Identification of Two Topologically Independent Domains in RAG1 and Their Role in Macromolecular Interactions Relevant to V(D)J Recombination*

Janeen L. Arbuckle‡, LeAnn J. Fauss‡, Rosemarie Simpson‡§, Leon M. Ptaszek¶, and Karla K. Rodgers‡***

From the ‡Department of Biochemistry and Molecular Biology, The University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma 73190 and the ¶Ruttenberg Cancer Center, Mount Sinai School of Medicine of New York University, New York, New York 10029

Received for publication, June 27, 2001
Published, JBC Papers in Press, July 30, 2001, DOI 10.1074/jbc.M105988200

V(D)J recombination is instigated by the recombination-activating proteins RAG1 and RAG2, which catalyze site-specific DNA cleavage at the border of the recombination signal sequence (RSS). Although both proteins are required for activity, core RAG1 (the catalytically active region containing residues 384–1008 of 1040) alone displays binding specificity for the conserved heptamer and nonamer sequences of the RSS. The nonamer-binding region lies near the N terminus of core RAG1, whereas the heptamer-binding region has not been identified. Here, potential domains within core RAG1 were identified using limited proteolysis studies. An iterative procedure of DNA cloning, protein expression, and characterization revealed the presence of two topologically independent domains within core RAG1, referred to as the central domain (residues 528–760) and the C-terminal domain (residues 761–980). The domains do not include the nonamer-binding region but rather largely span the remaining relatively uncharacterized region of core RAG1. Characterization of macromolecular interactions revealed that the central domain bound to the RSS with specificity for the heptamer and contained the predominant binding site for RAG2. The C-terminal domain bound DNA cooperatively but did not show specificity for either conserved RSS element. This domain was also found to self-associate, implicating it as a dimerization domain within RAG1.

The immune system displays remarkable specificity and diversity in its ability to recognize and eliminate foreign antigens. The basis for this immense diversity in many species is a complex rearrangement of the V (variable), D (diversity), and J (joining) gene segments that together encode the variable regions of T cell receptors and immunoglobulins (see Ref. 1 for review). This process, known as V(D)J recombination, requires the activity of a wide array of enzymes and is initiated by the lymphoid-specific recombination-activating proteins RAG1 and RAG2. The RAG proteins guide recombination events to conserved recombination signal sequences (RSSes) that flank the genomic regions to be rearranged. Each RSS consists of a conserved heptamer and nonamer sequence separated by a 12- or 23-base pair spacer, the sequence of which is poorly conserved. Efficient recombination occurs generally between an RSS containing a 12-base pair spacer (12RSS) and one containing a 23-base pair spacer (23RSS), a requirement referred to as the 12/23 rule.

The recombination process is often divided into two phases, the first phase of which consists of two distinct enzymatic steps catalyzed by the RAG proteins. The first step involves the binding of a RAG1-RAG2 complex to an RSS and the subsequent generation of a nick between the heptamer and its adjacent coding strand. The resulting 3′-OH group then performs a nucleophilic attack on the phosphodiester bond of the opposite strand. The primary products of this transsesterification reaction are a covalently sealed hairpin, referred to as the coding end, and a blunt-ended 5′ phosphorylated RSS, referred to as the signal end (2, 3). In the physiological reaction, coupled cleavage likely occurs on a 12- and 23RSS held in a precleavage complex by the RAG proteins. The hairpin formation step, in particular, seems to be highly restricted to the synaptic complex (4–6). The second phase of the reaction is governed by an array of enzymes that catalyze the opening and processing of the coding-end hairpins (1). Processed coding ends are then joined, most likely by DNA ligase IV with XRCC4, resulting in the ligation of two formerly distant regions of the genome (7, 8). The signal ends are also joined to form a precise heptamer/heptamer junction. Less is understood about the precise role played by the RAG proteins in the second phase of V(D)J recombination. Studies have shown that the RAG proteins remain bound to the signal and coding ends in a post-cleavage complex, possibly to stabilize and direct the appropriate joining of cleaved ends (9, 10). In in vitro studies, the RAG proteins have also been shown to open hairpins (11, 12) and remove 3′ overhangs (13), but the contributions of these catalytic activities to V(D)J recombination in vivo have not been established. Proteins that function in double-strand DNA-break repair through the nonhomologous end-joining pathway (e.g. Ku70, Ku80, DNA-dependent protein kinase catalytic subunit, and

---

* This work was supported by Research Project Grant RPG-00-032-01-CIM from the American Cancer Society, an Oklahoma Center for Advancement in Science and Technology award for project number HR99-040 and funds from the Presbyterian Health Foundation. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ Present address: Sangamo Biosciences, Inc., Richmond, CA 94804.

§ Present address: Section of Immunobiology, Yale University School of Medicine, New Haven, CT 06520.

** To whom correspondence should be addressed. Tel.: 405-271-2227, Ext. 1248; Fax: 405-271-3139; E-mail: Karla-Rodgers@ouhsc.edu.

