Supporting Information

Lovely et al. 10.1073/pnas.1503477112

SI Text

Calibrating the TPM Assay. To calibrate the TPM assay, we generated various DNA substrates ranging in length from 252 to 2,900 bp, tethered them to beads (Fig. S1*A*), and tracked their rms motion (Fig. S1*B*). We then performed a quadratic fit to the data for rms as a function of DNA length (Fig. S1*C*) and extracted an equation that makes it possible to interconvert between rms and DNA length.

Representative RAG1/2c Trajectories, Normalized Probability Distribution, and Absolute Reduction in DNA Length by RAG1c, RAG2c, RAG1/2c, and RAG Mutants. To determine the mean dwell times for RAG1/2c shown in Fig. 4*C*, we computed the time spent in the unbound and bound states over a range of concentrations, with representative trajectories shown in Fig. S2 *A* and *B*. We added the highest concentration of RAG1/2c used in the titration to DNA substrates with no, one, or two RSSs (Fig. S2*C*) and found that nearly all substrates were bound, which is shown in the composite trajectories and normalized probability distributions (Fig. S2 *D*–*I*). We also determined if RAG1/2c alone formed a paired complex using an intersignal spacing of roughly 250 bp between two RSSs and only observed a 100-bp reduction in DNA length (Fig. S2 *G*–*J*). This value is two times the reduction in the tether length for a single RSS but twofold less than the intersignal distance, indicating that we are observing RAG1/2c complexes bound to two RSSs on a single DNA substrate. In Fig. S2*J*, we show data illustrating the detection of individual RAGs and RAG mutants binding to one or two RSSs. We find that, at 50 nM, RAG1c can bind RSSs but also exhibits promiscuous binding as evidenced by the size of the error bars relative to RAG1/2c (Fig. S2*J*). Furthermore, we show that RAG1/2c in the presence of different divalent cations, RAG1/2c catalytic mutants, and maltose binding protein (MBP) -free RAG1/2c can shorten DNA in presence of RSSs (Fig. S2*J*). However, RAG2c and RAG1/2c lacking a nonamer binding domain [RAG1(Nonamer binding domain)/2c] do not bind either RSS.

Using Waiting Times to Compute K_d. As noted in the text, there are several different ways of obtaining K_d values, which in principle, should give the same results. The text used Gaussian fits to the main peaks. In this method, two Gaussians are fit to the two peaks, and the probability of being bound is then given as the ratio of the weight under the bound peak to the total weight under both peaks. Alternatively, one can compute the fraction of time that the system spends in the bound state as a function of concentration, and this technique should yield a fit to K_d. In Fig. S3, we show the data used to carry out this analysis. Fig. S3*A* shows the dwell times in both the bound and unbound states as a function of RAG1/2c concentration. In Fig. S3*B*, we plot the fraction of time that RAG1/2c is bound (circles) as a function of concentration in the presence of 12RSS (blue circles) or 23RSS (red circles). This value is calculated by taking the fraction of time that the system spends in the bound state vs. the total time of the trajectory of interest. Each binding probability curve (12RSS or 23RSS) is fit for K_d to a Langmuir isotherm (equivalent to a Hill function with $n = 1$),

$$p_{\text{Bound}} = \frac{\frac{c}{K_d}}{1 + \frac{c}{K_d}}, \qquad \text{[S1]}$$

which follows directly from the chemical equation for equilibrium binding/unbinding of a molecule to a target. We determined a $K_d = 18.5 \pm 2.9$ nM for the 12RSS, which is within error of the K_d determined using the other analysis method, and a $K_d = 13.9 \pm 4.7$ nM for RAG1/2c–12RSS, which is in reasonable accord with both previously reported literature values of 14.5 ± 2.4 nM (1) and 25.0 ± 5.0 nM (2) (Fig. S3). In addition, we found $K_d = 54.7 \pm 1.1$ nM for the 23RSS, which is larger than the $K_d = 44.4 \pm 6.5$ nM obtained using Gaussian fits to the bulk binding probability. The difference in K_d obtained using these two methods can be understood by considering how probability is assigned to the different states in these methods. The discrepancy is revealed in the low concentration limit. Even in the presence of very low concentrations of RAG1/2c protein (Fig. 4*A*, row 1), the distribution of measurements has a width that is comparable with the separation between the peaks corresponding to the bound and unbound state. As a result, simple thresholding will naturally assign some part of the tail of this distribution to the other state. As a result, we favor fitting Gaussians as the more accurate choice of analysis here.

Expected Dwell Time in Off State from Simple Diffusion. The flux of particles arriving at a target per unit time is

$$\frac{dn}{dt} = 4\pi D a c, \qquad \text{[S2]}$$

where a is the size of the absorbing target (the RSS site in this case), c is the concentration of the absorbing molecule (RAG1/2c in this case), and D is the diffusion constant of the absorbing molecule. The diffusion constant of a particle in a solvent can be written in terms of temperature (T), the solvent viscosity (η), and the particle size,

$$D = \frac{k_B T}{6\pi\eta a}, \qquad \text{[S3]}$$

where k_B is Boltzmann's constant; $k_B = 4.11 \times 10^{-21}$ J and $\eta \sim 10^3$ Pa·s for water at room temperature. The hydrodynamic radius of RAG1/2c is roughly 10 nm, and we will assume that the size of the target is roughly 1 nm. Therefore, we conclude that

$$\frac{dn}{dt} \approx 0.1 \left[\frac{\text{particles}}{\text{seconds nanomolar}}\right] c. \qquad \text{[S4]}$$

However, for our data, the leading constant is $(2.5 \pm 0.5) \times 10^{-4}$ (a difference of three orders of magnitude). However, we note that this calculation is made in the limit of a perfect absorber, meaning that, every time that a molecule arrives at the RSS, it will bind. This assumption completely neglects factors, such as the orientational effect, which requires the protein to arrive at its binding site in the correct orientation.

