

# Real-time analysis of RAG complex activity in V(D)J recombination

Jennifer Zigelbaum<sup>a</sup>, Noriko Shimazaki<sup>b,c,d,e</sup>, Zitadel Anne Esguerra<sup>b,c,d,e</sup>, Go Watanabe<sup>b,c,d,e</sup>, Michael R. Lieber<sup>b,c,d,e,1</sup>, and Eli Rothenberg<sup>a,1</sup>

<sup>a</sup>Department of Biochemistry and Molecular Pharmacology, Perlmutter Cancer Center, New York University School of Medicine, New York, NY 10016; <sup>b</sup>Department of Pathology, University of Southern California (USC) Keck School of Medicine, USC Norris Comprehensive Cancer Center, Los Angeles, CA 90033; <sup>c</sup>Department of Biochemistry & Molecular Biology, University of Southern California (USC) Keck School of Medicine, USC Norris Comprehensive Cancer Center, Los Angeles, CA 90033; <sup>d</sup>Department of Molecular Microbiology & Immunology, University of Southern California (USC) Keck School of Medicine, USC Norris Comprehensive Cancer Center, Los Angeles, CA 90033; and <sup>e</sup>Section of Molecular & Computational Biology, Department of Biological Sciences, University of Southern California (USC), Los Angeles, CA 90089

Edited by Rodney Rothstein, Columbia University Medical Center, New York, NY, and approved August 29, 2016 (received for review May 6, 2016)

**Single-molecule FRET (smFRET) and single-molecule colocalization (smCL) assays have allowed us to observe the recombination-activating gene (RAG) complex reaction mechanism in real time. Our smFRET data have revealed distinct bending modes at recombination signal sequence (RSS)-conserved regions before nicking and synapsis. We show that high mobility group box 1 (HMGB1) acts as a cofactor in stabilizing conformational changes at the 12RSS heptamer and increasing RAG1/2 binding affinity for 23RSS. Using smCL analysis, we have quantitatively measured RAG1/2 dwell time on 12RSS, 23RSS, and non-RSS DNA, confirming a strict RSS molecular specificity that was enhanced in the presence of a partner RSS in solution. Our studies also provide single-molecule determination of rate constants that were previously only possible by indirect methods, allowing us to conclude that RAG binding, bending, and synapsis precede catalysis. Our real-time analysis offers insight into the requirements for RSS–RSS pairing, architecture of the synaptic complex, and dynamics of the paired RSS substrates. We show that the synaptic complex is extremely stable and that heptamer regions of the 12RSS and 23RSS substrates in the synaptic complex are closely associated in a stable conformational state, whereas nonamer regions are perpendicular. Our data provide an enhanced and comprehensive mechanistic description of the structural dynamics and associated enzyme kinetics of variable, diversity, and joining [V(D)J] recombination.**

V(D)J recombination | recombination-activating gene | single-molecule FRET | immunoglobulin | antigen receptor gene

**V**ariable, diversity, and joining [V(D)J] recombination is essential for Ig and T-cell receptor (TCR) formation and occurs by rearrangement of variable exon regions during lymphocyte development. Complete exons are assembled from (V), (D), and (J) segments at the Ig and TCR loci. Cutting, rearranging, and recombining the few hundred preexisting (V), (D), and (J) segments can lead to millions of permutations and a highly diverse repertoire of antigen receptors and antibodies (1–4).

The recombination-activating genes 1 and 2 (RAG1/2) are lymphocyte-specific proteins that mediate recognition and splicing of (V), (D), and (J) segments. The RAG complex, which consists of RAG1/2 and the cofactor high mobility group box 1 (HMGB1), creates DNA double-strand breaks at recombination signal sequences (RSSs) that flank (V), (D), and (J) segments. All RSSs have conserved heptamer and nonamer regions separated by either 12 or 23 bp (12RSS or 23RSS) and act as specific targets for RAG1/2-mediated catalytic activity. It is thought that catalytic activity and initiation of the V(D)J pathway consists of three distinct steps. First, the RAG complex binds and nicks a single RSS substrate to create a single RSS complex (SC). Next, two partner RSSs, generally a 12RSS and 23RSS, synapse to form a paired complex (PC). Finally, RAG1/2 cleaves both RSSs through a transesterification reaction, resulting in hair pinned coding ends that are repaired by nonhomologous end joining (5–12). Whether

nicking occurs before or after synapsis is unclear, as is the stability of the synaptic complex (PC). Kinetic studies favor synapsis formation before nicking and a very high stability of the synaptic complex, once formed, but these concepts are based on bulk solution (ensemble) studies, and the specific steps of the reaction mechanism are uncertain due to inevitable averaging (13).

Recently, crystallography and electron microscopy techniques have been used to resolve the structures of the SC and PC (14, 15). Although these studies provide new insights into the mechanism of the RAG complex and its structural functionality, the explicit steps, related conformational dynamics, and molecular choreography involved in this intricate process remain undefined. Here we apply single-molecule fluorescence resonance energy transfer (smFRET) and single-molecule colocalization (smCL) microscopy techniques to determine the real-time activity of the RAG complex in V(D)J recombination (see *SI Introduction*). Single-molecule approaches are particularly advantageous in resolving important molecular features that are otherwise masked due to averaging in ensemble-based assays (16). We therefore probed the conformational dynamics of RSS DNA induced by the binding of RAG1/2 and HMGB1 using smFRET and revealed distinct bending modes of the RSS substrates at their conserved regions. The binding kinetics of fluorescently labeled RAG1/2 was derived from smCL assays, showing that binding stability is enhanced in the presence of

## Significance

**Recombination-activating genes 1 and 2 (RAG1/2) and high mobility group box 1 (HMGB1) create dsDNA breaks at recombination signal sequences (RSSs), initiating the variable, diversity, and joining [V(D)J] recombination pathway for antigen–receptor gene assembly. We have discovered that RAG complex binds and bends RSS DNA at the heptamer, with different binding modes for 12RSS and 23RSS substrates. Our experimentally derived binding kinetics have established that binding, bending, and synapsis precede catalysis. We have also directly observed the dynamics of the synaptic complex, revealing that RAG1/2 bending places the nonamers at nearly perpendicular orientations and that the RAG:12RSS:23RSS synaptic complex is very stable. We provide a kinetic model that integrates physical binding, bending, and synapsis steps of the RAG1/2 pathway.**

Author contributions: J.Z., M.R.L., and E.R. designed research; J.Z. performed research; N.S., Z.A.E., G.W., and M.R.L. contributed new reagents/analytic tools; G.W. generated structural models; J.Z. analyzed data; and J.Z., M.R.L., and E.R. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

<sup>1</sup>To whom correspondence may be addressed. Email: eli.rothenberg@nyumc.org or lieber@usc.edu.