1 The abbreviations used are: RSS, recombination signal sequence; 12- and 23RSS, RSS containing a 12- or 23-base pair spacer, respectively; MBP, maltose-binding protein; PAGE, polyacrylamide gel electrophoresis; GST, glutathione S-transferase; MALDI-TOF, matrix-assisted laser desorption ionization-time-of-flight; WT, wild type; MH, mutant heptamer.
the recently identified protein Artemis (14)) are also essential for the processing and joining reactions (reviewed in Ref. 15).

Although the first phase of V(D)J recombination has been well defined mechanistically, many questions remain concerning the RAG proteins and the nature of the protein-protein and protein-DNA complexes responsible for the DNA cleavage reaction. The investigation of these proteins was facilitated by the identification of the catalytically active regions of the RAG proteins, termed the core regions, which include residues 384–1008 (of 1040 total residues) in murine RAG1 and residues 1–387 (of 527 total residues) in murine RAG2 (1). These core regions, capable of achieving effective recombination when expressed together, are more soluble than their parent full-length proteins and therefore have been the focal points for the majority of subsequent research.

A multisubunit complex consisting of both RAG1 and RAG2 is required for cleavage of the RSS to yield first nicks and then hairpinned coding ends. Recognition and binding to the RSS seems to be largely mediated by RAG1. For instance, the RSS nonamer-binding site is localized to the N terminus of core RAG1 (residues 384–460) (16, 17). In addition, protein-DNA crosslinking studies showed that core RAG1 formed specific contacts with the RSS heptamer in the presence of RAG2 (18–20). Furthermore, electrophoretic mobility shift assays have demonstrated that core RAG1 bound to an isolated RSS in the absence of RAG2 and that this binding was specific for both the heptamer and the nonamer (21, 22). Although the region of RAG1 responsible for RSS nonamer recognition has been identified and shown to be homologous to the DNA-binding domain of Hin recombinase (16), the region of core RAG1 responsible for heptamer recognition has not yet been identified. In addition to its ability to recognize an isolated RSS, RAG1 has been shown recently to possess a triad of acidic residues (Asp-600, Asp-708, and Glu-962, known as the DDE motif) that are required for heptamer recognition has not yet been identified. In addition, the presence of the RSS heptamer and the nature of the protein-protein and protein-DNA interactions catalyzed by the RAG proteins (23–25). These residues are believed to coordinate 1–2 divalent metal cations, as is characteristic of other enzymes containing the DDE motif (26).

Interaction between RAG1 and RAG2 has been shown to occur in the absence of DNA, suggesting that the two proteins bind to the RSS as a preformed complex (27). Thus, regions within core RAG1 mediate protein-protein and protein-DNA interactions that are essential for the V(D)J recombination reaction. We propose that individual domains in the core region contribute each RAG1 binding activity and that these domains, as isolated modules, may retain their ability to form macromolecular interactions. To determine this possibility, we used limited proteolysis studies to identify and then characterize structural, or topologically independent, domains within core RAG1. The results from this study demonstrate that core RAG1 consists of multiple domains, each of which functions individually in one or more of the essential macromolecular interactions formed by the intact core protein.

**EXPERIMENTAL PROCEDURES**

**Fusion Protein Cloning**—Fragments of the murine RAG1 gene were amplified by polymerase chain reaction using primers that introduced a BamHI site at the 5′ end of the product and two stop codons and a SalI site at the 3′ end of the product. A gene encoding for an MBP fusion with RAG1 fragments 500–1008 and 760–1008) were determined by MALDI-TOF mass spectrometry. Each purified fusion protein was analyzed by western analysis of both MBP and GST proteins. After transfer, the Western analysis of both MBP and GST proteins.

**SDS-PAGE gels.** GST-core RAG2, expressed by transfection in 293T cells, was purified as described previously (16).

**MALDI-TOF Mass Spectrometry**—The lengths of the degradation products generated during the purification of MBP fusion proteins with RAG1 fragments 500–1008 and 760–1008) were determined by MALDI-TOF mass spectrometry. Each purified fusion protein was digested with 20 μM Tris-HCl, pH 8.0, 50 μM NaCl, 5 μM β-mercaptoethanol, and 10 μM ZnCl2 and analyzed at the National Science Foundation Experimental Program to Stimulate Competitive Research Oklahoma Laser Mass Spectrometry Facility. The fusion proteins were combined with sinapinic acid, spotted onto a grid, and analyzed by the Voyager Elite MALDI-TOF mass spectrometer (Applied Biosystems, Framingham, MA).

**RAG1 and RAG2 Binding Assays**—Glutathione-Sepharose 4B resin (Amersham Pharmacia Biotech) was washed with 1 mg/ml bovine serum albumin in interaction buffer (20 mM Tris-HCl, pH 8.0, 0.2 M NaCl, 10 μM ZnCl2, 10 μM EDTA, and 5 μM β-mercaptoethanol) for 30 min at 4 °C. The resin was washed three times with interaction buffer, and 200 ng of the appropriate MBP fusion plus 200 ng of GST-core RAG2 or GST in interaction buffer were added to the resin. The samples were incubated on the resin for 30 min at 4 °C. After three interaction buffer washes with interaction buffer, the bound protein was eluted from the resin with SDS loading buffer. The proteins were then resolved by SDS-PAGE (10% polyacrylamide gel) and electrotransferred to a polyvinylidene difluoride membrane. Two gels were run in parallel to enable Western analysis of both MBP and GST proteins. After transfer, the membrane was blocked for 2 h in 1% (w/v) bovine serum albumin in TBS (10 mM Tris-HCl, pH 7.5, 150 mM NaCl, and 0.1% Tween 20). Next, the membranes were incubated with the reactive primary antibody for 1 h (Monoclonal anti-MBP, Santa Cruz Biotechnology) followed by a biotinylated secondary antibody for 1 h and avidin-conjugated horseradish peroxidase for 45 min. Detection was done using enhanced chemiluminescence (ECL, Amersham Pharmacia Biotech) by exposure to Kodak X-Omat film.

