V(D)J Recombinase Binding and Cleavage of Cryptic Recombination Signal Sequences Identified from Lymphoid Malignancies

Ming Zhang and Patrick C. Swanson

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

V(D)J recombination is a process integral to lymphocyte development. However, this process is not always benign, since certain lymphoid malignancies exhibit recurrent chromosomal abnormalities, such as translocations and deletions, that harbor molecular signatures suggesting an origin from aberrant V(D)J recombination. Translocations involving LMO2, TAL1, Ttg-1, and Hox11, as well as a recurrent interstitial deletion at 1p32 involving SIL/SCL, are cited examples of illegitimate V(D)J recombination. Previous studies using extrachromosomal substrates reveal that cryptic recombination signal sequences (cRSSs) identified near the translocation breakpoint in these examples support V(D)J recombination with efficiencies ranging from about 30- to 20,000-fold less than bona fide V(D)J recombination signals. To understand the molecular basis for these large differences, we investigated the binding and cleavage of these cRSSs by the RAG1/2 proteins that initiate V(D)J recombination. We find that the RAG proteins comparably bind all cRSSs tested, albeit more poorly than a consensus RSS. We show that four cRSSs that support levels of V(D)J recombination above background levels in cell culture (LMO2, TAL1, Ttg-1, and SIL) are also cleaved by the RAG proteins in vitro with efficiencies ranging from 18 to 70% of a consensus RSS. Cleavage of LMO2 and Ttg-1 by the RAG proteins can also be detected in cell culture using ligation-mediated PCR. In contrast, Hox11 and SCL are nicked but not cleaved efficiently in vitro, and cleavage at other adventitious sites in plasmid substrates may also limit the ability to detect recombination activity at these cRSSs in cell culture.

The antigen binding domains of immunoglobulins and T cell receptors are encoded in germ line arrays of V, D, and J gene segments that are assembled into functional variable exons by V(D)J recombination during lymphocyte development (1). The site of recombination is directed by a recombination signal sequence (RSS) that flanks each receptor gene segment and consists of a conserved heptamer (consensus 5'-CACAGTG-3') and nonamer (consensus 5'-ACAAAAACC-3') separated by either 12 or 23 ± 1 nucleotides of more highly varied sequence. V(D)J recombination can be conceptually divided into two phases, a cleavage phase and a joining phase (2). In the cleavage phase, the lymphoid cell-specific RAG1 and RAG2 (recombination activating gene-1 and -2, respectively) proteins assemble a multiprotein synaptic complex with two RSSs (generally one 12-RSS and one 23-RSS) and introduce a DNA double-strand break at each RSS between the heptamer and adjacent coding segment via a nick-hairpin mechanism to yield a blunt 5'-phosphorylated signal end and a coding end terminating in a covalently sealed DNA hairpin structure. In the joining phase, the signal ends are generally joined precisely to form signal joints, and the coding ends are processed and joined to form coding joints that typically contain nucleotide additions or deletions at the junction. These processes are normally mediated by components of the nonhomologous end-joining DNA repair pathway.

Although V(D)J recombination is normally limited to antigen receptor loci, the RAG proteins may mediate illegitimate V(D)J recombination events outside antigen receptor loci (3–5). This may occur by any of several mechanisms, including mistakenly binding and cleaving a DNA sequence that resembles an authentic RSS by a standard nick-hairpin mechanism, by targeting non-B form DNA structures and cleaving them through the introduction of staggered nicks (6), by mobilizing signal ends or episomal signal joints generated after initial RSS cleavage and promoting their integration elsewhere in the genome (either by direct transposition, end donation, or trans-V(D)J recombination (7–10), or through repair failures that allow DNA breaks to become illegitimately joined to DNA ends produced by RAG-mediated cleavage at antigen receptor loci (5).

The first type of mechanism is commonly suggested whenever an RSS-like motif (a cryptic RSS (cRSS)) is identified in the germ line sequence of a proto-oncogene near the chromosomal breakpoint, particularly if recombination involves an immunoglobulin or T cell receptor locus. In these cases, the location of the cRSS has generally been based on the position of a CAC motif (the first three residues of the consensus heptamer) near-
cRSS Binding and Cleavage by the RAG Proteins

est the recombination breakpoint that contains the highest number of additional residues that match the consensus heptamer and nonamer. However, since even authentic RSSs exhibit some degree of sequence variability in the heptamer and nonamer, and since mutations in the heptamer, nonamer, and spacer motifs have position-dependent and possibly synergistic effects on recombination efficiency (11), determining whether a given cRSS can support V(D)J recombination is problematic in the absence of functional testing. Therefore, several laboratories have applied a well established extrachromosomal V(D)J recombination assay to assess the functionality of various cRSSs in cell culture (12–14), including those suspected of mediating chromosomal translocations involving LMO2 (t(11;14)(p13; q11)) (15, 16), TAL1 (t(1;14)(p34;q11)) (17), Ttg-1 (t(11; 14)(p15;q11)) (18, 19), and Hox11 (t(10;14)(q24;q11) (20, 21) as well as interstitial deletions involving SIL/SCL (1p32) (22, 23) and MTAP/p14 –16 (9p21) (13, 24). Perhaps not surprisingly, the putative cRSSs studied to date exhibit a spectrum of activities in these assays, supporting V(D)J recombination at levels from ~30-fold to over 20,000-fold less than a consensus RSS.

The molecular basis for these differences remains unclear, because the binding and cleavage of these cRSSs by the RAG proteins has not been formally investigated.

