2938–2946 *Nucleic Acids Research, 1999, Vol. 27, No. 14*

A dimer of the lymphoid protein RAG1 recognizes the recombination signal sequence and the complex stably incorporates the high mobility group protein HMG2

Karla K. Rodgers*, Isabelle J. Villey[1,2], Leon Ptaszek[1], Elizabeth Corbett[1,2], David G. Schatz[1,2] and Joseph E. Coleman

Department of Molecular Biophysics and Biochemistry, [1]Section of Immunobiology and [2]Howard Hughes Medical Institute, Yale University, New Haven, CT 06520-8114, USA

Received March 8, 1999; Revised and Accepted May 21, 1999

ABSTRACT

RAG1 and RAG2 are the two lymphoid-specific proteins required for the cleavage of DNA sequences known as the recombination signal sequences (RSSs) flanking V, D or J regions of the antigen-binding genes. Previous studies have shown that RAG1 alone is capable of binding to the RSS, whereas RAG2 only binds as a RAG1/RAG2 complex. We have expressed recombinant core RAG1 (amino acids 384–1008) in *Escherichia coli* and demonstrated catalytic activity when combined with RAG2. This protein was then used to determine its oligomeric forms and the dissociation constant of binding to the RSS. Electrophoretic mobility shift assays show that up to three oligomeric complexes of core RAG1 form with a single RSS. Core RAG1 was found to exist as a dimer both when free in solution and as the minimal species bound to the RSS. Competition assays show that RAG1 recognizes both the conserved nonamer and heptamer sequences of the RSS. Zinc analysis shows the core to contain two zinc ions. The purified RAG1 protein overexpressed in *E.coli* exhibited the expected cleavage activity when combined with RAG2 purified from transfected 293T cells. The high mobility group protein HMG2 is stably incorporated into the recombinant RAG1/RSS complex and can increase the affinity of RAG1 for the RSS in the absence of RAG2.

INTRODUCTION

V(D)J recombination assembles the antigen-binding regions of the T cell receptor and immunoglobulin genes from component gene segments known as V (variable), J (joining) and, sometimes, D (diversity) (1). The precise location of the initial cleavage reaction is directed by a conserved recombination signal sequence (RSS) flanking each gene segment. The RSS consists of two well-conserved sequence elements, a heptamer and an A/T-rich nonamer, separated by a spacer of only nominal conservation whose length is either 12 ± 1 or 23 ± 1 bp. These two types of signals are referred to as 12- or 23-RSSs. Efficient recombination requires one of each type of RSS, a phenomenon known as the 12/23 rule (1).

The initial cleavage reaction of V(D)J recombination is catalyzed by a combination of RAG1 and RAG2, proteins encoded by the recombination activating genes (2,3). The cleavage reaction consists of two steps. First, a nick is made between the coding gene segment and the heptamer of the RSS and then, by direct transesterification, the nucleophilic 3'-OH on the coding end is coupled to the phosphate group between the RSS and the opposite strand of the coding segment (3,4). Therefore, the initial products are a 5'-phosphorylated double-strand break at the signal end and a covalently closed hairpin at the coding end. The final phase of V(D)J recombination involves the activity of additional factors, including enzymes with DNA repair activities which process and join the two coding ends (5).

Recent studies using gel shift assays with crosslinked proteins have shown that the RAG proteins bind to the RSS (6,7). RAG1 plays a major role in recognition of the RSS, as a region of RAG1 has been found through surface plasmon resonance and a mammalian one-hybrid assay to bind specifically to the nonamer of the RSS (8,9). These results have been confirmed through chemical footprinting and gel shift assays (7,10,11). This region of RAG1 is referred to here as the nonamer-binding domain (NBD). However, there are a number of aspects of the interaction of the RAG proteins with the RSS that need to be clarified. First, RAG2 is required along with RAG1 for the initial cleavage reaction, but binding of RAG2 to the RSS in the absence of RAG1 has not been detected. RAG2 has been postulated to enhance the binding of RAG1 to the RSS or to require the presence of RAG1 in order to form specific contacts to the RSS (7,11). Second, while RAG1 binds to the RSS, there is conflicting evidence as to whether RAG1 recognizes the heptamer with any specificity and concerning the degree of specificity RAG1 has for binding the nonamer (7,10,11). Third, possible oligomeric forms of the RAG proteins bound to the RSS have not been determined. In the present studies a bacterially expressed form of RAG1 was used that shows the cleavage activity expected of RAG1 expressed in higher

*To whom correspondence should be addressed at present address: Department of Biochemistry and Molecular Biology, The University of Oklahoma Health Sciences Center, Oklahoma City, OK 73190, USA. Tel: +1 405 271 2227; Fax: +1 405 271 3139; Email: karla-rodgers@ouhsc.edu
Leon Ptaszek, MD PhD Student at Mount Sinai School of Medicine, City University of New York, New York, NY 10029, USA

eukaryotes. Using electrophoretic mobility shift assays (EMSA) we show that core RAG1 (residues 384–1008) has significant specificity for both the nonamer and heptamer, but has greater specificity for the nonamer. In addition, we demonstrate that core RAG1 is a homodimer in solution and, correspondingly, the homodimer is the initial species of core RAG1 that binds to the RSS. Finally, we demonstrate that HMG2 enters into a stable complex with RAG1 and the RSS and enhances the binding interaction with 23-RSS.

MATERIALS AND METHODS

Protein expression and purification

Murine core RAG1 (residues 384–1008), fused at its N-terminus to the maltose binding protein (MBP) and at its C-terminus to a six histidine tag, was encoded by the plasmid pCJM233. This plasmid was created by inserting the appropriate portion of the RAG1 gene into the *Bam*HI and *Xho*I sites of the polylinker of pMH6. pMH6 was derived by insertion of the following oligonucleotide into the *Sal*I and *Hin*dIII sites of pMal-c2 (New England Biolabs): d(TCGACCTCGAGCATCACCATCACCATCACTAATAGGCA) annealed to d(AGCTTGCCTATTAGTGATGGTGATGGTGATGCTCGAGG). The fusion protein was expressed in *Escherichia coli* BL21 cells grown in Luria broth at 25°C to an OD_{600} of ~1.0, 0.1 mM $ZnCl_2$ was added and protein expression induced by the addition of 1 mg/l IPTG. Cell growth was continued for an additional 15 h to an OD_{600} of 4–5.

The protein was released by sonication and purified over an amylose column using purification buffer (20 mM Tris–HCl, pH 8.0, 20 μM $ZnCl_2$, 5 mM 2-mercaptoethanol and 10% glycerol) plus 200 mM NaCl. The MBP–core RAG1 protein was eluted from the column with the above buffer plus 10 mM maltose. The protein was then purified through a heparin sulfate column eluting with a NaCl gradient from 0.1 to 1 M. As a final purification step the protein was concentrated and run through a Sepharose 4B gel filtration column. The protein was judged to be >95% pure based on Coomassie blue staining of SDS–PAGE gels. The GST–RAG1 (RAG1ΔN, amino acids 330–1040) and GST–RAG2 (RAG2ΔC, amino acids 1–388) proteins were expressed by transient transfection in 293T cells and purified on glutathione–agarose beads as described previously (8). Recombinant HMG2 protein, lacking the C-terminal acidic region, was expressed in bacteria and purified as described previously (12).

