Published online 20 February 2009 *Nucleic Acids Research, 2009, Vol. 37, No. 7* **2211–2226**
doi:10.1093/nar/gkp047

Full-length RAG1 promotes contact with coding and intersignal sequences in RAG protein complexes bound to recombination signals paired *in cis*

Sushil Kumar and Patrick C. Swanson*

Department of Medical Microbiology and Immunology, Creighton University Medical Center, Omaha, Nebraska, 68178, USA

Received November 13, 2008; Revised January 10, 2009; Accepted January 17, 2009

ABSTRACT

The RAG proteins initiate V(D)J recombination by mediating synapsis and cleavage of two different antigen receptor gene segments through interactions with their flanking recombination signal sequences (RSS). The protein–DNA complexes that support this process have mainly been studied using RAG–RSS complexes assembled using oligonucleotide substrates containing a single RSS that are paired *in trans* to promote synapsis. How closely these complexes model those formed on longer, more physiologically relevant substrates containing RSSs on the same DNA molecule (*in cis*) remains unclear. To address this issue, we characterized discrete core and full-length RAG protein complexes bound to RSSs paired *in cis*. We find these complexes support cleavage activity regulated by V(D)J recombination's '12/23 rule' and exhibit plasticity in RSS usage dependent on partner RSS composition. DNA footprinting studies suggest that the RAG proteins in these complexes mediate more extensive contact with sequences flanking the RSS than previously observed, some of which are enhanced by full-length RAG1, and associated with synapsis and efficient RSS cleavage. Finally, we demonstrate that the RAG1 C-terminus facilitates hairpin formation on long DNA substrates, and full-length RAG1 promotes hairpin retention in the post-cleavage RAG complex. These results provide new insights into the mechanism of physiological V(D)J recombination.

INTRODUCTION

The structural diversity in immunoglobulins (Igs) and T-cell receptors (TCRs) that enables antigen-specific recognition is generated by V(D)J recombination, a form of site-specific DNA rearrangement responsible for assembling Ig and TCR genes from variable (V), diversity (D) and joining (J) gene segments during lymphocyte development (1). Each gene segment is flanked by a recombination signal sequence (RSS) that contains a conserved heptamer (consensus: 5′-CACAGTG-3′) and nonamer (consensus: 5′-ACAAAAACC-3′) motif separated by either 12 or 23 bp of intervening spacer DNA (12- or 23-RSS, respectively). Two lymphoid cell-specific proteins, called RAG1 and RAG2 (recombination activating genes-1 and -2), initiate V(D)J recombination by preferentially assembling a synaptic complex with a 12-RSS and a 23-RSS adjoining two different gene segments (the '12/23 rule') and coordinately introducing a DNA double-strand break (DSB) at each RSS through a nick-hairpin mechanism, producing two blunt signal ends, and two covalently sealed (hairpin) coding ends (2,3). Pairs of signal ends and coding ends are subsequently processed and joined through the nonhomologous end-joining (NHEJ) repair pathway to create signal joints and coding joints, respectively (4,5).

Determining how the RAG proteins recognize the RSS and assemble a 12-RSS and a 23-RSS into a cleavage-competent synaptic complex is necessary to understand how the RAG proteins achieve site-specific and coordinated introduction of DNA DSBs and avoid illegitimate cleavage events that can lead to potentially oncogenic chromosomal rearrangements. To investigate these issues, we and others have examined RAG complexes bound to a single RSS or an RSS pair, and characterized the assembly, composition, organization, activity, and patterns of protein–DNA contacts in these complexes (6). In most cases, the RAG–RSS complexes characterized in these studies have been assembled using truncated, catalytically active 'core' forms of RAG1 and RAG2 and short oligonucleotide substrates (typically $<$70 bp) containing a single RSS, two of which can be paired *in trans* (as the RSSs are on separate DNA molecules) to form RAG synaptic complexes. This is unlike the *in vivo* situation in which RSSs undergoing rearrangement are typically found on the same DNA molecule (*in cis*).

*To whom correspondence should be addressed. Tel: +1 402 280 2716; Fax: +1 402 280 1875; Email: pswanson@creighton.edu

Forming RAG synaptic complexes on longer, more physiologically relevant RSS substrates *in vitro* presents certain challenges compared with oligonucleotide substrates. One challenge in long substrates is that the RAG proteins must contend with more nonspecific DNA flanking the RSSs, which may pose a significant barrier to synapsis due to charge repulsion of the flanking DNA sequence. In support of this possibility, Huye and Roth (7) demonstrated that the efficiency of RAG-mediated *in vitro* 12/23-regulated RSS cleavage *in trans* exhibits a length dependence: whereas short oligonucleotide substrates are cleaved efficiently, longer substrates (700–1000 bp) with more DNA flanking the RSS are not. This defect can be overcome by pairing the RSSs *in cis* to increase the local concentration of a complementary RSS. However, this solution introduces another challenge because the two RSSs must be sufficiently far apart to enable facile synapsis, as previous studies have documented that coupled RSS cleavage and V(D)J recombination efficiency is sensitive to intersignal distance (8,9). These studies, taken together, suggest that the requirements for RAG-mediated synapsis and cleavage of physiological RSS substrates are not necessarily reflected in the behavior of oligonucleotide substrates typically used to study RAG–RSS complexes. Understanding the basis of these differences would, in part, require being able to directly visualize and characterize discrete RAG complexes assembled on long DNA substrates containing RSSs paired *in cis*, which to our knowledge has not been described.

Toward this end, we describe here the detection and analysis of discrete RAG–RSS complexes assembled using purified core and full-length RAG1/2 and HMGB1 proteins and long (~500 bp) substrates containing an unpaired 12-RSS or appropriately paired (12/23) or mispaired (12/12 or 23/23) RSSs *in cis*. The RAG proteins in these complexes support DNA cleavage regulated by the 12/23 rule, and exhibit plasticity of RSS utilization that depends on partner RSS composition. Methylation interference footprinting of precleavage RAG–RSS complexes reveals subtle differences in RSS recognition by the core and full-length RAG proteins, and provides evidence that full-length RAG1 is more effective at stabilizing structural distortions in the RSS substrate than core RAG1. Pre- and post-cleavage RAG-RSS complexes probed by direct in-gel footprinting using 1,10-phenanthroline copper (CuOP) show evidence of RAG-induced structural distortions within the RSS as well as in regions outside the RSS, both in the coding flank and in intersignal sequence distal to the RSS nonamer. Interestingly, a previously identified joining-deficient RAG1 mutant (K980A) is found to exhibit a much greater defect in mediating hairpin formation on long DNA substrates containing RSSs paired *in cis* than is observed using oligonucleotide substrates paired *in trans*, possibly attributed to a failure to induce structural alterations distal to the RSS nonamer in long DNA substrates. Finally, full-length RAG1 is shown to promote coding end retention in a postcleavage RAG–RSS complex. The implications of these results are discussed.

MATERIALS AND METHODS

DNA constructs

Expression constructs encoding maltose binding protein (MBP) fused to murine wild-type (WT) or D600A core or full length RAG1 (WT or D600A cMR1 and FLMR1) and WT core or full-length RAG2 (cMR2 and FLMR2), or human polyhistidine-tagged full-length HMGB1 have been described previously (10). A mammalian expression construct encoding K980A FLMR1 was generated by recombination PCR as described in Supplementary Material.

Protein expression and purification

WT or mutant cMR1 or FLMR1 were co-expressed with cMR2 or FLMR2 in various combinations (see text) in HEK 293 cells as described previously (10), with minor modifications. In this study, 293 cells were plated in 150×25 mm cell culture dishes; five dishes were used to purify cMR1/cMR2 and cMR1/FLMR2 and eight dishes were used to purify FLM1/cMR2. Each dish was transfected with RAG expression constructs (30 μg/plate; 1:1 ratio cMR1:cMR2 or cMR1:FLMR2 and 3:2 ratio FLMR1:cMR2) using PEI (30 μg/ml) as described (10). Cells were harvested in cold PBS-EDTA 48 h posttransfection and proteins were purified immediately using the 'mild' purification protocol we recently described (11). HMGB1 was expressed and purified as described previously (10).

Oligonucleotide and PCR-generated RSS substrates

Oligonucleotide probes containing a consensus 12- or 23-RSS radiolabeled at the 5′-end were prepared as described (10). A substrate containing the 12-RSS sequence present in pJH299 (12RSSJH299) was similarly prepared by annealing the oligonucleotide 5′-GGGCTGG CAGGTCGACCACAGTGCTACGACTGGAACAAA AACCCTGACG-3′ to its complement (heptamer and nonamer sequences are underlined). Long (~500 bp) DNA substrates radiolabeled at the 5′-end of the top or bottom strand and containing a single 12-RSS (12 only) or a pair of RSSs *in cis* (12/23, 12/12 or 23/23) in the same orientation separated by ~250 bp were generated by PCR as described in Supplemental Material from the pJH299 plasmid V(D)J recombination substrate and its derivatives described elsewhere (12). For methylation interference footprinting experiments, PCR-generated substrates were modified with dimethylsulfate (DMS) and purified by agarose gel electrophoresis as described in Supplementary Material.

Electrophoretic mobility shift assay

Binding reactions to form RAG synaptic complexes on oligonucleotide substrates were assembled as previously described, except reactions were incubated on ice for 10 min (10). To assemble RAG complexes on PCR-generated RSS substrates, the RAG proteins (300 ng) were incubated with the radiolabeled DNA substrate (270 ng) in binding buffer (25 mM MOPS–KOH, 60 mM potassium acetate, 5 mM $MgCl_2$, 10% DMSO and 100 ng/μl BSA;

10 µl final volume) for 10 min on ice. Single-stranded salmon sperm DNA (ssDNA; 360 ng, 200 ng and 5.5 ng for reactions containing cMR1/cMR2, cMR1/FLMR2, and FLMR1/cMR2, respectively) was then added to the reaction mixture, incubated for an additional 15 min on ice, and then the binding reactions were fractionated by electromobility shift assay (EMSA) as previously described (10). The amount of ssDNA added to the reaction was determined empirically to enable detection of a discrete protein–DNA complex with a low background of nonspecific DNA binding.

In vitro and in-gel cleavage assays

In vitro cleavage assays to analyze the accumulation of nick and hairpin products formed over time using oligonucleotide or PCR-generated substrates were performed according to published procedures (12). The cleavage activity of discrete RAG–RSS complexes assembled on PCR-generated substrates and separated by EMSA was analyzed using an in-gel cleavage assay described in detail previously (10). Reaction products were fractionated on an 8% polyacrylamide (19:1 acrylamide/methylene(bis)acrylamide) sequencing gel (containing 7 M urea) for 2h at 2500 V and analyzed using a Storm860 PhosphorImager running the ImageQuaNT software.

