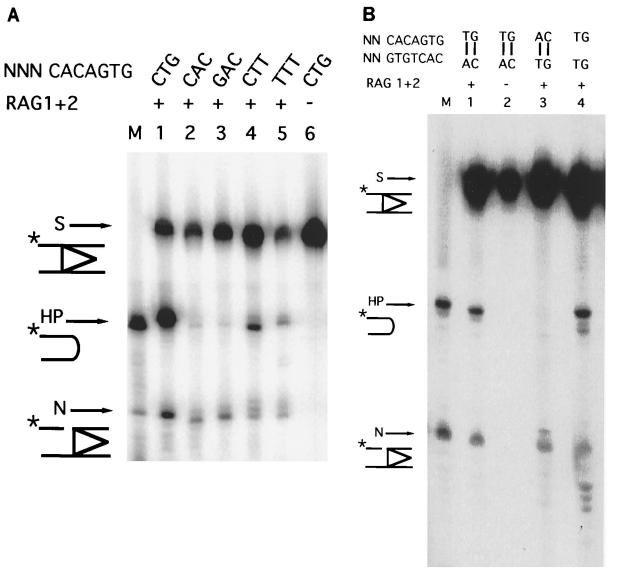

FIG. 4. Coding-end sequence differentially affects nicking and hairpin formation. (A) The effect of varying the sequence of three positions of the coding flank adjacent to the heptamer is shown. CTG is the sequence used in all previous experiments; this is a good flank sequence. The poly(T) tract of the coding flank in lane 5 extends for three more positions. (B) Base unpairing of bad flanks restores hairpin formation. The two positions of the coding flank that were varied are shown. Vertical lines indicate base pairing. The substrate in lane 4 contains the unpaired bad flank (with the bad flank on the bottom strand). The reactions in panel B were performed for 45 min at 37°C. Other symbols and markers are as for Fig. 2.

the ratio of nick to hairpin remained largely unchanged (Fig. 4A, lanes 4 and 5). In this case, we observed nicks at the normal coding/heptamer border as well as at the first two positions inside the heptamer. Thus, a poly(T) sequence at the coding border affects both the overall efficiency of cleavage and the specificity of nicking. These defects may be partially responsible for the large decreases in recombination frequencies that have been reported for similar substrates in transfection assays.

Formation of alternative hairpin structures. We wished to determine how the position of the nick in the top strand would affect product formation. We therefore generated prenicked substrates with the nick in the coding or heptamer sequence. Whereas a nick at the heptamer border led to a hairpin of the normal size, a nick two base pairs inside the coding flank yielded a hairpin two nucleotides smaller (Fig. 5A, lanes 1 and 2). Thus, the normal phosphodiester bond at the coding-heptamer border on the bottom strand was exclusively chosen as the target for the intramolecular strand transfer even though the attacking 3′ hydroxyl was no longer positioned opposite this site (diagram in Fig. 5B). When a nick was preformed within the heptamer sequence, a secondary nick was introduced at the correct position (at the 5′ border of the heptamer), after which hairpin conversion proceeded normally, resulting in a hairpin of standard size (Fig. 5A, lane 3). The site of cleavage on the bottom strand was confirmed by 5′ end labeling this strand (data not shown). Thus, even when the initial nick is not in the correct position relative to the RSS, it can still be converted to a hairpin.

Because the experiments with mismatched base pairs indicated that cleavage did not require fully duplex DNA at the coding/heptamer border, we asked what products, if any, would be formed when a gap of two or four nucleotides was introduced in the top strand of the coding sequence, such that a