Analytical gel filtration chromatography

MBP–core RAG1 in the purification buffer plus 200 mM NaCl was separated by chromatography through a Pharmacia Superdex 200 column using a BioRad Biologic chromatography system. Molecular weight standards from Sigma ranging in mass from 3 to 600 kDa were then separated in a run immediately after the RAG1 protein separation. The elution times of the standards were used to construct a standard linear curve from which the MBP–core RAG1 molecular mass was determined.

Cleavage assays

The activity of the different preparations of RAG proteins was assessed by cleavage of double-stranded oligonucleotides containing 12- and 23-RSS sequences. The 12-RSS and 23-RSS substrates were prepared as described (13). The sequences of the top strand oligonucleotides were as follows: 12-RSS, d(GATCTGGCCTGTCTTACACAGTGATACAGACCTTAACAAAAACCTGCACTCGAGCGGAG); 23-RSS, d(GATCTGGCCTGTCTTACACAGTGATGGAAGCTCAATCTGAACTCTGACAAAAACCTCGAGCGGAG). Cleavage reactions were performed for 1 h at 37°C, using 50 fmol of the ^{32}P-end-labeled gel-purified RSS substrate and 2 pmol of either RAG1 (GST–core RAG1 or MBP–core RAG1) and RAG2 (GST–core RAG2). The reactions were 50 μl in volume containing 2 mM $MnCl_2$, 5% glycerol, 50 mM NaCl, 0.1 mg/ml BSA, 20 mM Na-HEPES, pH 7.5, 10 μM $ZnSO_4$ and 2 mM DTT. The reactions were stopped by addition of 150 μl of 40 μg/ml proteinase K, 0.5% SDS, 5 mM EDTA, 20 mM Na-HEPES, pH 7.5, and incubated for 1 h at 55°C. The DNA was extracted once with an equal volume of phenol:chloroform (1:1) and precipitated with ethanol. The cleavage products were separated on an 8% denaturing polyacrylamide gel.

Coupled cleavage reactions were performed with the pC317-derived cleavage substrate for 2 h at 37°C, exactly as described previously (12) except that 5 mM $MgCl_2$ was substituted for 10 mM magnesium acetate.

Oligonucleotide DNA substrates for EMSA

The wild-type 12-RSS and 23-RSS substrates for EMSA were prepared by annealing the complementary oligonucleotides. The sequences of the top strand oligonucleotides were as follows: 12-RSS, d(CCGGACGCGTCGACGTCTTACACAGTGATACAGCCCTGAACAAAAACCGGATCC); 23-RSS, d(GATCCGGGCACAGTGGCCATGGTGGCTTGTCTGGCTGTACAAAAACCGTCGACA). The mutant 12-RSSs consisted of changes to the nonamer and heptamer sequences of the wild-type 12-RSS as outlined in Table 1. All of the duplexes were labeled with ^{32}P using [γ-^{32}P]ATP and T4 polynucleotide kinase.

EMSA

MBP–core RAG1 was incubated with 12- or 23-RSS for 30 min at room temperature prior to electrophoresis on 6% polyacrylamide gels using 45 mM Tris–borate as the electrophoresis buffer. The binding buffer contained 10 mM Tris–HCl, pH 7.5, 50 mM NaCl, 5 mM $MgCl_2$, 2 mM DTT and 6% glycerol. In addition, 0.5 μM 26 base single-stranded oligonucleotide, d(TTTGGTCGATATCCATATGGGGGGAC), was added as a non-specific competitor.

EMSA of Factor Xa-proteolyzed MBP–core RAG1 samples bound to the 12-RSS

Different amounts of Factor Xa, from 0.02 to 0.2 μg, were incubated with 20 μg of MBP–core RAG1 at 4°C for 15 h. The samples were immediately used in EMSA with the 12-RSS. The conditions for EMSA using the Factor Xa-proteolyzed samples were the same as listed above, except that the electrophoresis buffer was 89 mM Tris–borate and the binding buffer contained 150 mM NaCl.

Binding curve equations for gel shift titrations and competition assays

Gel shift titrations of MBP–core RAG1 to the RSSs were analyzed by phosphorimaging of dried gels. The fraction of

free DNA versus protein concentration was plotted and the data fitted to the following equation:

$$\text{fraction unbound DNA} = K_d^n/(K_d^n + P^n) \quad \mathbf{1}$$

where K_d is the dissociation constant, P is the protein concentration and n is the Hill coefficient.

Competition assays of MBP–core RAG1 were accomplished by adding 30 nM protein to a mixture of 1 nM labeled wild-type RSS and increasing concentrations of unlabeled competitor DNA. The intensity of the band corresponding to the protein–DNA complex formed in the presence of competitor DNA was compared to the complex formed in its absence. The ratio of ^{32}P-labeled RSS/protein complex in the absence of competitor DNA over that in the presence of competitor DNA was plotted versus the increasing concentration of unlabeled competitor DNA. The data were fitted to the following equation:

$$PD_+/PD_- = K_2(K_1 + P_T)/[K_2(K_1 + P_T) + K_1(C_T)] \quad \mathbf{2}$$

where PD_+ and PD_- are the amount of labeled complex in the presence and absence of competitor DNA, respectively, K_1 is the dissociation constant of the wild-type complex, K_2 is the dissociation constant of the competitor complex, P_T is the total protein concentration and C_T is the total concentration of the unlabeled competitor DNA.

Equation **2** was derived as follows. Without competitor DNA the dissociation constant of MBP–core RAG1 with the labeled wild-type RSS is $K_1 = [P_T(D_T - PD_-)]/PD_-$, since [free DNA] = $(D_T - PD_-)$ and D_T is the total labeled DNA. This can be rearranged to $PD_- = P_TD_T/(K_1 + P_T)$. In the assay the competitor complex is formed by titration of competitor DNA to the protein held at a constant concentration, such that $K_2 = (P_T - PC)(C_T)/PC$, where C_T is the total competitor DNA concentration and K_2 is the dissociation constant of the competitor complex. This can be rearranged to $PC = P_TC_T/K_2 + C_T$. In the presence of competitor DNA the dissociation constant of the labeled wild-type complex is $K_1 = (P_T - PC)(D_T - PD_+)/(PD_-)$, where D_T is the concentration of labeled wild-type RSS. Replacing PC with the expression above leads to $PD_+ = K_2(P_TD_T)/K_2(K_1 + P_T) + K_1C_T$. The ratio of the expressions for PD_+ over PD_- leads to equation **2**. However, with wild-type RSS as the competitor DNA, $K_1 = K_2$ and equation **2** simplifies to

$$PD_-/PD_+ = (K + P_T)/(K + P_T + C_T) \quad \mathbf{3}$$

The value for K determined with wild-type RSS as the competitor was used as the value of K_1 in equation **2** in experiments with mutant RSSs as competitors.