**Oligonucleotide Substrates for Electrophoretic Mobility Shift Assays**—Three various 12RSS substrates used were prepared by annealing complementary oligonucleotides. The sequence of the top strand of the WT12RSS is d(GATATGGCTGCTTTCAGACGTATAGCTCTTACAAACCCCTCAATCGCCGAGG). The MH12RSS oligonucleotide sequence is identical to that of the WT12RSS except that the heptamer sequence (CACAGTG) has been replaced by the sequence GAGAAGC. Similarly, the mutated nonamer 12RSS contains the sequence AGGC-ACAAAAACCTCCAATCGAGCGGAG (MH12RSS). Substrates were labeled with [γ-32P]ATP, where indicated, using T4 polynucleotide kinase.

**Electrophoretic Mobility Shift Assays**—Fusion proteins of each core RAG1 domain with MBP were incubated with [32P]-labeled WT12RSS and resolved on a 5% nondenaturing polyacrylamide gel as described (20). The protein-DNA complex was detected by autoradiography. The fusion proteins listed above as well as MBP fused to core RAG1 were expressed in Escherichia coli and released by sonication as described previously (21). The proteins were bound to an amynolose column in purification buffer (20 μM Tris-HCl, pH 8.0, 50 μM ZnCl2, 10 μM EDTA, and 5 μM β-mercaptoethanol) plus 500 mM NaCl, the column was washed with purification buffer plus 1.5 M NaCl, and the fusion proteins were eluted from the column in purification buffer plus 500 mM NaCl and 10 mM maltose. In some applications, the proteins were purified further through a Q-Heal matrix in fast flow column (20). In the absence of RAG2 and that this binding was specific for both

**Competition Assays—**Protein-DNA specificity assays, also referred to here as competition assays, were performed in the same binding buffer described above. The two domains were added to reactions containing 1 nM [32P]-labeled WT12RSS and 50–500 nM of the indicated unlabeled compounds.

**Downloaded from http://www.jbc.org/ by guest on July 25, 2018**

**Macromolecular Interactions Relevant to V(D)J Recombination**
Macromolecular Interactions Relevant to V(D)J Recombination

RESULTS

Partial Trypsin Digest of Core RAG1 Indicates Three Possible Domain Boundaries—To identify domain boundaries within core RAG1, a fusion protein between MBP and core RAG1 (MBP-core RAG1) was digested with increasing concentrations of trypsin under limiting conditions (Fig. 1A). By limiting the reaction conditions the initial cleavage events occur primarily at the more accessible regions of the protein. The initial products generated by tryptic cleavage are core RAG1 and MBP, consistent with the presence of the long flexible linker (including 10 consecutive asparagine residues) connecting the two tethered proteins. Products A–D are the result of further digestion of core RAG1, because this band is diminished with increasing amounts of trypsin, whereas the intensity of the MBP band remains relatively constant. N-terminal sequencing of these products indicated that cleavage had occurred primarily C-terminal to residues Arg-529, Arg-713, and Lys-777 (Fig. 1B). (All RAG1 residue numbers referred to in this study are from the full-length murine RAG1 sequence.) The molecular mass (and length) of each product was estimated by comparison of molecular mass standards with SDS-PAGE. With this combined information, the order of progression of the digestion of core RAG1 can be inferred from Fig. 1A. For instance, products A and B are formed readily at lower concentrations of trypsin. Product A appears to become further proteolyzed to product D, whereas product B does not appear to significantly degrade further with increasing concentrations of trypsin. Finally, product C only appears at the highest concentrations of trypsin, apparently as intact core RAG1 is further proteolized. Limited proteolysis of MBP-core RAG1 was also performed with thermolysin, yielding similar results to those shown in Fig. 1 (data not shown). This confirms that cleavage at the basic residues listed above is caused by increased accessibility in those regions of the protein and not strictly by enzyme specificity.