Effect of DNA Length Flanking RSS Sites on RAG–HMGB1–RSS Complex Formation and Percentage of DNA Compaction by HMGB1 as a Function of Concentration for a Range of DNA Lengths. As noted in the text, one concern in performing these TPM experiments is the proximity of the bead and the coverslip to the relevant RSSs. To explore the consequences of the bead for the protein–DNA complexes, we performed a series of experiments, in which the distance between the RSS and the bead was varied as shown in Fig. S4*A*. Indeed, we find that the distance between the RSS and the bead has to be sufficiently large to avoid restricted motion. As shown in Fig. S4*B*, increasing the distance between the bead and the RSS from 126 to 426 bp leads to a substantial increase in the rms motion. We also

find that the increase in rms motion is conserved for RAG1/2c and catalytically inactive RAG1(D708A)/2c in the presence of HMGB1, indicating that the reduction in length is independent of RAG nicking activity and by inference, facilitated by HMGB1 (Fig. S4*B*). We also determined the percentage of DNA compaction by HMGB1 alone from 5 to 225 nM for DNA substrates ranging in length from 539 to 2,900 bp. We found that the percentage of DNA compaction was relatively the same at a particular concentration of HMGB1 (Fig. S4*C*).

Computing the Predicted DNA Length for the Paired Complex. To compute the tether length of the predicted paired complex in the presence of 5 nM RAG1/2c and 80 nM HMGB1 for DNA substrates with an overall contour length of 2,900 bp and an intersignal distance of 1,200 or 1,800 bp, we subtract the intersignal distance of 1,200 or 1,800 from the contour length (2,900 bp) and obtain the expected tether length of 1,700 or 1,100 bp, respectively. Then, we subtract the 20% reduction in DNA length by 80 nM HMGB1 alone (Fig. S4*C*) from the remaining expected tether length (1,700 bp for the 1,200-bp intersignal distance or 1,100 bp for the 1,800-bp intersignal distance). This calculation results in a predicted tether length for the paired complex of 1,360 bp for the 1,200-bp intersignal distance or 880 bp for the 1,800-bp intersignal distance.

Composite Trajectories and Normalized Probability Distributions of 12RSS/23RSS–1,200 bp and 12RSS/23RSS–1,800 bp Bead Loss Events Preceded by Paired Complex Formation. We performed experiments to directly observe the paired complex on DNA substrates with an intersignal spacing of 1,200 and 1,800 bp (Fig. 6). We determined the mean dwell time of the trajectories resulting in paired complex formation before bead release on DNA substrates with an intersignal spacing of 1,200 or 1,800 bp (black bars in Fig. 6*C*). All of the trajectories that went into those two dwell time distributions are shown in Fig. S5.

Paired Complex Mean Dwell Time and Median Dwell Length Are Uncorrelated. To investigate whether there was a relationship between the paired complex mean dwell time and median dwell length, we plotted both parameters for each paired complex with or without bead loss for the 1,200- (Fig. S6*A*) and 1,800-bp (Fig. S6*B*) intersignal distances. We found that there was no correlation between the paired complex mean dwell time and median dwell length.

Catalytically Active RAG Proteins Are Required for Bead Release. To determine that bead release was dependent on catalytically active RAG, we first performed a bulk hairpin production assay and found that we only observed hairpin production when using RAG1/2c and not catalytically inactive RAG1(D708A)/2c (Fig. S7*A*). After establishing a single-molecule hairpin production assay that uses bead release as a readout, we performed another useful control to further confirm and clarify that bead loss is indicative of the actual RAG–HMGB1 dynamics. In particular, we used mutant proteins to confirm that catalytically active RAG1/2c was responsible for bead release. To that end, we added 50 nM RAG1/2c or catalytically inactive RAG1(D708A)/2c and 500 nM HMGB1 to DNA tethers with a 12RSS and a 23RSS. We only observe bead release when using RAG1/2c and not when using RAG1(D708A)/2c (Fig. S7 *B* and *C*). As an indirect control to determine that synapsis preceded hairpin production as represented by bead release in our assay, we reduced the intersignal spacing from 1,000 to 73 bp (Fig. S7*D*). The 73-bp intersignal spacing was previously shown to ablate hairpin production in a bulk assay, and we find that we also hamper bead release in our single-molecule assay (Fig. S7*E*) (3). Interestingly, we can rescue bead release by substituting Mn^{2+} for Mg^{2+}, which eliminates the requirement for synapsis to precede hairpin production (Fig. S7*F*) (4).

SI Methods

Large-Scale Purification of the RAG Proteins. RAG proteins were purified as described previously with a few modifications (5). WT and mutant RAG proteins were purified from suspension 293-E cells using CMV-driven pTT5–MBP–RAG1c (amino acids 384–1,040) and RAG2c (amino 1–387) expression vectors. We transformed pTT5–MBP–RAG1c or pTT5–MBP–RAG2c harboring ampicillin resistance in Z-competent cells (Zymoresearch) and plated it on LB agar containing 50 μg/mL carbenicillin (Sigma-Aldrich). A single colony was selected, and a large-scale plasmid purification was performed using the PowerPrep HP Plasmid Maxiprep Kit (Origene). Next, 1 L suspension 293-E cells were grown to a cell density of 1×10^6 cells/mL before transfection in Gibco Freestyle 293 Expression Medium (Life Technologies); 1 L 293-E cells were transfected with 450 μg pTT5–MBP–RAG1c, 550 μg pTT5–MBP–RAG2c, and 1.0 mL BioT (Bioland) brought to a final volume of 35 mL in 1× PBS (note: mix DNA and 1× PBS first). The transient transfection proceeds for 48 h, and the cells are harvested at 3,000 rpm ($700 \times g$) for 15 min. Then, the cell pellet was resuspended in 30 mL lysis Buffer A (10 mM sodium phosphate, pH 7.4, 500 mM NaCl, 1 mM DTT, 0.25% Tween-20, 10 μg/mL aprotinin, 10 μg/mL leupeptin, 10 μg/mL pepstatin, 100 μM PMSF) and microfluidized on ice. The cell lysate was centrifuged at $86,000 \times g$ for 40 min using the SW28 Rotor (Beckman). The supernatant was then added to 2 mL amylose resin (New England Biolabs) and inverted for 1.5 h. The amylose resin/supernatant was poured into a 10-mL polyprep gravity flow column (Bio-Rad). The amylose resin complex was washed with 20 mL Buffer A followed by 10 mL Buffer B (10 mM sodium phosphate, pH 7.4, 500 mM NaCl, 1 mM DTT) and 300 μL elution buffer (10 mM sodium phosphate, pH 7.4, 500 mM NaCl, 1 mM DTT, 10 mM maltose) to displace column void volume and finally, 2 mL elution buffer. The protein was dialyzed for 3 h in dialysis buffer (25 mM Tris·HCl, pH 8.0, 150 mM KCl, 2 mM DTT, 10% glycerol) using a 0.5–3 mL Slide-A-Lyzer (Pierce), snap-frozen, and stored at −80 °C.