Determination of protein concentration in EMSA

The concentration of MBP–core RAG1 was determined by measuring the absorbance at 280 nm using an extinction coefficient of 123 $mM^{-1}cm^{-1}$. The DNA-binding activity of proteins is often significantly less than 100% (14). To accurately quantitate binding equilibria it is necessary to determine percent binding activity of the protein. In these studies, the percent binding activity was determined using data from the competition assay with wild-type RSS as the competitor DNA. The summation $K + P_T$ from equation **3** can be found where P_T is the total active protein concentration. The relationship between K and P_T can be found from the midpoint of the gel shift titrations of MBP–core RAG1 to wild-type RSS, such that the variables in the summation above can be determined. The percent activity is the ratio of P_T determined in this manner over the concentration of protein determined from the absorbance at 280 nm. The percent binding activity for MBP–core RAG1 used in the EMSA experiments was determined to be 55%. The concentrations of protein noted in the gel shift assays have been adjusted accordingly.

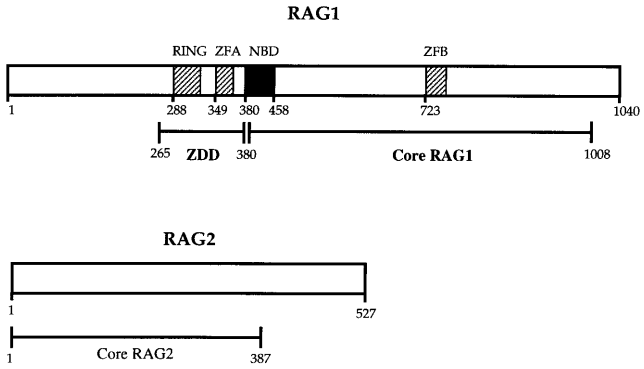

Figure 1. Schematic of full-length RAG proteins. A bar represents the RAG1 protein with the known domains highlighted. The minimal core region that is necessary for catalysis, as well as a zinc dimerization domain (ZDD), is marked. The DNA-binding domain that recognizes the nonamer of the RSS (NBD) is shown as a solid box. Hatched boxes represent zinc-binding motifs, including a RING finger and zinc finger (ZFA) in the ZDD region and a zinc finger (ZFB) within the core region of RAG1. The core region of RAG2 is marked.

Atomic absorption spectroscopy

The ratio of zinc ions bound to the MBP–core RAG1 was determined by atomic absorption spectroscopy using an Instrumentation Laboratory (Lexington, MA) IL157 spectrometer. Concentrations of protein used in these measurements ranged from 3 to 6 μM.

RESULTS

Catalytic activity of bacterially expressed RAG1

The core regions of RAG1 (residues 384–1008) and RAG2 (residues 1–387) are sufficient to catalyze the initial cleavage steps in the V(D)J recombination reaction (Fig. 1; 15–18). All previous studies have used RAG1 protein expressed in mammalian cell lines or in baculovirus-infected insect cells. There is no evidence of important post-translational modifications of RAG1 necessary for its activity, thus there is nothing obviously precluding expression of active protein in bacterial cells. We therefore expressed murine core RAG1 in *E.coli* as a fusion protein with MBP, yielding high levels of soluble protein.

We tested the catalytic activity of bacterially expressed core RAG1, in combination with RAG2 expressed in 293T cells, in two separate cleavage assays. First, cleavage activity was tested on an isolated 12- or 23-RSS in the presence of Mn^{2+} (13,19–21). With an isolated RSS as the substrate, both steps of the cleavage reaction were accomplished using bacterially

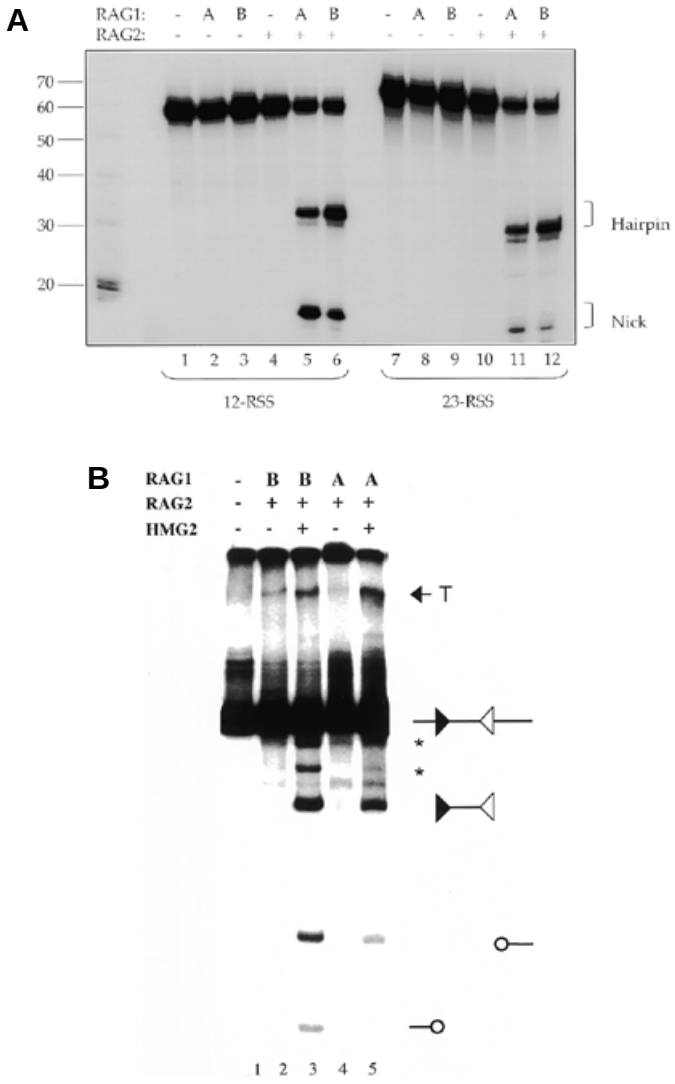

Figure 2. *Escherichia coli*-expressed RAG1 is active in cleavage assays with mammalian-expressed RAG2. (**A**) Isolated RSS cleavage reactions in Mn^{2+}. The substrates used were a 12-RSS (lanes 1–6) or 23-RSS (lanes 7–12). The cleavage products were analyzed on a denaturing polyacrylamide gel. A, MBP–RAG1 expressed in *E.coli*; B, GST–RAG1 expressed in 293T cells. The positions of the expected hairpin and nicked products are marked by brackets. The positions of molecular weight standards are indicated in nucleotides. Note that the labeled cleavage products produced from the two DNA substrates are the same size. Irregularities in the gel caused the 23-RSS cleavage products to migrate slightly more rapidly. (**B**) Coupled cleavage reactions in Mg^{2+}. The substrate, described in Agrawal *et al*. (12), contained a 12-RSS (black triangle) and a 23-RSS (white triangle). The reactions were analyzed on a native polyacrylamide gel. A and B as in (A). Cleavage by the RAG proteins, in the presence of HMG2 (lanes 3 and 5), generates a signal end/signal end fragment and two smaller fragments containing the hairpin coding ends (circles). Asterisks indicate the bands resulting from single cleavage at the 12- or 23-RSS. T represents the product of intramolecular transposition. The smallest cleavage product in lane 5 was visible with longer exposure times.