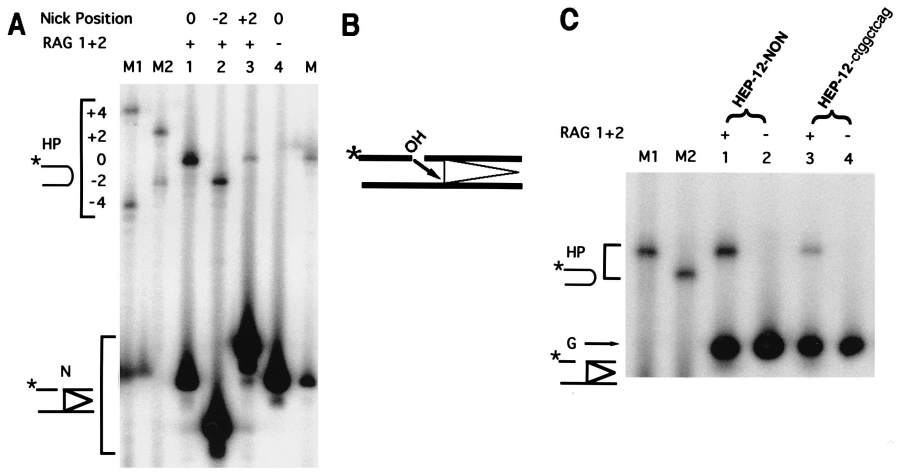

FIG. 5. Hairpinning of substrates with nicks and gaps. (A) Position of the nick does not affect the site of hairpin formation on the bottom strand. Substrates with nicks at either the coding/heptamer border (0) or a position two nucleotides into the coding flanks (−2) or inside the signal heptamer (+2) are shown. Size markers: lane M contains a standard hairpin, lane M1 contains markers for hairpins four longer or four shorter than standard, and lane M2 contains markers for hairpins two longer or shorter. Sequences are as in Materials and Methods. (B) Schematic of the position of strand transfer in a substrate with a nick at −2. (C) Hairpinning of a gapped substrate with a four-nucleotide gap in the top strand of the coding flank. Hairpinning proceeds with a wild-type RSS (lanes 1 and 2) and one lacking a nonamer (lanes 3 and 4). Marker for hairpins formed from gapped substrates (and distinct from those in panel A) are shown in lanes M1 and M2 and described in Materials and Methods. G, gapped substrate. Other symbols and markers are as for Fig. 2.

single bottom strand leftward (as drawn in Fig. 1) from the heptamer border was exposed. Hairpins were still efficiently formed and, by comparison with markers constructed of identical sequence, were exclusively directed to the 3′ end of the bottom-strand heptamer. These results, shown for a four-nucleotide gap in Fig. 5C, are consistent with our observations on prenicked substrates and indicate that great conformational flexibility must be available within the cleavage complex.

Utilization of a single-stranded RSS. Gaps of two and four nucleotides were also introduced rightward into the heptamer and served as efficient substrates for hairpinning (data not shown), suggesting the surprising possibility that a single-stranded heptamer might be recognized by the RAG proteins. Indeed, we observed RAG-mediated nicking and hairpinning in substrates with an entirely single-stranded RSS attached to a duplex coding sequence (Fig. 6). When the single-stranded RSS was on the top strand, nicks were introduced at the coding/heptamer border with an efficiency nearly as high as seen with a duplex RSS (Fig. 6A). Nicking in the presence of a single-stranded RSS was dependent on both RAG1 and RAG2 and required a heptamer sequence (Fig. 6A and data not shown). As expected, no hairpin formation was detected because the first nick would cut the RSS away from the coding flank.

On the other hand, when only the bottom strand of the RSS was retained (Fig. 6B and C), two hairpin products were