To address this issue, we have conducted extensive in vitro RAG cleavage and binding assays using cRSS substrates that have been characterized using extrachromosomal V(D)J recombination assays. We find that all intact oligonucleotide cRSS substrates tested show at least some level of RAG-mediated nicking at the 5′-end of the putative heptamer and some degree of RAG-mediated hairpin formation when the cRSS substrate is nicked. Interestingly, all intact cRSS substrates tested that support detectable levels of V(D)J recombination in cell culture support hairpin formation to levels ~18% of a consensus 23-RSS in an in vitro cleavage assay. In contrast, cRSSs that fail to support detectable levels of V(D)J recombination in vitro show only ~1% conversion to hairpins in the same assay. Electrophoretic mobility shift assays reveal that the RAG proteins bind all cRSSs tested more poorly than a consensus RSS, but there is little correlation between how well a given cRSS is cleaved by the RAG proteins and how well it is bound by them. We have also characterized RAG-mediated cleavage of plasmid V(D)J recombination substrates using ligation-mediated PCR (LM-PCR) to detect signal end breaks (SEBs) at a cRSS. SEBs were readily detected at cRSSs in vivo that support recombination efficiencies within ~500-fold of a consensus RSS. LM-PCR can also detect SEBs generated at even less efficient cRSSs in vitro, but competition with adventitious cRSSs elsewhere in the plasmid backbone probably reduces recombination activity to undetectable levels. We also show that a recently identified gain-of-function RAG1 mutant cleaves these cRSSs with greater efficiency than wild-type RAG1, raising the possibility that mutations in RAG1 or other factors that regulate RAG-mediated synopsis and cleavage may promote illegitimate V(D)J recombination in vivo. The predictive value of these in vitro assays in assessing the in vivo recombination potential of a cRSS compared with in silico methods is discussed.

EXPERIMENTAL PROCEDURES

Oligonucleotide and Plasmid Substrates—Oligonucleotides containing a consensus or cryptic RSS were synthesized and gel-purified using a commercial vendor (IDT Inc., Coralville, IA). Signal end oligonucleotides used to assemble nicked substrates included a 5′-phosphate introduced during chemical synthesis. Sequences for the substrates used in this study are shown in Table 1 (top strand only). To prepare intact or nicked duplex substrates, the top or bottom strand oligonucleotides were radiolabeled with [γ-32P]ATP using T4 polynucleotide kinase, annealed to their unlabeled complementary strand(s), and the duplex was purified on a native polyacrylamide gel as described previously (25). Plasmid V(D)J recombination substrates containing either consensus or cryptic recombination signals were generously provided by Michael Lieber and are indicated in Table 1 (14).

Protein Expression and Purification—Wild-type, catalytically inactive (D600A) or hyperactive (E649A) forms of core (residues 384–1040) RAG1 and core (residues 1–387) RAG2 were coexpressed in 293 cells as maltose-binding protein fusion proteins (cMR1 and cMR2, respectively) and purified as described previously (25). The yields of wild-type and mutant RAG proteins were all similar. In some experiments, preparations of core RAG1 and full-length RAG2 or full-length RAG1 and core RAG2 were used (cMR1/FLMR2 or FLMR1/cMR2, respectively) (26). Full-length HMGB1 (high mobility group box 1) was expressed in Escherichia coli strain BL21(DE3)pLysS and purified using a combination of immobilized metal affinity chromatography and ion exchange chromatography as described previously (25).

Oligonucleotide Cleavage and Binding Assays—RAG-mediated cleavage of oligonucleotide substrates was analyzed using an in vitro cleavage assay performed under permissive conditions containing Me2SO to promote RSS cleavage as described (25). Briefly, coexpressed wild-type or mutant RAG1/RAG2 proteins (100 ng) and HMGB1 (300 ng, where indicated) were incubated with substrate DNA (~0.02 pmol) in a 10-μl reaction containing sample buffer (25 mM MOPS-KOH (pH 7.0), 60 mM potassium glutamate, 100 μg/ml bovine serum albumin, 20% Me2SO, and 1 mM MgCl2). In some experiments, reactions were further supplemented with cold partner RSS (0.1, 1, or 10 pmol, as indicated). Samples were incubated at 37 °C for 1 h, and the reaction products were visualized and quantified from dried sequencing gels using a Molecular Dynamics Storm 860 PhosphorImager running the ImageQuant software. Direct RAG binding to cRSS substrates was analyzed by electrophoretic mobility shift assay (EMSA) using the same buffer conditions as those used in the in vitro cleavage assay, except that CaCl2 replaced MgCl2 in the binding reaction to avoid substrate cleavage (25). For competition experiments shown in Fig. 1, binding reactions were supplemented with unlabeled intact consensus or cRSS substrate (0.1, 1, or 10 pmol, as indicated) before the addition of the RAG proteins.

Cleavage of Plasmid Substrates in Vitro and in Vivo—RAG-mediated cleavage of plasmid V(D)J recombination substrates was analyzed using LM-PCR to detect signal end breaks at either the 12- or 23-RSS. For in vitro cleavage experiments, the
various plasmid substrates in Table 1 (100 ng; linearized with AatII or BglII) were incubated with the cMR1/cMR2 and HMGB1 proteins under the same reaction conditions as used to assess cleavage of oligonucleotide substrates. The cleaved DNA was ligated to linker DNA assembled from oligonucleotides DR19 and DR20 as described previously (27), and then the reaction was terminated by incubation at 65 °C for 30 min. Signal ends generated at the 12-RSS or 23-RSS positions were detected from 40 ng of recovered plasmid DNA using LM-PCR.