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1606721113/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1606721113/-DCSupplemental).

a second RSS. Finally, using smFRET capture assays, we characterized 12/23RSS synopsis and the internal dynamics of the PC. This showed a tight heptamer–heptamer association accompanied by RSS bending, thus placing the nonamers at a nearly perpendicular orientation. We find that RAG binding, bending, and synopsis precede catalysis and that the synaptic complex is very stable. Together, our findings provide a comprehensive mechanistic depiction of the structural dynamics and associated enzyme kinetics of RAG complex-mediated V(D)J recombination.

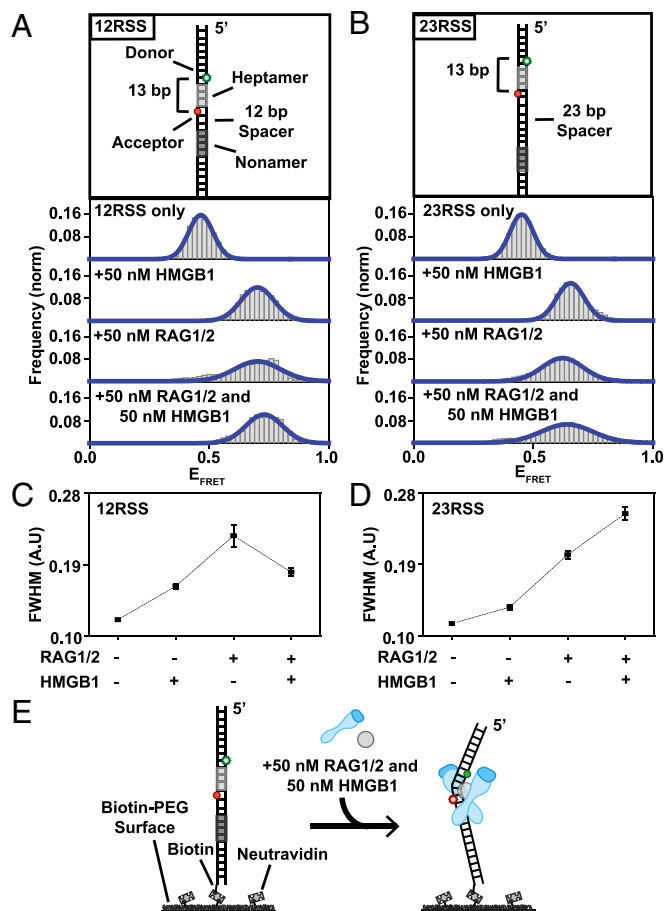
## Results

**RAG Complex Induces Conformational Changes in 12RSS and 23RSS DNA.** To probe RSS conformational changes associated with the binding of RAG complex to RSS sites (called single RSS complexes or SC formation), FRET probes were designed with donor (Cy3) and acceptor (Cy5) molecules flanking the heptamer region

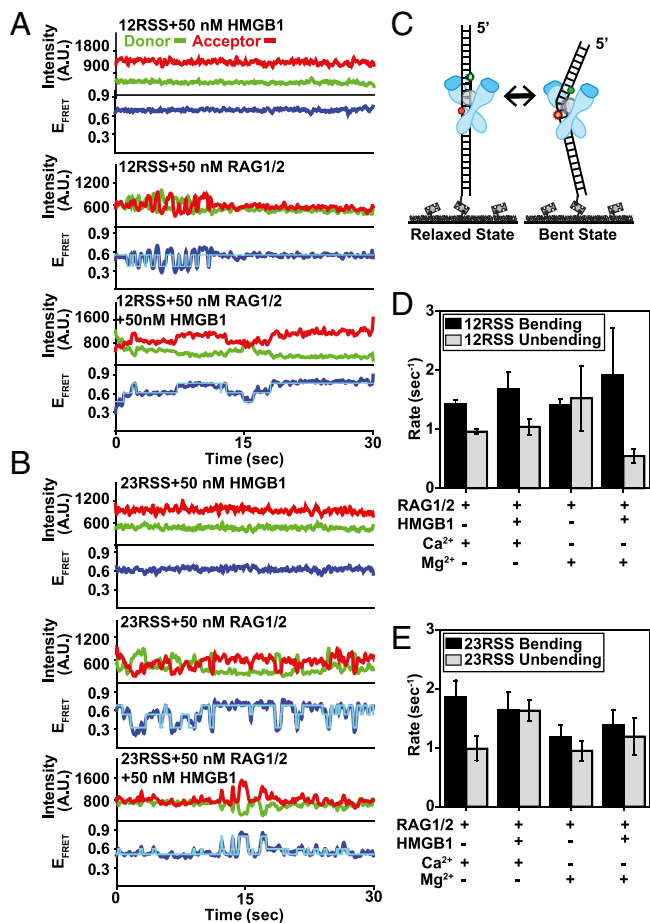
(Fig. 1 *A* and *B*). The placement of the donor and acceptor molecules in this configuration resulted in a FRET efficiency of  $\sim 0.45$  (DNA only) (Fig. S1*A*). Upon addition of 50 nM HMGB1, both the 12RSS and 23RSS substrates showed substantial increases in their FRET distribution, with a mean FRET value of  $\sim 0.75$ . This transition to high FRET reflects the DNA bending capability induced by HMGB1 even in the absence of RAG1/2 (17). Bending was not limited to RSS substrates and was also observed with dsDNA substrates that lacked heptamer and nonamer regions, highlighting the nonspecific nature of HMGB1-induced DNA bending (Fig. S1 *C–E*). We note that the dsDNA bending referred to throughout this study is broadly defined and consists of the coupled twisting and bending of the double helix, which would result in a change in FRET (18).