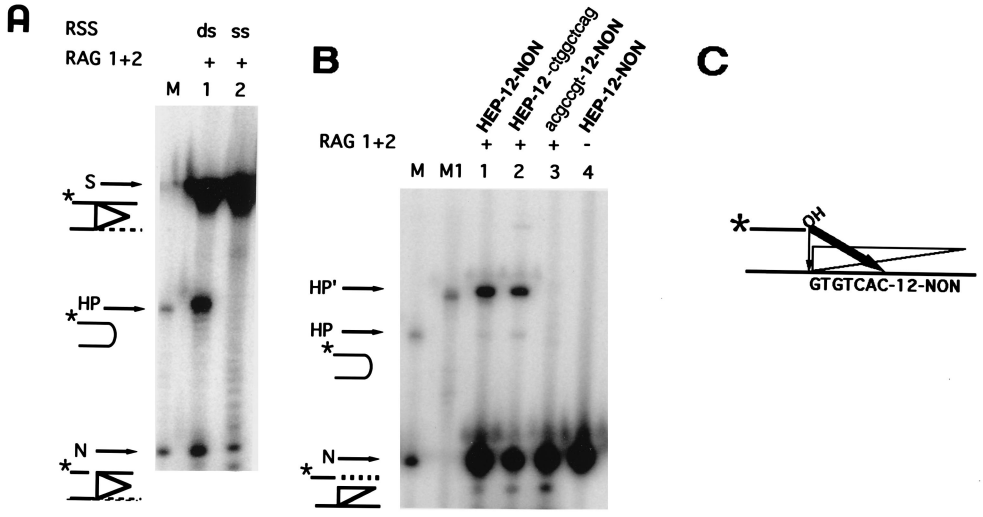

FIG. 6. A single-stranded RSS can direct specific cleavage. (A) A substrate containing a fully duplex RSS (ds) is compared to a substrate having duplex coding sequence but retaining only the top strand of the RSS (ss). The strand missing in the single-stranded substrates is shown as a dashed line in the diagram of S. (B) All substrates are duplex in the coding flank but contain only the bottom strand of the signal sequence (see panel C). The coding flank was labeled at the 5′ end of the top strand. The sequence of the signal is as indicated. HP′, hairpin specific to the single-stranded RSS. Lane M1 contains a marker specific for this HP′ sequence. (C) Diagram of the products from panel B. The sequence of the heptamer is written 3′ to 5′. Other symbols and markers are as for Fig. 2.

formed. The nature of the two products was determined by comparison with synthesized hairpin size markers. One product, HP′, was formed with efficiency equivalent to that seen for hairpin conversion with duplex DNA. It appears to result from joining the 3′ hydroxyl of the top coding strand to the 5′ end of the bottom-strand heptamer (diagrammed in Fig. 6C). The more conventional product expected for an attack at the 3′ end of the heptamer border was also observed (species HP), but at only 1/20 the amount of HP′. Formation of both hairpins was still dependent on both RAG1 and RAG2. Furthermore, substitution of the heptamer sequence blocked the formation of both hairpin species, but the absence of a functional nonamer did not affect the formation of either product. In these substrates, the heptamer can evidently be utilized independently. Labeling the 5′ end of the bottom strand revealed a product of 30 nucleotides, consistent with strand breakage at the 5′ end of the heptamer (diagrammed in Fig. 6C; data not shown). However, threefold more of this species than of the HP′ hairpin was present, suggesting that the remainder of the 30-nucleotide species resulted from direct nicking at the 5′ end of the heptamer. Therefore, an adjacent single-strand-to-double-strand transition is not required for the RAG proteins to introduce a nick. Thus, a single-stranded RSS with only the top or bottom strand retained can serve as an efficient recognition site for RAG-mediated strand scission or hairpin formation.

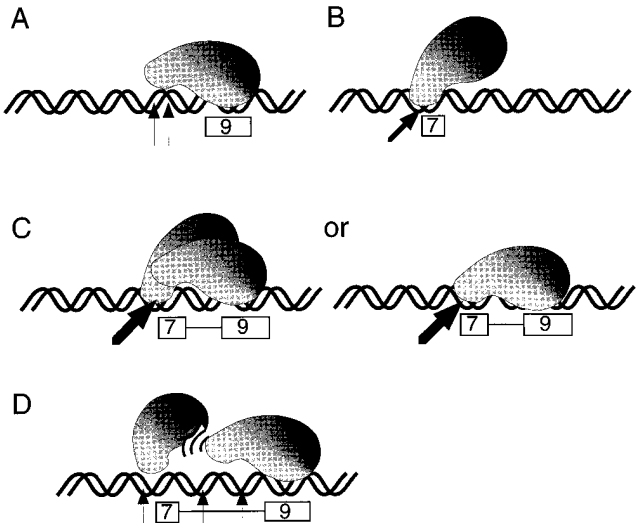

FIG. 7. Model for interaction of RAG proteins with variant RSS sequences. Shaded objects represent a complex of RAG1 and RAG2. Positions of heptamer and nonamer sites within the DNA are indicated as 7 and 9, respectively. Positions of arrows indicate cleavage sites, and arrow width indicates relative cleavage efficiency. (A) Interaction with a lone nonamer permits a low level of cleavage at multiple sites but with a preferred distance from the nonamer. (B) Cleavage at an isolated heptamer is specific and enhanced compared with that of a nonamer alone. (C) Cleavage of a consensus 12-spacer RSS. Two possibilities are presented for the interactions of the RAG proteins with the heptamer and nonamer. On the left, proper spacing of the motifs allows two complexes to interact synergistically through the heptamer and nonamer sites. Alternatively, on the right, a single RAG complex binds to both sites in such a way that specific cleavage is greatly enhanced. (D) When the heptamer and nonamer motifs are improperly spaced, the bound proteins clash, leading to a low level of nicking (and possibly hairpinning).

