Nucleic Acids Research, 1997, Vol. 25, No. 7 **1375–1382**

A complex of RAG-1 and RAG-2 proteins persists on DNA after single-strand cleavage at V(D)J recombination signal sequences

Ulf Grawunder and Michael R. Lieber*

Department of Pathology and Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, 660 South Euclid Avenue, St Louis, MO 63110, USA

Received December 9, 1996; Revised and Accepted February 10, 1997

ABSTRACT

The recombination activating gene (RAG) 1 and 2 proteins are required for initiation of V(D)J recombination *in vivo* and have been shown to be sufficient to introduce DNA double-strand breaks at recombination signal sequences (RSSs) in a cell-free assay *in vitro*. RSSs consist of a highly conserved palindromic heptamer that is separated from a slightly less conserved A/T-rich nonamer by either a 12 or 23 bp spacer of random sequence. Despite the high sequence specificity of RAG-mediated cleavage at RSSs, direct binding of the RAG proteins to these sequences has been difficult to demonstrate by standard methods. Even when this can be demonstrated, questions about the order of events for an individual RAG–RSS complex will require methods that monitor aspects of the complex during transitions from one step of the reaction to the next. Here we have used template-independent DNA polymerase terminal deoxynucleotidyl transferase (TdT) in order to assess occupancy of the reaction intermediates by the RAG complex during the reaction. In addition, this approach allows analysis of the accessibility of end products of a RAG-catalyzed cleavage reaction for N nucleotide addition. The results indicate that RAG proteins form a long-lived complex with the RSS once the initial nick is generated, because the 3′-OH group at the nick remains obstructed for TdT-catalyzed N nucleotide addition. In contrast, the 3′-OH group generated at the signal end after completion of the cleavage reaction can be efficiently tailed by TdT, suggesting that the RAG proteins disassemble from the signal end after DNA double-strand cleavage has been completed. Therefore, a single RAG complex maintains occupancy from the first step (nick formation) to the second step (cleavage). In addition, the results suggest that N region diversity at V(D)J junctions within rearranged immunoglobulin and T cell receptor gene loci can only be introduced after the generation of RAG-catalyzed DNA double-strand breaks, i.e. during the DNA end joining phase of the V(D)J recombination reaction.

INTRODUCTION

In vertebrates the antigen binding domains of immunoglobulin (Ig) and T cell receptor (TcR) molecules are assembled from numerous variable (V), joining (J) and sometimes diversity (D) gene segments that are separated in the germline (1). These gene segments are assembled in a random, yet site-specific manner during early lymphoid development by a process called V(D)J recombination, which generates the diversity of antibody and TcR molecules (2,3). The coding regions of all functional V, D and J gene segments are flanked by conserved recombination signal sequences (RSS) that consist of a palindromic heptamer, which is separated from an A/T-rich nonamer by a random spacer sequence of either 12 or 23 bp length. These RSSs are located directly adjacent to the coding region sequence. Site-specific recombination of V, D and J gene segments *in vivo* has been found to require the presence of two RSSs of different spacer length, which is generally referred to as the 12/23 bp rule (4).

Initiation of V(D)J recombination requires expression of the two recombination activating genes *RAG-1* and *RAG-2* (5,6). If either of the two *RAG* genes is non-functional, Ig and TcR loci remain in the germline configuration and mice carrying these mutations are devoid of peripheral B and T lymphocytes (7,8). Later stages of a complete V(D)J recombination reaction appear to involve factors that are also necessary to mediate the repair of DNA double-strand breaks, like the 470 kDa DNA-dependent protein kinase, which appears to be mutated in SCID mice (9,10), the DNA end binding proteins Ku70 and Ku86 (11–13) and the recently identified *XRCC-4* gene product (14).

Recently, a cell-free assay for the initiation of V(D)J recombination has been developed (15) and it has been demonstrated that both RAG proteins alone are sufficient to catalyze the initial DNA double-strand breaks at RSS (16). The cleavage reaction does not require a high energy cofactor and occurs by a two-step mechanism. First, a single-strand nick is introduced at the border of the heptamer of the RSS and the adjacent coding end, which is followed by nucleophilic attack of the *in situ* generated 3′-OH group at the nick on the phosphodiester bond of the opposite, complementary DNA strand (16). This second step occurs by a direct transesterification reaction, which does not involve a covalent protein–DNA intermediate and is therefore reminiscent of retroviral integration (17). The end products of a RAG-catalyzed

*To whom correspondence should be addressed. Tel: +1 314 362 4227; Fax: +1 314 362 0010; Email: lieber@pathology.wustl.edu

cleavage reaction are a hairpin coding end and a blunted, 5′-phosphorylated signal end retaining the RSS.

This reaction can occur at an isolated RSS when it is carried out in the presence of Mn^{2+} as divalent cation. Some limited degree of concerted cutting at presumably synapsed RSSs with different spacer lengths (i.e. following the 12/23 bp rule) is favored in the presence of Mg^{2+} (18,19). The degree of concerted cleavage at two signals is markedly lower than cleavage in crude extracts (19). Therefore, for the current study we have focused on the use of purified RAGs in the cleavage of a single RSS using Mn^{2+} as the metal cofactor. Despite the strong sequence specificity of the RAG-mediated cleavage reaction, documentation of binding of the RAG proteins to RSSs by conventional methods, including gel mobility shift assays or DNA–protein crosslinking experiments, has been difficult thus far (see Discussion). One possible explanation for this is the rapid on–off rates for RAG-1 and RAG-2 binding at the RSSs. This prompted us to analyze the accessibility of the reaction intermediate and end products of the RAG cleavage reaction for N nucleotide addition by the enzyme terminal deoxynucleotidyl transferase (TdT) in order to indirectly assess whether a more stable complex might form between RAG proteins and target DNA at any point during the cleavage reaction.