Methylation interference and in-gel CuOP footprinting

RAG binding reactions were assembled using DMS-treated or unmodified PCR-generated RSS substrates and protein–DNA complexes were fractionated by EMSA as described above. Precleavage RAG–RSS complexes assembled on unmodified DNA were subjected to direct in-gel CuOP footprinting following established protocols (13). To probe postcleavage RAG–RSS complexes with CuOP, precleavage RAG-RSS complexes were first fractionated by EMSA, and then the complexes were subjected to in-gel cleavage as described above. The resulting postcleavage RAG–RSS complexes were immediately probed by in-gel CuOP footprinting. Bound DNA was electrophoretically transferred to DEAE membrane and recovered as previously described (10). DMS-treated DNA was cleaved at methylated residues by incubation with 10% piperidine at 90°C for 30 min. Reaction products from piperidine- or CuOP-treated DNA were normalized and fractionated on a sequencing gel and analyzed as described above, except that sequencing gels used to separate DNA isolated from postcleavage complexes probed with CuOP additionally contained 40% formamide. Dried gels were scanned using a Storm860 PhosphorImager and line graphs were generated and analyzed using the ImageQuaNT software. Peaks were identified from the line graphs and peak areas were compared between free and bound DNA. In some cases, due to sample loading differences (despite attempts at normalization) that caused all bands in a given lane to appear underrepresented, peaks of interest were alternatively compared to an internal reference peak outside the RSS which displayed a similar abundance and distribution relative to nearby peaks in both bound and free DNA. Peaks showing >2-fold differences in these comparisons were denoted. Due to space constraints, line graphs are provided in Supplementary Material.

Postcleavage hairpin retention assay

Binding reactions containing PCR-generated RSS substrates were assembled as described above, incubated for 10 min at 37°C, placed on ice for 2 min, and protein–DNA complexes were immediately fractionated by EMSA. Bound DNA was recovered and analyzed as described for the in-gel cleavage assay.

RESULTS

Design and preparation of RAG proteins and RSS substrates

To determine if and how noncore portions of RAG1 and RAG2 influence the assembly and activity of RAG–RSS complexes formed on DNA substrates containing RSSs paired *in cis*, we prepared three different combinations of core and full-length RAG1 and RAG2 proteins (cMR1/cMR2, cMR1/FLMR2 or FLMR1/cMR2, respectively) using a mild purification protocol (11). Both wild-type proteins and their catalytically inactive counterparts were prepared (WT and D600A RAG1, respectively); all preparations show comparable yields (Supplementary Figure S1A).

Long RSS substrates used in this study were amplified by PCR from a plasmid V(D)J recombination substrate, called pJH299 (14) that contains a complementary pair of RSSs (12/23), or its derivatives generated previously (12-only, 12/12 and 23/23) (12). In these plasmids, the RSSs are arranged in the same orientation. PCR-generated substrates were radiolabeled at the 5′-end using a ^{32}P-labeled forward or reverse primer and are ~500 bp long, containing RSS pairs separated by ~250 bp with 100 bp of DNA flanking each RSS at the ends of the fragment (Figure 1A). We felt that including additional flanking DNA might be important for supporting long-range protein–DNA interactions that could help stabilize RAG complexes assembled on these long RSS substrates. We also felt that positioning the RSS pairs in the same orientation held certain advantages over substrates containing the RSSs in opposite orientations because this configuration enables us to examine nicking activity at each RSS on the top strand and also examine protein–DNA contacts at both coding ends and signal ends in the same substrate.

Discrete RAG–RSS complexes assembled on substrates containing RSSs paired *in cis* support 12/23-regulated cleavage *in vitro*

We first determined whether discrete RAG–RSS complexes assembled with PCR-generated RSS substrates could be detected by EMSA. In preliminary experiments, we found that RAG–RSS complexes assembled in the absence of competitor DNA failed to enter the gel matrix, but adding ssDNA after first incubating the RAG proteins with the radiolabeled substrate enabled discrete RAG–RSS complexes to be visualized by EMSA (data not shown). Under these conditions, we find that all three RAG protein preparations assemble discrete

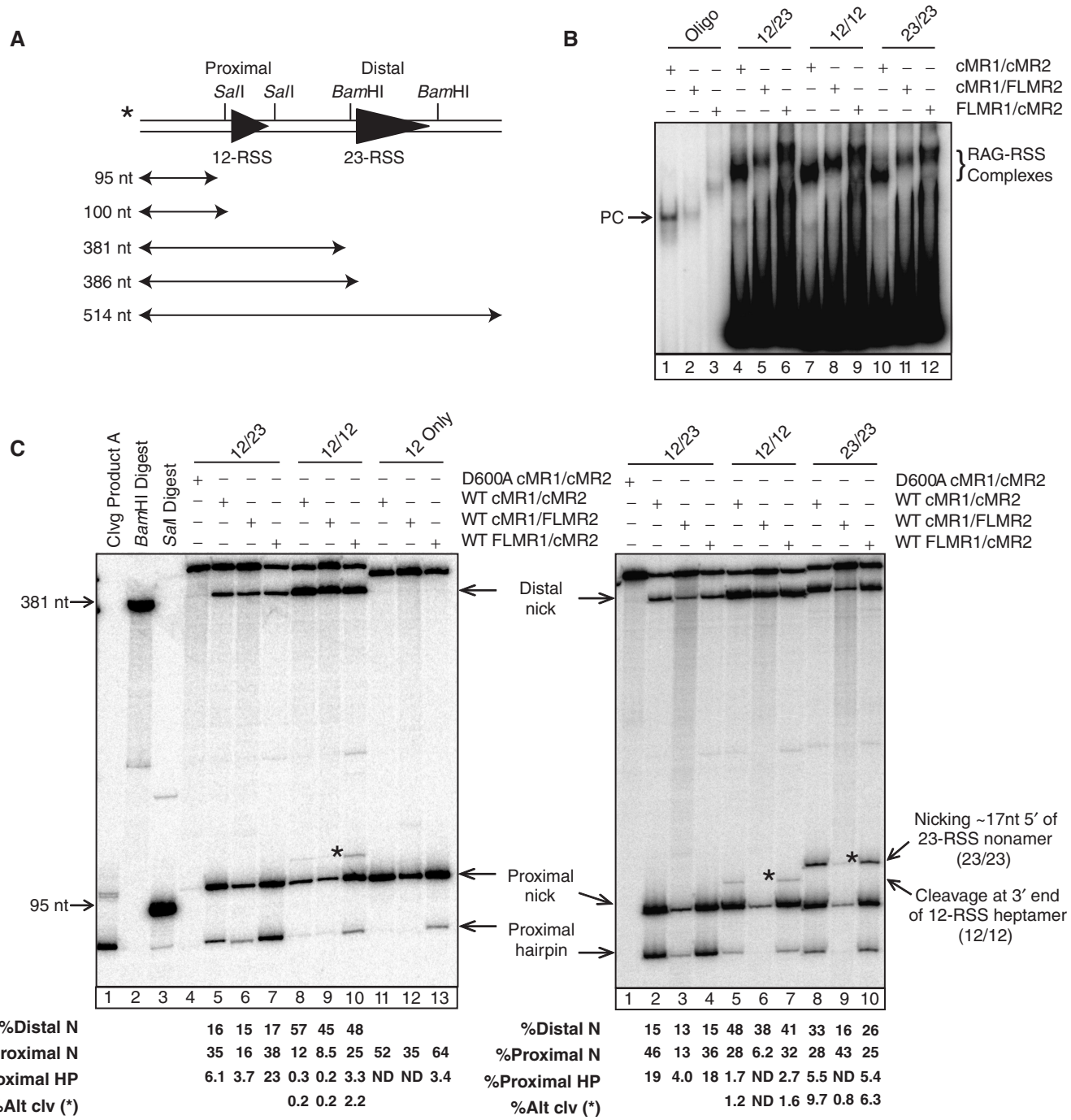

Figure 1. Discrete core and full-length RAG complexes mediate 12/23-regulated cleavage of DNA substrates containing RSS pairs *in cis*. (**A**) The 514 bp 12/23 substrate is radiolabeled on the top strand (asterisk) and contains a 12-RSS (small triangle) and 23-RSS (large triangle) in the same orientation. Sizing markers generated by SalI or BamHI endonuclease cleavage (95 nt or 381 nt, respectively), or RAG-mediated nicking at the 12- or 23-RSS (100 nt or 386 nt, respectively) are shown below the substrate. (**B**) WT RAG preparations were incubated with radiolabeled 12-RSS and cold 23-RSS oligonucleotide (oligo) substrates (see Supplementary Figure S3) or with radiolabeled 12/23, 12/12 or 23/23 substrates in the presence of Mg^{2+} and HMGB1 to promote synapsis, and protein–DNA complexes were fractionated by EMSA. Paired complexes (PC) and RAG–RSS complexes formed with oligo and PCR-generated substrates are indicated at left and right, respectively. (**C**) Discrete RAG–RSS complexes were subjected to an in-gel cleavage assay. Recovered reaction products were fractionated by denaturing gel electrophoresis and the percentage of appropriately sited nicks at the proximal or distal RSS (%proximal N and %distal N), hairpins at the proximal RSS (%proximal HP), and alternative cleavage products (%Alt clv, position denoted by asterisk) are shown below the gel. Sizing markers shown in (A) and a gel-isolated 100 bp hairpin product (see Supplementary Figure S1) were run in parallel and are indicated at left; alternative reaction products are indicated by an asterisk and identified at right (see Supplementary Figure S1).

protein–DNA complexes on all three substrates analyzed; the abundance of these complexes for a given RAG preparation is similar among tested substrates (Figure 1B). Notably, the RAG–RSS complexes formed with cMR1/cMR2, cMR1/FLMR2 and FLMR1/cMR2 exhibit progressively slower mobility by EMSA; all three sets of protein–DNA complexes migrate more slowly than their counterparts assembled on oligonucleotide RSS substrates (Figure 1B). We also notice that the abundance of cMR1/FLMR2 complexes formed on paired RSS substrates is consistently lower than counterparts assembled with cMR1/cMR2 and FLMR1/cMR2 for reasons that remain unclear.

We next used an in-gel cleavage assay to determine whether these RAG-RSS complexes support nicking and hairpin formation on substrate DNA. In these assays, a marker for hairpin formation at the proximal RSS and approximate markers for nicking at the proximal or distal RSS were run in parallel (Figure 1A; Supplementary Figure S1B and C). On the 12/23 substrate, we find that all WT RAG–RSS complexes, but not a D600A cMR1/cMR2 complex, support site-specific nicking and hairpin formation at the proximal 12-RSS, and lower levels of nicking at the distal 23-RSS, with little or no aberrant cleavage activity detected (Figure 1C). Complexes containing core or full-length RAG1 were comparably active in this assay, but complexes containing full-length RAG2 were routinely 2- to 4-fold less active for reasons that remain to be determined. Importantly, we observed that hairpin formation at the proximal 12-RSS is much greater when a 23-RSS is paired *in cis* compared to substrates in which the partner RSS is absent or contains a 12 bp spacer (~7- to ≥20-fold difference, depending on the RAG preparation being compared), suggesting that cleavage is regulated by the 12/23 rule. Hence, a significant fraction of precleavage RAG complexes assembled on the 12/23 substrate must be bound to both RSSs. However, determining the fraction of RAG complexes bound to one or both RSSs is inherently complicated when the RSSs are positioned *in cis*, because if the complement of RAG proteins is the same in both complexes, as has been previously suggested (15,16), the mobilities of the two types of complexes may be indistinguishable by EMSA. Indeed, the data in Figure 1B support this contention. On the other hand, extrapolating the abundance of synaptic complexes based on the level of hairpin formation observed after the in-gel cleavage reaction is also problematic because precleavage synaptic complexes may fail to complete both cleavage steps, either because the reaction is slow *in vitro* or because a fraction of the complexes dissociate during the course of the reaction. Thus, although analysis of the cleavage data would suggests that ~6–18% of RAG–RSS complexes assembled on the 12/23 substrate are synaptic, we consider it likely that this value underestimates the fraction of precleavage synaptic complexes actually present in the complexes visualized by EMSA in Figure 1B.