For in vivo cleavage experiments, each 10-cm dish of 293 cells was cotransfected with one of the plasmid V(D)J recombination substrates listed in Table 1 (5 μg) and pcDNA1 expression constructs encoding WT, D600A, or E649A cMR1 (2.5 mg/dish) and WT cMR2 (2.5 mg) using 30 μg of polyethyleneimine as described previously (28), no cleavage of intact RSS or cRSS substrates was observed in those that function as a 23-RSS (Ttg-1, TAL1, TAL1, and SIL), and others did not. Because the V(D)J recombination assay measures the culmination of both the cleavage and joining phases of V(D)J recombination, we wondered whether those cRSSs that did not support efficient recombination could be cleaved by the RAG proteins but not efficiently joined.

### Results

**Characterization of cRSS Cleavage by the RAG1-RAG2 Complex**—In previous studies (12, 14), the functionality of several cRSSs identified from lymphoid malignancies were tested using an extrachromosomal substrate assay (Table 1). Some cRSSs supported relatively robust recombination activity, whereas others did not. Because the V(D)J recombination assay measures the culmination of both the cleavage and joining phases of V(D)J recombination, we wondered whether those cRSSs that did not support efficient recombination may be cleaved by the RAG proteins but not efficiently joined.

The most straightforward way to begin testing this possibility is to compare RAG-mediated cleavage of consensus and cryptic RSSs in a standard in vitro cleavage assay. Toward this end, we prepared radiolabeled intact or nicked oligonucleotide substrates containing a consensus 12- or 23-RSS or a cRSS (LMO2, TAL1, Ttg-1, Hox11, SCL, and SIL) (Table 1). The substrates were incubated with a purified truncated “core” wild-type (WT) or catalytically inactive (D600A) RAG1-RAG2 complex (RAGs) in reaction buffer containing Mg²⁺ and Me₂SO for 1 h at 37 °C. The addition of Me₂SO enhances RAG cleavage activity, which is otherwise low in Mg²⁺; and models the stimulation afforded by synopsis (25) (supplemental Fig. 1A). Avoiding the need for adding partner RSS to promote synopsis simplifies the reaction scheme and is advantageous technically, because in most cases, the addition of a consensus partner RSS to reactions containing a labeled cRSS substrate reduces cRSS cleavage due to competitive inhibition of RAG binding (supplemental Fig. 1B; see below). Selected cleavage reactions were additionally supplemented with purified HMGB1, because this architectural DNA binding/bending protein is known to stimulate RAG binding and cleavage activity in vitro (particularly on 23-RSS substrates) and reduce aberrant nicking, even in the presence of Me₂SO (25). Reaction products generated from RAG-mediated cleavage in vitro were analyzed after fractionation on sequencing gels.

Based on previous studies (12, 14), the cRSSs examined here fall into three groups: those that function as a 12-RSS (LMO2), those that function as a 23-RSS (Ttg-1, TAL1, and SIL), and those for which functionality could not be unambiguously determined (Hox11 and SCL). As expected from previous studies (28), no cleavage of intact RSS or cRSS substrates was observed by the D600A RAGs (Fig. 1, A–D). In contrast, WT RAGs convert most of the intact 12- and 23-RSS substrates into nicked and hairpin products; supplementing the cleavage reaction with HMGB1 slightly suppresses RAG-mediated cleavage and selectively diminishes aberrant nicking of the 23-RSS under these conditions. In the absence of HMGB1, all intact cRSSs tested are nicked by WT RAGs to some degree at the 5′-end of the putative heptamer. In some of the cRSSs (e.g. TAL1, Ttg-1,
### Table

<table>
<thead>
<tr>
<th>Sequence</th>
<th>LMO2</th>
<th>TAL1</th>
<th>Hox11</th>
<th>SCL</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A.</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12-RSS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WT cMR1/cMR2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D600A cMR1/cMR2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HMGB1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nick Marker</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hairpin Marker</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% N</td>
<td>48 29</td>
<td>83 55</td>
<td>41 29</td>
<td>6.5 2.4</td>
</tr>
<tr>
<td>% HP</td>
<td>45 27</td>
<td>11 8.2</td>
<td>6.8 27</td>
<td>0.8 ND</td>
</tr>
<tr>
<td>% Abnick</td>
<td></td>
<td></td>
<td>22 4.1</td>
<td>9.3 3.8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sequence</th>
<th>LMO2</th>
<th>TAL1</th>
<th>Hox11</th>
<th>SCL</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>B.</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23-RSS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WT cMR1/cMR2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D600A cMR1/cMR2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HMGB1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nick Marker</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hairpin Marker</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% N</td>
<td>44 29</td>
<td>61 26</td>
<td>41 38</td>
<td>6.2 2.4</td>
</tr>
<tr>
<td>% HP</td>
<td>31 27</td>
<td>14 0.9</td>
<td>34 25</td>
<td>0.8 ND</td>
</tr>
<tr>
<td>% Abnick</td>
<td></td>
<td></td>
<td>2.1 9.0</td>
<td>3.5 8.0</td>
</tr>
</tbody>
</table>