Purification of HMGB1. HMGB1 was purified as described previously with slight modifications (5). HMGB1 was codon-optimized for bacterial expression by DNA2.0 and cloned into pJExpress vector. pJExpress–HMGB1 was transformed into BL21(DE3) cells and plated on LB agar with 50 μg/mL kanamycin (kan). A single clone was picked for overnight liquid culture in 50 mL LB-kan; the next day, 30 mL overnight culture was diluted 100-fold in 3 L LB-kan. The cells were grown to an OD_{600} of 0.7 and induced by 1 mM isopropyl β-D-1-thiogalactopyranoside at 30 °C for 4 h. The cells were harvested at 8,000 rpm for 30 min and resuspended in 30 mL binding buffer (40 mM Tris·HCl, pH 8.0, 500 mM KCl, 0.25% Tween-20, 50 mM imidazole, 10 μg/mL aprotinin, 10 μg/mL leupeptin, 10 μg/mL pepstatin A, 100 μM PMSF, 25 U/mL benzonase) for every 1 L cells. The resuspended cells were passed through a microfluidizer three consecutive times to lyse them. The cell lysate was centrifuged at $46,000 \times g$ for 40 min, and the cleared lysate was loaded onto a 5-mL nickel-nitrilotriacetic acid (Ni-NTA) column. The column was washed with 40 mL binding buffer without benzonase followed by 20 mL wash buffer (40 mM Tris·HCl, pH 8.0, 65 mM KCl, 50 mM Imidazole) and eluted with 30 mL elution buffer (40 mM Tris·HCl, pH 8.0, 65 mM KCl, 500 mM Imidazole). Next, sulphopropyl cation (SP) and quaternary ammonium (Q) anion exchange chromatography were used to eliminate truncated forms of HMGB1 using buffer solutions A (40 mM Tris·HCl, pH 8.0) and B (40 mM Tris·HCl, pH 8.0, 1 M KCl). The SP column was equilibrated in 8% solution B or 80 mM KCl, and the Ni-NTA eluate was loaded onto the column, washed with 6 mL 8% solution B, and eluted in 1-mL fractions with a gradient of 25 mL solution B from 8% to 65%. The fractions with HMGB1 were

pooled, diluted to 80 mM KCl, and loaded on a cation exchange Q-column equilibrated with 8% solution B. The column was washed with 7 mL 8% solution B and eluted in 1-mL fractions with a gradient of 40 mL solution B from 22% to 70%. The 1-mL fractions with HMGB1 were pooled, concentrated, snap-frozen, and stored at −80 °C.

Oligonucleotide and Long DNA Substrates. Consensus 12RSS (DAR39 and DAR40) and 23RSS (DAR61 and DAR62) oligonucleotide substrates were used. The sequence of the top strand of the 12RSS–DAR39 was 5′-GATCTG GCCTGTCTTACACAGTGCTACAGACTGGAACAAAAACCCTGCAG-3′, and the sequence of the top strand of the 23RSS–DAR61 was 5′-GATCTGGCCTGTCTTACACAGTGGTAGTACTCCACTGTCTGGCTGTACAAAAACCCTGCA-3′ (conserved heptamer and nonamer motif are underlined) (5). The long DNA substrates of various lengths used in bulk and single-molecule assays were created by assembly PCR and cloned into a Moloney murine leukemia viral vector pOOC (NoRSS, 12RSS, 23RSS, 12RSS/23RSS, 12RSS/12RSS, and 23RSS/23RSS). All DNA substrates used in TPM were generated using universal biotin 5′-AAAGGGAATAAGGGCGACAC-3′– and digoxigenin 5′-TATGGAAAAACGCCAGCAACG-3′–labeled primers.

In Vitro Cleavage Assay. The oligonucleotide RAG cleavage assay detects nicked and hairpin products and was performed as described previously (5). First, 125 nM RAG1/2c or RAG1 (D708A)/2c and 560 nM HMGB1 were preassembled in an 8-μL reaction in dilute dialysis buffer (DDB; 25 mM Tris·HCl, pH 8.0, 150 mM KCl, 2 mM DTT) and 0.1% glycerol for 5 min. Second, 12 μL EMSA reaction buffer [ERB; 25 mM Tris·HCl, pH 7.4, 65 mM potassium acetate, 5 mM $MgCl_2$, 100 μg/mL acetylated BSA (Sigma-Aldrich), 10% (vol/vol) DMSO] containing 2 nM 12RSS (hot) and 8 nM 23RSS (cold) oligonucleotide substrates was added to RAG1/2c and HMGB1, bringing the DDB and the ERB together to form the RAG reaction buffer (RRB) and bringing the reaction volume to 20 μL with 50 nM RAG1/2c, 225 nM HMGB1, 1 nM 12RSS, and 4 nM 23RSS. The reactions were incubated at 37 °C for 1, stopped with 10 mM EDTA in 95% formamide, and incubated at 95 °C for 2 min. The reaction products were fractionated on a 15% 19:1 polyacrylamide sequencing gel (National Diagnostics) and visualized with a phosphoimager.