Addition of 50 nM RAG1/2 led to an increase in FRET distribution to a mean value of  $\sim 0.75$  for the 12RSS and 23RSS substrates (Fig. 1 *A* and *B* and Fig. S2 *A* and *B*). The transitions to high FRET that occurred under noncatalytic conditions ( $\text{Ca}^{2+}$ ), when both nicking and hairpin formation are inhibited (5), prompted the conclusion that  $\text{Mg}^{2+}$  does not limit the bending step but rather modulates subsequent catalytic activity (Fig. S2 *A* and *B*). Transitions to high FRET for substrates in the presence of  $\text{Mg}^{2+}$  are consistent with RAG1/2-induced bending of both 12RSS and 23RSS substrates previously reported in ensemble (i.e., bulk solution) studies (Fig. 1 *A* and *B*) (19, 20). The observed bending induced by RAG1/2 differed from HMGB1 activity in two distinct ways: specificity and stability. Unlike HMGB1, RAG1/2 was unable to bend dsDNA lacking an RSS as efficiently as RSS DNA, suggesting an RSS-specific conformational change (Fig. S1*C*). Although there were some FRET fluctuations upon addition of 50 nM RAG1/2 to dsDNA, the broad range of dynamics and lack of distinct FRET peaks (other than the sustained DNA-only value) indicate that any transitions visualized are most likely nonspecific and minor interactions between RAG1/2 and non-RSS DNA. Additionally, full width at half maximum (FWHM) analysis of FRET histograms showed that RAG1/2 populations had a broader FRET distribution than HMGB1 populations, representing a less stable conformational change and rapid bending fluctuations (Fig. 1 *C* and *D*). To determine HMGB1's role as a cofactor, we conducted experiments with both 50 nM RAG1/2 and 50 nM HMGB1 in solution. Previous studies have hypothesized roles for HMGB1 in stabilizing the RAG1/2 nucleoprotein complex and increasing RAG1/2 binding affinity for RSS substrates (21, 22). We found that FRET distribution for the 12RSS substrate was narrowed in the presence of RAG–HMGB1 complex in comparison with RAG1/2-only conditions (Fig. 1*C*). Narrowing of the high-FRET peak may indicate less dynamic bending of the RSS heptamer region due to HMGB1-induced stabilization of the bent conformation (Fig. 1*E*). We note that changes in FRET distributions can also stem from a change in the distribution of values stably displayed by individual molecules. Our FRET value obtained for the 12RSS- and 23RSS-SC is in broad agreement with recent cryo-EM structures (Fig. S1 *G* and *H*) (14).

To further characterize the RAG–HMGB1 complex-induced bending conformation of the RSS heptamer region, we focused on features we observed in individual smFRET trajectories. Representative trajectories for 12RSS (Fig. 2*A*) and 23RSS (Fig. 2*B*) substrates were obtained in the presence of 50 nM HMGB1, 50 nM RAG1/2 ( $\text{Mg}^{2+}$ ), or both (Fig. 2 *A* and *B*, *Top*, *Middle*, and *Bottom*, respectively). The binding of HMGB1 led to persistent high FRET (Fig. 2 *A* and *B*, *Top*), indicative of stable bending of the heptamer region. In contrast, the presence of RAG1/2 and RAG–HMGB1 complex (RAG1/2+HMGB1) (Fig. 2 *A* and *B*, *Middle* and *Bottom*, respectively) resulted in rapid changes in FRET value, characteristic of dynamic transitions between bent and relaxed RSS conformations (Fig. 2*C*). Importantly, these dynamics were also prevalent under noncatalytic conditions in the presence of  $\text{Ca}^{2+}$  and therefore stem from RAG1/2



**Fig. 1.** RSS conformational changes in the SC. (*A* and *B*) Illustrations of RSS DNA FRET substrates are located above their respective histograms. Both 12RSS and 23RSS substrates display a clear shift to high FRET upon addition of 50 nM HMGB1, 50 nM RAG1/2, and 50 nM RAG–HMGB1 complex in the presence of  $\text{Mg}^{2+}$ . Histograms were generated after subtracting the zero FRET values and truncating photobleached portions of FRET trajectories. A minimum of 75 smFRET trajectories were used to generate each histogram. (*C* and *D*) Calculated FWHM values for RSS substrates using Gaussian fit: (*C*) 12RSS displayed an increase in FWHM upon addition of 50 nM RAG1/2, indicating more dynamic trajectories. The peak is narrowed upon addition of 50 nM RAG–HMGB1 complex, implying HMGB1 stabilization of the high-FRET conformation. (*D*) 23RSS substrate maintained broad histograms upon addition of 50 nM RAG1/2 and 50 nM RAG–HMGB1 complex, indicating a lack of HMGB1-stabilization of the RAG complex at the heptamer and different binding modes for the 12RSS and 23RSS substrates. (*E*) Illustration of RAG–HMGB1 complex-mediated bending of RSS substrates.



**Fig. 2.** Conformational dynamics of the SC and evaluation of HMGB1 stabilization. (A and B) Representative smFRET trajectories for 12RSS (A) and 23RSS (B) substrates in the presence of 50 nM RAG1/2, 50 nM HMGB1, and 50 nM RAG–HMGB1 complex (Mg<sup>2+</sup>) shown in the *Top*, *Middle*, and *Bottom* trajectories, respectively. For each smFRET trajectory, the top panel displays donor (green) and acceptor (red) intensities and the bottom panel displays the corresponding FRET efficiency ( $E_{FRET}$ ) in blue. HMM fit is in cyan. (C) Illustration of RSS substrate transitioning between the low-FRET (relaxed) state and the high-FRET (bent) state. (D and E) Calculated mean binding and dissociation rates for 12RSS (D) and 23RSS (E) substrates with different proteins and ions in solution (error bars, s.e.m.;  $n > 20$  for all measurements).

binding to DNA rather than any catalysis that might occur in the presence of Mg<sup>2+</sup> but not Ca<sup>2+</sup> (Fig. S2 A and B).