### Diagrams

#### A. 12-RSS
- LMO2: 
- TAL1: 
- Hox11: 
- SCL: 

#### B. 23-RSS
- SIL: 
- Ttg-1: 
- Hox11: 
- SCL: 

#### C. 12-RSS
- LMO2: 
- TAL1: 
- Hox11: 
- SCL: 

#### D. 23-RSS
- SIL: 
- Ttg-1: 
- Hox11: 
- SCL: 

#### E. 12-RSS
- LMO2: 
- TAL1: 
- Hox11: 
- SCL: 

#### F. 23-RSS
- SIL: 
- Ttg-1: 
- Hox11: 
- SCL: 

#### G. 12-RSS
- LMO2: 
- TAL1: 
- Hox11: 
- SCL: 

#### H. 23-RSS
- SIL: 
- Ttg-1: 
- Hox11: 
- SCL:
TABLE 2
Summary of cleavage and recombination data for consensus and cryptic cRSSs

<table>
<thead>
<tr>
<th>RSS or cRSS (functionality)</th>
<th>Heptamer</th>
<th>Nonamer</th>
<th>Relative recombination efficiency</th>
<th>RIC scorea</th>
<th>cRSS percentage (N+ HP)/23-RSS percentage (N+ HP) × 100</th>
<th>SEBs in vitro</th>
<th>SEBs in vivo</th>
</tr>
</thead>
<tbody>
<tr>
<td>12-RSS</td>
<td>CACAGTG</td>
<td>ACAAACC</td>
<td>20,000</td>
<td>-22.06 (P)</td>
<td>116</td>
<td>+ + + +</td>
<td>+ + + + +</td>
</tr>
<tr>
<td>23-RSS</td>
<td>CACAGTG</td>
<td>ACAAACC</td>
<td>20,000</td>
<td>-27.09 (P)</td>
<td>148</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
</tr>
<tr>
<td>LMO2 (12-RSS)</td>
<td>CACAGTA</td>
<td>AGGAAAC</td>
<td>760</td>
<td>-42.00 (P/F)</td>
<td>100</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
</tr>
<tr>
<td>TAL1 (23-RSS)</td>
<td>CACAGTG</td>
<td>CCTAGGA</td>
<td>5</td>
<td>-56.49 (F)</td>
<td>117 ± 11</td>
<td>40 ± 5.2</td>
<td>+ + + + +</td>
</tr>
<tr>
<td>Tgt-1 (23-RSS)</td>
<td>CACAGTG</td>
<td>ACTTGGA</td>
<td>38</td>
<td>-48.92 (P)</td>
<td>62 ± 72</td>
<td>18 ± 1.8</td>
<td>ND* ND</td>
</tr>
<tr>
<td>SIL (12-RSS)</td>
<td>CACTCGT</td>
<td>CAGTCAGT</td>
<td>&lt;0.8</td>
<td>-59.12 (F)</td>
<td>65 ± 21</td>
<td>70 ± 36</td>
<td>+ + + +</td>
</tr>
<tr>
<td>SIL (23-RSS)</td>
<td>CACTCCT</td>
<td>CAGTCAGT</td>
<td>27</td>
<td>-59.66 (F)</td>
<td>67 ± 26</td>
<td>26 ± 15</td>
<td>+ / +</td>
</tr>
<tr>
<td>SCL (12-RSS)</td>
<td>CACCTGC</td>
<td>CAGTCAGT</td>
<td>&lt;0.7</td>
<td>-52.97 (F)</td>
<td>8.5 ± 5.3</td>
<td>1.2 ± 1.2</td>
<td>+ / + + +</td>
</tr>
<tr>
<td>SCL (23-RSS)</td>
<td>CACCTGG</td>
<td>CAGTCAGT</td>
<td>&lt;1.2</td>
<td>-65.81 (F)</td>
<td>8.5 ± 5.3</td>
<td>1.2 ± 1.2</td>
<td>ND ND</td>
</tr>
<tr>
<td>Hox11 (12-RSS)</td>
<td>CACGTACG</td>
<td>CAGTCAGA</td>
<td>&lt;1</td>
<td>-59.84 (F)</td>
<td>8.5 ± 1.7</td>
<td>1.1 ± 0.9</td>
<td>– –</td>
</tr>
<tr>
<td>Hox11 (23-RSS)</td>
<td>CACGTATG</td>
<td>CAGTCAGA</td>
<td>&lt;0.9</td>
<td>-74.23 (F)</td>
<td>8.5 ± 1.7</td>
<td>1.1 ± 0.9</td>
<td>+ / + + +</td>
</tr>
</tbody>
</table>

aCalculated using the Recombination Signal Search program available on the World Wide Web by the Duke University Laboratory of Computational Immunology using DNA sequence provided in Table 1. RIC scores greater than threshold values established to indicate functionality (≥ -40 for 12-RSS and ≥ -60 for 23-RSS; see Ref. 32) are given a pass (P) designation; those below the threshold value and the lower limit for a mouse immunologic RSS (RIC12 = -48.16; RIC23 = -69.69) are given a pass/fail (P/F) designation; those below the lower limit for a mouse immunologic RSS are given a fail (F) designation.

bPercentage conversion to total nicked (N+) and/or hairpin (HP) products for cRSSs relative to consensus 23-RSS calculated from at least three independent experiments.