expressed RAG1 (Fig. 2A, lanes 5 and 11). As controls, the cleavage assay was done using core RAG1 expressed in 293T cells (lanes 6 and 12). No cleavage was observed when only

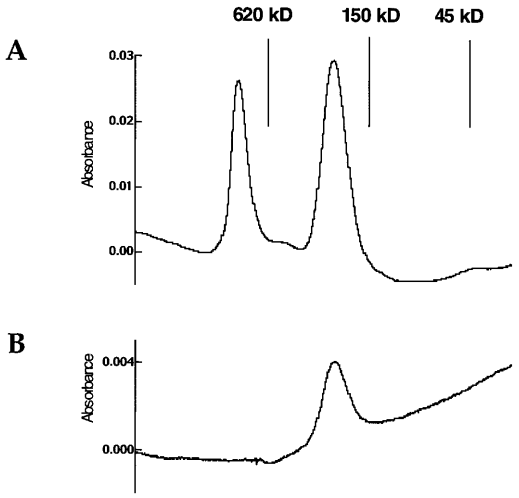

Figure 3. Elution profiles of MBP–core RAG1 from gel filtration chromatography. (**A**) The elution profile of ~30 µg of MBP–core RAG1 from gel filtration chromatography through a Pharmacia Superdex 200 column. MBP–core RAG1 is present in fractions corresponding to both peaks. The elution peaks of molecular weight standards are indicated in the figure. Protein corresponding to the first peak eluted in the void volume and most likely represents aggregated and/or misfolded protein. Protein from the second peak eluted at a molecular weight of 246 kDa, consistent with a dimer of MBP–core RAG1 (dimer molecular weight 228 kDa). (**B**) Protein that eluted in fractions corresponding to the second peak (as a dimer) were pooled and after 24 h loaded onto the same column.

one of the RAG proteins was present (lanes 2–4 and 8–10). Second, we found that recombinant RAG1 expressed in *E.coli* can replace mammalian RAG1 in a coupled cleavage assay with a substrate containing both a 12- and 23-RSS in the presence of Mg^{2+} (Fig. 2B). This reaction also required HMG2, as high mobility group proteins have previously been found to stimulate coupled cleavage (22,23). The coupled cleavage reaction demonstrates regulated cleavage activity that is more representative of physiological conditions than the reactions presented in Figure 2A. For both bacterial and mammalian expressed RAG1, formation of the coupled cleavage product is significantly enhanced in the presence of HMG2 over single cleavage at either RSS (lanes 3 and 5). Further evidence of coupled cleavage by the bacterially expressed RAG1 is the formation of an intramolecular transposition product (T in Fig. 2B), which is likely to form in the context of the signal end synaptic complex (12).

Core RAG1 is a dimer

The elution profile of MBP–core RAG1 fusion protein from analytical gel filtration chromatography on a Pharmacia Superdex 200 column yielded two peaks, demonstrating that over half of the sample eluted at a molecular weight consistent with a dimer (Fig. 3A). The remainder eluted in the void volume, most likely as an aggregated form. Rechromatography of the protein that eluted as a dimer yielded only the second peak, indicating that the dimeric protein is not in rapid equilibrium with a higher order oligomer or aggregate (Fig. 3B). Under these conditions of cell growth and purification, >50% of the protein exists as a stable dimer, with the remainder aggregated

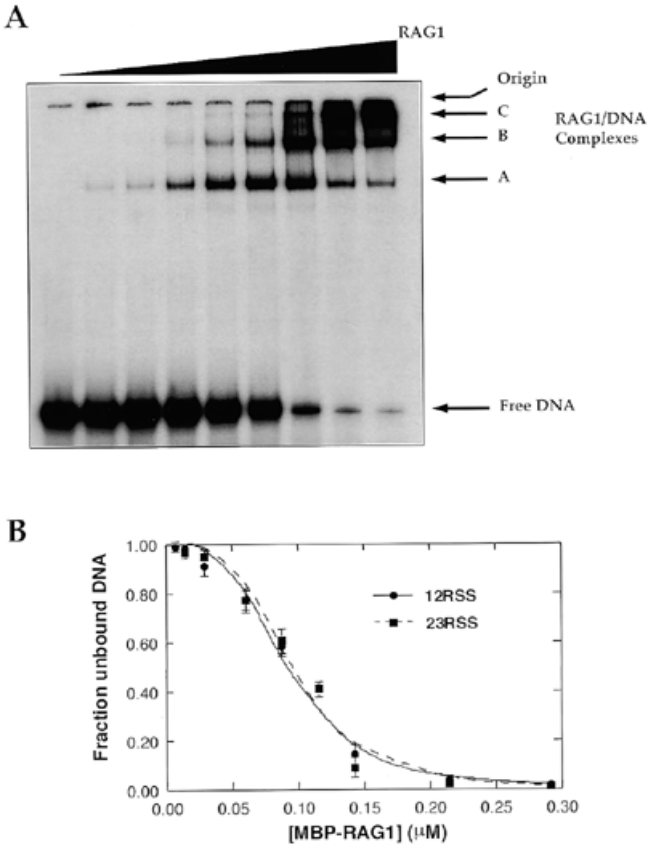

Figure 4. Gel shift titration of RAG1 to a single RSS. (**A**) 12-RSS was titrated with increasing concentrations of MBP–core RAG1 and subjected to electrophoresis on a 6% non-denaturing gel. The concentration of MBP–core RAG1 varied from 0 to 0.3 μM. Three specific complexes between MBP–core RAG1 are labeled A, B and C. (**B**) Quantitation of the gel shift titrations of MBP–core RAG1 to 12- versus 23-RSS. The fraction of DNA unbound is plotted versus concentration of MBP–core RAG1. The data is represented by circles for 12-RSS and squares for 23-RSS. Curve fitting to the data is as described in the text.

or misfolded. In the studies described here, gel filtration was performed with a Sepharose 4B column and the isolated dimer was employed in all the experiments.

Zinc content of core RAG1

The zinc content of core RAG1 was measured by atomic absorption spectroscopy and was found to bind 2 ± 0.2 zinc ions. Previously, we have called attention to a sequence (amino acids 723–756) in the RAG1 core, referred to as ZFB, which resembles the consensus sequence for a classical zinc finger (24). The potential zinc ligands are C^{727}, C^{730}, H^{744} and H^{750}. Assuming that one zinc is bound to ZFB, a site for co-ordination of the second zinc ion is not obvious, as an additional known zinc-binding motif is not clearly present in core RAG1. There are, however, 13 Cys and 19 His residues in murine core RAG1, excluding those in ZFB, some of which might participate as zinc ligands.