Strategy for Identification of Structural and Functional Domains—It is important to note that the tryptic fragments generated above may or may not represent topologically independent domains. Although the limited proteolytic reaction would restrict cleavage to the most accessible regions of the protein, the protease may also cleave at exposed loops in the protein that do not represent domain boundaries. To establish valid domain boundaries within core RAG1, we performed the following iterative procedure. First, regions of the core RAG1 gene were cloned based on the tryptic cleavage results. Second, the corresponding regions of the core RAG1 protein were expressed and purified as outlined under “Experimental Procedures.” Finally, whether the purified fragments of RAG1 represented structural and functional domains of the intact core protein was assessed. For clarity here, structural and functional domains represent regions that fold autonomously and perform functions similar to those characterized previously for the intact core RAG1 protein. The assignment of structural, or topologically independent, domains was based on the ability of the protein fragments to form discrete species (monomeric or dimeric) as determined by size-exclusion chromatography. The benchmark for assessing whether the protein regions represented functional domains of core RAG1 was based on the ability of the protein fragments to achieve one or more of the macromolecular interactions attributed to RAG1 in the V(D)J recombination reaction. As outlined in detail below, the purification of core RAG1 fragments fused to MBP twice resulted in proteolysis from endogenous E. coli proteases, which appeared to occur during cell lysis (data not shown). This proved to be advantageous in both instances, because the resulting products had an increased ability to form distinct monomeric or dimeric species (see following text). Although cleavage at the N terminus of each of the fusion proteins was possible in these cases, such cleavage would have removed significant portions of the MBP required to bind substrate (and the amylose column),
because an N-terminal loop including residues 1–20 contains important contacts to the substrate in the binding pocket (28). Had cleavage occurred at the N terminus of the fusion protein, the cleaved protein would not have been recovered from the purification protocol used in these studies. The degradation products were identified as described in the following section, and the procedure outlined above was repeated.

**Identification of Two Domains That Span the Active Site**—Based on the identified tryptic cleavage sites, two fragments of RAG1 (residues 528–1008 and 714–1008) were cloned as fusions to MBP. Purification of the MBP fusion with RAG1 residues 528–1008 (N528a) resulted in a degradation product that was topologically independent based on its elution from size-exclusion chromatography as a distinct single species. In contrast, the full-length protein (N528a) eluted entirely in the void volume indicative of misfolded and/or nonspecifically aggregated protein, perhaps resulting from solvent-exposed hydrophobic regions that are typically buried in the intact core protein. Analysis by MALDI-TOF mass spectrometry indicated that the degradation product was ~250 residues smaller than N528a (Fig. 2A), corresponding to truncation of the fusion protein at RAG1 residue 760. A fusion protein of MBP with RAG1 residues 528–760 (N528b) was then produced and, based on its ability to form a discrete species, appears to exist as a topologically independent domain.

With the identification of a structural domain that spans residues 528–760, we subsequently cloned the region encoding residues 761–1008 into the pMAL-c2 plasmid. In purifying the fusion protein of MBP to RAG1 residues 761–1008 (N761a) the full-length fusion protein was mostly misfolded and aggregated. However, as seen in the purification of N528a, proteolysis had occurred at the C terminus of the fusion protein, producing cleavage products that formed autonomously folded modules based on size-exclusion chromatography. MALDI-TOF mass spectrometry of these products (Fig. 2B) indicated that cleavage had occurred at the residues indicated. To be conservative in the establishment of our domain boundary, we chose the largest of the products to further characterize. Purification of a fusion protein of residues 761–980 (N761b) resulted in a proteolytically resistant, topologically independent domain capable of forming distinct monomeric and dimeric species (see the following text and Fig. 3).

The fragment consisting of RAG1 residues 714–1008 was cloned as a fusion to MBP, purified, and characterized as outlined above. The fusion protein was entirely aggregated with...
the formation of no obvious degradation products. Although these observations do not preclude Arg-713 as a valid domain boundary, we chose not to pursue this site as an N-terminal domain boundary.

Although tryptic cleavage at residue 530 suggested the presence of an N-terminal domain between residues 384 and 529 of RAG1, no protease-resistant fragments in this region of the core remained after limited tryptic digestion. As a result, we were unable to establish the boundaries of an N-terminal domain. The N-terminal region of the core contains a putative helix-loop-helix motif that has been shown to form specific interactions with the RSS nonamer (16, 17). It is possible that in the absence of DNA this region may be relatively unstructured and thus more susceptible to tryptic cleavage. Nevertheless, an isolated fragment of core RAG1 containing the nonamer-binding region has been shown previously to bind to glutathione-Sepharose resin (16), indicating that this domain persists predominantly as a monomer. The elution profile of the central domain resulted in one major peak. The molecular mass of this eluted species indicated that this domain is predominantly a monomer (Fig. 3A). In contrast, the elution profile of the C-terminal domain indicated the presence of two distinct species (Fig. 3B). When compared with the elution profiles of known standards, it is clear that the first of the two peaks represents a dimer, and the second represents a monomer. The C-terminal domain therefore represents a dimerization domain that most likely contributes to the self-association properties of core RAG1.

The Central Domain Contains the Predominant Binding Site for RAG2—To investigate the domains of core RAG1 responsible for interaction with RAG2, we performed the RAG1 and RAG2 interaction assay as outlined under “Experimental Procedures.” This assay tested the ability of the MBP-core RAG1 domains to associate with GST-core RAG2 bound to glutathione-Sepharose resin. GST-core RAG2 bound successfully to the resin for all samples, as shown by the band in each lane of the interest blot (Fig. 4, lanes 1–4). As expected, the α-MBP blot verified that MBP core RAG1 bound to GST-core RAG2 (lane 1). Significantly, the central domain effectively formed a complex with GST-core RAG2 (lane 2), whereas no significant interaction between the C-terminal domain and GST-core RAG2 was observed under the experimental conditions used here (lane 3).
The interaction is localized to the RAG1 and RAG2 portions of the fusion proteins as lane 4 shows no band for MBP, demonstrating that the MBP-RAG1 proteins do not interact with the resin or GST-core RAG2 via the MBP tag. As an additional control, we tested whether GST (in place of GST-core RAG2) could bind to the MBP-RAG1 proteins. Analysis of the α-MBP blot showed no bands corresponding to the MBP-RAG1 proteins, demonstrating that the GST tag did not interact with the MBP-RAG1 proteins (data not shown). From these results it can be concluded that core RAG2 interacts predominantly with the central domain of core RAG1.