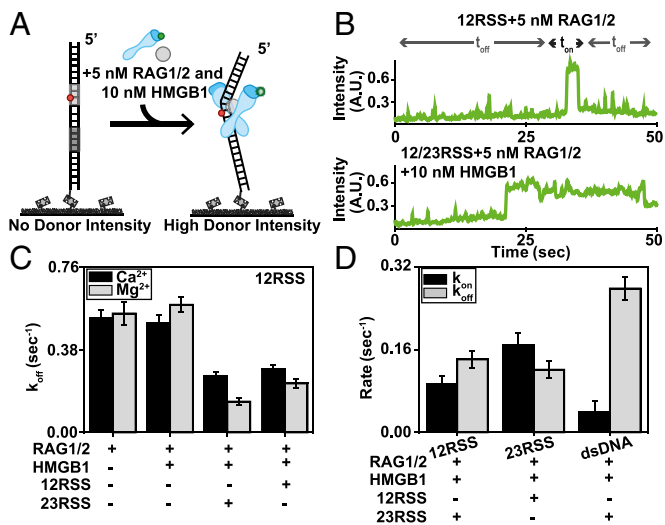
To quantify the dynamics of these trajectories, we performed Hidden Markov Model (HMM) analysis, yielding the transition rates between specific FRET states (23). Trajectories resulting from HMM analysis could be superimposed onto smFRET trajectories to demonstrate their correlation with the original trajectories and to exemplify the existence of well-defined states (Fig. 2 A and B and Fig. S2 A and B). Using the extracted frequency and transition rates for each FRET state, we constructed 2D transition density plots (TDPs) (Fig. S2 C and D). From these plots, we derived the mean rates of transition from low-to-high and high-to-low FRET states (termed bending and unbending, respectively) for the 12RSS and 23RSS substrates (Fig. 2 D and E). Addition of 50 nM RAG–HMGB1 complex in the presence of Mg<sup>2+</sup> significantly decreased the rate of 12RSS unbending at the heptamer (Fig. 2D). This highlights a longer dwell time in the high-FRET (bent) state and confirms stabilization of the SC at the heptamer by HMGB1 under catalytic conditions. These changes were not observed at the heptamer region of the 23RSS substrate, demonstrating that

HMGB1-bending stabilization of the heptamer is specific to 12RSS (Fig. 2E). Moreover, we found that RAG1/2-induced conformational changes of the 12RSS were maintained in the presence of 23RSS, consistent with similar bending patterns in the SC and PC (Fig. S1F).

We next sought to define the conformational changes occurring in the nonamer region and between the conserved regions in the presence of RAG complex proteins. To measure conformational changes in the nonamer region and between nonamer and heptamer regions, we used 12RSS DNA substrates with FRET pairs flanking the nonamer region or FRET pairs positioned at both heptamer and nonamer regions (Fig. S3 A and D). Similar to the behavior observed for the heptamer region (Figs. 1 and 2), the nonamer region displayed an increase in FRET or bending upon addition of RAG complex proteins (Fig. S3A). SmFRET trajectories confirmed that nonamer bending was persistent for HMGB1 but dynamic in the presence of 50 nM RAG1/2 and 50 nM RAG–HMGB1 complex (Fig. S3B). The combined effect of both HMGB1 and RAG1/2 at the nonamer resulted not in the high-FRET bending modes observed for either HMGB1 or RAG1/2 but rather in a high-FRET bent mode as well as an intermediate (nonbent) FRET mode. This result suggested that when RAG1/2 and HMGB1 bind together, their combined behavior may attribute to the reorganization of the complex and destabilization of bending at the nonamer. Although both the heptamer and nonamer are necessary for RAG1/2 bending at the RSS heptamer (Fig. S1 D and E), destabilization of bending at the nonamer by RAG–HMGB1 complex would allow for HMGB1-induced stabilization at the heptamer (Fig. S3C). HMGB1 could act as a cofactor by localizing RAG1/2 to its targeted nicking and cleavage site, which is 5' with respect to the heptamer region. We conclude that the heptamer is the pivotal point of bending. We also note that the combined bending induced by the RAG–HMGB1 complex within the RSS region is substantial enough to bring nonamer and heptamer regions within FRET range of each other (Fig. S3D), which is in agreement with previous ensemble measurements (19, 20).

**Binding Kinetics of the RAG Complex.** To monitor the kinetics of RAG1/2 binding to DNA substrates in real time, we developed a smCL assay using fluorescently labeled (Alexa Fluor 546) RAG1/2 (Fig. S4A). This enabled us to visualize both the RAG complex and the Cy5-labeled DNA substrate to which it binds (Figs. 3 and 4A). Surface-tethered DNA was labeled with a Cy5 fluorophore marker to confirm that RAG1/2<sup>AF546</sup> was binding to RSS substrates and to avoid interference from nonspecific reactions. To prevent background noise due to nonspecific surface binding of RAG1/2<sup>AF546</sup>, lower concentrations of RAG1/2 and HMGB1 were used. These concentrations were shown to have similar effects on the 12RSS substrate compared with the results of higher concentrations (Fig. S4B). Due to the limited evanescent field of total internal reflection fluorescence (TIRF) microscopy and the diffusion rates of proteins in solution, green intensity was only detected when RAG1/2<sup>AF546</sup> interacted with surface-tethered DNA substrates. For these experiments, we monitored fluctuations in green intensity due to RAG1/2<sup>AF546</sup> binding events rather than FRET. Fig. 3B and Fig. S5C show representative single-molecule trajectories, each with different proteins in solution, where both binding and dissociation events are observed. Trajectories display an increase in green intensity upon RAG1/2<sup>AF546</sup> binding and a drop back to baseline upon dissociation.

To derive RAG complex-binding kinetics, we measured the dwell times,  $t_{on}$  and  $t_{off}$ , in individual trajectories and generated dwell time histograms for 12RSS and 23RSS substrates (Fig. S4 C–F). Histograms were fitted with exponential decay curves, which enabled the determination of  $k_{on}$  and  $k_{off}$  rates (Fig. 3 D and E). For the 23RSS substrate, there was an increase in RAG1/2 dwell time in the presence of the HMGB1 cofactor (Fig. S5 E and F), which highlights HMGB1's role in stimulating



**Fig. 3.** Binding kinetics of the RAG complex. (A) Illustration of RAG1/2<sup>AF546</sup> smCL assay. Upon RAG1/2 binding, high donor intensity is detected. (B) Representative single-molecule binding trajectories for the 12RSS substrate in the presence of 5 nM RAG1/2, 10 nM HMGB1, and 1 nM 23RSS (Mg<sup>2+</sup>). Instantaneous increases in intensity indicate RAG1/2<sup>AF546</sup> binding events. (C)  $k_{off}$  rates for the 12RSS substrate in the presence of RAG complex proteins using the inverse of RAG1/2<sup>AF546</sup> dwell time ( $1/t_{on}$ ). Addition of 10 nM HMGB1 did not substantially change the dwell time for the 12RSS substrate, whereas addition of the 23RSS substrate to the solution decreased  $k_{off}$ . (D)  $k_{on}$  and  $k_{off}$  rates for 12RSS, 23RSS, and dsDNA substrates in the presence of 50 nM RAG–HMGB1 complex and 1 nM RSS (Mg<sup>2+</sup>).