Interestingly, we note that in complexes assembled on mispaired RSS substrates, alternative reaction products are observed at the proximal RSS. For example, on the 12/12 substrate, cleavage at the 3′-end of the heptamer is detected (Figure 1C, and Supplementary Figure S1C–D). This product is not detected when the distal 12-RSS is removed (12-only, Figure 1C). On the 23/23 substrate, aberrant nicking of the proximal 23-RSS in the 5′-end of the spacer is observed (Figure 1C). The nicked site is located about 17 nt from the 23-RSS nonamer (Supplementary Figure S1D), which is close to the position expected if the sequence is nicked as a 12-RSS. These results, taken together, suggest that how the RAG proteins perceive and cleave a given RSS is influenced by the composition of the partner RSS.

RAG–RSS complexes containing full-length RAG1 exhibit less sensitivity to substrate methylation than complexes containing core RAG1

Having demonstrated that the RAG–RSS complexes visualized by EMSA support cleavage activity at one RSS that is regulated by the composition of its partner, we were interested in investigating the protein–DNA contacts in these complexes. Although we acknowledge the likelihood that the complexes being analyzed are not homogeneous with respect to RSS occupancy, we felt that such an investigation would reveal important new insights into how long DNA substrates are recognized by the RAG proteins, and that by comparing patterns of protein–DNA contacts between substrates containing appropriately paired or mispaired RSSs and between core and full-length RAG proteins, we could identify interactions at one RSS that are influenced by RSS partner composition or by the presence of noncore portions of the RAG proteins. However, because the yield of RAG–RSS complexes assembled with cMR1/FLMR2 was low, we focused on analyzing and comparing protein–DNA contacts in complexes assembled with WT or D600A cMR1/cMR2 and FLMR1/cMR2. Initially, we probed the complexes using methylation interference footprinting because previous studies of RAG-RSS complexes assembled using core RAG proteins and oligonucleotide substrates demonstrate that RAG binding is inhibited by methylation of certain G residues in the consensus RSS, including the 7th residue of the heptamer (top strand) and the 2nd residue of the nonamer (bottom strand) (6). For the studies described here, RAG–RSS complexes were assembled in the presence of Mg^{2+} to support 12/23-regulated synapsis, but incubated on ice to prevent RSS cleavage. Methylation interference footprinting of D600A and WT cMR1/cMR2 complexes assembled on the 12/23 substrate labeled on the top or bottom strand reveals that the same G residues are important for stable complex formation (Figure 2A and Supplementary Figure S2A, 12/23 TOP, compare lanes 3–5 at +7 relative to reference position −10; Figure 2B and Supplementary Figure S2B, 12/23 BOT, compare lanes 3–5 at +32 relative to reference position +40). Methylation interference in the nonamer (bottom strand) is observed in WT and D600A cMR1/cMR2 complexes assembled on all three substrates tested, but interference at the heptamer (top strand) is only observed in complexes assembled on 12/23 and 23/23 substrates. Several A residues in the nonamer motif (+19 to 26) and in the spacer (+10) of the 12-RSS

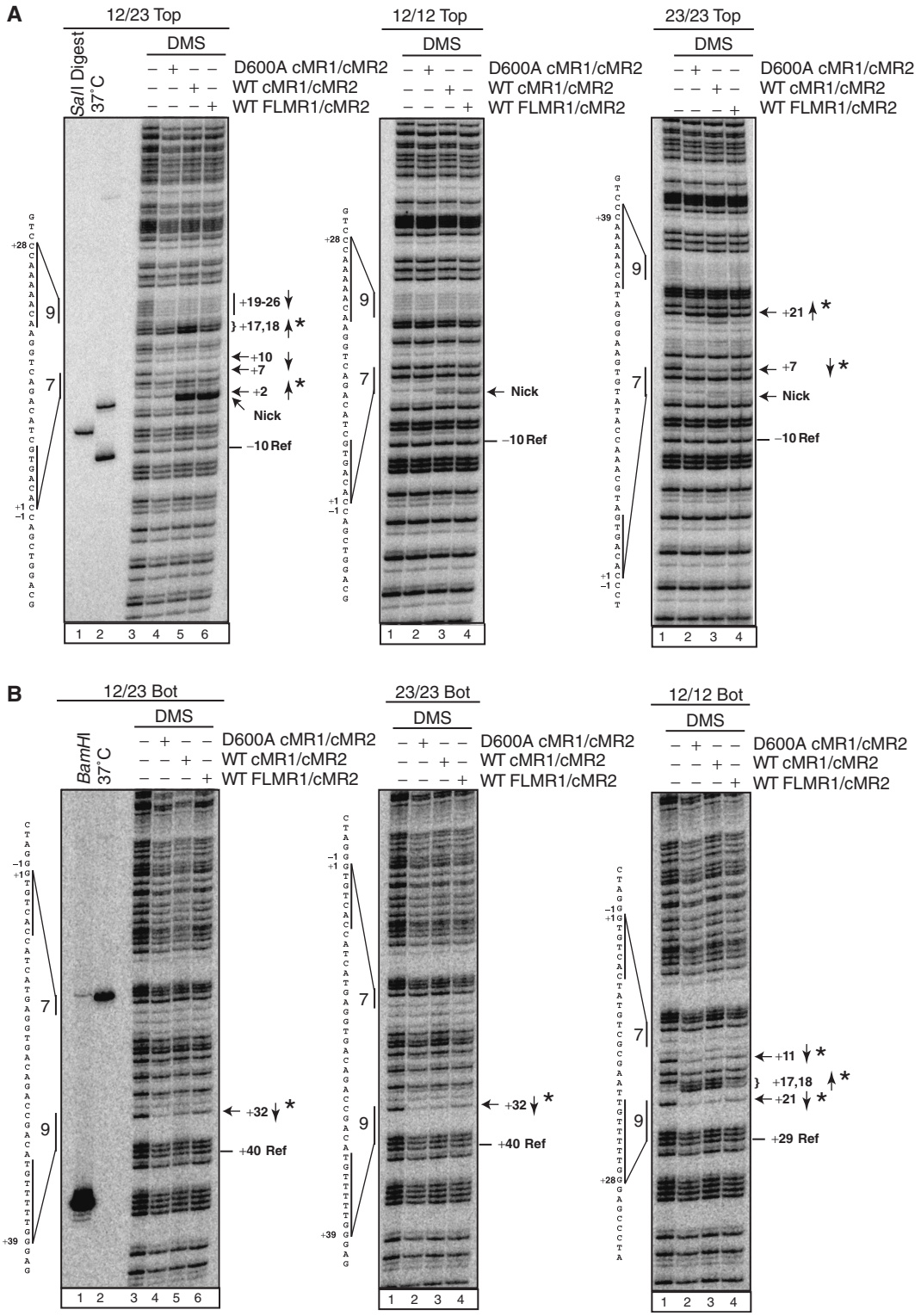

Figure 2. DMS footprinting detects sequence-specific interactions and differential sensitivity to substrate methylation between core and full-length RAG1. (**A** and **B**) DMS-modified 12/23, 12/12 or 23/23 substrates labeled on the top strand (A) or bottom strand (B) were incubated with the indicated RAG preparations, and protein–DNA complexes were fractionated by EMSA. Bound and free DNA was recovered, treated with piperidine and reaction products were fractionated on a sequencing gel. Positions showing increased or decreased representation relative to free DNA and/or an internal reference position on the top or bottom strand (−10 Ref, +40 Ref, or +29 Ref; noted at right) are indicated by up or down arrows, respectively, and numbered according to their location in the RSS shown at left. Positions showing differential representation based on the RAG preparation are indicated by an asterisk. Sizing markers shown in Figure 1 and reaction products from an *in vitro* RAG cleavage of the 12/23 substrate (37°C) were run in parallel. Note that all results for a given set of labeled substrates (e.g. all panel A gels) were obtained from samples prepared and processed in parallel, and are representative of independent experiments.

in the 12/23 substrate (top strand) also show evidence of methylation interference (Figure 2A and Supplementary Figure S2A), but this is not readily detected in the other substrates tested. Certain methylated residues in the 12-RSS heptamer (+2) and spacer (+17, 18) on the top strand exhibit preferential selection in WT cMR1/cMR2 complexes assembled with the 12/23 substrate, but not the 12/12 substrate; selection is tempered in complexes assembled with D600A cMR1/cMR2 (Figure 2A and Supplementary Figure S2A). A similar pattern is observed in the 23-RSS spacer on the top strand (+21). Interestingly, complexes assembled with FLMR1/cMR2 show less (albeit detectable) sensitivity to G residue methylation in the nonamer on the bottom strand and little evidence of preferential selection of modified G residues in the RSS heptamer and spacer on the top strand (Figure 2 and Supplementary Figure S2; 12/23 TOP, compare lanes 5 and 6 at +2, +17, +18; 12/23 BOT, compare lanes 5 and 6 at +32). Similar trends are also observed on the 12/12 and 23/23 substrates (e.g. Figure 2A, 23/23 TOP compare lanes 3 and 4 at +7, +21; Figure 2B, 12/12 BOT, compare lanes 3 and 4 at +11, +17, +18, +21). This outcome cannot be attributed to differential activity of the complexes, because under conditions used for complex assembly, WT cMR1/cMR2 and FLMR1/cMR2 nick the 12/23 substrate at very low, but comparable levels (Figure 2A; 12/23 TOP, compare lanes 5 and 6). It is worth pointing out that nicking activity under these conditions shows clear evidence of regulation by the 12/23 rule, as RAG complexes assembled on 12/12 or 23/23 substrates and analyzed in parallel show little or no evidence of nicking, a result consistent with previous reports (17).