Both Independent Core RAG1Domains Bind to DNA—Electrophoretic mobility gel shift assays were performed to determine the relative affinity of each domain (fused to MBP) for an oligonucleotide duplex containing a 12RSS (Fig. 5). The central domain forms a protein-DNA complex with the 12RSS at low micromolar concentrations of protein but only shifts a small fraction of the labeled probe (Fig. 5A, lanes 2–5). Although the slower mobility complex becomes more abundant with increasing concentrations of the central domain, the unbound 12RSS remains relatively constant in its intensity, indicating that the binding of the central domain to the 12RSS is of low affinity. In addition, the existence of a single band indicates that the central domain binds the RSS as a single species, likely as a monomer, although this has yet to be confirmed.

In contrast to the central domain, the C-terminal domain displays a significantly higher affinity for the 12RSS (Fig. 5B). Upon titration of increasing concentrations of the C-terminal domain, the majority of the unbound 12RSS is shifted to a slower mobility species consistent with complex formation between the 12RSS and the C-terminal domain. In addition, with low micromolar concentrations of the C-terminal domain the formation of a second slower mobility complex is observed (Fig. 5B, lanes 4–5), indicative of higher order C-terminal domain complexes with the 12RSS. The two complexes formed may represent a monomer and then a dimer of the C-terminal domain bound to the RSS, consistent with the observation of these oligomerization states of the C-terminal domain in the absence of DNA. Studies to determine the stoichiometry of the complexes between both the central and C-terminal domains to the RSS are underway currently.

The binding of the C-terminal domain to the 12RSS demonstrates positive cooperativity, because the range of protein concentration from the first appearance of the protein-DNA complex to conditions in which >90% of the DNA is bound occurs over a range of less than 10-fold. These results are similar to that observed previously with core RAG1, in which up to three complexes were formed between core RAG1 and the RSS over a narrow protein concentration range (21). These results suggest that the C-terminal domain may be responsible for the cooperativity in DNA binding that has been observed previously with intact core RAG1.

The results outlined above show that both isolated core RAG1 domains bind to the 12RSS; however, under the conditions of the experiments performed here, the degree of specificity of each domain to the conserved heptamer and nonamer sequences is not clear. We expect that the domains do not bind specifically to the RSS nonamer, because mutation of the nonamer-binding region at the N-terminal end of intact core RAG1 eliminates specificity for the RSS nonamer (17). However, previous results also demonstrated that MBP-core RAG1 binds with detectable specificity to the RSS heptamer (21, 22, 29). Thus, it is reasonable to question whether either the central or C-terminal domains of core RAG1 demonstrate sequence-specific binding to the RSS heptamer.

The Central Domain Contains the Heptamer-binding Region—To determine whether either the central or C-terminal domains form specific contacts with the RSS heptamer, competition assays were performed. In these experiments each domain (fused to MBP) was incubated in the presence of radiolabeled WT12RSS and varying concentrations of unlabeled WT12RSS or mutant heptamer (MH) 12RSS. With the central domain, increasing concentrations of unlabeled WT12RSS readily competed with the radiolabeled WT12RSS for complex formation with the protein, whereas increasing concentrations of unlabeled MH12RSS were significantly less efficient (Fig. 6). In contrast, there was no significant difference in the ability of the unlabeled WT12RSS versus the unlabeled MH12RSS to compete the labeled WT12RSS for binding to the C-terminal domain, indicating that this domain does not show specificity for the RSS heptamer under the conditions used here (data not shown).

The interaction between each core RAG1 domain and the WT- versus MH12RSS competitor was quantitated as done previously (21). These results indicate that the central domain shows a 3-fold specificity for the RSS heptamer over nonspecific DNA. Although this specificity is not large, it is similar to the magnitude of the specificity previously observed for the intact MBP-core RAG1 (21). Finally, to confirm that neither domain forms significant sequence-specific contacts with the RSS.
nonamer, both domains complexed with the radiolabeled WT12RSS were competed with an unlabeled 12RSS that contained a mutated nonamer. The unlabeled mutated nonamer 12RSS competed the labeled WT12RSS for binding to both domains at a level comparable with that of unlabeled WT12RSS, indicating that neither domain shows specificity for the RSS nonamer (data not shown). Our results outlined here indicate that the central domain contains the RSS heptamer-binding site, whereas the C-terminal domain represents a non-specific DNA-binding domain.