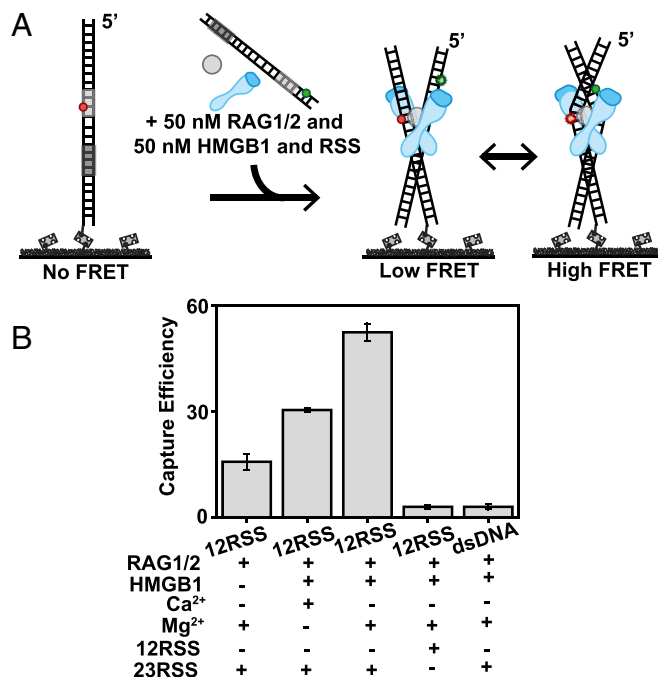
RAG1/2 binding (5). This may also be reflected in their kinetics where  $k_{off}$  transitioned from  $0.20 \pm 0.01 \text{ s}^{-1}$  for RAG1/2-only conditions to  $0.18 \pm 0.02 \text{ s}^{-1}$  upon addition of RAG–HMGB1 complex proteins. Reflecting the trend observed in trajectories, there was also a clear increase in dwell time when both 12RSS and 23RSS were included in solution with RAG–HMGB1 complex (Mg<sup>2+</sup>). This change was more drastic under catalytic conditions than in the presence of Ca<sup>2+</sup> (Fig. S4C). Overall, a decrease in  $k_{off}$  highlighted an increase in RAG1/2 binding stability upon PC formation (Fig. 3C). This increase in stability was observed when either a 12RSS or 23RSS was in solution with the surface-bound 12RSS, implying an attempt at synapsis despite the 12/23 pair rule. From our data, we experimentally determined that the  $k_{off}$  of RAG1/2 is  $0.14 \pm 0.02 \text{ s}^{-1}$  and the  $k_{on}$  is  $0.09 \pm 0.02 \text{ nM}^{-1}\text{s}^{-1}$  (Fig. 3D and Fig. S5A and B). This provides an experimental  $K_D$  of  $1.52 \pm 0.31 \text{ nM}$ , which is close to previous kinetic models of RAG1/2 enzyme binding ( $4.7 \pm 0.8 \text{ nM}$ ) (Fig. S5E) (13). The metrics were also similar to the  $k_{off}$  and  $k_{on}$  rates measured for synapsis reactions with surface-tethered 23RSS and 12RSS in solution, which yielded the following:  $k_{off} = 0.12 \pm 0.02 \text{ s}^{-1}$  and  $k_{on} = 0.17 \pm 0.02 \text{ nM}^{-1}\text{s}^{-1}$  (Fig. 3D and Fig. S5A and B). Additionally, RAG1/2 showed high substrate specificity. RAG1/2 had fleeting interactions with dsDNA lacking an RSS with a measured  $k_{off}$  of  $>0.50 \pm 0.06 \text{ s}^{-1}$  and  $k_{on}$  of  $0.04 \pm 0.02 \text{ nM}^{-1}\text{s}^{-1}$  (Fig. 3D and Fig. S5D). These are in agreement with previous studies showing that RAG1/2 has a decreased affinity for non-RSS substrates (24).

**Visualizing RAG1/2-Mediated RSS Pairing and Synapsis in Real Time.**

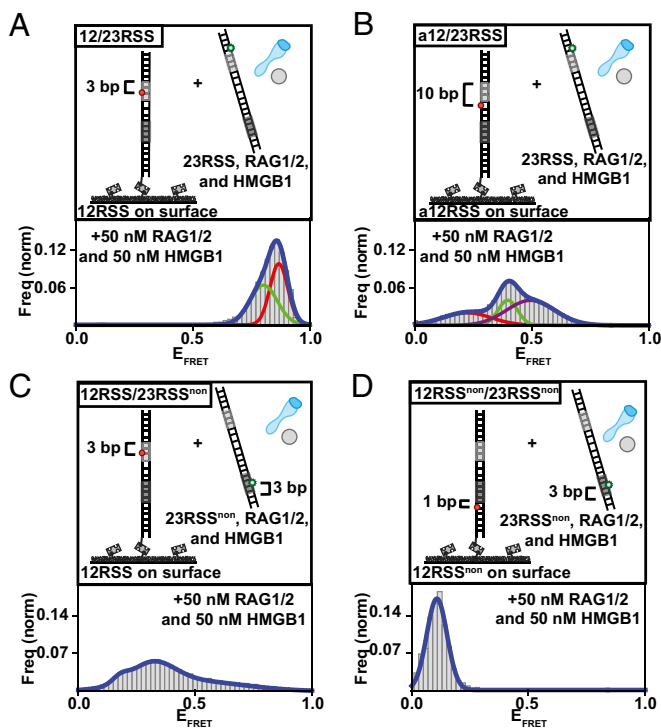
To investigate the requirements for synapsis and to monitor PCs in real time, we established a smFRET capture assay (Fig. 4A). This assay used an acceptor-labeled RSS substrate tethered to the surface and a donor-labeled RSS substrate with RAG complex proteins in solution. Because only donor molecules were being excited, no signal was produced until donor- and acceptor-labeled substrates synapsed. By positioning donor and acceptor