Core RAG1 binds the RSS

To characterize the DNA-binding properties of RAG1 with the RSS, oligonucleotides containing a 12- or 23-RSS were titrated with MBP–core RAG1 and the bound complexes resolved by EMSA. Three distinct complexes, A, B and C, of MBP–core RAG1 bound to RSS are seen to shift relative to the free DNA (Fig. 4A). The slower migrating complexes occurring at higher protein concentrations in the titration correspond to the disappearance of the faster migrating complexes and appear to represent higher order RAG1 complexes bound to the RSS. The titration assays were done in the presence of a 26 base non-specific single-stranded oligonucleotide, as done previously (6,7), at a concentration of 250-fold (per base) over the labeled RSS. Similar results were seen using the double-stranded competitor poly(dI-dC) (data not shown). In the absence of the non-specific oligonucleotide, RAG1 formed additional complexes with the RSS resulting in higher order adducts trapped in the well. The buffer used in the titrations included 5 mM Mg^{2+}, although we also found equivalent binding of the MBP–core RAG1 to the RSS in the absence of divalent metal cations (data not shown). The binding curve in Figure 4B, determined using equation **1**, illustrates that RAG1 binds to both 12- and 23-RSS with the same affinity, K_d = 91 ± 7 nM for the 12-RSS versus K_d = 95 ± 6 nM for the 23-RSS. The Hill coefficient, n, is 3 ± 0.3 in each case (equation **1**).

Core RAG1 binds as a dimer to the RSS as shown by heterodimer formation

To determine the stoichiometry of the RAG1/RSS complexes in Figure 4, we performed EMSA using complexes formed with mixtures of MBP–core RAG1 and core RAG1. In this case the number of shifted bands in the gel shift assay is $n + 1$, where n is the number of protein molecules bound to the DNA. Core RAG1 was generated using the Factor Xa cleavage site located between the MBP and core RAG1 regions of the fusion protein. By varying the concentration of Factor Xa incubated with the fusion protein, we obtained samples that were 50 to nearly 100% proteolyzed (Fig. 5A). The 12-RSS was incubated with the unproteolyzed and the two proteolyzed MBP–core RAG1 samples and the complexes resolved by EMSA (Fig. 5B). The concentration of protein used was that which results in formation of complex A as the major species (Fig. 4A, lane 5).

The unproteolyzed fusion protein yields a shifted band at a similar migration position as that of complex A in Figure 4 (Fig. 5B, lane 1). The 100% proteolyzed sample (lane 3) yields a complex with a significantly faster mobility than MBP–core RAG1 (lane 1), consistent with a lower molecular weight for the protein after removal of MBP. Finally, the 50% proteolyzed sample shows the presence of three shifted complexes in EMSA (lane 2), with an intermediate band between the bands present in lanes 1 and 3 as expected if a heterodimer of MBP–core RAG1/core RAG1 is bound to the RSS. Thus, the three bands bound to the RSS in Figure 5B are assigned as the MBP–core RAG1 homodimer, the MBP–core RAG1/core RAG1 heterodimer and the core RAG1 homodimer. These data fit with the assignment of the fastest migrating complex in Figure 4A as a homodimer of MBP–core RAG1 bound to the RSS. The stoichiometries of the slower migrating complexes in Figure 4 are difficult to determine, as multiple bands from proteolysis of MBP–core RAG1 would overlap between the different complexes. As there is no evidence of a monomeric form of RAG1 complexed to RSS, a plausible model is that the

Table 1. The affinity of MBP–core RAG1 for wild-type and mutant RSSs

12-RSS	Heptamer sequence	Nonamer sequence	K_d (nM)[a]	Difference from wild-type
WT	CACAGTG	ACAAAAACC	100	–
MH	GAGCAGT	Wild-type sequence	310	3.1-fold
MN1	Wild-type sequence	ACAAGGACC	820	8.2-fold
MN2	Wild-type sequence	AGTCTCTGA	880	8.8-fold
M(N+H)	GAGCAGT	ACAAGGACC	720	7.2-fold

[a]The standard error is within ±12% for all values.

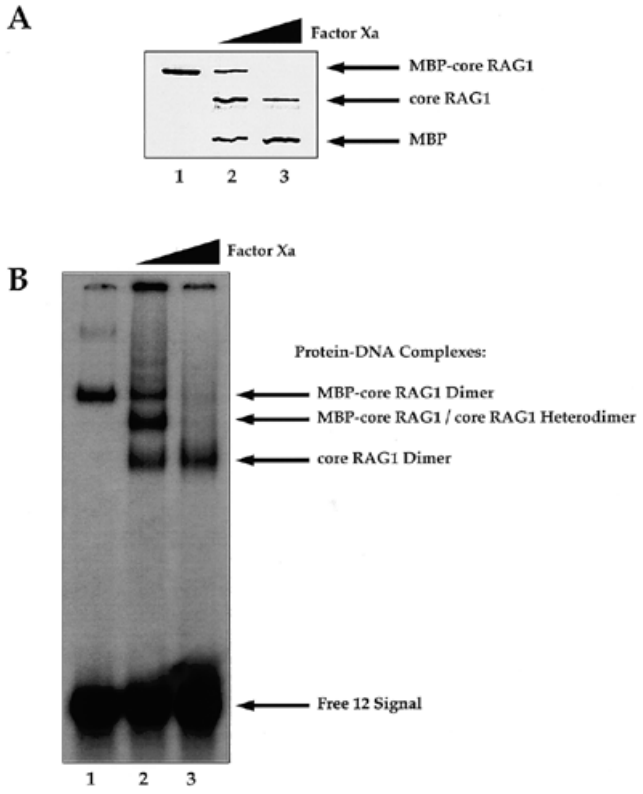

Figure 5. Determination of the stoichiometry in the protein–DNA complex. (**A**) SDS–PAGE of the MBP–core RAG1 that had been proteolyzed with no or differing concentrations of Factor Xa. Lane 1, unproteolyzed MBP–core RAG1; lane 2, ~50% proteolyzed MBP–core RAG1; lane 3, >95% proteolyzed MBP–core RAG1. The identities of the bands are indicated as either MBP–core RAG1, core RAG1 or MBP. (**B**) EMSA of 12-RSS with the Factor Xa proteolyzed samples of MBP–core RAG1 shown in (A). The protein samples used in the different lanes correspond to the lanes in the SDS–PAGE gel shown in (A).

three complexes in Figure 4 are one, two and three dimers of MBP–core RAG1 bound to the RSS, respectively.

In these experiments it was necessary to use higher ionic strengths to observe the core RAG1 complex with the RSS. This may be due to decreased solubility of core RAG1 in the absence of the N-terminal MBP region at low ionic strength. Increased ionic strength does not significantly affect the binding affinity of MBP–core RAG1 for the RSS, since the MBP–core RAG1/RSS complex is relatively insensitive to changes in ionic strength in the range used here (data not shown).

RAG1 binds specifically to both the nonamer and heptamer sequences of the RSS

The RSS contains two conserved regions, including a dyad-symmetric heptamer, CACAGTG, and an A/T-rich nonamer, ACAAAAACC. It has been previously established that a putative helix–turn–helix sequence near the N-terminal portion of the RAG1 core, with homology to the Hin invertase DNA-binding domain, binds specifically to the nonamer of the RSS (8,9). To confirm these results for the bacterially expressed RAG1 and to determine if there is any specificity for the heptamer, we compared formation of the wild-type RSS/core RAG1 complex in the presence of unlabeled wild-type versus mutant sequences of the RSS (Fig. 6 and Table 1). The protein was added last to a mixture of labeled wild-type and unlabeled wild-type or mutant RSS to minimize effects from exchange of protein–DNA complexes and the protein concentration adjusted to form one dimer of RAG1 bound to the RSS as the major complex. The labeled wild-type RSS/MBP–core RAG1 complex forms more readily in the presence of a mutated unlabeled RSS nonamer sequence than in the presence of unlabeled wild-type RSS (Fig. 6A).