RSS structural distortions and coding flank contacts in RAG–RSS precleavage complexes are modulated by active site status and the N-terminus of RAG1

To complement methylation interference footprinting assays discussed above, RAG–RSS complexes assembled on paired RSS substrates were probed directly by in-gel CuOP footprinting. We find that inactive D600A cMR1/cMR2 complexes assembled on 12/23 and 12/12 substrates exhibit CuOP hypersensitivity in the 12-RSS spacer region on the top strand (Figure 3A and Supplementary Figure S3A; 12/23 and 12/12 TOP, compare lanes 5 and 6 [12/23 TOP] and lanes 2 and 3 [12/12 TOP] at positions +12 and +13). Compared to an internal reference position (−36), the degree of hypersensitivity is much higher in the 12/23 substrate compared with the 12/12 substrate, which suggests that the structural distortion imparting CuOP hypersensitivity at this location is regulated by the 12/23 rule. Interestingly, spacer hypersensitivity is markedly lower (but still detectable) in WT cMR1/cMR2 and FLMR1/cMR2 complexes at position +13. Nonamer hypersensitivity at position +23 in the 12-RSS top strand is also clearly evident in WT FLMR1/cMR2 complexes assembled on 12/23 and 12/12 substrates, but is less discernable in D600A or WT cMR1/cMR2 complexes (Figure 3A and Supplementary Figure S3A, 12/23 and 12/12 TOP, compare lanes 6–8 [12/23 TOP] and lanes 3–5 [12/12 TOP]). No sites of top strand hypersensitivity are evident in any RAG complexes assembled on the 23/23 substrate.

In the coding flank, we detect modest protection at the −14 position in all RAG complexes assembled on 12/23 substrates, but protection is less evident on 12/12 or 23/23 substrates (Figure 3A and Supplementary Figure S3A). Curiously, however, in all RAG complexes analyzed, we observe a distinctive shift of CuOP reactive sites in the coding flank on the top strand relative to free DNA, spanning a region from about −18 to −31 (Figure 3A and Supplementary Figure S3A). The degree of shifting differs slightly based on the composition of the RAG proteins, but is consistent among the substrates tested. No comparable shift is observed on the bottom strand (Figure 3B), although substrate length may preclude observation of this subtle effect. However, patterns of bottom strand protection and hypersensitivity are detected in these complexes: protection is observed in the spacer region (Figure 3B and Supplementary Figure S3B; +11, +18, +19, +29, +30 in the 23 RSS; +8 and +18 in the 12-RSS) and hypersensitivity is observed in the heptamer region (Figure 3B and Supplementary Figure S3B; +2, +3 and +6 in the 23-RSS, and +2 and +3 in the 12-RSS) and the spacer (Figure 3B and Supplementary Figure S3B; +13 and +14 in both the 12- and 23-RSS).

RAG-mediated RSS cleavage is accompanied by 12/23-regulated distortion of the intersignal sequence

We were interested in determining how core and full-length RAG proteins interact with DNA in the postcleavage RAG–RSS complex. To address this issue, precleavage D600A and WT cMR1/cMR2 and FLMR1/cMR2 complexes assembled on 12/23, 12/12 and 23/23 substrates were subjected to in-gel cleavage after EMSA, and then the resulting RAG postcleavage RAG-RSS complexes were immediately probed by in-gel CuOP footprinting (Figure 4 and Supplementary Figure S4). Unexpectedly, D600A cMR1/cMR2 complexes showed little evidence of RSS spacer hypersensitivity in the 12-RSS and less protection of the 23-RSS compared to the same complexes probed immediately after EMSA (e.g. compare Figure 3, lane 6 to Figure 4, lane 5 in 12/23 TOP at positions +12, +13; and compare Figures 3 and 4, lane 5, in 12/23 BOT at positions +29, +30). We speculated that complexes formed with D600A cMR1/cMR2 are unstable to prolonged incubation at 37°C. Consistent with this hypothesis, we find that D600A cMR1/cMR1 complexes incubated at 37°C for 1 h and fractionated by EMSA are present in much lower yield than their counterparts assembled following the standard procedure (Supplementary Figure S5, compare lanes 2–4 to lanes 5–7).

As expected from data in Figure 1, DNA recovered from WT but not D600A RAG–RSS complexes shows substantial substrate nicking on the top strand (Figure 4A) and cleavage of the bottom strand (Figure 4B), precluding analysis of protection beyond these breakpoints. Interestingly, postcleavage complexes

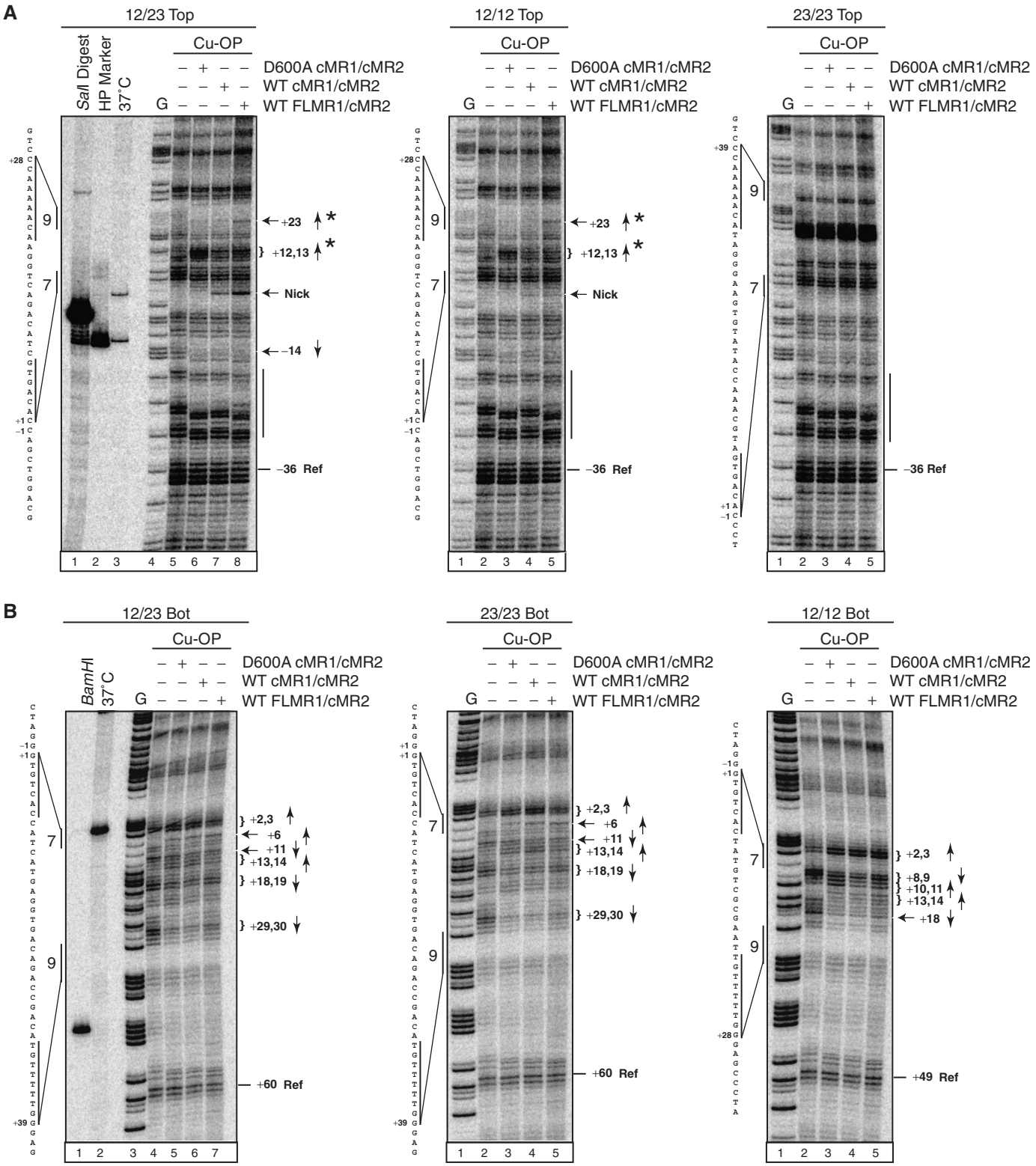

Figure 3. CuOP footprinting of precleavage RAG–RSS complexes reveals evidence for structural distortion of the RSS and coding flank. (**A** and **B**) Unmodified 12/23, 12/12 or 23/23 substrates labeled on the top strand (A) or bottom strand (B) were incubated with the indicated RAG preparations. Protein–DNA complexes were separated by EMSA, subjected to in-gel CuOP footprinting, and reaction products recovered from bound and free DNA were fractionated on a sequencing gel along with sizing markers described in Figures 1 and 2 and a 'G' ladder. Results are displayed as in Figure 2; note the shift in the position of reaction products in the coding flank, indicated by a vertical line at right. Internal reference positions on the top and bottom strand are also shown (−36 and +60, respectively).

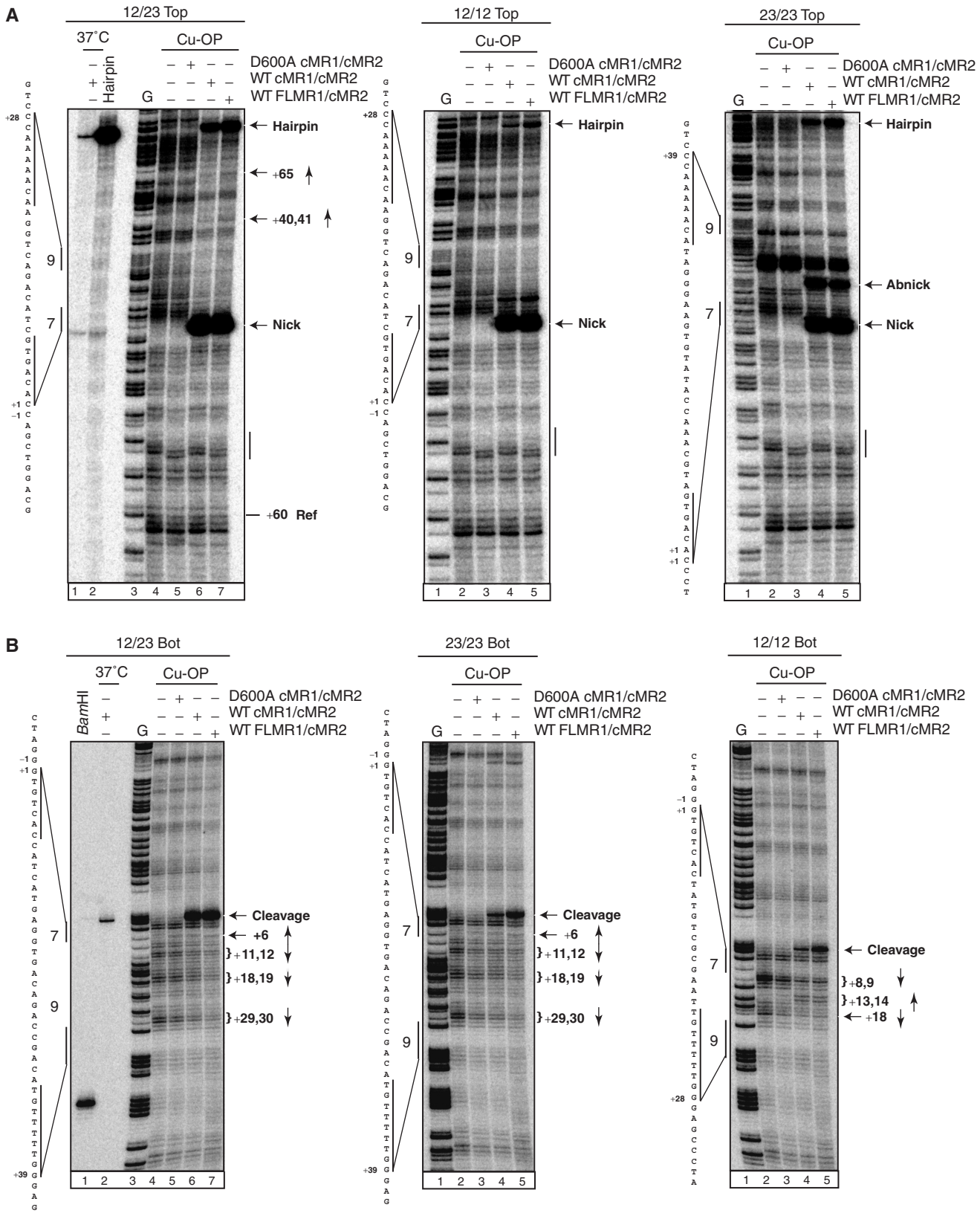

Figure 4. CuOP footprinting of postcleavage RAG–RSS complexes show evidence of temperature-dependent and 12/23-regulated structural distortion of the intersignal sequence. (**A** and **B**) RAG–RSS complexes assembled and fractionated by EMSA were subjected to in-gel cleavage as in Figure 1 and the resulting postcleavage RAG–RSS complexes were immediately probed by in-gel CuOP footprinting as in Figure 3A and B. Subsequent sample processing and data display are as described in Figure 3, except that sequencing gels contained 40% formamide to denature the hairpin product. Note hypersensitivity at positions +40, +41 and +65 in the top strand (arrows, at right) in WT RAG complexes assembled on 12/23 substrates, but not 12/12 and 23/23 substrates.