molecules near the heptamer region of the synapsed RSS substrates, we were able to visualize the heptamer–heptamer dynamics of the PC. To visualize a substantial number of molecules despite low synapsis efficiency, proteins were incubated for 90 min for all reactions (Fig. S6A). To ensure that the selected molecules exhibited comparable conformational changes at 20-min and 90-min time points, we compared their smFRET histograms. Similar peaks confirmed that we were not visualizing highly stable complexes that were unable to undergo hairpin formation but were rather visualizing highly stable complexes that were very similar populations despite the time difference (Fig. S6B and C). By quantifying the number of PCs formed on the surface (Fig. S6D), we obtained the efficiency of synapsis as a function of DNA substrate and added proteins. This revealed that both HMGB1 and Mg<sup>2+</sup> are important in increasing synapse efficiency (Fig. 4B). Although there was 12/23RSS synapsis in the presence of Ca<sup>2+</sup>, it occurred with lower efficiency than in the presence of Mg<sup>2+</sup>. Adding HMGB1 to the solution increased 12/23RSS synapsis by over threefold. Conversely, both dsDNA/23RSS and 12/12RSS reactions led to unsuccessful PC formation, confirming the 12/23 pair rule (25).

To determine the position and dynamics of the conserved regions within the 12/23RSS PC, we carried out synapsis experiments with 12/23RSS DNA substrates, probing different portions of RSS substrates (Fig. 5A–D and Fig. S7E). These substrates enabled us to monitor FRET changes between the heptamer–heptamer, nonamer–nonamer, and heptamer–nonamer regions within the paired 12/23RSS complex. The FRET histograms for these substrates revealed substantial high and medium FRET populations for the 12/23RSS heptamer regions (Fig. 5A and B), indicating a close association between them. The PC also showed distinctive medium–low FRET interaction between the 23RSS nonamer and 12RSS heptamer regions (Fig. 5C), but little to no FRET was observed between nonamer regions (Fig. 5D). We note that these findings are consistent with



**Fig. 4.** Synapsis requirements and visualization of the PC. (A) Illustration of smFRET capture assay for RSS/RSS synapsis. (B) Capture efficiency was measured as the average number of smFRET pairs per imaged area (error bars, s.e.m.;  $n > 3$ ). The RSS listed directly below the graph represents the RSS tethered to the surface, and those in the subsection represent RSSs in solution.



**Fig. 5.** Analysis of the internal organization and conformational dynamics of the PC. (A–D) Illustrations of RSS substrates used are located above their respective histograms. The 12RSS (A) and a12RSS (B) substrates differ in the location of the acceptor molecule. FRET histograms for both 12RSS and a12RSS displayed broad FRET distributions corresponding to multiple FRET states. This indicates internal dynamics within the PC. To emphasize the different states, histograms were fit with multiple-peak Gaussian fits. (C) FRET histogram for 12RSS<sup>non</sup> and 23RSS<sup>non</sup> substrates displayed broad FRET distribution corresponding to multiple intermediate-low FRET states. This indicates that bending of 23RSS brings its nonamer region within FRET distance of the 12RSS heptamer region. (D) Unlike at the heptamer–heptamer and nonamer–heptamer regions, there is little to no FRET between nonamer regions within the PC. A minimum of 75 smFRET trajectories were used to generate each histogram.

the recently published PC structure (14) and the coordination of the conserved regions within it (Fig. S7F). The observed FRET populations indicate that PCs retained not a single state but rather multiple conformations. This is evident in the smFRET trajectories of individual 12/23RSS synaptic complexes where transitions between different conformational states in the heptamer–heptamer (Fig. S6E and F) and nonamer–heptamer (Fig. S6G) regions can be observed. HMM analysis was used to resolve their internal dynamics and extract their transition rates (Fig. S7B–D). Importantly, the obtained rates are within the range of the bending rates of the conserved regions and likely correspond to the respective bending of the heptamers within the PCs. Despite being a dynamic process, the stability of PCs increased over time. After 20 min, 54% of 12/23RSS complexes displayed FRET behavior for the entire trajectory, and after 90 min, 73% of complexes displayed FRET behavior for the entire trajectory (Fig. S7A). Our data confirm increased stability and decreased dissociation of RSS substrates upon PC formation (13). Taken together, our analysis reveals that the PC is stabilized in a conformation where the heptamer regions are in close association, which would be an optimal configuration for cleavage (Fig. S8).

## Discussion

In the work reported here, we used single-molecule assays to define the steps and dynamics of the RAG1/2 reaction mechanism

in real time. smFRET analysis of the RAG complex as it binds to its target (RSS) DNA site revealed that bending occurs at the heptamer, even under noncatalytic conditions (Fig. 1 and Fig. S2). This bending is substantial and comparable to conformational changes in the synaptic complex (Fig. S1). Additionally, our assays allowed us to determine HMGB1's role as a cofactor and how this protein stabilizes RAG1/2-induced bending at the 12RSS heptamer (Fig. 2). HMGB1 also destabilizes RAG1/2-induced bending at the nonamer, highlighting different functionalities of the conserved regions in binding, bending, and catalytic activity (Fig. S3). Using smCL assays, we have derived the DNA binding kinetics of individual RAG1/2–12RSS, –23RSS, and –non-RSS substrates, values that were previously based on inferential assays (Fig. 3). Our data revealed that HMGB1 stimulates RAG1/2 binding to 23RSS and that RAG1/2 dwell time is longer upon synaptic complex formation. We have also derived the  $k_{on}$  and  $k_{off}$  values of RAG1/2 on dsDNA, showing exceedingly short dwell times and highlighting RAG1/2 substrate specificity. The internal conformational dynamics and architecture of the conserved regions within the synaptic complex were characterized with smFRET capture assays, resolving stable association of the heptamer regions and extreme bending of the 23RSS substrate in the 12/23RSS synaptic complex. These assays also revealed that 12/23RSS synapsis can occur before catalytic activity (i.e., nicking) and that the stability of the synaptic complex increases over time (Figs. 4 and 5). Our results provide an integrated structural and kinetic description of RAG1/2 complex activity and the formation and coordination of the synaptic complex in initiation of V(D)J recombination (SI Discussion).