TPM: Flow Cell Assembly, Data Acquisition, Data Analysis, and Calibration. TPM flow cell assembly began by first drilling 12 holes into a precleaned glass slide with a 1.5-mm diamond-tipped drill bit (CRLaurence) to create inputs and outputs for six channels. The drilled glass slide (Corning) and a partner coverslip (Fisher) were sonicated (Model 75T; Aquasonic) in 100% EtOH for 30 min, rinsed with ddH_2O, sonicated in 1 M KOH for 30 min, rinsed with ddH_2O, sonicated in 5% HCl, rinsed with ddH_2O, and functionalized with epoxysilane [116 mL Isopropanol, 1 mL ddH_2O, 3 mL 3-glycidoxypropyldimethoxymethylsilane (Sigma), 120 μL *N*,*N*-Dimethylbenzylamine (Sigma)] at room temperature for 1.5 h. The drilled glass slides and coverslips were washed in Isopropanol, dried with N_2 gas, and baked at 100 °C for 30 min. Next, 0.02-in i.d./0.06-in o.d. tygon microbore tubing (Cole-Parmer) was placed in each of 12 holes of the drilled glass slide, designating an input and an output for all six channels. Then, 5-min epoxy (Devcon) was applied at the tubing–glass interface, and a piece of double-sided tape (0.12-mm thick; Grace Bio-Labs) with six holes designating the channels of the flow cell was placed on the drilled glass slide and sandwiched between a coverslip. After the flow cell was assembled, polyclonal sheep antidigoxigenin (antidig; Roche) was added to each of six channels to a final concentration of 40 ng/μL. The antidig solution was incubated overnight at room temperature to allow covalent coupling of antidig to epoxy groups within the channels of the flow cell. Next, TPM assembly buffer (TPM-AB; 20 mM Tris-acetic acid, pH 8.0, 130 mM KCl, 1 mM $MgCl_2$, 0.1 mM DTT, 0.1 mM EDTA, 20 μg/mL acetylated-BSA, 3 mg/mL biotin-free casein) was added to the flow cell's channels and incubated for 1 h to flush out uncoupled antidig and quench active epoxy groups. Then, biotin- and digoxigenin-labeled DNA was added to each channel to a final concentration of 5 pg/μL and incubated for 20 min. The excess DNA was washed away with TPM-AB and 0.49 μm streptavidin-coated polystyrene beads (Bangs Laboratories) diluted in TPM-AB to a final concentration; 3×10^{10} beads/mL were added to each channel and incubated for 10 min. The excess beads were washed away with TPM-AB and RRB, completing the assembly of DNA-tethered beads. Each channel has ∼2.5×10^6 DNA-tethered beads. The DNA-tethered beads were imaged using bright-field microscopy (because the polystyrene beads scatter light) on inverted Olympus IX71 Microscopes with either a 100× oil objective or a 60× oil objective with a 1.6× magnifier. A Basler A602f Camera was used to acquire images at 30 Hz. DNA-tethered beads were tracked by cross-correlating each frame with the initial frame in a time series for each bead. This particle-tracking technique generated raw x and y positions of each bead's excursion relative to its anchor point. Drift was corrected from these raw data by subtracting the results of a low-pass first-order Butterworth filter with a cutoff frequency $f_{cB} = 0.05$ Hz. The rms motion was obtained by applying a Gaussian filter with a −3 dB frequency of $f_{cG} = 0.0326$ Hz corresponding to a 4-s or 120-frame SD of the filter to the mean-squared displacement of the data ($\vec{x}^2 + \vec{y}^2$). The rms motion of the bead is the square root of the result of the convolution of ($\vec{x}^2 + \vec{y}^2$) and the Gaussian filter.

TPM was calibrated by tracking the rms of DNA molecules of various lengths (252, 539, 736, 946, 1,124, 1,316, 1,521, 1,717, 1,910, 2,077, 2,280, and 2,900 bp) and performing a quadratic fit to data to extract an equation that relates rms to DNA length in base pairs. To define a good DNA-tethered bead, we used the symmetry-of-motion and the length-of-motion checks used previously (6). We then tracked the good beads in RRB; then, we stopped tracking the beads, added a given amount of RAG1/2c or RAG1/2c–HMGB1, and continued tracking the same beads. This scheme allowed for precision measurements of changes to the rms motion of DNA-tethered beads by RAG1/2c and RAG1/2c–HMGB1. Bad beads that passed the initial checks were discarded if the beads stuck to the surface and the bead's initial rms was not within error to the previously measured rms for a particular DNA length.

Statistical Mechanical Model of RAG–RSS and RAG–HMGB1–RSS Binding. We fit the single-molecule RAG–RSS or RAG–HMGB1–RSS trajectories to two Gaussians to determine the probability of being in the bound or unbound state. Then, we built a statistical mechanical model using Boltzmann's distribution to compute p_{Bound} and p_{Unbound} and fit the equation for p_{Bound} to the data, namely

$$p_{\text{Bound}} = \frac{\frac{R}{R_0} e^{-\beta \Delta\varepsilon_{12\text{RSS or }23\text{RSS}}}}{1 + \frac{R}{R_0} e^{-\beta \Delta\varepsilon_{12\text{RSS or }23\text{RSS}}}}, \qquad \text{[S5]}$$

where $\beta = 1/k_BT$, R = RAG1/2c concentration, $R_0 = 1$ M, $\Delta\varepsilon_{12RSS} = -18.09\ k_BT$, and $\Delta\varepsilon_{23RSS} = -16.93\ k_BT$. In the presence of 25 nM HMGB1, we find $\Delta\varepsilon_{12RSS} = -19.31\ k_BT$ and $\Delta\varepsilon_{23RSS} = -18.54\ k_BT$.

Kinetic Theory of RAG–RSS Binding. To analyze the transition rates and dwell times in TPM data, bead trajectories were first thresholded based on rms value to be in a bound (smaller rms) or an unbound (higher rms) state or one of two spurious states. The first spurious state corresponds to a transient sticking event, where the rms of the bead plummets as its movement is halted. The second

state occurs when the tracking program has a tracking error, and the rms of the bead jumps dramatically. To identify these two spurious states, we also applied upper and lower thresholds; these thresholds were chosen conservatively and were unambiguous: both sticking and tracking errors produced enormous changes in the rms of the bead compared with the small changes observed from protein binding/unbinding. The threshold between the bound and unbound states was chosen by eye from a histogram of rms values in a sample where the bead spends roughly one-half of its time in either of the two states (bound and unbound). Next, to reduce the influence of noise in the trajectories, short states were removed. This analytical technique corresponded to removing states that did not last longer than four times the dead time of the Gaussian filter applied to the data, corresponding to roughly 21 s in this analysis (7). These states were removed by assigning the first one-half of the spurious state to the previous state of the bead and the second one-half of the spurious state to the next state of the bead. If a short state began (or ended) a trace, the entire duration was assigned to the next (previous) state. Spurious states were handled in a similar manner; spurious states occurring in the middle of the trace were assigned in the same way (the first one-half to the previous state and the second one-half to the next state), and spurious states that began a trace were assigned to the next state. However, spurious states that ended a trace were simply discarded. Statistics were then gathered on the traces produced. When calculating the mean dwell time in the bound state (Fig. 3), we applied a correction, which accounts for the fact that, even in the absence of protein, the bead occasionally dipped below the threshold and gave a spurious on state. To correct for this, we calculated the probability per unit time of such a transition, p_{false}, from our collection of no protein traces by counting the total number of false transitions observed divided by the total time of calibration across all observed beads. We also calculated the mean dwell time of one of these false transitions, τ_{false}. From here, we could calculate the expected number of false transitions observed, n_{false}, for each concentration of RAG1/2c, which was p_{false} times the total time that the bead was observed in the off state. The measured mean dwell time in the on state is

$$<\tau_{\text{obs}}> = \omega <\tau_{\text{bound}}> + (1-\omega)<\tau_{\text{false}}>, \qquad \textbf{[S6]}$$

where ω is the fraction of transition events measured that were expected to be real events (total number subtracted by n_{false} divided by the total number). Thus, the mean dwell time in the on state was calculated by

$$<\tau_{\text{bound}}> = \frac{<\tau_{\text{obs}}> - (1-\omega)<\tau_{\text{false}}>}{\omega}. \qquad \textbf{[S7]}$$