probed by CuOP show a more restricted region of shifted CuOP sensitivity on the top strand (−24 to −26) compared to precleavage complexes; protection of the −14 position is also not clearly evident (Figure 4A and Supplementary Figure S4A). One other notable difference is that in 12/23 WT RAG postcleavage complexes, particularly those containing FLMR1, CuOP hypersensitivity is observed well into the intersignal sequence, distal to the nonamer on the top strand (Figure 4A; 12/23 TOP, lanes 6 and 7; positions +40, +41 and +65). Hypersensitivity at these sites has three important features: first, it is regulated by the 12/23 rule because hypersensitivity is not detected in comparable complexes assembled on 12/12 and 23/23 substrates; second, it occurs before nicking, because hypersensitivity at these sites could not be visualized on nicked substrates (since they are 3′ of the heptamer); and third, it is 'activated' by incubation at 37°C, since hypersensitivity at these sites is not observed in RAG–RSS complexes assembled at 4°C and probed by CuOP immediately after EMSA (compare Figure 3A, lanes 7 and 8, to Figure 4A, lanes 6 and7 in 12/23 TOP). Although we considered the possibility that these sites represent targets of aberrant nicking by the RAG complex, three lines of evidence argue against this scenario: first, nicking at these sites is not detected in the in-gel cleavage assays shown in Figure 1C; second, they do not occur at a sequence that contains an obvious target of RAG-mediated nicking (e.g. a 5′-CAC sequence); and third, cleavage at these sites is not evident in comparable complexes containing K980A RAG1, which supports 12-RSS nicking at levels comparable to WT RAG1 (see below).

On the bottom strand, the CuOP reactivity profile for a given WT RAG–RSS postcleavage complex is quite similar to its counterpart precleavage complex, except that hypersensitivity at the 5′-end of the 12- and 23-RSS heptamer (+2, +3) and the 23-RSS spacer (+13, +14) is not as evident in the postcleavage complex (compare these positions in Figures 3B and 4B), suggesting CuOP hypersensitivity at these sites is relieved by substrate nicking.

Hairpin retention in RAG–RSS postcleavage complexes is promoted by the amino-terminus of RAG1

Previous studies suggest that after RSS synapsis and cleavage, the RAG proteins retain the coding ends, at least transiently, in a postcleavage synaptic complex or 'cleaved signal complex' (CSC) (18–20). Since those studies relied on core RAG1 proteins for experimental analysis, we wondered whether full-length RAG1 might be more efficient than core RAG1 in retaining coding ends in a postcleavage RAG–RSS complex. To test this possibility, RAG-RSS complexes were assembled on ice on a 12/23 substrate using WT cMR1/cMR2 and WT and D600A FLMR1/cMR2, and incubated for 10 min at 37°C to initiate cleavage. An aliquot of this reaction was retained as 'input DNA' and the remaining sample was fractionated by EMSA. Bound DNA was isolated and analyzed for the ratio of nicks and hairpin products compared to input DNA. If nicked and hairpin DNA are retained equivalently in the postcleavage complex, this ratio should remain the same between input DNA and DNA isolated from the CSC. As an additional control, we tested a FLMR1/cMR2 preparation containing a K980A RAG1 mutation reported previously to impair coding joint formation (21). We included this control based on a previous study suggesting that joining-deficient RAG1 mutants may exhibit less stable interactions with the RSS coding flank (22). In testing the K980A FLMR1/cMR2 preparation in our *in vitro* RSS binding and cleavage assays, we found that, consistent with previous results (21), K980A FLMR1/cMR2 binds and cleaves an oligonucleotide substrate with a 'good' coding flank *in vitro* under conditions favoring synapsis at levels comparable to WT FLMR1/cMR2 (Supplementary Figure S6). When we analyzed the *in vitro* cleavage activity of the 12-RSS derived from pJH299, which contains a coding flank that is less permissive for cleavage (23,24), we find that under similar conditions, K980A FLMR1/cMR2 nicks the oligonucleotide substrate as efficiently as WT FLMR1/cMR2, but exhibits a very modest (∼1.5-fold) defect in hairpin formation (Figure 5A and B). Surprisingly, the level of impairment becomes much more severe (∼8- to 10-fold) when the length of DNA flanking the 12-RSS is increased in the context of the 12/23 substrate (Figure 5C and D). Despite this defect, we continued to include the K980A FLMR1/cMR2 preparation in our further experiments.

When the various CSCs (Figure 6A) were analyzed for the presence of DNA hairpin products, complexes containing FLMR1/cMR2 were found to retain about 7- to 10-fold more hairpin than complexes containing cMR1/cMR2, despite the two RAG preparations supporting comparable levels of hairpin formation during the *in vitro* cleavage reaction (Figure 6B, compare lanes 4 and 5 to lanes 9 and 10). Thus, these data suggest that full-length RAG1 facilitates retention of coding ends in the CSC. We note, however, that the retained hairpin product represents only a small fraction of total hairpin DNA produced during the *in vitro* cleavage reaction, most of which has dissociated from the CSC (Figure 6A). Attempts to analyze hairpin retention in CSCs assembled with K980A FLMR1/cMR2 were complicated by the low level of hairpin formation supported by this RAG1 mutant. Although nick-to-hairpin ratios between input DNA and DNA recovered from WT and K980A FLMR1/cMR2 complexes appears to be similar, the data do not allow us to unambiguously determine whether the K980A mutation specifically disrupts the ability of the RAG1 N-terminus to stabilize hairpin retention in the CSC.

A K980A RAG1 mutant exhibits defects in promoting RSS structural alterations correlated to productive hairpin formation

Since K980A FLMR1/cMR2 does not exhibit an apparent deficiency in maintaining coding ends in the CSC, but rather fails to support efficient hairpin formation, we wondered whether this defect might be manifested by an altered pattern of protein–DNA contacts in CSCs assembled with K980A FLMR1/cMR2 compared to WT

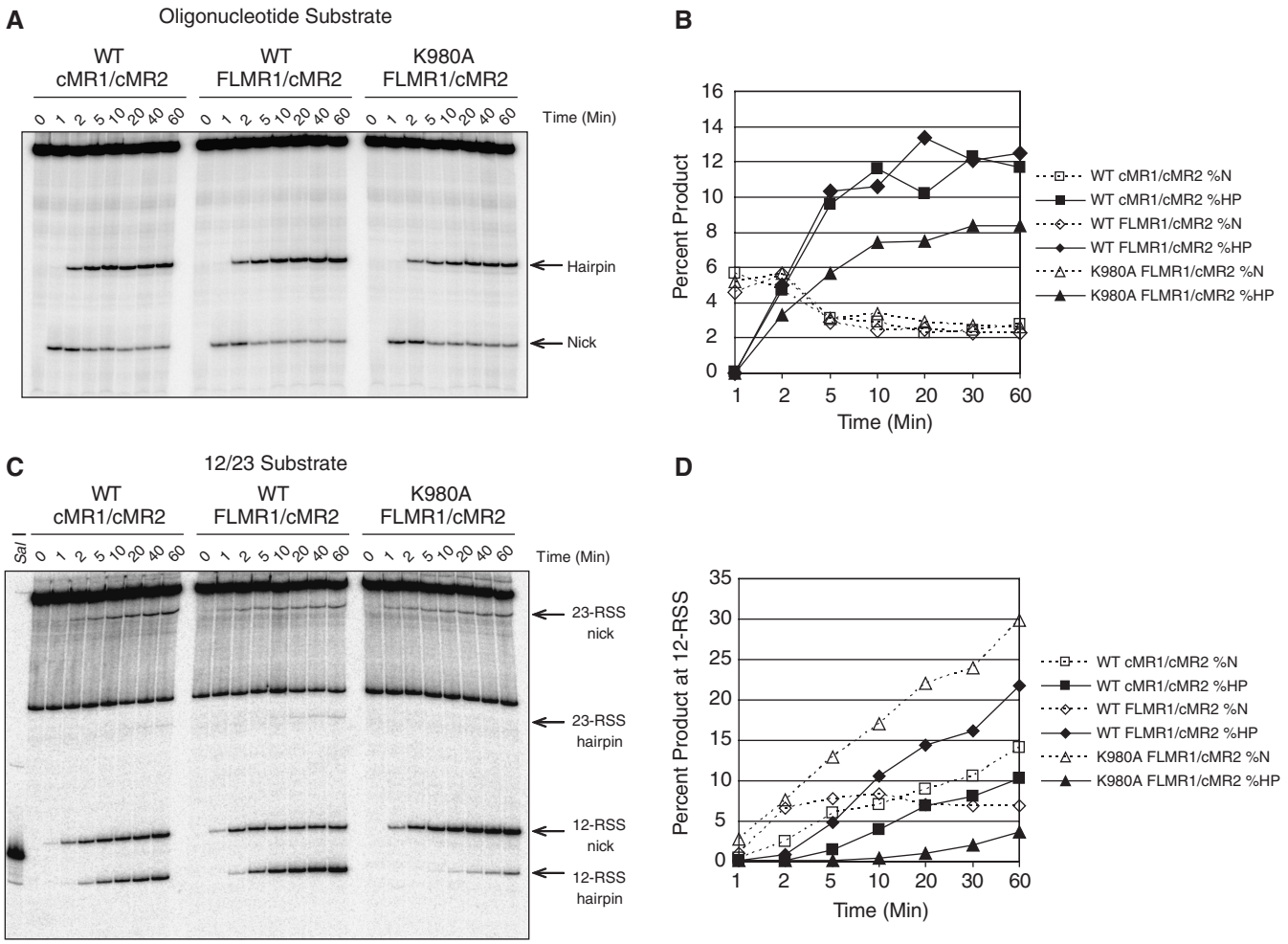

Figure 5. K980A RAG1 exhibits a defect in hairpin formation dependent on RSS substrate length. (**A–D**) WT cMR1/cMR2, WT FLMR1/cMR2 and K980A FLMR1/cMR2 were incubated with HMGB1 and either the radiolabeled oligonucleotide 12RSSJH299 substrate along with cold partner 23-RSS (A and B), or the radiolabeled 12/23 substrate (C and D). (A and C) Samples removed from preparative *in vitro* cleavage reactions at the indicated times were analyzed on a sequence gel. (B and D) Nicked (N) and hairpin (HP) products visualized in (A) or (C), respectively, were quantified for each RAG preparation and plotted as a percentage of the total reaction product at each time point. Note that these reactions were performed simultaneously using RAG proteins expressed and purified in parallel, and are representative of independent experiments.