The effect of RAG1/2 complex binding to RSS DNA was previously addressed using biochemical and biophysical techniques (19, 20, 22, 26) and more recently visualized within the PC in structural analysis (see SI Discussion) (14). These studies indicated that binding of HMGB1 and RAG1/2 induced bending in both the 12RSS and 23RSS DNA substrates as well as in each substrate within the PC. To examine these, we used smFRET analysis and monitored the bending and conformational dynamics of the heptamer and nonamer regions within the RSS substrate. Our measurements revealed that HMGB1 induces persistent bending of DNA substrates that is not RSS sequence-specific, whereas binding of RAG1/2 to DNA substrates results in dynamic bending that is specific for the RSS sequence (Figs. 1 and 2 and Fig. S1). These bending modes were observed around the nonamer and heptamer regions in both the 12RSS and 23RSS substrates (Fig. S3). Additionally, we found that in the presence of HMGB1, RAG1/2-induced transient bending is stabilized at the heptamer but destabilized at the nonamer for the 12RSS substrate (Figs. 1 and 2 and Fig. S3). For the 23RSS substrate, we did not find any change in stabilization by HMGB1 at the heptamer but rather did find an increased RAG1/2 dwell time in the presence of its cofactor. Together our findings demonstrate that the RAG1/2 complex induces dynamic bending within the RSS substrates but with distinctive bending modes exhibited within the 12RSS and 23RSS in the PC.

The DNA substrate specificity of the RAG complex is fundamental to the V(D)J recombination process, uniquely designating the role of RSS regions (5, 6). This specificity, along with the 12/23 rule, has been the subject of several ensemble studies as well as a recent single-molecule study (22). Although these data provided important characteristics of RAG1/2 DNA binding activity, they were mostly derived from inferential assays. Here, we used smCL assays to directly monitor the binding kinetics and substrate specificity of fluorescently labeled RAG1/2 in real time. Our measurements reveal that RAG1/2 binding to RSS is specific and dramatically increases in the presence of both 12RSS and 23RSS and Mg<sup>2+</sup> (Fig. 3). The binding kinetics derived from our experiments provide a basis for a description of the progression of the different steps in the recombination process, including RAG1/2-mediated binding, nicking, and

synapsis, as recently outlined in molecular models (13). Lower boundaries that were predicted for  $k_{on}$  and  $k_{off}$  for 12RSS binding (designated  $a_{12}$  and  $b_{12}$ , respectively, in supplemental table 1 of ref. 13) are reasonably close to our measured values here. Specifically, their predicted lower boundary is  $a_{12} = 1.55 \text{ nM}^{-1} \cdot \text{min}^{-1}$  or  $0.0258 \text{ nM}^{-1} \cdot \text{s}^{-1}$ , and our measured  $k_{on}$  here is  $0.09 \pm 0.02 \text{ nM}^{-1} \cdot \text{s}^{-1}$ . Their predicted lower boundary is  $b_{12} = 6.9 \text{ min}^{-1}$  or  $0.115 \text{ s}^{-1}$ , and our measured  $k_{off}$  is  $0.14 \pm 0.02 \text{ s}^{-1}$ . Given the  $k_{off}$  and  $k_{on}$  values, we conclude that RAG1/2 binds to each RSS substrate sequentially before any catalytic reactions (nicking or hairpinning). Subsequently, RAG1/2 will nick both RSSs in the PC before proceeding to hairpin formation. This kinetic model is illustrated in Fig. S8A.

Beyond the conformation of the RSS substrate and RAG complex binding kinetics, we also determined the efficiency of pairing and the architecture and association of the substrates within the 12/23RSS synaptic complex (Figs. 4 and 5). Consistent with recent cryo-EM studies of the 12/23RSS RAG1/2 synaptic complex (14), our data show a stable and close association between the heptamer regions (Fig. S8B), whereas the nonamer regions are perpendicular (crossed or X conformation), such that the RSS nonamer ends are positioned further away from each other (Fig. S7F). Additionally, the extreme bending ( $\sim 120^\circ$ ) of the 23RSS substrate brings its nonamer region within FRET distance of the heptamer region in the 12RSS substrate. The observed bending dynamics of the RSS substrates (Figs. 1 and 2) are also evident within the synaptic complex (Fig. 5), where the rates of FRET transitions measured between opposing heptamer

regions are within the range of bending transitions (Fig. S7D). Although the heptamer regions have minor variations in FRET value, the large dynamics at locations farther away from the site of cleavage could be associated with the flexible nonamer-binding domain (NBD) (14, 27).

Overall, the unique single-molecule assays we developed in the course of this work provide a powerful platform for probing RAG1/2-mediated V(D)J recombination. The assays we describe provide a basis for the future determination of the role of RAG complex in disease and how these proteins are modulated by other factors. For example, experiments with RSS-like motifs of cryptic RSSs (cRSSs) and mutated RAG1/2 would offer an opportunity to decipher faulty molecular mechanisms and potential targets for therapeutic intervention.

## Methods

**Protein Expression and Purification.** Detailed protein production, purification, and labeling protocols are provided in *SI Methods*. Murine RAG1 (core region: amino acids 384–1008) or RAG2 (full-length: amino acids 1–527, or core region: amino acids 1–383) vectors were designed as previously described (28). Fluorescently labeled RAG proteins carried a SNAP tag. Proteins were purified as described in ref. 29. Mouse recombinant C-terminal truncated HMGB1 was expressed in bacteria and purified as previously described (30, 31).

**SmFRET.** A full description of the DNA preparation and smFRET and smCL assays is provided in *SI Methods*. All oligonucleotides (Table S1) were purchased from Integrated DNA Technology (IDT). SmFRET experiments were carried out as previously described (32, 33).