**DISCUSSION**

We have identified two topologically independent domains within core RAG1. The central domain (residues 528–760) is the predominant binding site for core RAG2 and displays specificity for the RSS heptamer. The C-terminal domain (residues 761–980) is capable of dimerization and binds DNA cooperatively. It is important to note that core RAG1 displays a combination of these properties and that the ability of each domain to carry out a portion of these properties is further evidence that they are both structural and functional domains in core RAG1 (Fig. 7A).

Investigation of the RAG proteins has been guided by their apparent similarity in catalytic activities to members of the transposase and integrase families of enzymes. In addition to containing core regions that possess a DDE active site, a number of these enzymes are also similar to RAG1 in that they consist of 2–3 topologically independent domains (30). In contrast to RAG1, the DDE triad of most of these enzymes is localized in the central domain. However, there are a few exceptions including Tn10 transposase, which splits the active site residues between two domains with the third active site residue separated in the C-terminal domain (31). Although sequence-specific DNA binding is often a function of the N terminus of a number of the DDE motif enzymes, both the N-terminal and central domains of core RAG1 are capable of binding RSS elements with specificity. Another notable differ-

---

**FIG. 6.** Electrophoretic mobility gel shift assays of the central domain/RSS complexes in the presence of unlabeled wild-type (WT) or mutant heptamer (MH) RSSs. Each lane contains 1.0 μM of the central domain with 1 nM radiolabeled 12RSS. Lane 1 contains no competitor DNA. Lanes 2–6 contain increasing concentrations of unlabeled WT12RSS competitor. Lanes 7–11 contain increasing concentrations of unlabeled MH12RSS competitor. The concentration of competitor DNA ranged from 0 to 50 nM.

**FIG. 7.** Domain model of core RAG1. A, a diagram of core RAG1 domains and their properties. The labels of the motifs and active site residues are the same as described in the Fig. 2C legend. The macromolecular interaction properties of each domain are listed beneath its location in core RAG1. B, a model of RAG1 domains bound to the 12RSS. The 12RSS was modeled using the program NAMOT (46). Subunits of the dimeric core RAG1 are labeled A and B.
ence between RAG1 and the majority of DDE motif enzymes includes the exceptionally large spacing between the active site residues in RAG1 (23–25).

Although interaction between RAG1 and RAG2 has been shown, there have been conflicting reports indicating the location within each protein that these contacts occur. A recent study has shown that a region containing the zinc finger motif of RAG1 (residues 692–758) is sufficient to interact with core RAG2 (32). In support of these findings, we show here that the topologically independent central domain of core RAG1 is the predominant binding site for core RAG2. In addition, trypsin and endogenous E. coli proteases tend to cleave first on either side of the zinc finger motif (Figs. 1B and 2A), indicating that this region of the protein may be relatively exposed and therefore more accessible for binding to RAG2. Although earlier studies suggest that residues in the C terminus of core RAG1 are involved in binding core RAG2 (33), the C-terminal domain did not display the ability to bind core RAG2 under the stringent conditions used in the assay described here. Although it is possible that core RAG2 makes contacts with the C-terminal domain of core RAG1, we suggest that this domain plays a secondary role in the interaction of RAG1 and RAG2.

The degree of specificity with which RAG1 binds the RSS has been a matter of some debate. We have demonstrated previously that core RAG1 alone binds specifically to both elements of the RSS, with less specificity for the heptamer than the nonamer (21). Although RAG2 has been shown to enhance the specificity with which core RAG1 binds the RSS, it alone is incapable of binding specifically to the RSS (34, 35). This has led to the proposal that RAG2 induces conformational changes in RAG1, allowing it to bind the RSS with higher specificity and orienting the catalytic residues to the site of cleavage (21, 34). Regarding the interactions of the core RAG1 domains with the RSS, we show here that the central domain of core RAG1 alone binds specifically to the RSS. Finally, although the C-terminal domain did not bind specifically to the RSS, it did bind with significantly higher affinity to DNA than the central domain. This then raises the possibility that RAG2 may increase the specificity of RAG1 for the RSS by modulating the nonspecific DNA binding properties of the C-terminal domain in intact core RAG1 instead of (or in addition to) directly affecting binding of RAG1 to the RSS elements.

It has been shown that core RAG1 self-associates predominantly to form a dimer in solution (21), but the regions responsible for dimerization have yet to be identified. In the studies presented here, we show that the C-terminal domain is capable of self-association in the absence of DNA, implicating this region of the protein as a viable candidate for the dimerization domain of core RAG1. However, because we have yet to identify a structural domain in the N terminus of core RAG1, we cannot exclude the possibility that this region may also be capable of self-association. In addition, a domain within the noncore region of RAG1 (referred to as the zinc-binding dimerization domain) can also dimerize (36, 37), suggesting that the self-association properties of RAG1 may be fairly complex. Further investigation is necessary to assess the relative contribution of each self-associating domain to that of the full-length protein.