*ND, not determined.

and Hox11), adventitious nicks are also introduced elsewhere in the substrate and accumulate with reaction kinetics similar to the appropriately sited nick (data not shown). Interestingly, hairpin products are observed in cleavage reactions containing LMO2, TAL1, SIL, and Tgt-1 substrates but are virtually undetectable in comparable reactions containing Hox11 and SCL substrates. The efficiency of hairpin formation using LMO2, TAL1, SIL, and Tgt-1 substrates compared with the 23-RSS ranges from 18 to 70% in the absence of HMGB1, whereas for Hox11 and SCL, these values are around 1% (Table 2). The addition of HMGB1 has variable effects on nicking and hairpin formation, depending on the substrate tested. In some cases (e.g. TAL1), aberrant nicking is reduced and hairpin formation is stimulated; in others (e.g. SIL), hairpin formation is suppressed. The molecular basis for these distinct outcomes remains unclear but, in the case of SIL, may be correlated with altered cRSS binding by the RAGs in the presence of HMGB1 (see below).

We also tested whether core and full-length RAG proteins exhibited comparable cRSS cleavage activity (supplemental Fig. 2). Interestingly, RAG preparations containing full-length RAG2 supported less nicking and hairpin formation on all cRSSs tested except LMO2 when compared with an all core RAG preparation, despite similarly cleaving the consensus RSS substrates. Whether this trend also holds true for full-length RAG1 could not be unambiguously determined, because this preparation is substantially less active than an all-core RAG preparation in an in vitro cleavage assay.

In principle, the poor efficiency of hairpin formation on intact Hox11 and SCL substrates may be due to poor nicking, which is a rate-limiting step in this reaction. To address this possibility, we examined RAG-mediated hairpin formation using nicked RSS and cRSS substrates (Fig. 1, C and D). As one might predict, nicked LMO2, TAL1, Tgt-1, and SIL substrates support conversion to hairpin products. Interestingly, a nicked SCL substrate supports almost as much hairpin formation as its SIL counterpart. However, hairpin formation, although detected using a nicked Hox11 substrate, remains inefficient. Nevertheless, taken together, these data suggest that all cRSSs tested are capable of supporting RAG-mediated cleavage through the standard nick-hairpin mechanism.

Characterization of cRSS Binding by the RAG1-RAG2 Complex—In principle, the differences observed in RAG-mediated cleavage of the various cRSS substrates could be attributed to differences in how well the individual cRSS substrates are bound by the RAGs. To examine this possibility, we performed direct and competition DNA binding experiments. In the first set of experiments, the RAGs were incubated in a binding reaction containing a radiolabeled intact RSS or cRSS substrate with or without added HMGB1, and then protein-DNA complexes were fractionated using an EMSA (Fig. 1, E and F; supplemental Fig. 3). As expected from previous studies (28), two major RAG-RSS complexes are detected by EMSA when the RAGs are incubated with a consensus RSS in the absence of HMGB1, termed SC1 and SC2. Previous work in our laboratory suggests that SC1 and SC2 contain a RAG1 dimer and either one (SC1) or two (SC2) subunits of RAG2 (28). The addition of HMGB1 super-shifts both complexes (HSC1 and HSC2, respectively). When the cRSS substrates were similarly analyzed, none of the cRSSs were found to support SC1 or SC2 formation to levels comparable with a consensus RSS. The addition of HMGB1 to the

FIGURE 1. The RAG proteins variably cleave but similarly bind oligonucleotide cRSS substrates in vitro. A–D, radiolabeled intact (A and B) or nicked (C and D) oligonucleotide RSS or cRSS substrates described in Table 1 were incubated with purified WT or D600A RAG proteins in the absence or presence of HMGB1 in an in vitro cleavage reaction containing Mg2+ as indicated above the gel. Samples were fractionated on a sequencing gel along with the predicted nick and hairpin products as markers. The percentage of appropriately sited nicked (N+) and hairpin (HP) products as well as aberrantly nicked products (Ab nick) includes products denoted by asterisks in each lane was calculated using a Storm 860 PhosphorImager running the ImageQuant software and is shown below the gel. E and F, radiolabeled intact RSS or cRSS substrates were incubated with purified RAG proteins in the absence or presence of HMGB1 in binding reactions containing Ca2+ as indicated above the gel. Protein-DNA complexes were fractionated by EMSA and visualized from dried gels using a Storm 860 Phosphor-Imager. G and H, purified RAG proteins were incubated with a radiolabeled intact 12-RSS (G) or 23-RSS (H) substrate in the absence or presence of HMGB1 in binding reactions containing Ca2+ and supplemented with increasing amounts of unlabeled oligonucleotide RSS or cRSS substrate, as indicated above the gel (0.1, 1, or 10 pmol). Protein-DNA complexes were fractionated by EMSA and visualized as in E and F.
binding reaction promoted RAG binding to SIL almost as well as 23-RSS but had little effect on RAG binding to other cRSS substrates tested.

To confirm these results, we incubated the RAGs with HMGB1 and a radiolabeled intact 12- or 23-RSS in binding reactions supplemented with increasing concentrations of cold cRSS: Hox11 or SCL.
RSS or cRSS substrate as a competitor (Fig. 1, G and H) and analyzed protein-DNA complex formation by EMSA. We find that all cRSSs tested are poorer competitors for RAG binding compared with a consensus RSS, requiring at least 10-fold greater cRSS substrate than a consensus RSS to achieve similar level of competitive inhibition of RAG binding to a consensus RSS. These results are largely consistent with data obtained from direct DNA binding experiments, with the exception of SIL. This apparent discrepancy may be explained if HMGB1 exhibits a selective preference for stably integrating into a RAG complex assembled on a consensus RSS but remains capable of promoting RAG binding to certain cryptic sequences in the absence of competitor DNA.