Similarly, this correction was done to the off-time measurement $<\tau_{\text{unbound}}>$. We calculated the average time in the off state by taking the total time that we measured the bead in that state divided by the total number of transitions observed in that time. We correct for false transitions in these data by removing the expected number of false transitions from the total number and adding the expected total time of those transitions to the total time. The dwell time histograms in Fig. S3 have not had the above correction applied.

Kinetic Analysis of Paired Complex Formation. Trajectories are segmented exactly as described above for RAG1/2c binding and unbinding with a few caveats. In this analysis, the bound state now corresponds to the formation of a paired complex, and we keep track of states that terminate in bead loss. For these data, we discard states as spurious if they do not last for at least 60 s; this time is set to avoid including clear spurious events in the data. Finally, tethers that do not begin in an unlooped state with an rms consistent with that expected for the given DNA length and HMGB1 concentration are discarded. With this set of conditions, we see no paired complex transitions in either control (NoRSS or single 12RSS).

TPM: RAG1/2c–RSS, RAG1/2c–HMGB1–RSS Binding, Hairpin Production, and Paired Complex Detection Assays. Single-molecule experiments performed at saturating concentrations of RAG1/2c were assembled by first diluting the RAG proteins to 200 nM in 50 μL DDB; then, 50 μL ERB was added followed by 100 μL RRB, yielding a 200 μL final volume of 50 nM RAG1/2c, which was added to the TPM flow cell by gravity flow. The rms motion of the DNA-tethered beads was tracked for 700 s before and after RAG1/2c was added. To acquire the experimental data for determining K_{d} values, RAG1/2c was titrated (1, 2.5, 5, 10, 25, and 50 nM) in the presence of a single RSS, and the DNA-tethered beads were tracked for 4,000 s. RAG1/2c–HMGB1 experiments performed at saturation were first preincubated with 50 nM RAG1/2c and 25 nM HMGB1 in DDB for 10 min. Then, the proteins were diluted in RRB to a 200 μL final volume and added to DNA-tethered beads. The rms motion of the DNA-tethered beads was tracked for 700 s before and after the addition of RAG1/2c–HMGB1. The hairpin production assay was performed from $t = 0$ h to $t = 2$ h, in which we counted the beads initially at $t = 0$ h to ensure that we started with 70–80 beads. Next, we added RAG1/2c and HMGB1 and monitored bead loss after $t = 2$ h or every 20 min for $t = 2$ h to capture the dynamics of hairpin production. The percentage of hairpin production for the 12RSS /23RSS was determined as % loss $= f_{\text{RSS}} - f_{\text{NoRSS}}$, where f_{RSS} is the fraction of beads lost with RSSs present and f_{NoRSS} is the fraction of beads lost with no RSSs present. The tethers that were not cleaved in the hairpin production assay might represent RAG- and HMGB1-bound dead-end states, HMGB1 alone-bound complexes, or complexes that still may go on to cleavage. The percentage of nonspecific DNA compaction is derived from the change in rms at different HMGB1 concentrations on the NoRSS DNA substrate. Paired complex detection assays were performed by tracking DNA tethers with a 2,900-bp contour and 1,200- or 1,800-bp intersignal spacing for 1 h in the presence of 5 nM RAG1/2c and 80 nM HMGB1. The trajectories were then inspected for a signature of paired complex formation. Single-molecule binding, hairpin production, and paired complex detection assays were conducted at room temperature.

1. Wang G, Dhar K, Swanson PC, Levitus M, Chang Y (2012) Real-time monitoring of RAG-catalyzed DNA cleavage unveils dynamic changes in coding end association with the coding end complex. *Nucleic Acids Res* 40(13):6082–6096.
2. 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.
3. Eastman QM, Leu TM, Schatz DG (1996) Initiation of V(D)J recombination *in vitro* obeying the 12/23 rule. *Nature* 380(6569):85–88.
4. van Gent DC, Ramsden DA, Gellert M (1996) The RAG1 and RAG2 proteins establish the 12/23 rule in V(D)J recombination. *Cell* 85(1):107–113.
5. Bergeron S, Anderson DK, Swanson PC (2006) RAG and HMGB1 proteins: Purification and biochemical analysis of recombination signal complexes. *Methods Enzymol* 408: 511–528.
6. Han L, et al. (2009) Concentration and length dependence of DNA looping in transcriptional regulation. *PLoS ONE* 4(5):e5621.
7. Manzo C, Finzi L (2010) Quantitative analysis of DNA-looping kinetics from tethered particle motion experiments. *Methods Enzymol* 475:199–220.

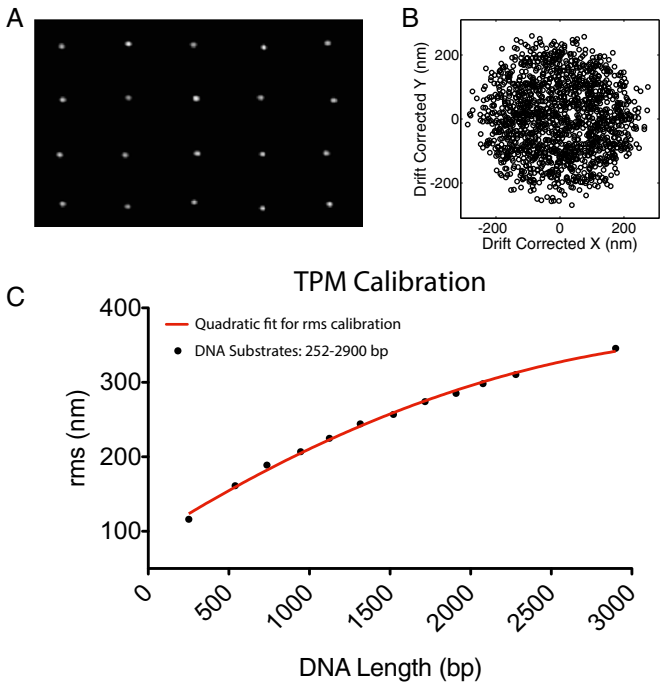

Fig. S1. TPM acquisition and analysis. (*A*) DNA-tethered beads assembled with a single DNA molecule (this image assembles an array of images of qualified beads). (*B*) Drift-corrected positions (black circles) generated by a DNA-tethered bead. (*C*) The rms for the Brownian distribution as a function of length in base pairs (black dots) is plotted, and a quadratic fit is performed of the form $y = Ax^2 + Bx + C$ to extract a calibration equation (red line), where $A = (-1.77 \pm 0.3) \times 10^{-5}$, $B = 0.14 \pm 0.01$, and $C = 89.9 \pm 4.6$. The errors on *A*, *B*, and *C* were derived from the fit to the rms as a function of DNA length.