FLMR1/cMR2. To test this possibility, we compared the CuOP footprints of postcleavage RAG–RSS complexes assembled on all three paired RSS substrates with WT, D600A, or K980A FLMR1/cMR2 (Figure 7 and Supplementary Figure S7). We find that the CuOP footprints of RAG–RSS complexes containing D600A FLMR1/cMR2 resemble those containing D600A cMR1/cMR2. WT and K980A FLMR1/cMR2 complexes show a similar pattern of coding end contacts; better resolution of reaction products near the cleavage site also reveals evidence of hypersensitivity at the −3 and −4 positions (Figure 7 and Supplementary Figure S7). Strikingly, unlike WT FLMR1/cMR2 complexes, 12/23-regulated hypersensitivity in the intersignal sequence (+40, +41 and +65) is not detected in RAG–RSS complexes assembled with K980A FLMR1/cMR2 (Figure 7, 12/23 TOP, compare lanes 6 and 7). Because both WT and K980A FLMR1/cMR2 complexes support comparable levels of nicking, hypersensitivity at these sites must not be required for catalyzing hydrolysis at the heptamer-coding junction. The implications of these findings are discussed further below.

DISCUSSION

Previous studies suggest that the requirements for RAG-mediated synapsis and cleavage of oligonucleotide substrates commonly used in assays of RAG activity are not necessarily mirrored on longer, more physiological substrates containing RSSs paired *in cis* (7). To partly address this issue, we describe here the first direct detection and characterization of discrete RAG–RSS complexes containing various combinations of core and full-length RAG1 and RAG2 assembled on long PCR-generated substrates containing RSSs positioned *in cis*. These results significantly extend previous work characterizing RAG–RSS complexes assembled on oligonucleotide substrates and reveal many novel features of RAG–RSS complex assembly and activity not observed using oligonucleotide substrates. Moreover, since the PCR-generated

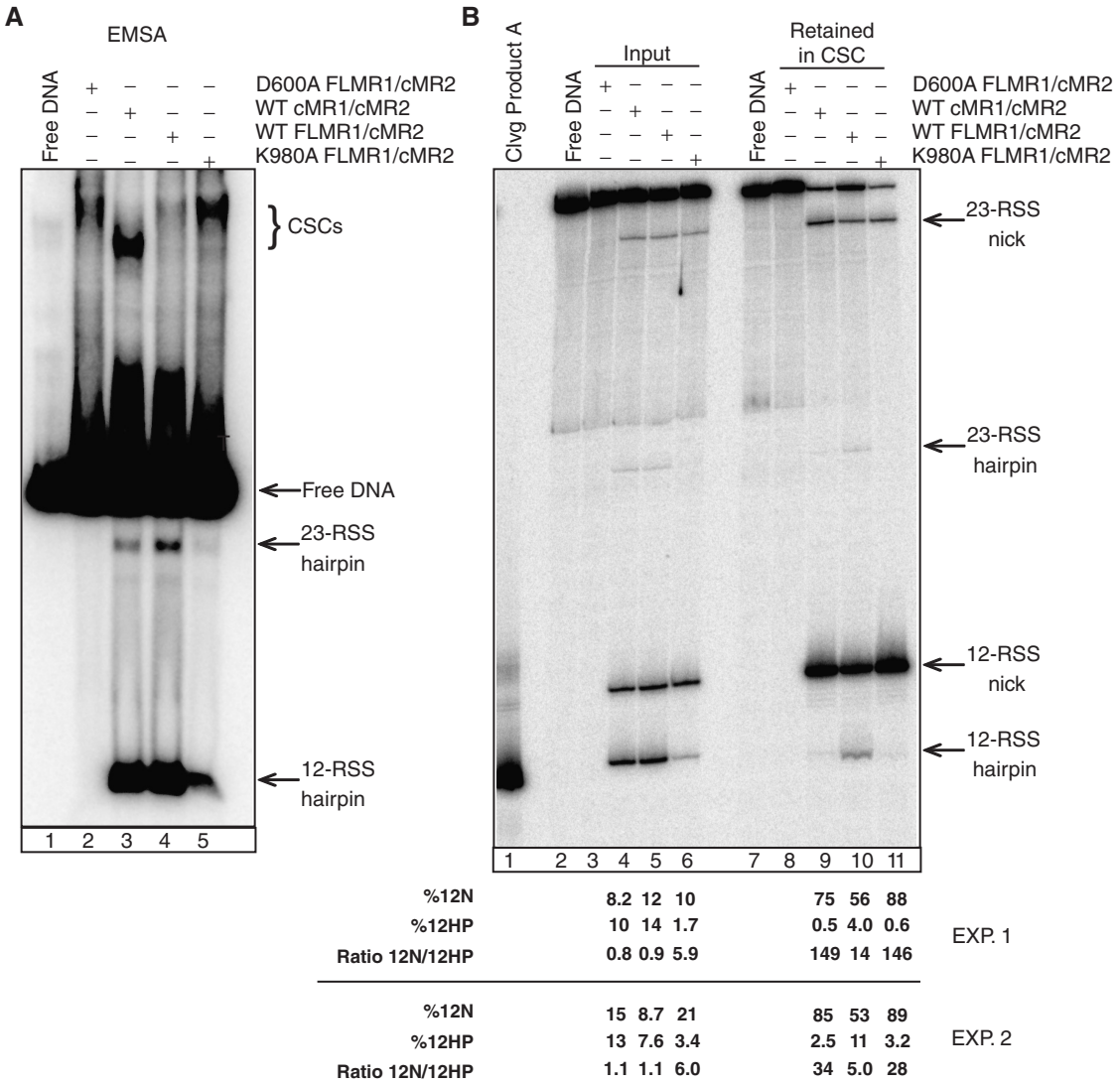

Figure 6. Full-length RAG1 promotes hairpin retention in a postcleavage RAG–RSS complex. (**A**) The 12/23 substrate was incubated with RAG preparations indicated above the gel in an *in vitro* cleavage reaction at 37°C for 10 min. Bound, free and cleaved DNA was separated by EMSA. Note that the K980A FLMR1/cMR2 complex binds the 12/23 substrate, but cleaves it poorly. (**B**) A sample of the *in vitro* cleavage reaction was fractionated on a sequencing gel (Input) in parallel with DNA recovered and normalized from the discrete CSCs visualized by EMSA in (A). For each sample, the percentage of 12-RSS nick and hairpin products (%12N and %12HP, respectively), and the ratio of the two products (ratio 12N/12HP), are quantified below the gel. Results from two independent experiments (EXP. 1 and EXP. 2) are shown. Note the differences in 'Input' and 'CSC' 12N/12HP ratios between RAG complexes containing WT core and full-length RAG1.

substrates used in our experiments have been amplified from a plasmid used to assay V(D)J recombination in cell culture, our results are meaningful for understanding how the RAG proteins mediate this process *in vivo*. Although beyond the scope of the present work, future comparative studies could be envisioned to characterize how RSS recognition by the RAG proteins is influenced by variations in the length or composition of the coding flank, the orientation or sequence of the RSS (e.g. endogenous RSS pairs), or intersignal distance.

Evidence for plasticity in RSS utilization by the RAG complex

We previously demonstrated that a derivative of the plasmid V(D)J recombination substrate pJH299 containing two 12-RSSs (12/12) in the same orientation *in cis* supports aberrant recombination in cell culture between the 3′-end of one 12-RSS heptamer and the 5′-end of the other 12-RSS heptamer (the 'proximal' and the 'distal' RSS, respectively, in the 12/12 substrate shown in Supplementary Figure S1) (12). As a result, the intervening sequence is deleted, rather than being inverted as would be expected when the RSSs are complementary (12/23). We speculated that this might be due to the presence of a cryptic nonamer sequence in the coding flank that enables the 'proximal' 12-RSS to be 'converted' to a 23-RSS in the opposite orientation. Data presented here showing that 12/12 substrates amplified from this plasmid, but not 12-only or 12/23 substrates, are subjected to aberrant cleavage at the 3′-end of the proximal 12-RSS