Here we show that the internal 3′-OH group of the nicked intermediate, which is generated during the first step of a RAG-catalyzed cleavage reaction and which remains unconverted for at least 3 h during the assay, is resistant to N nucleotide addition by TdT, indicating that the RAG proteins form a complex with the substrate DNA that persists after the single-strand cleavage step (nicking) has occurred and until completion of the strand transfer reaction (hairpin formation).

MATERIALS AND METHODS

Insect cell culture and expression of RAG-1 and 2 proteins

Recombinant baculoviridae for expression of maltose binding protein (MBP)-tagged, truncated mouse RAG-1 and RAG-2 proteins were kindly provided by Drs Dik van Gent and Martin Gellert (NIH). MBP-tagged RAG proteins were expressed in Hi-5 insect cells (Invitrogen, San Diego, CA) grown at 28°C in either suspension cultures on a shaker set to 100 r.p.m. or as adherent cells in 150 mm dishes in Ex-Cell 400 serum-free insect medium (JRH Biosciences, Woodland, CA). Expression of recombinant RAG proteins was generally carried out by co-infecting adherent Hi-5 cells for 48–72 h with recombinant RAG-1 and RAG-2 baculovirus stocks at ratios between 1:3 and 1:5 with a multiplicity of infection ranging between 5 and 10.

Primer sequences and cleavage substrates

DNA oligomers were synthesized by the Protein and Nucleic Acid Chemistry Laboratory (PNACL) at Washington University School of Medicine (St Louis, MO). Crude oligomers were gel purified by dentauring PAGE and isolated by a standard crush and soak procedure. Aliquots of 10 pmol were used for 5′-labeling with [γ-^{32}P]ATP, which were subsequently annealed to 10 pmol cold complementary DNA by heating to 96°C in 25 mM Tris–HCl, pH 8.0, 100 mM NaCl for 10 min, followed by slow cooling from 65°C to room temperature over a time period of 1–2 h. Samples of 1 µl (0.1 pmol) annealed substrate were used in the standard cleavage reaction.

Primer sequences. Sequences for substrate B (79mer substrate): UG078 (top strand), 5′-ATCAGGATGTGGTGATCCACAGTGTGATCCCTCCTCACAAAAACCGCAGGTCTTCAGTT-3′; UG079 (bottom strand), 5′-AACTGAAGACCTGCGGTTTTTGTGAGGAGGGATCACACTGTGGATCACCACATCCTGAT-3′. Sequences for substrate C (79mer substrate), UG211 (top strand): 5′-ATCAGGATGTGGTACAGTGTGATCCCTCCTCACAAAAACCGCAGGTCTTCAGTT-3′; UG079 (bottom strand), 5′-AACTGAAGACCTGCGGTTTTTGTGAGGAGGGATCACACTGTGCATCACCACATCCTGAT-3′. Sequences for substrate A (50mer) and its pre-nicked derivatives: UG071 (top strand), 5′-TTGCATCGCATGCCTCCACAGTGATCATCCTCGAGACAAAAACCTGCAAC-3′, UG072 (bottom strand), 5′-GTTGCAGGTTTTTGTCTCGAGGATGATCACTGTGGAGGCATGCGATGCAA-3′; UG073 (top strand/coding end), 5′-TTGCATCGCATGCCTC-3′; UG074 (top strand/signal end), 5′-CACAGTGATCATCCTCGAGACAAAAACCTGCAAC-3′; UG075 (top strand/coding end, gap –1), 5′-TTGCATCGCATGCCT-3′; UG076 (top strand/coding end, flap +1), 5′-TTGCATCGCATGCCTCA-3′; UG077 (top strand/coding end, flap +5), 5′-TTGCATCGCATGCCTCAAGTG-3′.

Purification of recombinant truncated RAG-1 and RAG-2 proteins

MBP-tagged RAG proteins were isolated from insect cells essentially as described by McBlane *et al*. (16). In brief, infected insect cells were sonicated for 2 min on ice in binding buffer (25 mM Tris–HCl, pH 7.5, 500 mM NaCl, 20 mM imidazole, 2 mM 2-mercaptoethanol), cell debris was removed by centrifugation at 12 000 r.p.m., 4°C, 15 min and cleared supernatants were incubated with Ni–NTA agarose beads (Qiagen, Hilden, Germany) in batch mode at 4°C by end-over-end rotation for 1 h. The beads were washed three times with binding buffer and RAG proteins were batch-eluted with binding buffer containing 1000 mM imidazole. The eluate was 5-fold diluted in amylose resin binding buffer (25 mM Tris–HCl, pH 8.0, 500 mM NaCl, 0.25% Tween 20, 2 mM dithiothreitol) and was then incubated with amylose resin (NEB Biolabs, Beverly, MA) in batch mode at 4°C by end-over-end rotation for 1 h. The amylose beads were washed twice with binding buffer and twice with binding buffer without Tween 20 and RAG proteins were eluted with 25 mM Tris–HCl, pH 8.0, 500 mM NaCl, 2 mM dithiothreitol, 20 mM maltose. Proteins were finally dialyzed for 3 h at 4°C against 25 mM Tris–HCl, pH 8.0, 150 mM KCl, 10% glycerol, 2 mM dithiothreitol. Protein samples were analyzed by SDS–PAGE followed by SYPRO-orange™ (Molecular Probes, Eugene, OR) staining and visualized on either a regular UV light box or using a STORM Phosphorimager equipped with ImageQuant[a] software (version 1.1) (Molecular Dynamics, Mountain View, CA). Proteins were further visualized by silver staining of the protein gels using a BioRad silver stain kit (BioRad, Hercules, CA) according to the recommendations of the manufacturer. Based on SYPRO-orange and silver staining, all RAG protein samples had a purity of >95%.