Testing Requirements for RAG-mediated DNA Double Strand Break Induction at Hox11 and SCL—Since isolated Hox11 and SCL substrates were cleaved poorly by the RAG proteins, we wondered whether synapsis might promote cleavage of these cRSSs or whether these sequences might be cleaved through an alternative mechanism. To test the first possibility, we examined whether the addition of appropriate cold partner DNA (12- or 23-RSS) enhances RAG-mediated cleavage of intact Hox11 and SCL in vitro. We find that the abundance of reaction products obtained after incubation of intact Hox11 or SCL substrates with the RAGs decreases with increasing concentration of either cold 12-RSS or 23-RSS partner RSS, regardless of whether the partner RSS is added before or after the RAG proteins (Fig. 2, A and B). This outcome is not unique to these substrates, since similar results were obtained for Ttg-1; however, LMO2 cleavage by the RAG complex was found to be stimulated slightly (~3-fold) when partner RSS was present at the lowest concentration tested (supplemental Fig. 1B).

Since the RAG proteins bind Hox11 and SCL poorly (Fig. 1, E and H), this result is explained by competitive inhibition of RAG binding to the labeled cRSS substrate by the cold partner RSS. We performed a similar experiment using nicked Hox11 or SCL substrates in the absence or presence of HMGB1. In contrast to results obtained using intact substrates, RAG-mediated cleavage of both cRSS substrates is stimulated when cold nicked partner RSS is added at low concentration, but higher concentrations of partner RSS inhibit cRSS cleavage (Fig. 2, C and D). This outcome is observed both in the absence and presence of HMGB1 (albeit to a lesser extent) and does not depend on the composition of the partner RSS. These data suggest that upon nicking, Hox11 and SCL can support RAG-mediated synapsis with a nicked partner independent of the 12/23 rule and undergo cleavage via a nick-hairpin mechanism.

Although all cRSS substrates tested here support standard V(D)J-type cleavage under at least some conditions in vitro, evidence suggests that the RAG proteins can cleave DNA through an alternative pathway whereby the RAG proteins introduce two nicks in close proximity on opposite DNA strands (29, 30). To determine whether this “nick-nick” mechanism may contribute to cleavage of these cRSSs, particularly Hox11 and SCL, the RAG proteins were incubated with a consensus 23-RSS or various cRSS substrates radiolabeled on the bottom strand (Fig. 2E). As expected from previous results, substrates that support detectable levels of hairpin formation in vitro, including the 23-RSS, LMO2, Ttg-1, and TAL1, yield a reaction product corresponding to cleavage of the bottom strand after hairpin formation (note that the top strand signal end oligonucleotide used to prepare the nicked substrate was run as a marker that does not precisely comigrate with the predicted product because its sequence is the complement of the product being detected). Interestingly, the RAGs introduce adventitious nicks on the bottom strand of all substrates tested in the absence of HMGB1, most of which are suppressed in the presence of HMGB1 and are therefore unlikely to contribute to cleavage via the nick-nick mechanism, especially in competition with the nick-hairpin mechanism, where it occurs robustly (e.g. 23-RSS and LMO2). Nevertheless, in some cases, nicking of the bottom strand persists (e.g. TAL1 and SIL) or is even enhanced (e.g. Ttg-1) in the presence of HMGB1, raising the possibility that RAG-mediated cRSS cleavage via the nick-nick mechanism may be occurring at low levels or possibly contributing to generating overhanging DNA ends before or after cRSS cleavage by the standard nick-hairpin mechanism.

Detection of RAG-mediated cRSS Cleavage in Plasmid Substrates in Vitro and in Vivo—Because the functionality of the cRSSs tested in this study were evaluated in cells in the context of a plasmid, we wished to determine whether cRSSs that support cleavage in an oligonucleotide substrate are also cleaved when embedded in plasmid DNA. Toward this end, we linearized plasmid constructs containing consensus RSSs or cRSSs in various configurations (14) (see Table 1) and incubated them with WT or D600A RAGs and HMGB1 in the presence of Mg2+.

The abundance of SEBs produced by RAG-mediated cleavage at the consensus RSS or cRSS was determined by LM-PCR (schematically depicted in Fig. 3, A and B). We find that SEBs are readily detected at a consensus RSS by LM-PCR after in vitro cleavage by WT RAGs but not D600A RAGs (Fig. 3, C and D). Consistent with results obtained using oligonucleotide substrates, SEBs are detected after in vitro cleavage of plasmid substrates containing LMO2 (as a 12-RSS), and Ttg-1 (as a 23-RSS). Cleavage at SIL as a 23-RSS is more readily detected than SIL as a 12-RSS, and cleavage at Hox11 (as a 23-RSS) is also detected at very low levels, but adventitious RAG-mediated DNA double strand breaks occurring elsewhere in the plasmid backbone are also evident at similar or higher levels in these...
cases. We were unable to detect cleavage of SCL when paired with a consensus 23-RSS or with SIL.