DISCUSSION

The DNA sequence requirements for V(D)J recombination in extrachromosomal substrates have been extensively explored, and the optimal RSS and coding-flank preferences are thus well defined (reviewed in reference 12). The ability to carry out RSS-dependent cleavage in vitro now permits a greater understanding of the function of these sequence elements and their role in the mechanism of cleavage and recombination. The RSS must serve as a binding site for the RAG proteins and as a target site for the nick and strand transfer reactions that result in double-strand breakage at the heptamer border. Here we find that although the maximal specificity and efficiency of nick and hairpin formation occurs with the canonical RSS, surprisingly large variations are tolerated. Furthermore, these variations can have differential effects on the two cleavage steps.

An isolated heptamer directs accurate cutting, both nicking and hairpinning, but with reduced efficiency compared with a canonical RSS. Unexpectedly, the nonamer also can serve to mediate both nicking and hairpinning, but much less efficiently than the heptamer alone. Nonamer-mediated cutting is also less accurate than cleavage in the presence of a heptamer, with several nick sites including a prominent band one base closer to the nonamer. Both nicking and hairpinning at either an isolated heptamer or nonamer are dependent on the action of both RAG1 and RAG2. These results suggest that one or both RAG proteins can bind to DNA independently through a heptamer or a nonamer sequence and that cutting can proceed in either case (Fig. 7A and B).

V(D)J recombination in vivo is a coupled reaction requiring one 12-spacer signal and one 23-spacer signal (26). This requirement suggests that the nucleoprotein complexes formed on the two signals are distinguishable. The only distinction we note in this regard is that a 23-spacer signal is a somewhat less efficient substrate than a 12-spacer signal and displays reduced specificity. It will be interesting to see if greater discrimination between signals of different spacer length will occur in the coupled reaction such that a spacer length mutation that still serves as a substrate for single cleavage will not be cleaved under coupled reaction conditions.

The experiments described here, performed in the absence of constraints imposed by coupled cleavage, reveal the importance of proper spacing even for activating efficient cleavage at one RSS by the RAG proteins. Spacer lengths of 12 and 23 bp are preferred over others. Altering the spacer length reduces the overall efficiency and specificity of nicking and can reduce hairpin formation to a level lower than seen with an isolated heptamer. It thus appears that with improper spacing, binding of the RAG proteins at one site (heptamer or nonamer) interferes with the interaction at the second site (Fig. 7C and D). Because neither part of the signal is as effective as a complete RSS, interaction of the RAG proteins with both the heptamer and the nonamer separated by the correct spacing activates nicking and intramolecular strand transfer. Because the naturally occurring spacer lengths of 12 and 23 bp are offset from each other by about one helical turn, the spacer length requirement is likely to reflect a physical constraint on the positioning of RAG proteins on the helical surface of the DNA such that proper interaction between proteins bound at the heptamer and the nonamer is facilitated.

In recombination in vivo, the first several nucleotides of coding sequence immediately adjacent to the heptamer can have large effects on recombination efficiency. One sequence which reduces recombination is a T homopolymer adjacent to the CAC of the heptamer. Recombination frequencies are reduced 10- to 1,000-fold in vivo, depending on the number of adjacent T nucleotides and whether one or both RSSs are flanked by them (2, 6, 8). An examination of cleavage by purified RAG proteins of an RSS with a flanking TT or TTT

TTT sequence shows that both nicking and hairpinning are reduced comparably, by roughly a factor of 5. Moreover, the specificity of nicking is disturbed; nicks one and two nucleotides into the heptamer sequence accumulate to nearly the same levels as those at the heptamer border. Thus, in contrast to the bad flanks discussed below, poly(T) flanks affect both the specificity and efficiency of nick and hairpin formation. However, the relatively small effects on overall hairpin formation do not appear to fully explain the in vivo sensitivity to poly(T) flanks. Perhaps a larger effect will be seen in coupled cleavage or at a later stage in recombination.