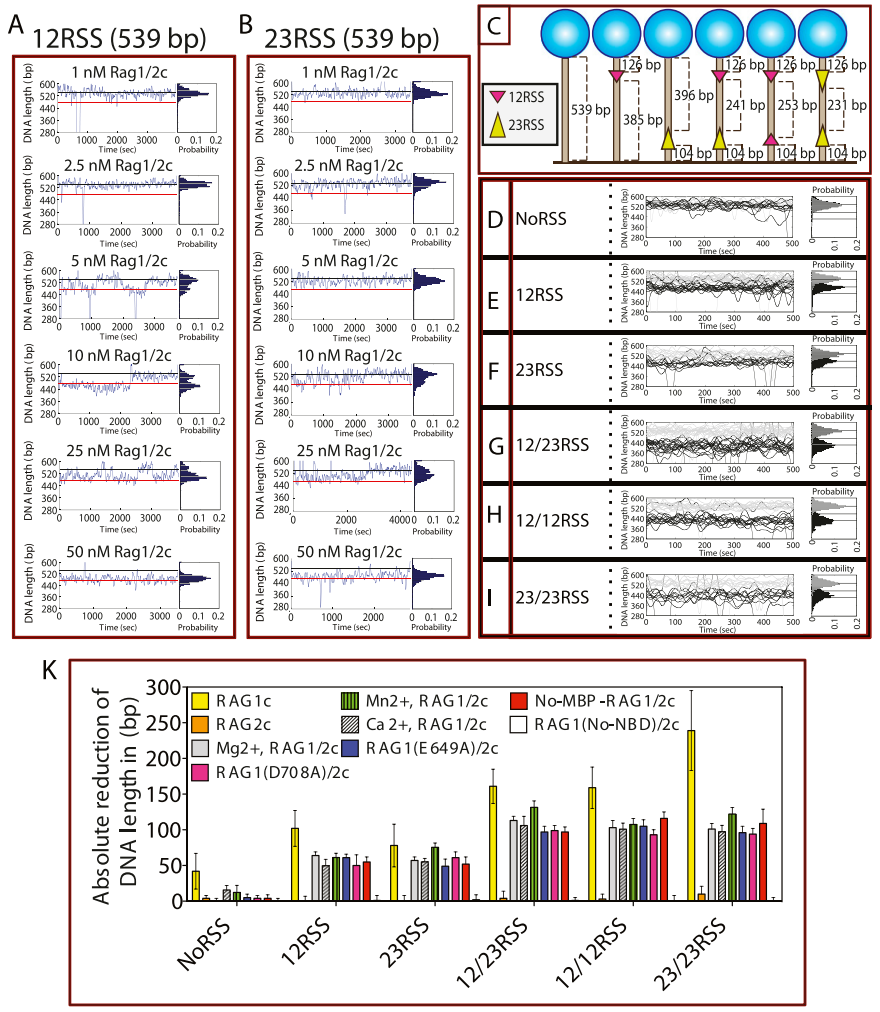

Fig. S2. Dynamic RAG trajectories, individual RAGs, and RAG mutants shortening one or two RSSs on a single DNA molecule. (*A* and *B*) RAG trajectories switching between the bound (red line) and the unbound (blue line) states. Histograms on the right are the resultant distribution of the single trajectories shown on the left. (*C*) A schematic of DNAs with NoRSS, 12RSS (magenta), 23RSS (yellow), 12RSS/23RSS combination, 12RSS/12RSS combination, and 23RSS/23RSS combination. (*D–I*) Trajectories with no protein (gray) and WT RAG1/2c at saturation (50 nM) in the presence of no RSS or 12RSS, 23RSS, 12RSS/23RSS, 12RSS/12RSS, and 23RSS/23RSS (black trajectories). (*J*) Individual core RAGs and RAG1/2c mutants in the presence of one or two RSSs; 50 nM RAG1c (yellow), RAG2c (orange), Mg^2 RAG1/2c (gray), Ca^{2+} RAG1/2c (green and black striped), Mn^{2+} RAG1/2c (white and black hatched), paired complex independent hairpin mutant RAG1(E649A)/2c (blue), RAG1(D708A)/2c (magenta), No-MBP–RAG1/2c (red), and RAG1[No-Nonamer binding domain (NBD)]/2c (white) were added.