By considering the results presented here in the context of other data, we can contribute additional insight to a developing model for the RAG1-RSS complex (Fig. 7B). In this model, the N-terminal region of the protein contacts the RSS nonamer while the central domain binds the RSS heptamer. Other studies have shown that mutations within the central domain of core RAG1 result in sensitivity to the coding flank sequence, suggesting that this region of the protein may interact with the coding sequence of the RSS (38, 39). Our results do not preclude this possibility as this domain’s ability to bind to the RSS heptamer places it in close contact with the coding flank. However, the C-terminal domain showed no specificity for either conserved RSS element. To place the three active site residues at the site of cleavage, this domain most likely binds nonspecifically to the coding flank near the RSS heptamer. This is corroborated further by a recent study in which a C-terminal fragment of core RAG1 was crosslinked to the coding sequence (40). Finally, the domains in the core RAG1 dimer are placed in a trans configuration in the model, with the N-terminal domain from one subunit contacting the RSS nonamer and the central and C-terminal domains from the second subunit contacting the heptamer and coding flank, respectively, of the same RSS. This latter point is based on recent results that indicate that the nonamer-binding domain is contributed in trans to the active site of core RAG1 (41). Although not presented in this model, RAG2 would also be placed at the heptamer-coding border because of results from protein-DNA crosslinking studies (19, 20, 34) as well as its ability to interact with the core RAG1 central domain.

All three domains of the dimeric core RAG1 interacting with the RSS in a trans configuration are analogous to the complex formed between Tn5 transposase and the pair of transposon ends. In the crystal structure of the synaptic Tn5 transposase complex, a dimer of the protein is bound to its two DNA recognition sites orienting the active sites for coupled cleavage (42). In Fig. 7B, placement of a 23RSS contacting the remaining core RAG1 domains not bound to DNA would result in a similar synaptic complex. However, the stoichiometry of protein to DNA in the RAG1-RAG2-RSS synaptic complex is not yet known. In addition, synaptic complex formation in vitro requires the presence of the high mobility group proteins, HMG1 or HMG2 (43, 44). A role for the HMG proteins in synaptic complex formation is proposed to include enhancement and stabilization of the DNA-bending activity of RAG1 and RAG2 (22).

The minimal region of RAG1 that has been found to be required for an in vitro endonucleolytic activity with RAG2, namely removal of 3’ overhangs, includes residues 510–1008 (13). Thus, although the residues of the DDE triad are contained within the domains identified here, they may not contain all regions required for catalytic activity. Nevertheless, it would be interesting to determine whether the catalytic activity of RAG1 could be reconstituted from component domains, as has been achieved previously with Tn10 transposase (45). In conclusion, identification of topologically independent domains has provided further insight into the macromolecular interactions of RAG1 and introduces useful tools for the further characterization of this protein and its role in V(D)J recombination.

Acknowledgments—We thank David Schatz for many useful discussions and a critical reading of the manuscript.

REFERENCES
1. Fugmann, S. D., Lee, A. I., Shokett, P. E., Villey, I. J., and Schatz, D. G. (2000) Annu. Rev. Immunol. 18, 495–527.
2. 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.
3. van Gent, D. C., Mizuuchi, K., and Gellert, M. (1996) Science 271, 1592–1594.
4. Eastman, Q. M., and Schatz, D. G. (1997) Nucleic Acids Res. 25, 4370–4378.
5. West, R. B., and Lieber, M. R. (1998) Mol. Cell. Biol. 18, 6408–6415.
6. Yu, K., and Lieber, M. R. (2000) Mol. Cell. Biol. 20, 7914–7921.
7. Grawunder, U., Zimmer, D., Fugmann, S., Schwarz, K., and Lieber, M. R. (1998) Mol. Cell 2, 477–484.
8. Grawunder, U., Zimmer, D., Kulesza, P., and Lieber, M. R. (1998) J. Biol. Chem. 273, 24708–24714.
9. Agrawal, A., and Schatz, D. G. (1997) Cell 89, 43–53.
10. Hiom, K., and Gellert, M. (1998) Mol. Cell 1, 1011–1019.
11. Besmer, E., Mansilla-Soto, J., Cassard, S., Sawchuk, D. J., Brown, G., Sadovsky, M., Lewis, S. M., Nussenzweig, M. C., and Cortes, P. (1998) Mol. Cell Biol. 2, 817–828.
12. Shokett, P. E., and Schatz, D. G. (1999) Mol. Cell. Biol. 19, 4158–4166.
13. Santagata, S., Besmer, E., Villa, A., Bonzi, F., Allingham, J. S., Sobaechi, C.,
Macromolecular Interactions Relevant to V(D)J Recombination

Haniford, D. B., Vezzoni, P., Nussenzweig, M. C., Pan, Z.-Q., and Cortes, P. (1999) *Mol. Cell* **4**, 935–947

14. Moshous, D., Callebaut, I., de Chasseval, R., Corneo, B., Cavazzana-Calvo, M., Le Deist, F., Tzcan, I., Sanal, O., Bertrand, Y., Philippe, N., Fischer, A., and de Villartay, J.-P. (2001) *Cell* **105**, 177–186

15. Grawunder, U., and Harfst, E. (2001) *Curr. Opin. Immunol.* **13**, 186–194

16. Spanopoulou, E., Zaitseva, F., Wang, F.-H., Santagata, S., Baltimore, D., and Panayotou, G. (1996) *Cell* **87**, 263–276

17. Difilippantonio, M. J., McMahan, C. J., Eastman, Q. M., Spanopoulou, E., and Schatz, D. G. (1996) *Cell* **87**, 253–262