Cleavage reactions with recombinant RAG proteins and TdT tailing assays

RAG protein-catalyzed cleavage reactions were generally carried out in 25 mM MOPS–KOH, 7.5 mM Tris–HCl, 10 mM NaCl, 30 mM KCl, 60 mM potassium glutamate, 2.4 mM dithiothreitol,

1 mM $MnCl_2$, 2% glycerol, pH 7.2 (including the components from the dialysis buffer and from the buffer used to anneal the substrates) for 3 h at 37°C with 100 fmol labeled DNA substrate and recombinant RAG-1 and RAG-2 in the range 50–100 fmol. Cleavage and hairpin activity was not detectable if RAG-1 or RAG-2 were used individually.

TdT tailing assays were carried out in the same buffer (whenever TdT tailing was performed in the absence of RAG proteins, the reaction volume was adjusted with dialysis buffer to achieve identical ionic conditions) supplemented with 10 µM dNTPs and 2 µM ddNTPs for 3 h at 37°C in a 10 µl volume. The dideoxynucleotides were added in order to limit the degree of N nucleotide addition at 3′-OH groups. TdT was obtained from Boehringer Mannheim Biochemicals.

Cleavage and/or TdT tailing reactions were stopped on ice by addition of 10 µl formamide, 1 µl 100 mM EDTA and 1 µl 1% SDS. Samples were denatured at 99°C for 10 min, cooled on ice and immediately fractionated by 20% denaturing PAGE containing 8 M urea using a regular sequencing unit. Specific signals were visualized by overnight autoradiography at –80°C and/or phosphorimager analysis. Quantification of signals was carried out using the ImageQuant 1.1 software package (Molecular Dynamics).

RESULTS

Expression of recombinant truncated mouse RAG proteins

Recombinant core RAG-1 (amino acids 384–1004) and RAG-2 (amino acids 1–387) proteins were expressed in Hi-5 insect cells as MBP fusion proteins (with the MBP tag fused to the N-terminus) containing an additional His_9 tag and three myc epitopes at the C-terminus for immunodetection and further purification (16). After two affinity purification steps using Ni–NTA–agarose and amylose resin (see Materials and Methods), RAG proteins were obtained that were >95% pure as judged by SYPRO-orange™ protein staining (data not shown). In order to obtain active RAG proteins, insect cells had to be co-infected with recombinant RAG-1 and RAG-2 baculovirus stocks at ratios ranging between 1:5 and 1:3 (data not shown).

Coding end sequence affects the efficiency of site-specific cleavage at recombination signal sequences

The efficiency of endonucleolytic cleavage at recombination signal sequences by these RAG proteins was strongly dependent on the coding end sequence of the substrate if the reaction was carried out in the presence of Mn^{2+} as divalent cation (17,20). To illustrate this finding, Figure 1 shows the kinetics of a cleavage reaction for two different substrates that differ by only 1 bp at the terminal position of the coding end, i.e. directly adjacent to the palindromic heptamer of the RSS. If the top strand of a cleavage substrate is labeled at the 5′-end, the nicked intermediate of the cleavage reaction can be detected by denaturing polyacrylamide gel electrophoresis as a 17 nt DNA fragment, which can be discriminated from the covalent hairpin coding end migrating at twice the size, i.e. 34 nt (Fig. 1). The hairpin results from nucleophilic attack of the newly generated free 3′-OH at the nick on the phosphodiester bond of the opposite complementary DNA strand, thereby leading to a double-strand break at the border of the RSS with the coding end (17).

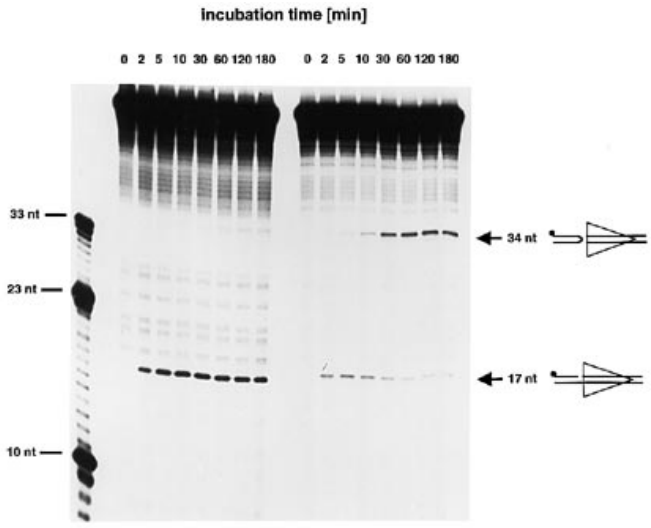

Figure 1. Kinetics of RAG-catalyzed cleavage with two different substrates. Cleavage reactions were carried out in the presence of Mn^{2+} as divalent cation and were stopped after various time points as indicated. The left panel shows the kinetics of cleavage with 79mer substrate B (UG078+UG079, …GATC-heptamer-12-nonamer) and the right panel with 79mer substrate C (UG211+UG212, …GATG-heptamer-12-nonamer). Substrate B displays a strong preference for nicking, whereas substrate C efficiently converts the nicked intermediate into hairpin coding ends that accumulate to near plateau levels (3.6%) after 30 min incubation. The left-most lane contains labeled DNA oligonucleotides that were used as size markers. The anticipated structures of the nicked intermediate (17 nt) and the cleaved substrate with the 34 nt hairpin coding end are schematically displayed at the right side of the figure. The position of the label is depicted as a black dot at the 5′-end of the top strand. The open triangle represents a RSS containing a 12 bp spacer with the flat end at the heptamer side.