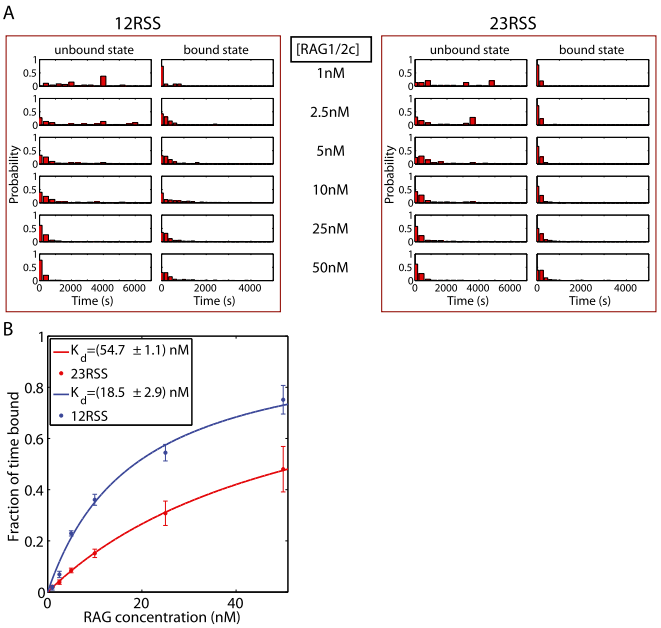

Fig. S3. Using fractional occupancy to determine K_d values. (*A*) Histograms of individual dwell times measured with (*Left*) 12RSS and (*Right*) 23RSS for RAG1/2c concentrations between 1 and 50 nM. (*B*) The fractional occupancy of RAG1/2c as a function of RAG concentration (blue, 12RSS; red, 23RSS) is calculated by summing all of the time spent in the on state over every bead and dividing by the total time of measurement. To extract the K_d for a given binding site, the fractional occupancy curves are fit vs. the concentration of RAG1/2c (c) to a function of the form $p_{Bound} = (c/K_d)/(1 + (c/K_d))$ (a Langmuir isotherm or Hill function with $n = 1$). These fits gave $K_d = 18.5 \pm 2.9$ nM for 12RSS and $K_d = 54.7 \pm 1.1$ nM for 23RSS.

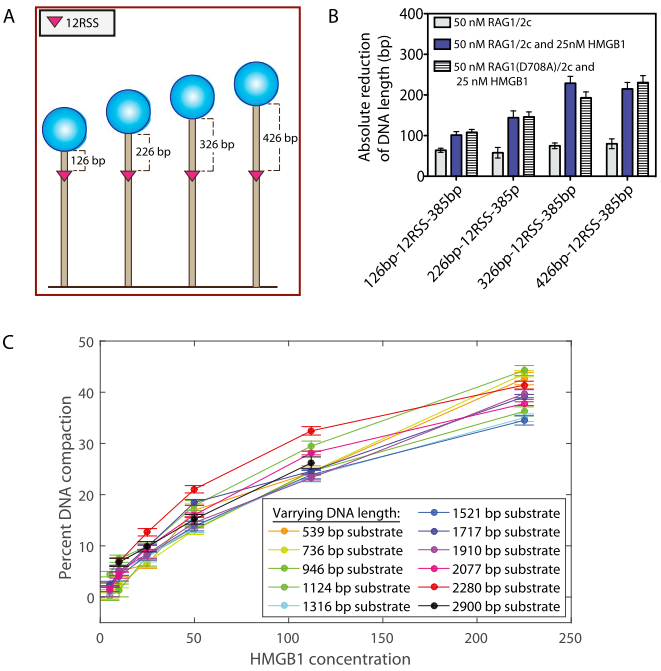

Fig. S4. Length dependence of RAG–HMGB1–RSS complex formation. (*A*) A schematic of DNA substrates to calibrate the optimal DNA length flanking RSSs to observe complete RAG and HMGB1 bends. The length of DNA between the 12RSS (magenta) and the slide was fixed at 385 bp, whereas the length of the DNA between the RSS and the bead was varied from 126 to 426 bp. (*B*) Effect of DNA length on RAG1/2c–RSS and RAG1/2c–HMGB1–RSS complex formation; 50 nM RAG1/2c (gray), 25 nM HMGB1 (blue), and 50 nM RAG1(D708A)/2c and 25 nM HMGB1 (white stripes) were added to the DNA substrates shown in *A* to measure the reduction in DNA length. (*C*) The presence of HMGB1 altered the measured effective length of DNA tethers in a concentration-dependent manner. This effect was observed in the absence of RAG1/2c and found to depend on the overall length of the DNA substrate.

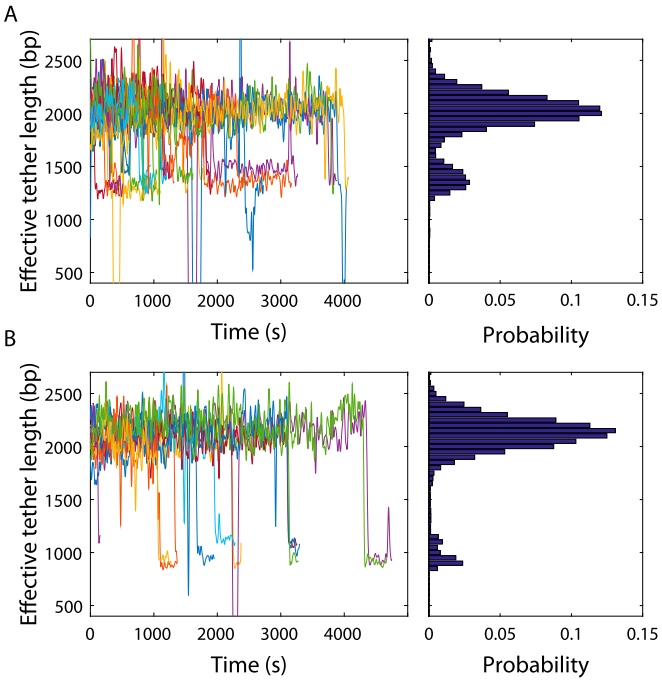

Fig. S5. Composite trajectories and normalized probability distributions of 12RSS/23RSS–1,200 bp and 12RSS/23RSS–1,800 bp DNA-tethered beads. Trajectories and distributions for the (*A*) 12RSS/23RSS–1,200 bp and (*B*) 12RSS/23RSS–1,800 bp beads showing the paired complex resulting in bead loss in the presence of 5 nM RAG1/2c and 80 nM HMGB1.