18. Swanson, P. C., and Desiderio, S. (1998) *Immunity* **9**, 115–125

19. Eastman, Q. M., Villey, I. J., and Schatz, D. G. (1999) *Mol. Cell. Biol.* **19**, 3788–3797

20. Mo, X., Bailin, T., and Sadofsky, M. J. (1999) *J. Biol. Chem.* **274**, 7025–7031

21. Rodgers, K. K., Villey, I. J., Ptaszek, L., Corbett, E., Schatz, D. G., and Coleman, J. E. (1999) *Nucleic Acids Res.* **27**, 2938–2946

22. Aidinis, V., Ronaldi, T., Beltrame, M., Santagata, S., Bianchi, M. E., and Spanopoulou, E. (1999) *Mol. Cell. Biol.* **19**, 6532–6542

23. Kim, D. R., Dai, Y., Mundy, C. L., Yang, W., and Oettinger, M. A. (1999) *Mol. Cell. Biol.* **19**, 3788–3797

24. Fugmann, S. D., Villey, I. J., Ptaszek, L. M., and Schatz, D. G. (2000) *Mol. Cell* **5**, 97–107

25. Landree, M. A., Wibbenmeyer, J. A., and Roth, D. B. (1999) *Genes Dev.* **13**, 3509–3569

26. Rice, P. A., and Baker, T. A. (2001) *Nat. Struct. Biol.* **8**, 302–307

27. Santagata, S., Aidinis, V., and Spanopoulou, E. (1998) *J. Biol. Chem.* **273**, 16325–16331

28. Spurlino, J. C., Lu, G. Y., and Quiocho, F. A. (1991) *J. Biol. Chem.* **266**, 5202–5219

29. Nagawa, F., Ishiguro, K.-I., Tsuboi, A., Yoshida, T., Ishikawa, A., Takemori, T., Otsuka, A. J., and Sakano, H. (1998) *Mol. Cell. Biol.* **18**, 655–663

30. Haren, L., Ton-Hoang, B., and Chandler, M. (1999) *Annu. Rev. Microbiol.* **53**, 245–281

31. Bolland, S., and Kleckner, N. (1996) *Cell* **84**, 223–233

32. Aidinis, V., Diaz, D. C., Gomez, C. A., Bhattacharyya, D., Spanopoulou, E., and Santagata, S. (2000) *J. Immunol.* **164**, 5826–5832

33. Moshous, D., Callebaut, I., de Chasseval, R., Corneo, B., Cavazzana-Calvo, M., Le Deist, F., Tzcan, I., Sanal, O., Bertrand, Y., Philippe, N., Fischer, A., and de Villartay, J.-P. (2001) *Cell* **105**, 177–186

34. Swanson, P. C., and Desiderio, S. (1999) *Naturwissenschaften* **86**, 3674–3683

35. Akamatsu, Y., and Oettinger, M. A. (1998) *Mol. Cell. Biol.* **18**, 4670–4678

36. Rodgers, K. K., Pu, Z., Fleming, K. G., Schatz, D. G., Engelman, D. M., and Coleman, J. E. (1996) *J. Mol. Biol.* **260**, 70–84

37. Bellon, S. F., Rodgers, K. K., Schatz, D. G., Coleman, J. E., and Steitz, T. A. (1997) *Nat. Struct. Biol.* **4**, 586–591

38. Sadofsky, M. J., Hesse, J. E., van Gent, D. C., and Gellert, M. (1995) *Genes Dev.* **9**, 2183–2199

39. Kale, S. B., Landree, M. A., and Roth, D. B. (2001) *Mol. Cell. Biol.* **21**, 459–466

40. Mo, X., Bailin, T., and Sadofsky, M. J. (2001) *Mol. Cell. Biol.* **21**, 2038–2047

41. Swanson, P. C. (2001) *Mol. Cell. Biol.* **21**, 449–458

42. Davies, D. R., Goryshin, I. Y., Reznikoff, W. S., and Rayment, I. (2000) *Science* **289**, 77–85

43. Sawchuk, D. J., Weis-Garcia, F., Malik, S., Besmer, E., Bustin, M., Nussenzweig, M. C., and Cortes, P. (1997) *J. Exp. Med.* **185**, 2025–2032

44. van Gent, D. C., Hiom, K., Paul, T. T., and Gellert, M. (1997) *EMBO J.* **16**, 2665–2670

45. Kwon, D., Chalmers, R. M., and Kleckner, N. (1995) *Proc. Natl. Acad. Sci. U. S. A.* **92**, 8234–8238

46. Tung, C. S., and Soumpasis, D. M. (1995) *J. Biomol. Struct. Dyn.* **13**, 577–582
Identification of Two Topologically Independent Domains in RAG1 and Their Role in Macromolecular Interactions Relevant to V(D)J Recombination

Janeen L. Arbuckle, LeAnn J. Fauss, Rosemarie Simpson, Leon M. Ptaszek and Karla K. Rodgers

J. Biol. Chem. 2001, 276:37093-37101.
doi: 10.1074/jbc.M105988200 originally published online July 30, 2001

Access the most updated version of this article at doi: 10.1074/jbc.M105988200

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 46 references, 24 of which can be accessed free at http://www.jbc.org/content/276/40/37093.full.html#ref-list-1