The substrate containing a dC·dG base pair coding end (with deoxycytidine on the top strand; substrate B in Fig. 1, left side) appears to be quickly and efficiently nicked, as evidenced by the emergence of the 17 nt DNA fragment after 2 min incubation with RAG proteins. However, this substrate allows only slow and inefficient conversion of the nicked substrate into the hairpin coding end and the blunted and 5′-phosphorylated signal end. Even after 180 min incubation at 37°C, only 10.4% of the nicked intermediate had been converted into the hairpin coding end (as determined by quantification using a phosphorimager). A change of this dC·dG base pair at the coding end to a dG·dC base pair (now with dG adjacent to the CACAGTG heptamer sequence on the top strand; substrate C in Fig. 1, right side) had a strong effect on the kinetics of the overall cleavage rection. Although the nicked intermediate of the cleavage reaction became equally detectable within 2 min incubation with RAG proteins, it did not accumulate to levels comparable with the previous substrate, but was rather efficiently converted into the hairpin coding end. However, the combined percentage of nicked intermediate and hairpin coding end was comparable for both substrates (on average 4–6% conversion in several independent experiments). The apparent lower conversion rate for substrate C (Fig. 1) is the result of lower amounts of total radioactivity loaded per lane for substrate C as compared with substrate B (on average 30% less radioactivity per lane). In summary, the terminal nucleotide of the coding end can have a significant impact on the efficiency of

conversion of the nicked intermediate into the hairpin product. These results confirm and extend recent results, which do not include these particular coding ends, but demonstrate the concept of coding end sequence effects (18,20,21). Coding end effects have been seen in cellular V(D)J recombination studies for wild-type (22) and mutant RAG proteins (23). In the studies below, we utilize these substrate differences to maximize probing of the intermediates for occupancy by the RAG protein complex.

Nicked DNA double-strand molecules are substrates for TdT

One way to detect whether the intermediates are stably bound by RAG proteins throughout the cleavage reaction is to investigate to what extent the *in situ* generated internal 3′-OH group at the initial nick might be accessible for N nucleotide addition catalyzed by TdT. In order to demonstrate that a 3′-OH group at a nick can serve as a substrate for TdT, three DNA oligomers were designed, 5′-phosphorylated and annealed to generate the nicked intermediate structure. The sequence of this substrate was the same as the RSS substrate, displaying a slow rate of hairpin formation. To confirm that the three DNA oligomers would anneal in the predicted fashion (i.e. form the nicked double-stranded DNA molecule), the bottom strand was 5′-labeled, annealed with the two complementary oligomers and digested with restriction enzymes that would cut upstream (*Sph*I and *Fok*I) and downstream (*Mbo*I and *Xho*I) of the internal nick. In each case >95% of the substrate was specifically cleaved by these restriction enzymes (data not shown), confirming that the DNA oligomers did anneal in the predicted fashion to generate the nicked double-stranded DNA substrate.

TdT-catalyzed N nucleotide addition at such an internal nick can be detected if the DNA oligomer upstream of the nick is labeled at its 5′-end. As shown in Figure 2, N nucleotide addition was readily detectable at the internal nick, beginning at TdT concentrations as low as 0.2 U/1 µl reaction (20 U/ml) and increasing in intensity as the TdT concentration was raised to 25 U/10 µl reaction (2500 U/ml), a concentration at which >98% of the nicked DNA substrate could be tailed by TdT. All TdT tailing reactions were carried out in the presence of RAG cleavage buffer (see Materials and Methods), with Mn^{2+} as the divalent cation in order to maintain reaction conditions identical to those chosen for RAG-mediated cleavage reactions. This degree of N nucleotide tailing was identical to that observed with a blunt-ended DNA substrate (data not shown).

From these experiments we infer that a 3′-OH group at a DNA nick can serve as a substrate for TdT.

The nicked intermediate generated during a RAG-catalyzed cleavage reaction is protected from N nucleotide addition by TdT

We analyzed the extent to which N nucleotides could be added by TdT to either the nicked intermediate, the newly generated blunted signal end or the 3′-OH group at the distal end of the hairpin. All these products became detectable during a RAG-catalyzed cleavage reaction at comparable levels if the two different substrates with either a propensity for nicking or hairpin formation were used. Titration of various amounts of TdT into the RAG-catalyzed cleavage reaction demonstrates that N nucleotides can be added to the signal end after completion of the cleavage reaction (Fig. 3, right panel), as evidenced by a similar

Figure 2. N nucleotide addition at an internal nick catalyzed by TdT. A pre-nicked DNA double-strand 50mer was generated by annealing labeled primer UG073 (top strand, coding end, 16mer) with cold primers UG074 (top strand, signal end, 34mer) and UG072 (complementary bottom strand, 50mer). This substrate was incubated with various concentrations of TdT as indicated. The percentage of elongated products ($n \geq 16$) was determined by phosphor-imager quantification to be 0.4, 43.5, 98.7 and 98.1% in the reactions containing 0.2, 1.0, 5.0 and 25.0 U TdT per 10 µl reaction respectively. Percentages of N nucleotide addition at these TdT concentrations were practically identical for blunt-ended substrates (data not shown). Restriction enzyme digestion of the same annealed substrate but 5′-labeled on the bottom strand (UG072) with either *Sph*I, *Fok*I, *Mbo*I or *Xho*I resulted in nearly complete site-specific digestion (>95%) of the substrate, thereby confirming that the primers annealed in the predicted fashion to generate DNA double-strand substrate with an internal nick.

degree of TdT tailing observed at the 3′-OH group of the hairpin (Fig. 3, middle panel), which should be readily accessible for TdT since it is located 17 bp from the RSS. However, no N nucleotide addition was detectable at the 3′-OH group that results from single-strand nicking at the RSS as the first step of a RAG-catalyzed cleavage reaction (see Fig. 3, left panel). In fact, at least 90% of the 3′-OH groups of the nick in the reaction intermediate were completely resistant to TdT tailing at concentrations of up to 5 U/10 µl reaction, which resulted in >95% tailing in the case of a control nicked substrate (Fig. 2).