The ratio of labeled complex in the presence of unlabeled competitor DNA over that in the absence was plotted versus concentration of unlabeled competitor DNA (Fig. 6B). The data were fitted assuming simple binding equilibria using equation **2**. Quantitation of the labeled complex in the presence of the various mutant RSSs shows that both the mutant nonamer and heptamer sequences are less effective than the wild-type sequence in competing with the labeled wild-type complex. For example, the K_d value for binding of RAG1 to the mutant nonamer sequences is ~8- to 9-fold greater than to the wild-type RSS (Table 1). MN2 and MN1 yield similar dissociation constants, even though the MN2 mutation is more disruptive of the nonamer sequence. The changes in MN1 are therefore sufficient to disrupt specific binding of RAG1 to the nonamer. The K_d for the heptamer mutation is 3-fold greater than for the wild-type, showing that RAG1 binds specifically to the heptamer sequence, although not with as high a specificity as for the nonamer sequence. Interestingly, the combined mutations of the nonamer and heptamer, M(N+H), results in a similar K_d value as MN1 and MN2. Instead, one might have

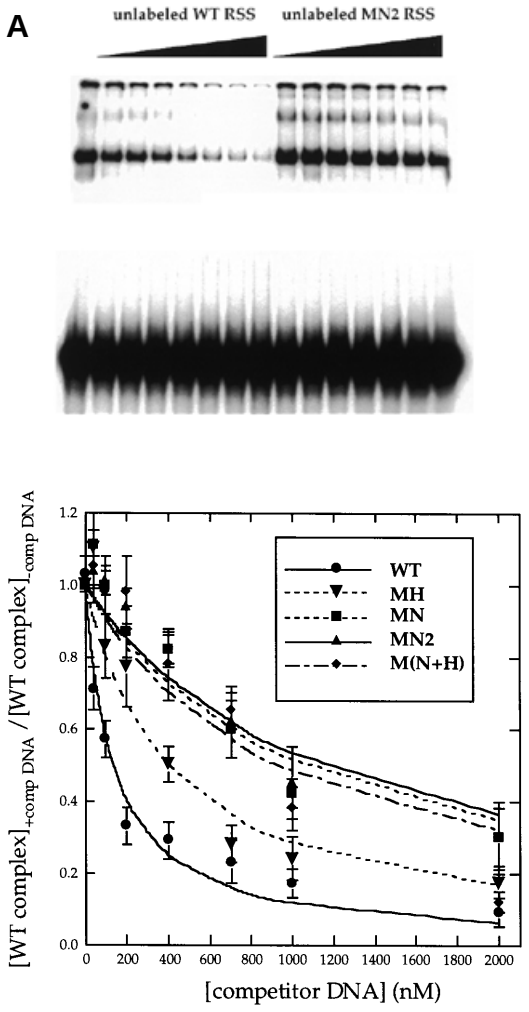

Figure 6. EMSA of MBP–core RAG1 to wild-type 12-RSS in the presence of mutant competitor RSSs (competition assays). (**A**) A representative competition assay. Each sample contains 40 nM MBP–core RAG1 bound to 1 nM labeled wild-type 12-RSS. Lane 1, no unlabeled competitor DNA; lanes 2–8; increasing amounts of unlabeled wild-type 12-RSS; lanes 9–15, increasing amounts of unlabeled 12-RSS where the nonamer sequence had been mutated (Table 1). The amounts of unlabeled DNA ranged from 40 to 2000 nM in both cases. (**B**) Plot of the ratio of RAG1 bound to wild-type 12-RSS in the presence over that in the absence of competitor DNA versus concentration of competitor DNA. The sequences of the mutant RSSs are listed in Table 1. Curve fitting to the data was done as described in the text.

expected an additive effect from the nonamer and heptamer, resulting in RAG1 binding with the weakest affinity to M(N+H) RSS. Several factors could account for this, including cooperative binding effects or a preference of RAG1 to bind to the nonamer prior to the heptamer.

HMG2 increases the binding affinity of core RAG1 for the 23-RSS

Although RAG1 and RAG2 are the only lymphoid-specific factors required for the initial stages of V(D)J recombination, recent studies have shown that high mobility group proteins, HMG1 or HMG2, enhance the cleavage activity on an isolated 23-RSS by up to 10-fold, with no significant effects observed for an isolated 12-RSS (22,23). In the presence of HMG proteins a mixture of RAG1/RAG2 bound with increased affinity to a 23-RSS, with no corresponding increase for binding to the 12-RSS (22). We asked whether HMG2 had similar effects on the affinity of RAG1 binding to a 12- versus a 23-RSS, since RAG1 alone can bind to the RSSs with specificity for both the nonamer and heptamer. To examine the effect of HMG2 on the affinity of MBP–core RAG1 to a 12- versus a 23-RSS, a gel shift titration was conducted in the presence of HMG2 (Fig. 7A). The presence of HMG2 results in a slight supershift of all bands, particularly evident for the fastest migrating complex, indicative that HMG2 is incorporated into the RAG1/23-RSS complex (Fig. 7A) as well as into the RAG1/12-RSS complex (data not shown). In the presence of HMG2 the titration curves illustrate that the affinity of RAG1 for a 23-RSS is increased by 2-fold over that to a 12-RSS, (K_d = 121 ± 6 nM for 12-RSS versus 56 ± 8 nM for 23-RSS) (Fig. 7B). Thus, although RAG1 binds with an equal affinity to a 12- versus a 23-RSS (Fig. 4B), the presence of HMG2 leads to a differentiation in the binding of RAG1 to each RSS. To confirm that the supershift is due to HMG2, we titrated pre-formed RAG1/RSS (either 12- or 23-RSS) complexes with HMG2 (Fig. 7C). The supershifted species appeared only with the addition of HMG2 and the intensity of the band increased with increasing amounts of HMG2. The results in Figure 7 demonstrate that HMG2 exerts at least some portion of its effects on RSS binding by influencing the RAG1/RSS interaction, independent of the presence of RAG2.

DISCUSSION

We have shown that recombinant RAG1 expressed in *E.coli* exhibits similar DNA cleavage activity, when combined with RAG2, as RAG1 purified from mammalian cell lines. The activity of the purified recombinant RAG1 was similar to that of RAG1 purified from transiently transfected 293T cells both in cleavage of substrates containing isolated RSSs in Mn^{2+} and coupled cleavage of substrates containing both 12- and 23-RSSs in the presence of HMG2 and Mg^{2+} (Fig. 2). These are the first results that show that a recombinant RAG protein expressed in bacterial cells retains catalytic activity.