Another effect of coding-flank sequences is revealed when the cleavage of bad flanks (those that decrease recombination with the RAG1 D32 mutant) is considered. Remarkably, we find here that hairpin formation with the purified wild-type RAG proteins is sensitive to the same coding-flank sequences that diminish recombination with the D32 mutant in vivo (21). With purified RAG proteins, nicking is not affected by bad flanks but hairpin formation is dramatically reduced. It is possible, but unlikely, that a bad flank sequence decreases the affinity of the RAG proteins for the target site and that this decreased affinity affects only the ability of RAG to catalyze the second step of the cleavage reaction. A more plausible explanation for the effect of the bad flank is that the sequence itself has a limited ability to exist in an altered structural state that may be required for the intramolecular nucleophilic attack and strand transfer. Because the hairpins are formed by one-step transesterification, the 3′ hydroxyl group liberated by the initial nick must directly attack the phosphodiester bond on the opposite DNA strand (30). Thus, considerable distortion of the DNA helices is required. In this regard, our observation that heteroduplex DNA with a bad flank in either the top or bottom strand serves as an efficient substrate indicates that the dinucleotide of the bad flank does not create an uncleavable site. Rather, it seems likely that the RAG proteins cannot distort the duplex structure formed by a bad flank sufficiently for transesterification but that this difficulty can be overcome by unpairing the flanking DNA.

Recent work shows that the effect of a bad flank on cleavage at an individual RSS is not seen when cleavage is carried out under coupled cleavage conditions (31). Thus, the formation of a synaptic complex may allow the RAG proteins to alter the structure of homoduplex bad flank sequence sufficiently to catalyze strand exchange. It is possible, then, that the mutation in D32 affects not the binding of RAG1 to the RSS and adjacent coding sequence but rather the ability of RAG1 to properly form a synaptic complex or to unwind the DNA at the cleavage site.

As has been noted previously (7), the RSS sequence itself has unusual structural features. The heptamer, in addition to its dyad symmetry, largely consists of alternating purine and pyrimidine residues (CACA). Good flanks, such as 5′TG, tend to allow this alternation to extend beyond the heptamer, while bad flanks (e.g., 5′AC) in general disrupt it. This finding again suggests that particular DNA structural features are important for RAG function. In addition, structural studies have shown that the CACA sequence has an intrinsically abnormal structure in which normal base pairing is disrupted (4, 17, 28). This abnormal structure may facilitate further the DNA distortion (see below) that is likely to be critical for the reaction; the formation of a perfect hairpin requires that the DNA at the tip of the hairpin be unpaired, and the intramolecular transesterification would require a kink on the bottom strand at the signal-coding border.

The unusual ability of the RAG proteins to recognize and cleave either double-stranded or single-stranded DNA in a sequence-specific manner suggests that a possible role for the RAG proteins in the cleavage reaction is to alter the structure of the RSS and adjacent coding flank, possibly unwinding the DNA to make it accessible for cleavage. If unwound DNA is indeed the normal substrate for cleavage, this could explain the ability of single-stranded DNA to efficiently serve as a target for the RAG proteins. This idea is consistent with the observation that unpaired DNA at the heptamer border still allows the reaction to occur and more importantly can restore the ability of bad flank sequences to be hairpinned. Local strand separation might allow recognition and nicking of the top strand, followed by a second independent recognition step and hairpinning to the bottom strand.

A role for the RAG proteins in distorting DNA to facilitate the strand transfer reaction is similar to the role proposed for the MuA protein in transposition (14). Additional observations serve to underscore the similarities between the cleavage reaction catalyzed by the RAG proteins and the family of reactions involved in transposition. In all of these reactions, strand exchange proceeds via direct transesterification (reference 30 and references therein). As with RAG cleavage, unpairing at the site of cleavage has also been shown to facilitate cutting by MuA (22). Like MuA and human immunodeficiency virus integrase, which require a terminal CA at the site of cleavage (3, 11, 27), the RAG-mediated reaction requires the terminal CA of the heptamer, but whether these requirements reflect similar roles for the CA dinucleotide is unclear.