Because most of the nicked intermediate was not converted into the hairpin coding end by the RAG proteins in the assay time period of 3 h, we conclude that the RAG proteins form a persistent complex with the substrate DNA during the course of the cleavage reaction which appears to remain stable under the conditions of the assay and prevents TdT from adding N nucleotides at the 3′-OH of the nick. The conclusion that the nicked intermediate is protected from TdT addition by the RAGs applies to substrates that are efficiently converted to the cleavage (hairpin) product (Fig. 3, middle panel, nicked species at 17 nt) as well as to the less efficiently cleaved substrates (Fig. 3, right panel).

We investigated whether binding of the RAG proteins to the nicked intermediate would be detectable as a decrease in overall TdT tailing efficiency if pre-nicked substrates were used for RAG-catalyzed cleavage assays. N nucleotide addition by TdT to nicked and gapped substrates was measured using increasing amounts of TdT in either the presence or absence of RAG proteins. Despite easily detectable hairpin forming activity that resulted in up to 18% conversion of substrates, no significant decrease in the degree of N nucleotide addition at the various TdT concentrations was detectable when RAG proteins were present

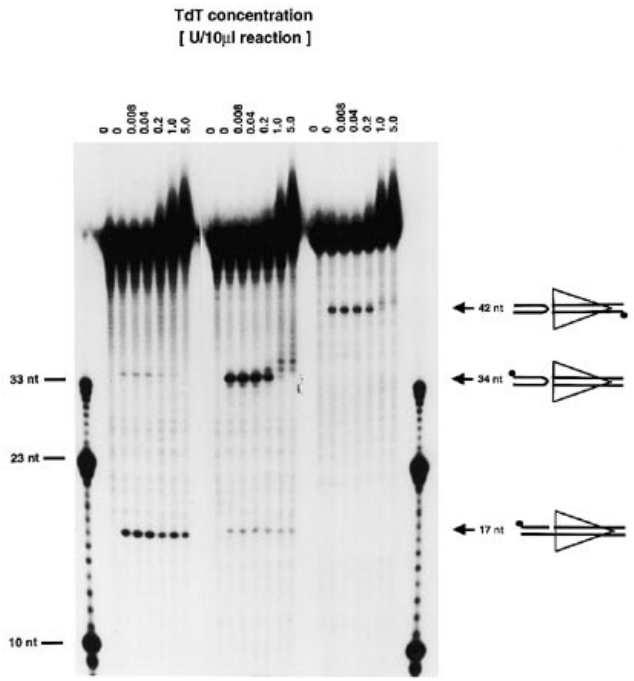

Figure 3. Titration of TdT into RAG-catalyzed cleavage reactions. Standard cleavage reactions were carried out with substrate B (left panel), 5′-labeled on the top strand, or substrate C, 5′-labeled on either the top (middle panel) or the bottom strand (right panel). The first lane of each set of experiments shows substrates incubated in buffer alone; all other lanes represent RAG-catalyzed cleavage reactions in the presence of increasing amounts of TdT (as indicated). Products of the cleavage reaction and postion of the label are schematically indicated on the right side of the autoradiograph. At increasing concentrations of TdT the signals for the hairpin coding end (migrating at 34 nt) and the signal end (migrating at 42 nt) disappear due to increased N nucteotide addition by TdT, whereas the signal for the nicked intermediate (17 nt) remains practically unaffected even at elevated concentrations of TdT. The left- and right-most lanes contain a mixture of labeled DNA oligonucleotides serving as size standards.

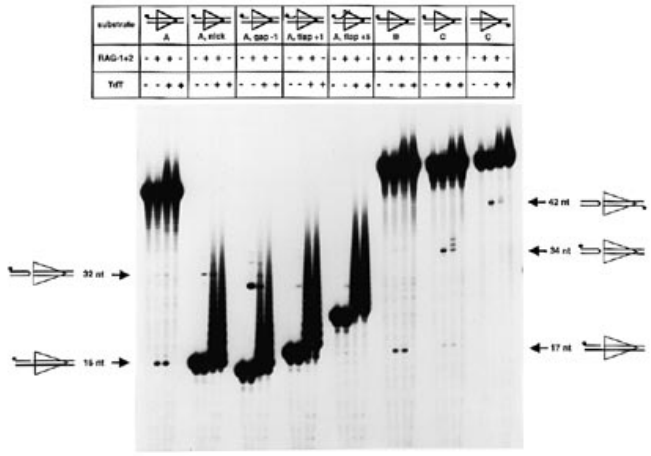

Figure 4. TdT tailing in the absence and presence of RAG proteins with different pre-nicked and normal substrates. The various substrates were incubated either in buffer without any proteins added (first lane), with RAG proteins in the absence (second lane) or presence (third lane) of 0.5 U TdT/10 µl reaction or with TdT alone (same concentration, fourth lane), as indicated at the top of the figure. Sizes and structures of reaction intermediates and end products of cleavage reactions for substrate A and substrates B and C are schematically indicated on the left and right respectively. The series of pre-nicked substrates had identical coding and signal end sequences, except that their top strands were nicked and varied in the length of the coding end primer, which either lacked the 3′ base (in the case of the gap −1 substrate UG075+074+072), had one or five additional bases at the 3′-end with no complementary sequence on the bottom strand [the flap +1 (UG076+074+072) and flap +5 substrates (UG077+074+072) respectively] or had a simple nick in the top strand (UG073+074+072). Substrates B and C were identical to those in Figure 1. The position of the label in the schematic drawings of the various substrates at the top of the figure is indicated as a dot.

in the tailing reaction (not shown). This analysis was extended by comparing the tailing efficiency in the presence and absence of RAG proteins at a given TdT concentration (0.5 U/10 µl reaction) by including other pre-nicked substrates with either merely a nick at the RSS or with a 1 nt or a 5 nt 3′ overhang (3′-flap) (Fig. 4). Again, no substantial difference in tailing efficiency could be detected for any of the different pre-nicked substrates, whether or not RAG proteins were included in the assay. The conclusion from these experiments is that apparently only a small fraction of the pre-nicked substrates is complexed and converted by RAG proteins which cannot therefore be detected by the indirect TdT tailing assay under the conditions used in the previous sections. However, once the RAG proteins associate with the RSS and initiate the cleavage reaction by creating a nick, they remain bound to the DNA until hairpin formation and double-strand cleavage are completed.