Initial characterization of the bacterially expressed core RAG1 shows that it binds two zinc ions. One zinc ion would be expected to bind to the ZFB zinc finger, whereas the location of the second zinc ion is not certain. These sites combined with those present in the N-terminal non-core region (25) illustrate that full-length RAG1 contains at least six zinc-binding sites whose direct contribution to V(D)J recombination remains to be determined. Within the non-core region, zinc sites have been proposed to exist within the first 250 amino acid region of the full-length protein (26), although this has yet to be experimentally determined. In addition, the crystal structure for the zinc dimerization domain (ZDD), just N-terminal to the core region, reveals that this domain coordinates four zinc ions (25).

Using gel filtration chromatography we found that the core of RAG1 is capable of self-associating to form a stable homodimer (Fig. 3). The region of the core that mediates dimer formation has yet to be identified. Previously the ZDD of RAG1 was found to dimerize readily in solution and in a crystal form (24,25). The formation of a compact dimeric structure by

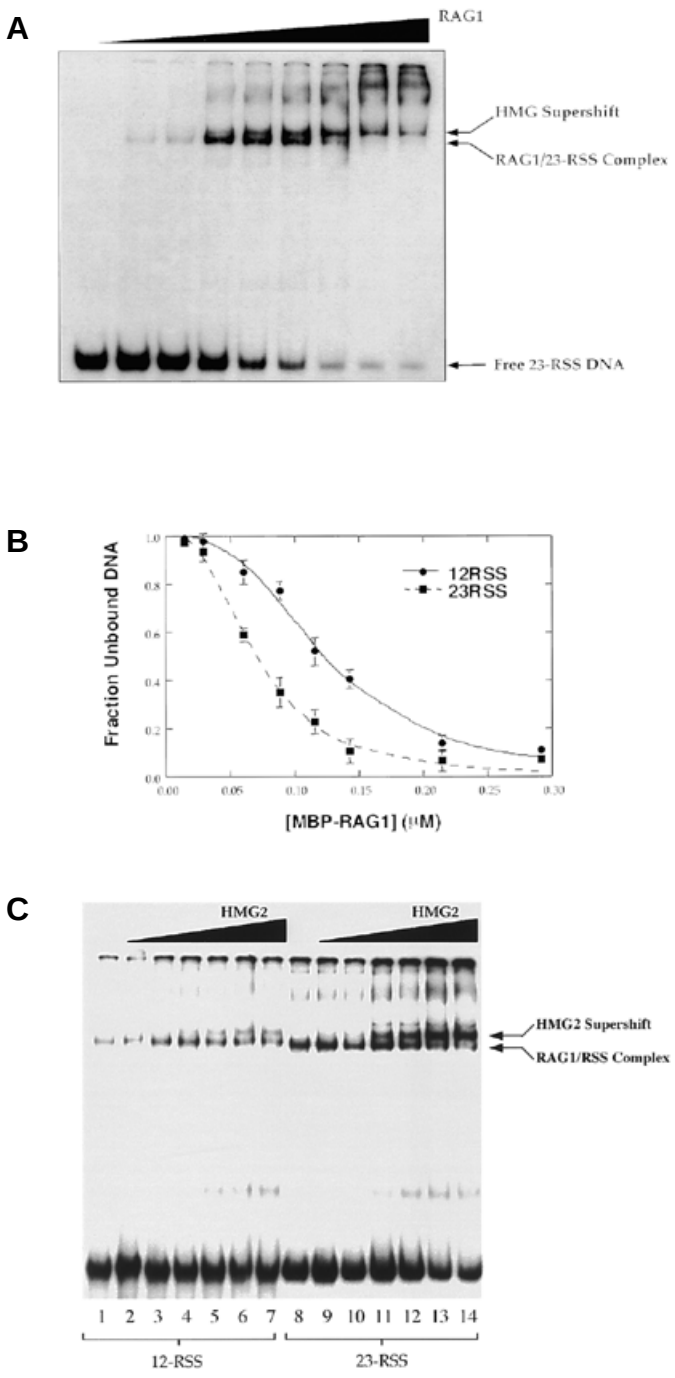

Figure 7. The effect of HMG2 on the affinity of MBP–core RAG1 for 12- and 23-RSSs. (**A**) EMSA of MBP–core RAG1 binding to 23-RSS in the presence of HMG2. 23-RSS (1 nM) was titrated with MBP–core RAG1 (0–0.3 μM) in the presence of 0.2 μM HMG2. A supershift arising from the incorporation of HMG2 into the MBP–core RAG1/23-RSS complex is marked by an arrow. (**B**) Titration curves of MBP–core RAG1 to 12- versus 23-RSS in the presence of HMG2. The data was fitted to equation **1**. (**C**) Titration of HMG2 (50 nM–0.5 μM) into samples containing 40 nM MBP–core RAG1 and 1 nM 12- or 23-RSS. Lanes 1 and 8 contain no HMG2. Bands resulting from the formation of a RAG1/RSS complex and supershift due to the incorporation of HMG2 are marked by arrows.

the ZDD (residues 264–380) requires the immediately adjacent residues of the two core regions to exit in opposite directions (25). This raises interesting structural questions about how the ZDD relates to the core region and the configuration of the dimeric full-length protein.

We determined several properties of the RAG1/RSS complex, including affinity and specificity of binding, as well as stoichiometry of the protein–DNA complex. Using EMSA, we observed multiple species of RAG1 bound to a single 12- or 23-RSS. In addition, complete titration curves were obtained, as nearly all the DNA was complexed at the higher concentrations of RAG1. Furthermore, RAG1 was found to bind with similar affinities to the 12- and the 23-RSS (Fig. 4B), which is in contrast to the results seen using both RAG proteins, where binding appeared to be tighter to the 12- over the 23-RSS (6). This difference may be explained by the possibility that RAG2 increases the affinity of RAG1 for the 12-RSS, but not the 23-RSS. Previously, *in vivo* mammalian one-hybrid experiments yielded results consistent with this possibility (9).

A homodimer of core RAG1 is the initial species that binds to the RSS, as determined by partial proteolysis of the MBP–core RAG1 protein with Factor Xa (Fig. 5), and is consistent with very recent findings using co-expressed core RAG1 molecules fused to one or two MBP molecules (27). This is in contrast to previous results where two shifted complexes were proposed to be due to monomer and then dimer binding to the RSS (13). These results were inferred from the appearance of two bands in a native gel after crosslinking core RAG1 with glutaraldehyde. We believe that our method more directly determined the stoichiometry of the protein–DNA complex.

Besides the specificity of the NBD of RAG1 for the RSS nonamer, no other regions of either RAG protein has been shown to demonstrate specific binding to the RSS. For instance, there are contradictory results concerning whether RAG1 binds specifically to the heptamer sequence of the RSS. Chemical modification interference assays showed no evidence of RAG1 interactions with the heptamer (11). In contrast, DNase I footprinting experiments, where RSS was titrated with increasing amounts of RAG1, clearly showed protection of the heptamer region (10). Consistent with the latter results, EMSA experiments revealed a decrease in binding of RAG1 to mutant over wild-type heptamer sequences (7). Here we show, using competition assays, that RAG1 binds specifically to both regions of the RSS, with greater specificity for the nonamer and a smaller, but significant, specificity for the heptamer (Fig. 6). This is supported by recent UV crosslinking experiments that reveal RAG1 in close proximity to several residues of the heptamer and coding flank (21). As there is good evidence for the location of the nonamer DNA-binding domain of RAG1, the question now arises as to the location of the heptamer-binding region within core RAG1.