To test whether signal end intermediates could be detected at the consensus RSS or cRSS in plasmid V(D)J recombination assays performed in cell culture, we cotransfected the plasmid substrates with either WT or D600A RAG1 and RAG2 expression constructs in 293 cells. After 72 h, plasmid DNA was recovered from harvested cells and subjected to LM-PCR (Fig. 3, E and F). Consistent with our in vitro cleavage experiments, SEBs generated at LMO2 and Ttg-1 are detectable in plasmid DNA recovered from cells by LM-PCR, albeit at lower levels than their counterpart consensus RSS. SEBs at SIL, Hox11, or SCL
were essentially undetectable above background. Curiously, although sequence analysis of TA-cloned LM-PCR products confirmed that LMO2 and SIL support cleavage at the predicted heptamer, we were unable to obtain LM-PCR products reflecting cleavage at the predicted heptamer in Ttg-1. Rather, linker ligation occurred at several alternative sites within the Ttg-1 sequence, three of which include CAC or CAT motifs (supplemental Fig. 4). This result may not be entirely surprising, given the abundance of CAC motifs in the Ttg-1 sequence and evidence that Ttg-1 supports RAG-mediated nicking at several alternative sites on both the top and bottom strand in vitro (Fig. 1).

Enhanced cRSS Cleavage by a Gain-of-function RAG1 Mutant—We recently identified an E-649A RAG1 mutant that exhibits increased cleavage activity relative to WT RAGs in the presence of Mg$^{2+}$ by promoting hairpin formation, possibly due to a defect in sensing 12/23-regulated synopsis (27). We hypothesized that this RAG1 mutant also supports enhanced cleavage of cRSS substrates. To test this hypothesis, we compared the cleavage activities of WT and E649A RAGs on intact or nicked 23-RSS and SIL and SCL substrates (supplemental Fig. 5A). Consistent with our previous studies (27), the total abundance of reaction products generated after in vitro cleavage of an intact 23-RSS by WT and E649A RAGs is quite similar, but E649A RAGs more readily convert nicks to hairpin products, regardless of the presence of HMG1. E649A RAGs also support higher levels of hairpin formation using a nicked 23-RSS substrate. Interestingly, both intact and nicked SIL and SCL substrates are cleaved more efficiently by E649A RAGs relative to WT RAGs, both in the absence and presence of HMG1. This outcome is particularly striking for SCL, where in vitro cleavage by the E649A RAGs yields over 10-fold more hairpin product than comparable cleavage reactions containing WT RAGs. The enhanced cleavage activity observed with E649A RAGs is not due to increased binding of the target substrate, since WT and E649A RAGs assemble protein-DNA complexes to similar levels with all substrates tested by EMSA (supplemental Fig. 5B). The enhanced cleavage activity of SIL and SCL by the E649A RAGs observed using oligonucleotide substrates extends to plasmid substrates in vitro and in cell culture V(D)J recombination assays for SIL but not SCL (supplemental Fig. 5C). Taken together, these studies raise the possibility that mutations in the RAG proteins or other factors that impair sensing of 12/23-regulated synopsis may contribute to increased genomic instability by promoting illegitimate cRSS cleavage.

**DISCUSSION**

Certain forms of leukemia and lymphoma exhibit recurrent chromosomal translocations or deletions that have been attributed to illegitimate V(D)J recombination involving sequence motifs in proto-oncogenes that resemble an authentic RSS. The functionality of several of these cRSSs have been tested in cell culture using plasmid V(D)J recombination assays and have been shown to support V(D)J recombination at levels that are 30-fold to greater than 20,000-fold less than a consensus RSS (Table 1). Here we show that four cRSSs that support levels of V(D)J recombination above background levels (LMO2, TAL1, Ttg-1, and SIL) are also cleaved into hairpin products in vitro by purified RAG proteins to levels that are ≥18% of a consensus 23-RSS when cleavage is assessed using an intact oligonucleotide substrate (Table 2). LMO2, Ttg-1, and SIL also support RAG-mediated cleavage when embedded in a plasmid substrate, but the site preference for Ttg-1 cleavage is more diverse than observed using oligonucleotide substrates. In contrast, two other putative cRSSs that do not support V(D)J recombination above background levels (Hox11 and SCL) are also not cleaved efficiently by the RAG proteins in vitro using intact oligonucleotide substrates (≤1% of a consensus 23-RSS) or plasmid substrates. Thus, these data argue that the poor recombination efficiency observed for Hox11 and SCL is attributed primarily on their inability to support RAG-mediated cleavage rather than supporting inefficient joining. Furthermore, our experimental results from in vitro cleavage assays suggest that if a putative intact oligonucleotide cRSS substrate is cleaved by the core RAG proteins to levels approaching ~20% of a consensus RSS, the cRSS should be expected to support detectable levels of V(D)J recombination in cell culture. However, in vitro cleavage assays performed using only purified RAG proteins cannot accurately predict how well a given cRSS supports V(D)J recombination compared with other cRSSs. For example, LMO2 and Ttg-1 substrates are cleaved comparably by the core RAG proteins in vitro, yet LMO2 supports about 20-fold more recombination than Ttg-1 in the plasmid V(D)J recombination assay. In this case, additional assays reveal potential explanations that differentiate the functionality of the two cRSSs at the level of substrate cleavage. First, RAG cleavage of LMO2, but not Ttg-1, is slightly stimulated by synopsis (supplemental Fig. 1B). Second, unlike the core RAG proteins, RAG complexes containing full-length RAG2 cleave LMO2 more efficiently than Ttg-1, suggesting the C-terminal “noncore” portion of RAG2 influences the binding site specificity of the RAG complex (supplemental Fig. 2), as others have argued (31). We also acknowledge the possibility that distinct cRSSs may be similarly bound and cleaved by the RAG complex but support joining at different levels due to sequence variations (32).