The protected nicked intermediate becomes fully sensitive upon phenol extraction

If the nicked intermediate is not susceptible to TdT-catalyzed addition because of RAG complex occupancy, then phenol extraction would be expected to reverse this. This was tested by incubating the RAG core proteins with the substrate to generate the nicked species and a small amount of hairpin product (Fig. 4). Half of the reaction mix was phenol extracted and half was not. The TdT tailing assay shows that while the phenol-extracted nicked species could readily be tailed, the 3′-OH is protected in the unextracted reaction (Fig. 5).

DISCUSSION

Recombination signal sequences that flank coding regions of TcR and Ig gene segments are highly conserved during vertebrate evolution. Although they are an indispensable requirement for the recognition and site-specific cleavage catalyzed by RAG proteins, direct binding of RAG proteins to these sequences has not yet been reported by standard methods. Recently, RAG-1 has been shown to interact with the nonamer (24,25) and, while this manuscript was under review, RAG-1 and RAG-2 have been shown to form a specific complex with a complete single signal sequence when stabilized by glutaraldehyde (26). We have used an indirect approach to assess RAG protein–DNA interaction by using the template-independent DNA polymerase TdT in order to probe the intermediates and end products of the RAG cleavage reaction for accessibility to N nucleotide tailing. In order to be able to compare the extent of TdT tailing on either the reaction intermediates or the end products, we took advantage of the finding that in cleavage buffer with Mn^{2+} as divalent cation, different coding ends strongly affect the outcome of the

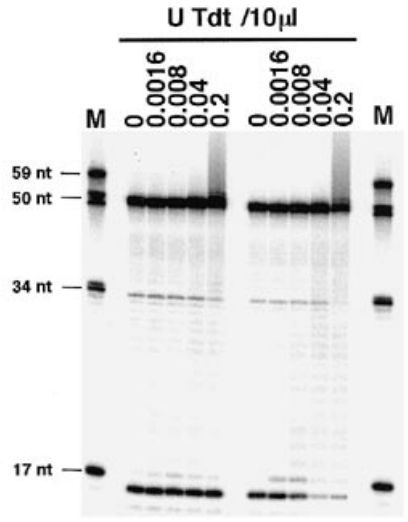

Figure 5. TdT tailing after but not before phenol extraction indicates that protection at the nicked intermediate can be removed by organic extraction. Two pmol recombination signal substrate were reacted with ~2 pmol each RAG-1 and RAG-2 core proteins in a 200 µl volume in the standard Mn^{2+} buffer at 37°C for 180 min. Half of the total volume was kept on ice and the other half was phenol/chloroform extracted followed by ethanol precipitation. The pellet was dissolved in 100 µl 1× Mn^{2+} buffer. Aliquots of 10 µl (0.1 pmol substrate) of both the phenol-extracted and the non-extracted mixtures were assayed for TdT reaction. After addition of dNTP/ddNTP solution, TdT in Mn^{2+} buffer and 100 µg/ml BSA (see Materials and Methods) was added and the reaction mixture was incubated for 30 min at 37°C. The reaction was then terminated with stop solution and heating to 95°C for 10 min. Reaction products were resolved on a 20% denaturing polyacrylamide gel in 8 M urea. Substrate A had a strong preference for nicking, since it had the same coding end sequence as substrate B (Fig. 1). (It was generated by annealing labeled oligonucleotide UG071 with unlabeled UG072.) The outside lanes (M) contain the markers. The nicked product is 16 nt; the hairpin product is 32 nt. The first five lanes on the left are without phenol extraction. They demonstrate that TdT is not able to tail at the 3′-OH of the nick; this is indicated by the fact that the 16 nt product is not converted to higher molecular weight species. The five lanes on the right show that the nick can be readily tailed after phenol extraction; this is indicated by the fact that the 16 nt product is converted to higher molecular weight species with increasing amounts of TdT.

RAG-catalyzed cleavage reaction. Hence, different substrates would yield comparable levels of either intermediate or end products.

The results from these analyses are that the nicked intermediate that is formed as a first step of a RAG-mediated cleavage reaction is resistant to N nucleotide tailing by TdT (despite the fact that the same structure is a good substrate for TdT under identical assay conditions). Some of our substrates result in substantial conversion to double-strand cleavage product (15–20% in Fig. 5) and yet the nicked intermediate remains completely protected by RAGs, here as well as when conversion to cleavage product is much lower. In contrast, the 3′-OH group generated at the signal end after completion of the reaction is readily accessible for TdT tailing. These results indicate that RAG proteins form a persistent complex with the substrate DNA once the reaction has been initiated by nicking at the heptamer/coding end border.

It should be noted that our analysis provides information only after the nicking step. We cannot infer anything about binding of RAG proteins prior to this step because the modifying agent, TdT, is only capable of adding nucleotides after the nick is created.

Recent results have provided information on the V(D)J reaction pathway (15,17,18) and our results allow additional detail to be included in such a reaction scheme for a single recombination signal (Fig. 6). None of these steps is ATP dependent (15,17,18). The initial binding step is reversible. The strand cleavage step (Fig. 6, step II) is depicted as irreversible; a reverse reaction would represent the formation of a phosphodiester bond in the absence of an energy source. Once the nicked intermediate is formed (Fig. 6, structure C), our results indicate that TdT has no significant access to the 3′-OH of the nick. TdT access can only be detected at the signal end after the hairpin is formed (Fig. 6, structure D). This indicates that RAG complex dissociation from the nicked intermediate (Fig. 6, structure C) to generate an unbound nick (Fig. 6, structure E) can only occur to a relatively small extent and this must be below the detection limit of the TdT assay used here. The fact that an unbound nicked intermediate (E) can be converted to hairpin product (D) in the presence of RAG proteins indicates the an equilibrium between C and E exists, as shown here and by McBlane *et al.* (16). However, our data indicate that the equilibrium between structures C and E is shifted heavily in favor of C. Hence, the pathway from A→B→C→D, without diversion to E, is likely to be the favored one.