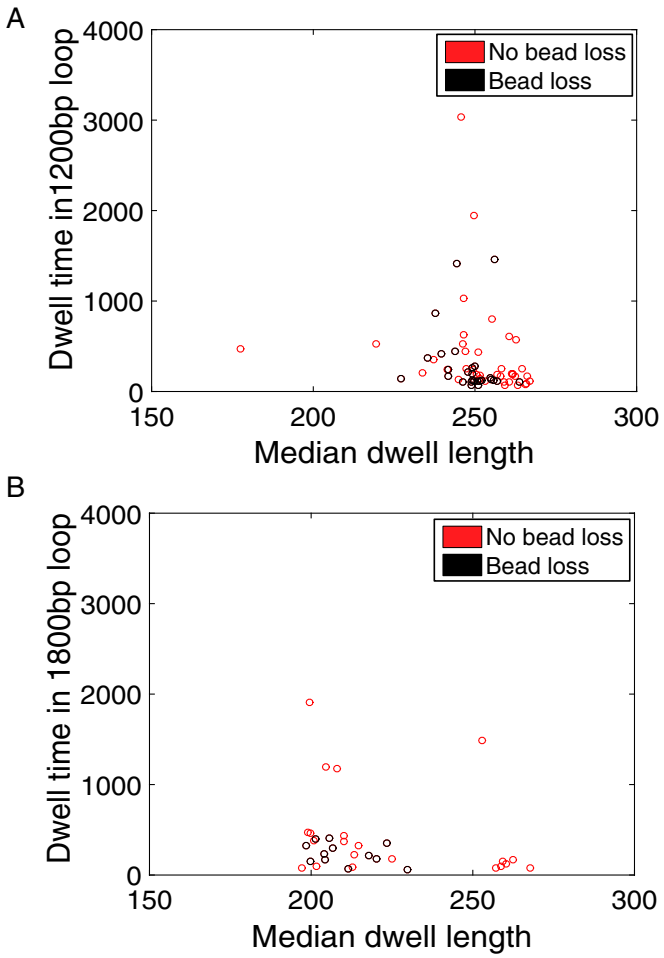

Fig. S6. Paired complex mean dwell time as a function of median dwell length. The mean dwell time and median dwell length for the (*A*) 1,200-bp intersignal distance and (*B*) 1,800-bp intersignal distance with (black circles) or without (red circles) bead loss.

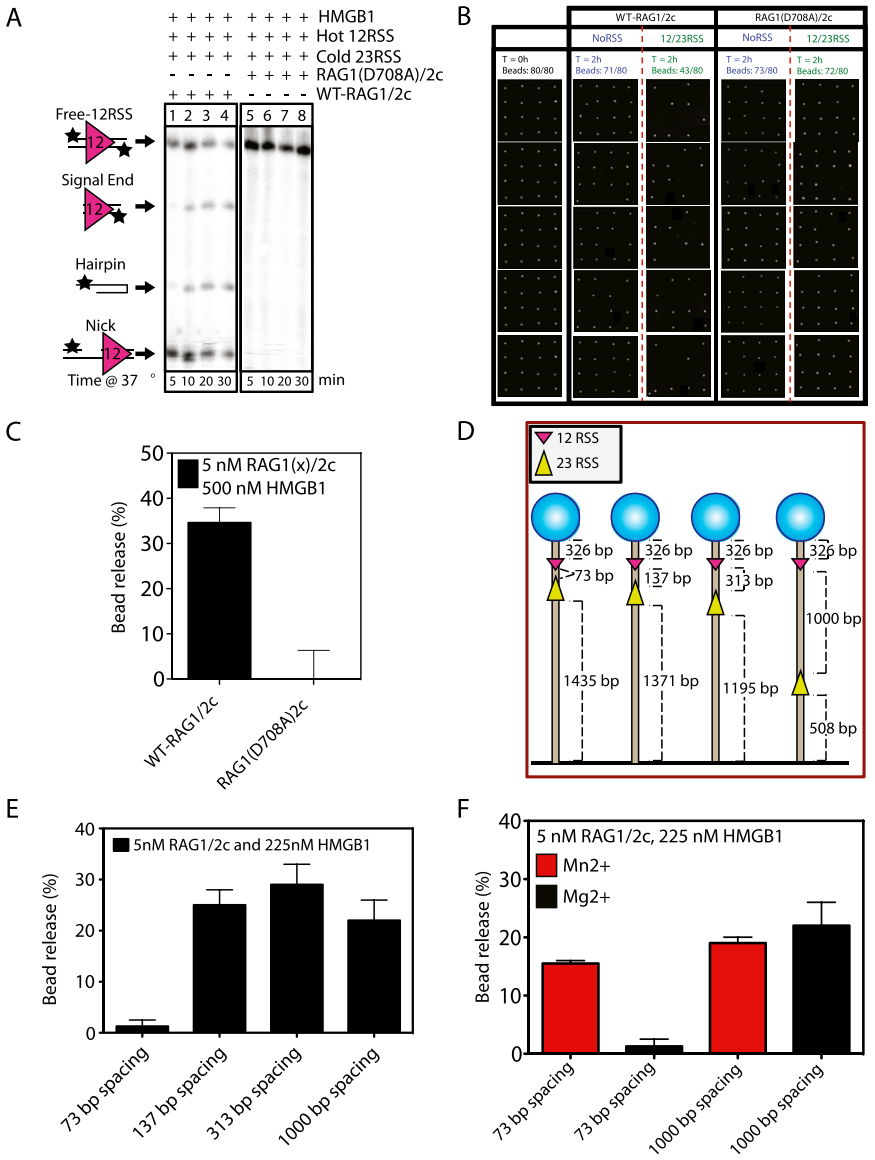

Fig. S7. RAG and HMGB1 ensemble and single-molecule in vitro cleavage assays. (*A*) Bulk hairpin production assay to test catalytically active and inactive RAG1/2c; 50 nM RAG1/2c (lanes 1–4) or RAG1(D708A)/2c (lanes 5–8) and 225 nM HMGB1 were added to 1 nM radiolabeled 12RSS oligo (50 bp), and then, 4 nM 23RSS oligo (65 bp) was added *in trans*. The reaction was incubated at 37 °C, and samples were removed at 5, 10, 20, and 30 min. The samples were fractionated on a sequencing gel to detect nicks and 12/23 rule-regulated hairpin production. (*B*) In the single-molecule assay for hairpin production, we added 5 nM RAG1/2c or RAG1(D708A)/2c and 500 nM HMGB1 to NoRSS (blue) or 12RSS/23RSS (red). We determined the percentage of bead release as % loss = $f_{RSS} - f_{NoRSS}$, where f_{RSS} is the fraction of beads lost with RSSs present and f_{NoRSS} is the fraction of beads lost with no RSSs present. (*C*) The percentage of bead release for the experiment shown in *B* for RAG1/2c or RAG1(D708A)/2c (black bars). (*D*) A schematic of DNA substrates with 12RSS and 23RSS spacings of 73, 137, 313, and 1,000 bp. (*E*) Length dependence of intersignal distance on 12/23 rule-regulated bead release; 5 nM RAG1/2c and 225 nM HMGB1 were added to the four DNA substrates shown in *D* and assayed for bead release. (*F*) Effect of divalent cation type and intersignal distance on bead release; 5 nM RAG1/2c and 225 nM HMGB1 were added in the presence of 12RSS/23RSS substrates with 73- or 1,000-bp intersignal spacing in the presence of Mn^{2+} (red) or Mg^{2+} (black).