The ability of the RAG proteins to attack sequentially both the top and bottom strands of DNA suggests that the bound RAG proteins make contacts on both strands. However, because a single-stranded heptamer in either the top or bottom strand can also be recognized and cleaved, the recognition of the two strands may not normally be simultaneous. Further, the effects of flanking sequence on cleavage suggest that the active RAG complex must make direct contact with at least one or two nucleotides outside the RSS. Analysis of substrates with nicks and gaps in the coding sequence reveals that the attacking 3′ OH from the top strand always mediates strand transfer to the phosphodiester bond at the heptamer border on the bottom strand. This observation suggests that the RAG proteins, when bound to the RSS, make only this bond accessible to nucleophilic attack. In addition, the ability of the cleavage reaction to utilize substrates in which the initial nick is not at the heptamer border suggests that some of the nucleotide insertions and deletions seen in vivo may result from the cleavage reaction itself, not later modifications.

If unwinding is important for efficient target recognition and cleavage, this would provide one means by which the cell could regulate target choice. By affecting local DNA structure, the ability of RAG to access or unpair an RSS could be controlled. Perhaps some of the RSS-binding proteins that have been identified could play a role in such regulation, along with DNA helicases and other factors which could affect the local chromatin structure.

ACKNOWLEDGMENTS

We are grateful to Chip Romeo for his expert advice and assistance throughout this project. We thank Tania Baker and members of the Gellert laboratory for valuable discussions and members of the Oettinger laboratory for advice and critical reading of the manuscript.

This work was supported by a Ryan Fellowship (C.A.C.), National Institutes of Health grant GM48026, and the Pew Scholars Program.

ADDENDUM IN PROOF

Similar results have recently been reported by D. A. Ramsden, J. F. McBlane, D. C. van Gent, and M. Gellert (EMBO J. **15:**3197–3206, 1996).