- Jung D, Giallourakis C, Mostoslavsky R, Alt FW (2006) Mechanism and control of V(D)J recombination at the immunoglobulin heavy chain locus. *Annu Rev Immunol* 24:541–570.
- Kumari G, Sen R (2015) Chromatin interactions in the control of immunoglobulin heavy chain gene assembly. *Adv Immunol* 128:41–92.
- Jung D, Alt FW (2004) Unraveling V(D)J recombination; insights into gene regulation. *Cell* 116(2):299–311.
- Swanson PC (2002) Fine structure and activity of discrete RAG-HMG complexes on V(D)J recombination signals. *Mol Cell Biol* 22(5):1340–1351.
- Schatz DG, Swanson PC (2011) V(D)J recombination: Mechanisms of initiation. *Annu Rev Genet* 45:167–202.
- Gellert M (2002) V(D)J recombination: RAG proteins, repair factors, and regulation. *Annu Rev Biochem* 71:101–132.
- Jones JM, Gellert M (2002) Ordered assembly of the V(D)J synaptic complex ensures accurate recombination. *EMBO J* 21(15):4162–4171.
- Swanson PC (2002) A RAG-1/RAG-2 tetramer supports 12/23-regulated synapsis, cleavage, and transposition of V(D)J recombination signals. *Mol Cell Biol* 22(22):7790–7801.
- Swanson PC, Kumar S, Raval P (2009) Early steps of V(D)J rearrangement: Insights from biochemical studies of RAG-RSS complexes. *Adv Exp Med Biol* 650:1–15.
- Ropars V, et al. (2011) Structural characterization of filaments formed by human Xrcc4-Cernunnos/XLF complex involved in nonhomologous DNA end-joining. *Proc Natl Acad Sci USA* 108(31):12663–12668.
- Lescale C, et al. (2016) RAG2 and XLF/Cernunnos interplay reveals a novel role for the RAG complex in DNA repair. *Nat Commun* 7:10529.
- Roy S, et al. (2012) XRCC4's interaction with XLF is required for coding (but not signal) end joining. *Nucleic Acids Res* 40(4):1684–1694.
- Askary A, Shimazaki N, Bayat N, Lieber MR (2014) Modeling of the RAG reaction mechanism. *Cell Reports* 7(2):307–315.
- Ru H, et al. (2015) Molecular mechanism of V(D)J recombination from synaptic RAG1-RAG2 complex structures. *Cell* 163(5):1138–1152.
- Kim MS, Lapkouski M, Yang W, Gellert M (2015) Crystal structure of the V(D)J recombination RAG1-RAG2. *Nature* 518(7540):507–511.
- Cornish PV, Ha T (2007) A survey of single-molecule techniques in chemical biology. *ACS Chem Biol* 2(1):53–61.
- Sánchez-Giraldo R, et al. (2015) Two high-mobility group box domains act together to underwind and kink DNA. *Acta Crystallogr D Biol Crystallogr* 71(Pt 7):1423–1432.
- Iqbal A, et al. (2008) Orientation dependence in fluorescent energy transfer between Cy3 and Cy5 terminally attached to double-stranded nucleic acids. *Proc Natl Acad Sci USA* 105(32):11176–11181.
- Ciobotaru M, et al. (2015) The architecture of the 12RSS in V(D)J recombination signal and synaptic complexes. *Nucleic Acids Res* 43(2):917–931.
- Ciobotaru M, et al. (2013) RAG and HMGB1 create a large bend in the 23RSS in the V(D)J recombination synaptic complexes. *Nucleic Acids Res* 41(4):2437–2454.
- Thomas JO, Travers AA (2001) HMG1 and 2, and related 'architectural' DNA-binding proteins. *Trends Biochem Sci* 26(3):167–174.
- Lovely GA, Brewster RC, Schatz DG, Baltimore D, Phillips R (2015) Single-molecule analysis of RAG-mediated V(D)J DNA cleavage. *Proc Natl Acad Sci USA* 112(14):E1715–E1723.
- McKinney SA, Joo C, Ha T (2006) Analysis of single-molecule FRET trajectories using hidden Markov modeling. *Biophys J* 91(5):1941–1951.
- Zhao S, Gwyn LM, De P, Rodgers KK (2009) A non-sequence-specific DNA binding mode of RAG1 is inhibited by RAG2. *J Mol Biol* 387(3):744–758.
- Hiom K, Gellert M (1998) Assembly of a 12/23 paired signal complex: A critical control point in V(D)J recombination. *Mol Cell* 1(7):1011–1019.
- Ciobotaru M, Kriatchko AN, Swanson PC, Bright FV, Schatz DG (2007) Fluorescence resonance energy transfer analysis of recombination signal sequence configuration in the RAG1/2 synaptic complex. *Mol Cell Biol* 27(13):4745–4758.
- Yin FF, et al. (2009) Structure of the RAG1 nonamer binding domain with DNA reveals a dimer that mediates DNA synapsis. *Nat Struct Mol Biol* 16(5):499–508.
- Shimazaki N, Tsai AG, Lieber MR (2009) H3K4me3 stimulates the V(D)J RAG complex for both nicking and hairpinning in trans in addition to tethering in cis: Implications for translocations. *Mol Cell* 34(5):535–544.
- Bergeron S, Anderson DK, Swanson PC (2006) RAG and HMGB1 proteins: Purification and biochemical analysis of recombination signal complexes. *Methods Enzymol* 408:511–528.
- West RB, Lieber MR (1998) The RAG-HMG1 complex enforces the 12/23 rule of V(D)J recombination specifically at the double-hairpin formation step. *Mol Cell Biol* 18(11):6408–6415.
- Bianchi ME, Falciola L, Ferrari S, Lilley DM (1992) The DNA binding site of HMG1 protein is composed of two similar segments (HMG boxes), both of which have counterparts in other eukaryotic regulatory proteins. *EMBO J* 11(3):1055–1063.
- Chatterjee S, et al. (2014) Mechanistic insight into the interaction of BLM helicase with intra-strand G-quadruplex structures. *Nat Commun* 5:5556.
- Reid DA, et al. (2015) Organization and dynamics of the nonhomologous end-joining machinery during DNA double-strand break repair. *Proc Natl Acad Sci USA* 112(20):E2575–E2584.
- Roy R, Hohng S, Ha T (2008) A practical guide to single-molecule FRET. *Nat Methods* 5(6):507–516.
- Friedman LJ, Gelles J (2012) Mechanism of transcription initiation at an activator-dependent promoter defined by single-molecule observation. *Cell* 148(4):679–689.
- Cisse II, Kim H, Ha T (2012) A rule of seven in Watson-Crick base-pairing of mismatched sequences. *Nat Struct Mol Biol* 19(6):623–627.
- Rothenberg E, Trakselis MA, Bell SD, Ha T (2007) MCM forked substrate specificity involves dynamic interaction with the 5'-tail. *J Biol Chem* 282(47):34229–34234.
- Rothenberg E, Ha T (2010) Single-molecule FRET analysis of helicase functions. *Methods Mol Biol* 587:29–43.
- DeLano WL (2002) *The PyMOL Molecular Graphics System* (DeLano Scientific, San Carlos, CA).