A DNA-bending protein, HMG2, differentially affects the affinity of bacterially expressed core RAG1 for the 23- over the 12-RSS, as is seen with both RAG proteins. It has been proposed that binding to the 23-RSS for the RAG1/RAG2 complex is less favorable than for the 12-RSS, due to the extra helical turn between the heptamer and nonamer regions, and may require bending of the 23-RSS to achieve maximum interactions between the protein and DNA. The association of HMG proteins would stabilize the bent DNA, thereby increasing the affinity of the RAG proteins for the longer RSS (22). In the case of RAG1 alone, HMG2 leads to a differentiation between the affinity of RAG1 for the two RSSs. This is despite the fact

that HMG2 can be incorporated into either RSS complex. Importantly, these data reveal for the first time the stable inclusion of HMG protein in the precleavage RAG/RSS complex and demonstrate that HMG2 can operate solely through RAG1 to influence DNA binding. In a close parallel with our findings, HMG1 and HMG2 have been shown to enhance sequence-specific DNA binding by a number of proteins, many of which are involved in regulating transcription (reviewed in 28). Interestingly, in all cases examined the HMG proteins exert their influence by interacting directly with the DNA-binding domain of these factors. This appears to also be the case for RAG1. A recent study reveals that enhanced binding of the RSS by RAG1 occurs as a result of a direct interaction between the nonamer-binding domain of RAG1 and the HMG boxes of HMG1 or HMG2 (29).

Despite substantial understanding of the specific bases contacted by the RAG proteins in the RAG/RSS complex (7,10,11), little has been established concerning the configuration of the RAG proteins complexed to the RSS. This has gained more relevance as the RAG proteins have recently been found to bind both 12- and 23-RSSs in a single complex (30). There exist several possibilities for the RAG proteins bound to only a single RSS. For instance, a single protein molecule, most likely RAG1, could bridge the heptamer and nonamer. Alternatively, multiple protein complexes could form on a single RSS with each forming specific contacts to either the heptamer or nonamer. From these investigations on the interaction of RAG1 with the RSS we have begun to delineate the contributions of the individual components of the complex which leads to V(D)J recombination. Future analysis of additional interactions, such as the stoichiometry of the RAG1/RAG2 complex and the configuration of both proteins complexed to one or two RSSs, will be essential to revealing the nature of the catalytically competent macromolecular assembly and the basis for the 12/23 rule.

ACKNOWLEDGEMENTS

We thank Q. Eastman for assistance with activity assays. This work was supported in part by grants from the National Institutes of Health to J.E.C. (DK09070) and to D.G.S. (AI32524) and a Presidential Faculty Fellows Award to D.G.S. from the National Science Foundation. I.J.V. was supported by an INSERM post-doctoral fellowship. D.G.S. is an associate investigator of the Howard Hughes Medical Institute.

REFERENCES

1. Lewis,S.M. (1994) *Adv. Immunol.*, **56**, 27–150.
2. van Gent,D.C., McBlane,J.F., Ramsden,D.A., Sadofsky,M.J., Hesse,J.E. and Gellert,M. (1995) *Cell*, **81**, 925–934.
3. 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.
4. van Gent,D.C., Mizuuchi,K. and Gellert,M. (1996) *Science*, **271**, 1592–1594.
5. Smider,V. and Chu,G. (1997) *Semin. Immunol.*, **9**, 189–177.
6. Hiom,K. and Gellert,M. (1997) *Cell*, **88**, 65–72.
7. Akamatsu,Y. and Oettinger,M.A. (1998) *Mol. Cell. Biol.*, **18**, 4670–4678.
8. Spanopoulou,E., Zaitseva,F., Wang,F.-H., Santagata,S., Baltimore,D. and Panayotou,G. (1996) *Cell*, **87**, 263–276.
9. Difilippantonio,M.J., McMahan,C.J., Eastman,Q.M., Spanopoulou,E. and Schatz,D.G. (1996) *Cell*, **87**, 253–262.
10. 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.
11. Swanson,P.C. and Desiderio,S. (1998) *Immunity*, **9**, 115–125.
12. Agrawal,A., Eastman,Q.M. and Schatz,D.G. (1998) *Nature*, **394**, 744–751.
13. Santagata,S., Aidinis,V. and Spanopoulou,E. (1998) *J. Biol. Chem.*, **273**, 16325–16331.
14. Liu-Johnson,H.N., Gartenberg,M.R. and Crothers,D.M. (1986) *Cell*, **47**, 995–1005.
15. Silver,D.P., Spanopoulou,E., Mulligan,R.C. and Baltimore,D. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 6100–6104.
16. Sadofsky,M.J., Hesse,J.E., McBlane,J.F. and Gellert,M. (1993) *Nucleic Acids Res.*, **21**, 5644–5650.
17. Cuomo,C.A. and Oettinger,M.A. (1994) *Nucleic Acids Res.*, **22**, 1810–1814.
18. Sadofsky,M.J., Hesse,J.E. and Gellert,M. (1994) *Nucleic Acids Res.*, **22**, 1805–1809.
19. van Gent,D.C., Ramsden,D.A. and Gellert,M. (1996) *Cell*, **85**, 107–113.
20. Eastman,Q.M. and Schatz,D.G. (1997) *Nucleic Acids Res.*, **25**, 4370–4378.
21. Eastman,Q.M., Villey,I.J. and Schatz,D.G. (1999) *Mol. Cell. Biol.*, **19**, 3788–3797.
22. van Gent,D.C., Hiom,K., Paull,T.T. and Gellert,M. (1997) *EMBO J.*, **16**, 2665–2670.
23. 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.
24. Rodgers,K.K., Bu,Z., Fleming,K.G., Schatz,D.G., Engelman,D.M. and Coleman,J.E. (1996) *J. Mol. Biol.*, **260**, 70–84.
25. Bellon,S.F., Rodgers,K.K., Schatz,D.G., Coleman,J.E. and Steitz,T.A. (1997) *Nature Struct. Biol.*, **4**, 586–591.
26. Roman,C.A.J., Cherry,S.R. and Baltimore,D. (1997) *Immunity*, **7**, 13–24.
27. Swanson,P.C. and Desiderio,S. (1999) *Mol. Cell. Biol.*, **19**, 3674–3683.
28. Bianchi,M.E. and Beltrame,M. (1998) *Am. J. Hum. Genet.*, **63**, 1573–1577.
29. Aidinis,V., Bonaldi,T., Beltrame,M., Santagata,S., Bianchi,M.E. and Spanopoulou,E. (1999) *Mol. Cell. Biol.*, **19**, in press.
30. Hiom,K. and Gellert,M. (1998) *Mol. Cell*, **1**, 1011–1019.