REFERENCES

1. **Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (ed.).** 1989. Current protocols in molecular biology. Greene Publishing Associates and Wiley-Interscience, New York.
2. **Boubnov, N. V., Z. P. Wills, and D. T. Weaver.** 1993. Coding sequence composition flanking either signal element alters V(D)J recombination efficiency. Nucleic Acids Res. **23:**1060–1067.
3. **Burlingame, R. P., M. G. Obukowicz, D. L. Lynn, and M. M. Howe.** 1986. Isolation of point mutations in bacteriophage Mu attachment regions cloned in a lambda::mini-Mu phage. Proc. Natl. Acad. Sci. USA **83:**6012–6016.
4. **Cheung, S., K. Arndt, and P. Lu.** 1984. Correlation of lac operator DNA imino proton exchange kinetics with its function. Proc. Natl. Acad. Sci. USA **81:**3665–3669.
5. **Eastman, Q. M., T. M. Leu, and D. G. Schatz.** 1996. Initiation of V(D)J recombination in vitro obeying the 12/23 rule. Nature (London) **380:**85–88.
6. **Ezekiel, U. R., P. Engler, D. Stern, and U. Storb.** 1995. Asymmetric processing of coding ends and the effect of coding end nucleotide composition on V(D)J recombination. Immunity **2:**381–389.
7. **Gellert, M.** 1992. Molecular analysis of V(D)J recombination. Annu. Rev. Genet. **22:**425–446.
8. **Gerstein, R. M., and M. R. Lieber.** 1993. Coding end sequence can markedly affect the initiation of V(D)J recombination. Genes Dev. **7:**1459–1469.
9. **Hesse, J. E., M. R. Lieber, K. Mizuuchi, and M. Gellert.** 1989. V(D)J recombination: a functional definition of the joining signals. Genes Dev. **3:**1053–1061.
10. **Kirch, S. A., P. Sudarsanam, and M. A. Oettinger.** 1996. Regions of RAG1 protein critical for V(D)J recombination. Eur. J. Immunol. **26:**886–891.
11. **La Femina, R. L., P. L. Callahan, and M. G. Cordingley.** 1991. Substrate specificity of recombinant human immunodeficiency virus integrase protein. J. Virol. **65:**5624–5630.
12. **Lewis, S. M.** 1994. The mechanism of V(D)J joining: lessons from molecular, immunological, and comparative analyses. Adv. Immunol. **56:**27–150.
13. **McBlane, J. F., D. C. van Gent, D. A. Ramsden, C. Romeo, C. A. Cuomo, M. Gellert, and M. A. Oettinger.** 1995. Cleavage at a V(D)J recombination signal requires only RAG1 and RAG2 proteins and occurs in two steps. Cell **83:**387–395.
14. **Mizuuchi, K.** 1992. Transpositional recombination: mechanistic insights from studies of Mu and other elements. Annu. Rev. Biochem. **61:**1011–1051.
15. **Mombaerts, P., J. Iacomini, R. S. Johnson, K. Herrup, S. Tonegawa, and V. E. Papaioannou.** 1992. RAG-1-deficient mice have no mature B and T lymphocytes. Cell **68:**869–877.
16. **Oettinger, M. A., D. G. Schatz, C. Gorka, and D. Baltimore.** 1990. RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. Science **248:**1517–1523.
17. **Patel, D. J., L. Shapiro, and D. Hare.** 1987. NMR-distance geometry studies of helical errors and sequence dependent conformations of DNA in solution, p. 115–161. *In* R. D. Wells and S. C. Harvey (ed.), Unusual DNA structures. Springer, New York.
18. **Ramsden, D. A., K. Baetz, and G. E. Wu.** 1994. Conservation of sequence in recombination signal sequence spacers. Nucleic Acids Res. **22:**1785–1796.
19. **Roth, D. B., J. P. Menetski, P. B. Nakajima, M. J. Bosma, and M. Gellert.** 1992. V(D)J recombination: broken DNA molecules with covalently sealed (hairpin) coding ends in scid mouse thymocytes. Cell **70:**983–991.
20. **Roth, D. B., P. B. Nakajima, J. P. Menetski, M. J. Bosma, and M. Gellert.** 1992. V(D)J recombination in mouse thymocytes: double-strand breaks near T cell receptor δ rearrangement signals. Cell **69:**41–53.
21. **Sadofsky, M., J. E. Hesse, D. C. van Gent, and M. Gellert.** 1995. RAG-1 mutations that affect the target specificity of V(D)J recombination: a possible direct role of RAG-1 in site recognition. Genes Dev. **9:**2193–2199.
22. **Savilahti, H., P. A. Rice, and K. Mizuuchi.** 1995. The phage Mu transpososome core: DNA requirements for assembly and function. EMBO J. **14:** 4893–4903.
23. **Schatz, D. G., M. A. Oettinger, and D. Baltimore.** 1989. The V(D)J recombination activating gene (RAG-1). Cell **59:**1035–1048.
24. **Schlissel, M., A. Constantinescu, T. Morrow, M. Baxter, and A. Peng.** 1993. Double-strand signal sequence breaks in V(D)J recombination are blunt, 5′-phosphorylated, RAG-dependent, and cell cycle regulated. Genes Dev. **7:**2520–2532.
25. **Shinkai, Y., G. Rathbun, K.-P. Lam, E. M. Oltz, V. Stewart, M. Mendelsohn, J. Charron, M. Datta, F. Young, A. M. Stall, and F. W. Alt.** 1992. RAG-2-deficient mice lack mature lymphocytes owing to inability to initiate V(D)J rearrangement. Cell **68:**855–867.
26. **Steen, S. B., L. Gomelsky, and D. B. Roth.** The 12/23 rule is enforced at the cleavage step of V(D)J recombination in vivo. Genes Cells, in press.
27. **Surette, M. G., T. Harkness, and G. Chaconas.** 1991. Stimulation of the Mu A protein-mediated strand cleavage reaction by the Mu B protein, and the requirement of DNA nicking for stable type 1 transpososome formation. J. Biol. Chem. **266:**3118–3124.
28. **Timsit, Y., E. Vilbois, and D. Moras.** 1991. Base-pairing shift in the major groove of $(CA)_n$ tracts by B-DNA crystal structures. Nature (London) **354:** 167–170.
29. **van Gent, D. C., J. F. McBlane, D. A. Ramsden, M. J. Sadofsky, J. E. Hesse, and M. Gellert.** 1995. Initiation of V(D)J recombination in a cell-free system. Cell **81:**925–934.
30. **van Gent, D. C., K. Mizuuchi, and M. Gellert.** 1996. Initiation of V(D)J recombination: similarities to transposition and retroviral integration. Science **271:**1592–1594.
31. **van Gent, D. C., D. A. Ramsden, and M. Gellert.** 1996. The RAG1 and RAG2 proteins establish the 12/23 rule in V(D)J recombination. Cell **85:** 107–113.