The results here would suggest that N nucleotide addition is unlikely prior to nucleolytic opening of the hairpins. This is based on the fact that the bound nicked intermediate (Fig. 6) structure C continues to D rather than E, based on our observed TdT inaccessibility. Even for the small fraction of molecules for which nicks might be exposed after dissociation of the RAG complex (Fig. 6, E), TdT extension of the nick would produce a 3′-flap that appears to be a poor direct substrate for hairpin formation (Fig. 4). Interestingly, the low level of hairpin product that does arise from 3′-flap substrates has an identical mobility to the hairpin generated from a 1 nt gapped substrate. This suggests that the RAG complex can cleave 1 and 5 nt 3′-flaps at the nicked intermediate, converting them to 1 nt gaps that can form hairpins. This may provide a salvage pathway for failed reactions that abort at the nicked stage.

The finding that RAG proteins remain bound to recombination signal sequences throughout the initiation of V(D)J recombination suggests that N region diversity at V(D)J gene junctions must primarily only be introduced during the DNA end joining phase of the V(D)J recombination reaction. Without protection of the nicked intermediate by an active RAG protein complex, it is conceivable that coding end nucleotides on the top strand might be lost or added due to an exonuclease or TdT respectively. This is especially the case in the light of the finding that recombination substrates with a 1 (see Fig. 4), 2 or even 3 nt gap at the coding flank, as well as substrates with mismatched coding end sequence, can efficiently be converted by RAG proteins (20). This raised the possibility that N region diversity could be introduced during the first RAG-mediated events of V(D)J recombination. However, our results exclude such a possibility as the major pathway. Protection of the nicked intermediate during RAG-catalyzed initiation of V(D)J recombination appears to prevent N nucleotide tailing of the intermediate in TdT-expressing precursor lymphocytes. These results were obtained under buffer conditions where Mn^{2+} is the metal cofactor. In the cell it is likely that Mg^{2+} is the necessary cofactor. We expect that our results regarding conversion from nicked intermediate to cleaved hairpin product will hold when it is possible to conduct similar studies with Mg^{2+} as the cofactor.

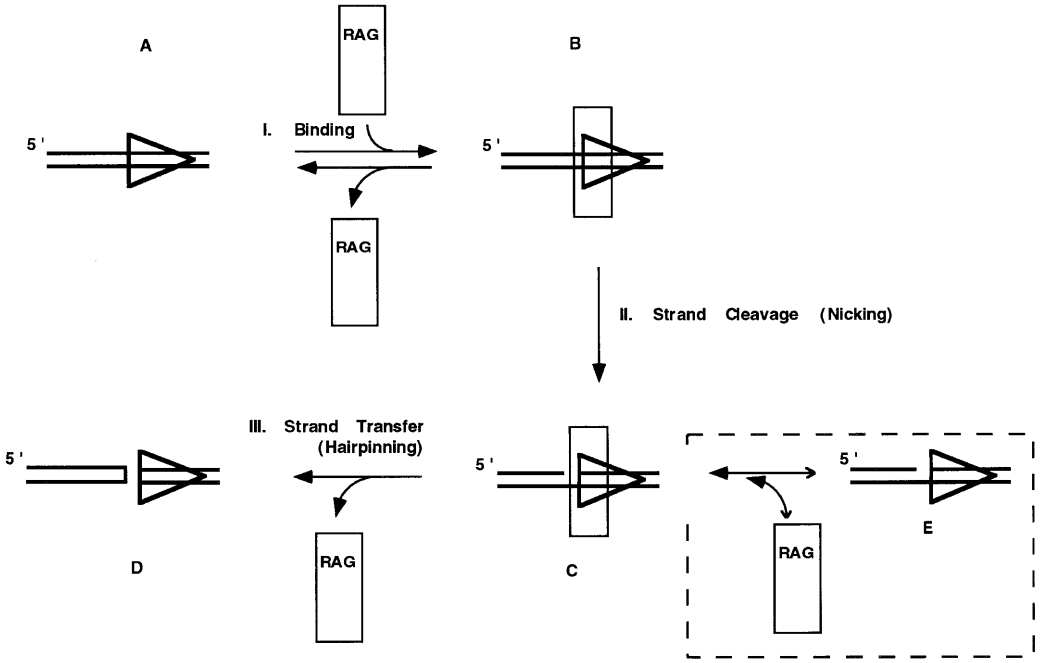

Figure 6. Biochemical model for the early steps of V(D)J recombination. DNA and DNA–protein complexes are designated by the letters A–E. Reaction steps are designated by Roman numerals. The RAG-1–RAG-2 protein complex is designated by a rectangle. The triangle and parallel lines represent a single 12 base spacer recombination signal. According to this model, structures A→B→C→D represent the normal pathway, with structure E representing an infrequent alternative (as designated by the surrounding dashed line). The binding step (I) is designated as reversible. The strand cleavage step (II, nicking) is designated as irreversible because energy would be required to restore a phosphodiester bond. Based on the results here, we propose that once the nick is made, the RAG complex usually remains bound until hairpin formation is completed. At this point, the RAG complex releases from the signal end. However, we have no data bearing on release of the RAG complex from the coding end.

Concerted cleavage at synapsed recombination signal sequences in the presence of Mg^{2+} seems to increase the overall conversion rate of substrates by RAGs (18,19), nevertheless, it is tempting to speculate that additional cofactors are required to allow the reaction to occur more efficiently. The kinetics of cleavage at either isolated signals or synapsed RSSs, which appear to reach steady-state levels below a conversion rate of 50% (16,18,19), raise the possibility that RAG-catalyzed cleavage may be a one turnover reaction.