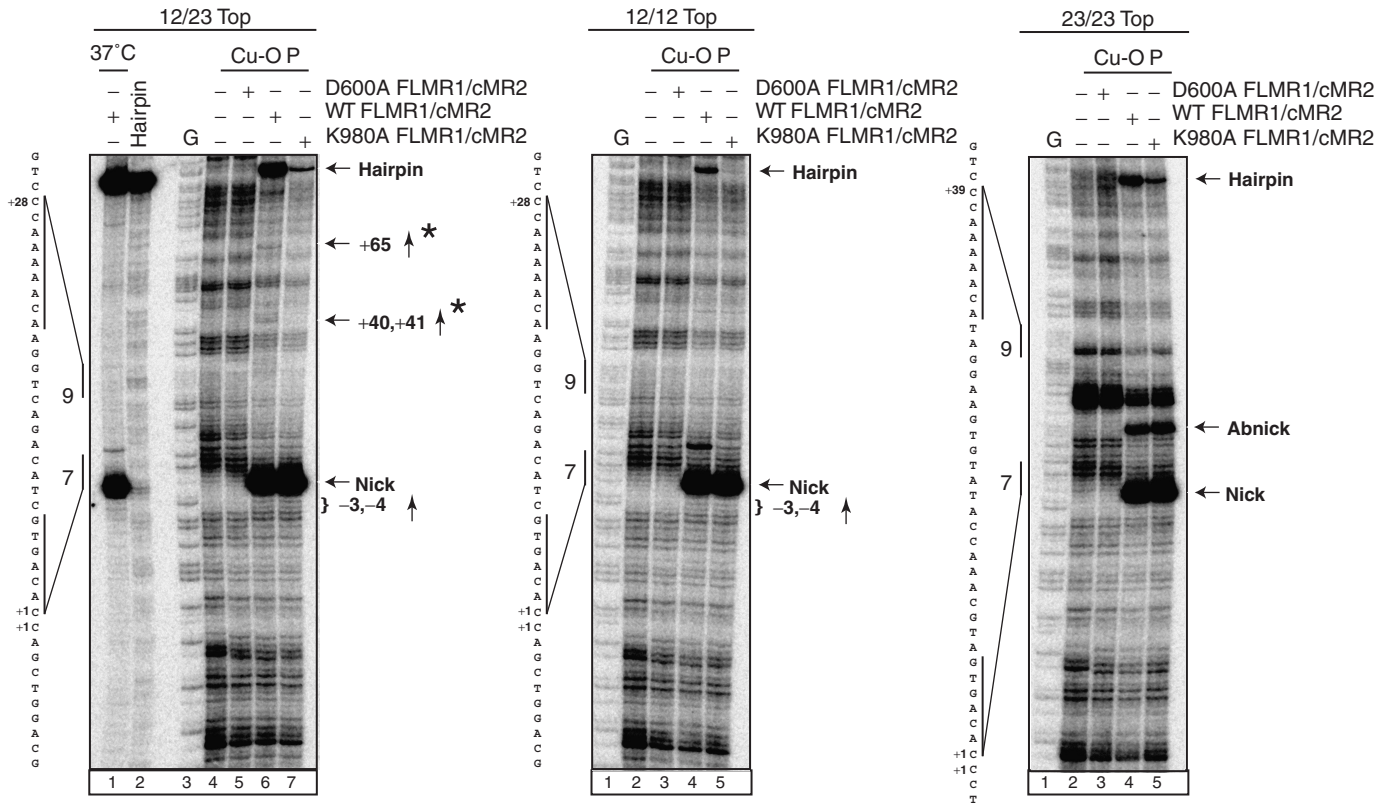

Figure 7. K980A RAG1 fails to support temperature-dependent structural distortions in a RAG–RSS complex. D600A, WT and K980A FLMR1/cMR2 postcleavage RAG–RSS complexes were prepared and subjected to in-gel CuOP footprinting as in Figure 4A. Data are presented as in Figure 4A. Note the absence of CuOP hypersensitivity at positions +40, +41 and +65 in complexes containing K980A FLMR1 compared to WT FLMR1 (asterisk).

heptamer provides additional experimental support for this idea, and further suggests that the way the RAG complex perceives a given RSS is somewhat flexible and dependent on the composition of the available partner RSS. Such plasticity would generally go undetected using oligonucleotide substrates, because the length of coding flank sequence present in such substrates is usually quite limited.

Comparison of RAG-mediated RSS recognition in oligonucleotide and long PCR-generated substrates

RAG–RSS complexes assembled on oligonucleotide substrates have been subjected to extensive DNA footprint analysis using various chemical and enzymatic probes (6). To facilitate analysis and comparison to earlier data, footprinting results from this study have been summarized in Figure 8 (12-RSS) and Supplementary Figure S8 (23-RSS). Methylation interference footprinting has been conducted here and in previous studies of RAG–RSS–HMGB1 complexes (25), enabling the data to be directly compared. This is particularly instructive for the 12-RSS, because its sequence is identical in both the 12/23 substrate used here and in an oligonucleotide substrate used previously (25). Inspection of the methylation interference patterns between the two substrates reveals similarities and subtle differences. For example, while methylation interference is detected in both substrates on the top strand in the heptamer (+7), methylation in the spacer at positions +17 and +18 is preferentially selected by the core RAG proteins in the 12/23 substrate, whereas methylation of these positions in an oligonucleotide 12-RSS substrate either does not affect RAG binding (+17) or interferes with binding (+18). That methylation interference is also detected on the bottom strand in the nonamer (+21) in both DNA substrates suggests that RAG-mediated recognition of the heptamer and nonamer elements involves well-defined base-specific contacts, whereas RAG interactions with the major groove in the 3′-end of the spacer region are more flexible and context-dependent, as one might expect from the lack of sequence conservation in this region (26). In contrast, RAG contacts with the 5′-end of the spacer appear to more consistently involve base-specific interactions, as methylation interference is detected on the bottom strand at position +11 (G residue), as previously observed using an oligonucleotide substrate (25), and on the top strand at position +10 (A residue). Both of these residues exhibit much higher levels of sequence conservation (62% and 58%, respectively) than residues at the 3′-end of the spacer (26).

Shifts in coding sequence sensitivity to CuOP in RAG–RSS complexes assembled here on paired RSS substrates were not previously detected in RAG complexes bound to oligonucleotide RSS substrates (27,28), possibly due to the

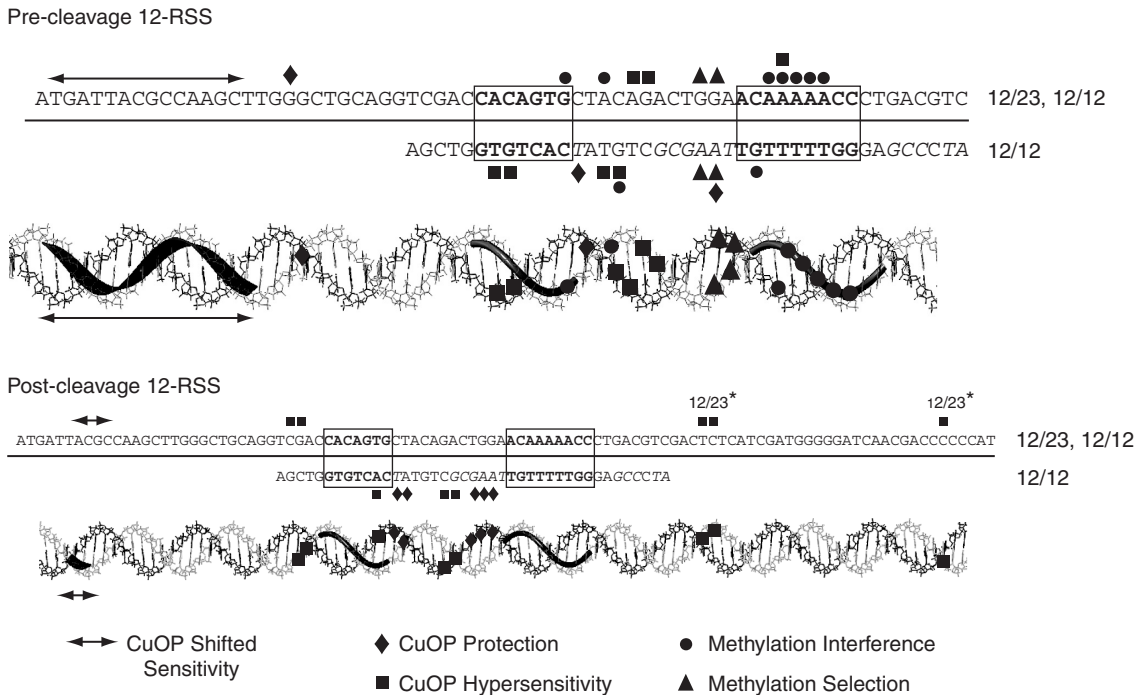

Figure 8. Summary of 12-RSS contacts in pre- and post-cleavage RAG–RSS complexes. Sites of 12-RSS methylation interference (filled circles) or selection (filled triangles), and protection (filled diamonds), hypersensitivity (filled squares) or shifted reactivity (line with double arrows) toward CuOP-mediated cleavage in precleavage (top) and postcleavage (bottom) RAG–RSS complexes are indicated above or below the corresponding sites in the top or bottom strand sequence, respectively, or shown on the corresponding residue in a 12-RSS modeled as a B-form DNA duplex using Sybyl 7.1 (Tripos Inc., St Louis, MO, USA). In the linear diagram, the top strand sequence is from the proximal 12-RSS in the 12/23 and 12/12 substrates, whereas the bottom strand sequence is from the distal 12-RSS in the 12/12 substrate, which differs slightly from the top strand complementary sequence (italics). The model was generated using the top strand DNA sequence (light gray) and its complement (dark gray). Positions of shifted CuOP reactivity are shown by a shaded ribbon. Heptamer and nonamer sequences are shown in bold font and boxed, or indicated by tubes in the 12-RSS helical model. CuOP hypersensitivity strongly regulated by the 12/23 rule is indicated (12/23*).

short-coding DNA sequence (16 bp) in the substrates used in those studies. CuOP is well known for its sensitivity for detecting subtle changes in minor groove configuration in DNA bound by proteins (29). We speculate that RAG interactions near the cleavage site (possibly near position −14) propagate a structural distortion in the minor groove along the coding sequence, shifting the CuOP reactivity pattern toward the 5′-end of the DNA substrate. Substrate nicking and/or cleavage appears to partly relieve this distortion and restore the minor groove conformation, because postcleavage RAG–RSS complexes exhibit a more restricted shift in CuOP reactivity compared to precleavage RAG–RSS complexes (Figure 8). This outcome appears correlated with the loss of protection at position −14, suggesting that the RAG proteins loosen their hold on the coding flank after cleavage.

The RAG1 N-terminus stabilizes structural distortions in RAG–RSS complexes and promotes coding end retention in the postcleavage complex

DNA footprinting experiments show that FLMR1/cMR2 exhibits less preference for binding intact paired RSS substrates methylated in the spacer region than cMR1/cMR2 (Figure 2), and induces greater CuOP sensitivity of the nonamer and intersignal sequence (Figures 3 and 4). These data suggest that, compared to core RAG1, full-length RAG1 is more efficient at inducing and/or stabilizing structural distortions in the RSS. In addition, we find that CSCs containing full-length RAG1 retain more coding ends than those containing core RAG1. However, despite this increase, the association of the hairpin with the CSC remains quite labile *in vitro*, as the vast majority of coding ends dissociate after cleavage. It is possible that RAG association with other cellular DNA binding factors, such as modified histones (30,31) or 53BP1 (32), may further contribute to coding end retention in the CSC. Nevertheless, these results, taken together, may partly explain why full-length RAG1 supports more efficient cellular V(D)J recombination than core RAG1 (33,34).

A K980A RAG1 mutant exhibits a DNA length-dependent defect in hairpin formation associated with loss of intersignal 'activation-induced' CuOP hypersensitivity

Because the N-terminus of RAG1 promotes coding end retention in the postcleavage complex, we wondered whether RAG1 mutations associated with defects in coding joint formation impair coding end retention. We tested a K980A RAG1 mutant based on a previous study showing that this mutant exhibits a defect in signal and coding joint formation, with the latter more impaired than the former (21). Interestingly, this mutant was also shown to exhibit a mild cleavage defect on plasmid V(D)J recombination substrates in cell culture ($\leq$10-fold relative to WT

RAG1), but the basis for this defect remained unclear because purified WT and K980A RAG1 (as the core protein) supported *in vitro* cleavage of oligonucleotide substrates with almost comparable efficiency (21). We have confirmed this finding (Supplementary Figure S6), but show here that the K980A RAG1 mutant exhibits a striking DNA length-dependent defect in hairpin formation in an *in vitro* cleavage assay using the 12/23 substrate (Figure 5). The ~8- to 10-fold reduction in hairpin formation observed in our work is comparable to the ≤10-fold decrease in plasmid cleavage activity detected previously by ligation-mediated PCR (21). These data clearly illustrate the differential requirements for RAG-mediated cleavage of RSS substrates of varying length, and specifically argue that coding flank length poses a constraint on hairpin formation that must be overcome by the RAG proteins. The C-terminus of RAG1 that includes Lys-980 has been shown previously to bind nonspecifically to DNA and interact with the coding flank of the RSS (35,36). The data presented here provides a plausible role for this interaction: namely, to orient a physiological coding flank into a position favorable for hairpin formation. This conformational change may be temperature-dependent and accompanied by structural distortions in the intersignal sequence, as CuOP hypersensitivity at positions +40, +41 and +65 is evident in 'activated' RAG complexes containing full-length WT RAG1, but not K980A RAG1 (Figure 7). It is noteworthy that these hypersensitive sites appear on nearly opposite sides of the DNA duplex (Figure 8). It is tempting to speculate that hypersensitivity is caused by RAG-mediated bending of the intersignal sequence to facilitate synapsis and align the coding ends into a position favorable for hairpin formation.