**Mechanisms of cRSS Cleavage**—Two general models to explain chromosomal translocations in lymphoid malignancies mediated by mistakes in V(D)J recombination have emerged from structural analysis of derivative chromosomes and functional studies of cRSSs in cell culture (4, 5). In one model, the RAG proteins assemble a synaptic complex containing an authentic RSS and a functional cRSS in a proto-oncogene and subsequently mediate cRSS cleavage through the standard nick-hairpin mechanism. A subset of translocations involving LMO2 and TAL1, among others are thought to arise through this mechanism (4, 5). Data presented here are consistent with this hypothesis, since both substrates are cleaved by the RAG proteins through a nick-hairpin mechanism in an in vitro cleavage assay. In a second model, broken DNA ends fortuitously introduced in a proto-oncogene infiltrate or are captured by a postcleavage RAG complex generated as an intermediate during normal V(D)J recombination. Whereas end joining in the first model produces two joining products (the equivalent of one signal joint and one coding joint), end joining in the second scenario produces three joining products in various possible
cRSS Binding and Cleavage by the RAG Proteins

configurations. Translocations involving Hox11 and the Bcl2 major breakpoint region, among others, are thought to occur by the second model (4, 5). The origin of the DNA break introduced in the proto-oncogene may or may not be RAG-mediated. In the case of the Bcl2 major breakpoint region, recent evidence suggests that the RAGs, acting as a structure-specific endonuclease, target a non-B DNA structure that this region adopts and introduce staggered nicks that result in DNA breaks with overhanging DNA ends (6). In other cases, the source of DNA breaks is less clear. Here we present evidence that cRSSs that fail to support efficient V(D)J recombination in cell culture (e.g. Hox11 and SCL) are nevertheless nicked at the 5′-end of putative cRSS heptamer and can be cleaved inefficiently through a nick-hairpin mechanism in the presence of nicked partner DNA. In other cases, cRSSs that can undergo RAG-mediated cleavage by a nick-hairpin mechanism, like TAL1, Ttg-1, and SIL, may also be nicked on the bottom strand proximal to the cleavage site. Thus, the RAG proteins may take two different routes to generate DNA breaks in a cRSS, possibly creating heterogeneous DNA ends that may be repaired by different mechanisms. This scenario may explain why translocations involving some cRSSs, like TAL1, appear to arise through both standard V(D)J-type recombination and repair failures (5).

Comparison of in Vitro and in Silico Methods to Determine cRSS Functionality—Based on analysis of illegitimate V(D)J rearrangements observed in plasmids, Lewis et al. (33) estimated that ~10 million cRSSs are present in the human genome that support V(D)J recombination to Ξ1% of a consensus RSS. However, other than the presence of a CAC motif at the 5′-end of the heptamer, the degree of sequence variation observed among individual cRSSs known to support detectable levels of V(D)J recombination is too great to allow accurate de novo prediction of cRSS functionality. To address this issue, a computational algorithm to determine the “recombination information content” (RIC) within a given sequence to predict its potential to support V(D)J recombination was developed by Cowell et al. (34). Using RIC scores set to physiological threshold levels of Ξ40 for a 12-RSS and Ξ60 for a 23-RSS, the authors concluded that the number of functional cRSSs in the human genome may be 10-fold less than reported by Lewis et al. (33) but are nevertheless highly abundant. We were interested in determining how well RIC scores correlated with levels of RAG-mediated cleavage and V(D)J recombination activity of the cRSSs studied here. When RIC scores were calculated for all of the cRSSs analyzed in this study, we find that only Ttg-1 scored higher than the thresholds established to indicate functionality, although the RIC12 score for LMO2 was very close to the threshold (Table 2). However, all four of the six cRSSs that support detectable levels of V(D)J recombination in cell culture are cleaved by the RAG proteins in vitro. Thus, the in vitro cleavage assay appears to more accurately predict functionality of a cRSS than the in silico approach. This conclusion is underscored by the comparison of RIC23 scores obtained for SIL and SCL. By the RIC23 score, SCL more closely resembles a functional 23-RSS than SIL, but in cell culture (14) and in in vitro cleavage assays (Fig. 1), SIL supports more efficient RAG cleavage and recombination than SCL. These data suggest that further refinements to mathematical algorithms used to evaluate RSS functionality are required to enhance the predictive capability of these powerful computational tools so that they can more accurately determine recombination potential of cRSSs with RIC scores that are indicated to be subthreshold for functionality.

Acknowledgments—We thank Sathees Raghavan and Michael Lieber for generously providing the plasmid V(D)J recombination substrates listed in Table 1 and Dirk Anderson and Sarina Coles for technical assistance. This investigation was conducted in a facility constructed with support from the Research Facilities Improvement Program of the National Institutes of Health National Center for Research Resources (C06 RR17417-01).

REFERENCES
cRSS Binding and Cleavage by the RAG Proteins

V(D)J Recombinase Binding and Cleavage of Cryptic Recombination Signal Sequences Identified from Lymphoid Malignancies
Ming Zhang and Patrick C. Swanson

doi: 10.1074/jbc.M710301200 originally published online January 9, 2008

Access the most updated version of this article at doi: 10.1074/jbc.M710301200

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

Supplemental material:
http://www.jbc.org/content/suppl/2008/01/10/M710301200.DC1

This article cites 34 references, 21 of which can be accessed free at
http://www.jbc.org/content/283/11/6717.full.html#ref-list-1