In studies utilizing single signals and Mn^{2+} as a cofactor for the RAG proteins, there is a strong effect of the sequence of the DNA adjacent to the signal (18,20,21; this study). This DNA corresponds to the coding segment end in the genome. The Gellert laboratory has also described coding end effects in Mn^{2+} (20), but they indicate that these are less apparent in Mg^{2+}-dependent reactions, in which they found some degree of synapsis (18). In the cellular V(D)J recombination assay, coding end effects can be quite large (22). These initial cellular studies used several different transformed pre-B cells from wild-type mice and, hence, the RAG proteins are full-length and the ionic conditions are physiological. In these cellular assay studies, coding end sequence variations could reduce signal joint formation and coding joint formation equivalently. The simplest explanation for this is that the coding end sequence effect occurs at the initial cleavage step of V(D)J recombination (22). The observation of coding end effects with Mn^{2+} may reflect those seen in the cellular assay. However, coding end effects of a different spectrum can be observed in the cellular assay when employing mutant versions of RAG-1 protein (23) and these are reported to be more closely related to the types of coding end preferences seen when using Mn^{2+} as cofactor (18). Additional studies are aimed at creating a DNA oligomer-based RAG cleavage assay with synapsed RSSs using Mg^{2+} as cofactor, in order to test whether the same rules regarding occupancy of the RSS apply. Those studies will provide interesting insights into the relationship between the cofactor, alterations in the RAG proteins and coding end sequence.

ACKNOWLEDGEMENTS

We would like to thank Drs Dik van Gent and Martin Gellert for the generous gift of the baculovirus stocks that were required for expression of the maltose binding protein tagged RAG-1 and RAG-2 proteins. We would further thank them for communication of results prior to publication and their interest in the progress of this study. U.G. is a post-doctoral fellow of the Boehringer Mannheim Foundation.

REFERENCES

1 Tonegawa,S. (1983) *Nature*, **302**, 575–581.
2 Alt,F.W., Oltz,E.M., Young,F., Gorman,J., Taccioli,G. and Chen,J. (1992) *Immunol. Today*, **13**, 306–314.
3 Lewis,S.M. (1994) *Adv. Immunol.*, **56**, 27–150.
4 Lieber,M.R. (1991) *FASEB J.*, **5**, 2934–2944.
5 Schatz,D.G., Oettinger,M.A. and Baltimore,D. (1989) *Cell*, **59**, 1035–1048.
6 Oettinger,M.A., Schatz,D.G., Gorka,C. and Baltimore,D. (1990) *Science*, **248**, 1517–1523.

7 Shinkai,Y., Rathbun,G., Lam,K.P., Oltz,E.M., Stewart,V., Mendelsohn,M., Charron,J., Datta,M., Young,F., Stall,A.M. and Alt,F.W. (1992) *Cell*, **68**, 855–867.
8 Mombaerts,P., Iacomini,J., Johnson,R.S., Herrup,K., Tonegawa,S. and Papaioannou,V.E. (1992) *Cell*, **68**, 869–877.
9 Blunt,T., Finnie,N.J., Taccioli,G.E., Smith,G.C.M., Demengeot,J., Gottlieb,T.M., Mizuta,R., Varghese,A.J., Alt,F.W., Jeggo,P.A. and Jackson,S.P. (1995) *Cell*, **80**, 813–823.
10 Peterson,S.R., Kurimasa,A., Oshimura,M., Dynan,W.S., Bradbury,E.M. and Chen,D.J. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 3171–3174.
11 Rathmell,W.K. and Chu,G. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 7623–7627.
12 Smider,V., Rathmell,W.K., Lieber,M.R. and Chu,G. (1994) *Science*, **266**, 288–291.
13 Taccioli,G.E., Gottlieb,T.M., Blunt,T., Priestley,A., Demengeot,J., Mizuta,R., Lehmann,A.R., Alt,F.W., Jackson,S.P. and Jeggo,P.A. (1994) *Science*, **265**, 1442–1445.
14 Li,Z., Otevrel,T., Gao,Y., Cheng,H.-L., Seed,B., Stamato,T.D., Taccioli,G.E. and Alt,F.W. (1996) *Cell*, **83**, 1079–1089.
15 van Gent,D.C., McBlane,J.F., Ramsden,D.A., Sadofsky,M.J., Hesse,J.E. and Gellert,M. (1995) *Cell*, **81**, 925–934.
16 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.
17 van Gent,D.C., Mizuuchi,K. and Gellert,M. (1996) *Science*, **271**, 1592–1594.
18 van Gent,D.C., Ramsden,D.A. and Gellert,M. (1996) *Cell*, **85**, 107–113.
19 Eastman,Q.M., Leu,T.M.J. and Schatz,D.G. (1996) *Nature*, **380**, 85–88.
20 Ramsden,D.A., McBlane,J.F., van Gent,D.C. and Gellert,M. (1996) *EMBO J.*, **15**, 3197–3206.
21 Cuomo,C. and Oettinger,M. (1996) *Mol. Cell. Biol.*, **16**, 5683–5690.
22 Gerstein,R.M. and Lieber,M.R. (1993) *Genes Dev.*, **7**, 1459–1469.
23 Sadofsky,M., Hesse,J.E., van Gent,D.C. and Gellert,M. (1995) *Genes Dev.*, **9**, 2193–99.
24 Difilippanio,M., McMahan,C., Eastman,Q., Spanopoulou,E. and Schatz,D. (1996) *Cell*, **87**, 253–262.
25 Spanopoulou,E., Zaitseva,F., Wang,F.-H., Santagata,S., Baltimore,D. and Panayoutou,G. (1996) *Cell*, **87**, 263–276.
26 Hiom,P. and Gellert,M. (1997) *Cell*, **88**, 65–72.