Coding joints generated by K980A RAG1 are characterized by a high frequency of excessive deletions, short sequence homologies and abnormally long P nucleotide insertions, leading to speculation that the K980A mutation either causes coding ends to be prematurely released from the CSC, or impairs recruitment of NHEJ factors to the coding ends, rendering them more susceptible to extensive processing than normal and/or enabling them to be repaired by an alternative DNA end-joining pathway (21). The first scenario seems unlikely because WT and K980A RAG1 retain a comparable fraction of coding ends in the CSC (after normalizing for differences in hairpin formation; Figure 6). If the second scenario is true, it is unlikely that Ku recruitment is specifically impaired, since WT and K980A FLMR1/cMR2 similarly support formation of a higher order precleavage RAG–RSS complex containing Ku70/Ku80 (11) (Supplementary Figure S6B), although we cannot rule out the possibility that Ku transfer to the coding ends is less efficient with K980A RAG1. A plausible alternative, suggested previously based on the high frequency of long P nucleotides observed in coding joints formed with K980A RAG1 (21), is that K980A RAG1 fails to efficiently recruit the Artemis/DNA–PKcs complex to the coding ends, rendering the ends accessible to hairpin opening by other mechanisms.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank Dr Leo Kinarsky for assistance in RSS model preparation.

FUNDING

National Institutes of Health (R01AI055599 and 2R56AI055599-06A1 to P.C.S.); the Research Facilities Improvement Program of the National Institutes of Health National Center for Research Resources (C06 RR17417-01); UNMC Bioinformatics Core Facility National Institutes Health (P30 CA36727) in part; the Nebraska Research Initiative. Funding for open access charge: NIH 2R56AI055599-06A1.

Conflict of interest statement. None declared.

REFERENCES

1. Bassing,C.H., Swat,W. and Alt,F.W. (2002) The mechanism and regulation of chromosomal V(D)J recombination. *Cell*, **109(Suppl)**, S45–S55.
2. Fugmann,S.D., Lee,A.I., Shockett,P.E., Villey,I.J. and Schatz,D.G. (2000) The RAG proteins and V(D)J recombination: complexes, ends, and transposition. *Annu. Rev. Immunol.*, **18**, 495–527.
3. Gellert,M. (2002) V(D)J recombination: RAG proteins, repair factors, and regulation. *Annu. Rev. Biochem.*, **71**, 101–132.
4. Rooney,S., Chaudhuri,J. and Alt,F.W. (2004) The role of the non-homologous end-joining pathway in lymphocyte development. *Immunol. Rev.*, **200**, 115–131.
5. Lieber,M.R., Lu,H., Gu,J. and Schwarz,K. (2008) Flexibility in the order of action and in the enzymology of the nuclease, polymerases, and ligase of vertebrate non-homologous DNA end joining: relevance to cancer, aging, and the immune system. *Cell Res.*, **18**, 125–133.
6. Swanson,P.C. (2004) The bounty of RAGs: recombination signal complexes and reaction outcomes. *Immunol. Rev.*, **200**, 90–114.
7. Huye,L.E. and Roth,D.B. (2000) Differential requirements for cis and trans V(D)J cleavage: effects of substrate length. *Nucleic Acids Res.*, **28**, 4903–4911.
8. Eastman,Q.M., Leu,T.M. and Schatz,D.G. (1996) Initiation of V(D)J recombination in vitro obeying the 12/23 rule. *Nature*, **380**, 85–88.
9. Sheehan,K.M. and Lieber,M.R. (1993) V(D)J recombination: signal and coding joint resolution are uncoupled and depend on parallel synapsis of the sites. *Mol. Cell Biol.*, **13**, 1363–1370.
10. Bergeron,S., Anderson,D.K. and Swanson,P.C. (2006) RAG and HMGB1 proteins: purification and biochemical analysis of recombination signal complexes. *Methods Enzymol.*, **408**, 511–528.
11. Raval,P., Kriatchko,A.N., Kumar,S. and Swanson,P.C. (2008) Evidence for Ku70/Ku80 association with full-length RAG1. *Nucleic Acids Res.*, **36**, 2060–2072.
12. Kriatchko,A.N., Anderson,D.K. and Swanson,P.C. (2006) Identification and characterization of a gain-of-function RAG-1 mutant. *Mol. Cell Biol.*, **26**, 4712–4728.
13. Kuwabara,M.D. and Sigman,D.S. (1987) Footprinting DNA-protein complexes in situ following gel retardation assays using 1,10-phenanthroline-copper ion: Escherichia coli RNA polymerase-lac promoter complexes. *Biochemistry*, **26**, 7234–7238.
14. Hesse,J.E., Lieber,M.R., Mizuuchi,K. and Gellert,M. (1989) V(D)J recombination: a functional definition of the joining signals. *Genes Dev.*, **3**, 1053–1061.

15. Mundy,C.L., Patenge,N., Matthews,A.G. and Oettinger,M.A. (2002) Assembly of the RAG1/RAG2 synaptic complex. *Mol. Cell Biol.*, **22**, 69–77.
16. Swanson,P.C. (2002) A RAG-1/RAG-2 tetramer supports 12/23-regulated synapsis, cleavage, and transposition of V(D)J recombination signals. *Mol. Cell Biol.*, **22**, 7790–7801.
17. Eastman,Q.M. and Schatz,D.G. (1997) Nicking is asynchronous and stimulated by synapsis in 12/23 rule-regulated V(D)J cleavage. *Nucleic Acids Res.*, **25**, 4370–4378.
18. Agrawal,A. and Schatz,D.G. (1997) RAG1 and RAG2 form a stable postcleavage synaptic complex with DNA containing signal ends in V(D)J recombination. *Cell*, **89**, 43–53.
19. Tsai,C.L., Drejer,A.H. and Schatz,D.G. (2002) Evidence of a critical architectural function for the RAG proteins in end processing, protection, and joining in V(D)J recombination. *Genes Dev.*, **16**, 1934–1949.
20. Hiom,K. and Gellert,M. (1998) Assembly of a 12/23 paired signal complex: a critical control point in V(D)J recombination. *Mol. Cell*, **1**, 1011–1019.
21. Huye,L.E., Purugganan,M.M., Jiang,M.M. and Roth,D.B. (2002) Mutational analysis of all conserved basic amino acids in RAG-1 reveals catalytic, step arrest, and joining-deficient mutants in the V(D)J recombinase. *Mol. Cell Biol.*, **22**, 3460–3473.
22. Nagawa,F., Hirose,S., Nishizumi,H., Nishihara,T. and Sakano,H. (2004) Joining mutants of RAG1 and RAG2 that demonstrate impaired interactions with the coding-end DNA. *J. Biol. Chem.*, **279**, 38360–38368.
23. Ramsden,D.A., McBlane,J.F., van Gent,D.C. and Gellert,M. (1996) Distinct DNA sequence and structure requirements for the two steps of V(D)J recombination signal cleavage. *EMBO J.*, **15**, 3197–3206.
24. Cuomo,C.A., Mundy,C.L. and Oettinger,M.A. (1996) DNA sequence and structure requirements for cleavage of V(D)J recombination signal sequences. *Mol. Cell Biol.*, **16**, 5683–5690.
25. Swanson,P.C. (2002) Fine structure and activity of discrete RAG-HMG complexes on V(D)J recombination signals. *Mol. Cell Biol.*, **22**, 1340–1351.
26. Ramsden,D.A., Baetz,K. and Wu,G.E. (1994) Conservation of sequence in recombination signal sequence spacers. *Nucleic Acids Res.*, **22**, 1785–1796.
27. Swanson,P.C. and Desiderio,S. (1998) V(D)J recombination signal recognition: distinct, overlapping DNA-protein contacts in complexes containing RAG1 with and without RAG2. *Immunity*, **9**, 115–125.
28. Akamatsu,Y. and Oettinger,M.A. (1998) Distinct roles of RAG1 and RAG2 in binding the V(D)J recombination signal sequences. *Mol. Cell Biol.*, **18**, 4670–4678.
29. Sigman,D.S., Kuwabara,M.D., Chen,C.H. and Bruice,T.W. (1991) Nuclease activity of 1,10-phenanthroline-copper in study of protein-DNA interactions. *Methods Enzymol.*, **208**, 414–433.
30. Matthews,A.G., Kuo,A.J., Ramon-Maiques,S., Han,S., Champagne,K.S., Ivanov,D., Gallardo,M., Carney,D., Cheung,P., Ciccone,D.N. *et al.* (2007) RAG2 PHD finger couples histone H3 lysine 4 trimethylation with V(D)J recombination. *Nature*, **450**, 1106–1110.
31. Liu,Y., Subrahmanyam,R., Chakraborty,T., Sen,R. and Desiderio,S. (2007) A plant homeodomain in RAG-2 that binds Hypermethylated lysine 4 of histone H3 is necessary for efficient antigen-receptor-gene rearrangement. *Immunity*, **27**, 561–571.
32. Difilippantonio,S., Gapud,E., Wong,N., Huang,C.Y., Mahowald,G., Chen,H.T., Kruhlak,M.J., Callen,E., Livak,F., Nussenzweig,M.C. *et al.* (2008) 53BP1 facilitates long-range DNA end-joining during V(D)J recombination. *Nature.*, **456**, 529–533.
33. Sadofsky,M.J., Hesse,J.E., McBlane,J.F. and Gellert,M. (1993) Expression and V(D)J recombination activity of mutated RAG-1 proteins. *Nucleic Acids Res.*, **21**, 5644–5650.
34. Kirch,S.A., Sudarsanam,P. and Oettinger,M.A. (1996) Regions of RAG1 protein critical for V(D)J recombination. *Eur. J. Immunol.*, **26**, 886–891.
35. Mo,X., Bailin,T. and Sadofsky,M.J. (2001) A C-terminal region of RAG1 contacts the coding DNA during V(D)J recombination. *Mol. Cell Biol.*, **21**, 2038–2047.
36. Arbuckle,J.L., Fauss,L.A., Simpson,R., Ptaszek,L.M. and Rodgers,K.K. (2001) Identification of two topologically independent domains in RAG1 and their role in macromolecular interactions relevant to V(D)J recombination. *J. Biol. Chem.*, **276**, 37093–37101.