Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

# Structural gymnastics of RAG-mediated DNA cleavage in V(D)J recombination

Heng Ru, Pengfei Zhang and Hao Wu

A hallmark of vertebrate immunity is the diverse repertoire of antigen-receptor genes that results from combinatorial splicing of gene coding segments by V(D)J recombination. The $(RAG1\text{-}RAG2)_2$ endonuclease complex (RAG) specifically recognizes and cleaves a pair of recombination signal sequences (RSSs), 12-RSS and 23-RSS, via the catalytic steps of nicking and hairpin formation. Both RSSs immediately flank the coding end segments and are composed of a conserved heptamer, a conserved nonamer, and a non-conserved spacer of either 12 base pairs (bp) or 23 bp in between. A single RAG complex only synapses a 12-RSS and a 23-RSS, which was denoted the 12/23 rule, a dogma that ensures recombination between V, D and J segments, but not within the same type of segments. This review recapitulates current structural studies to highlight the conformational transformations in both the RAG complex and the RSS during the consecutive steps of catalysis. The emerging structural mechanism emphasizes distortion of intact RSS and nicked RSS exerted by a piston-like motion in RAG1 and by dimer closure, respectively. Bipartite recognition of heptamer and nonamer, flexibly linked nonamer-binding domain dimer relatively to the heptamer recognition region dimer, and RSS plasticity and bending by HMGB1 together contribute to the molecular basis of the 12/23 rule in the RAG molecular machine.

**Address**
Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, and Program in Cellular and Molecular Medicine, Boston Children's Hospital, Boston, MA 02115, United States

Corresponding authors: Ru, Heng (heng.ru@childrens.harvard.edu), Wu, Hao (hao.wu@childrens.harvard.edu)

**Current Opinion in Structural Biology** 2018, **53**:178–186

This review comes from a themed issue on **Protein-nucleic acid interactions**

Edited by **Eric Westhof** and **Dinshaw Patel**

For a complete overview see the Issue and the Editorial

Available online 23rd November 2018

**https://doi.org/10.1016/j.sbi.2018.11.001**



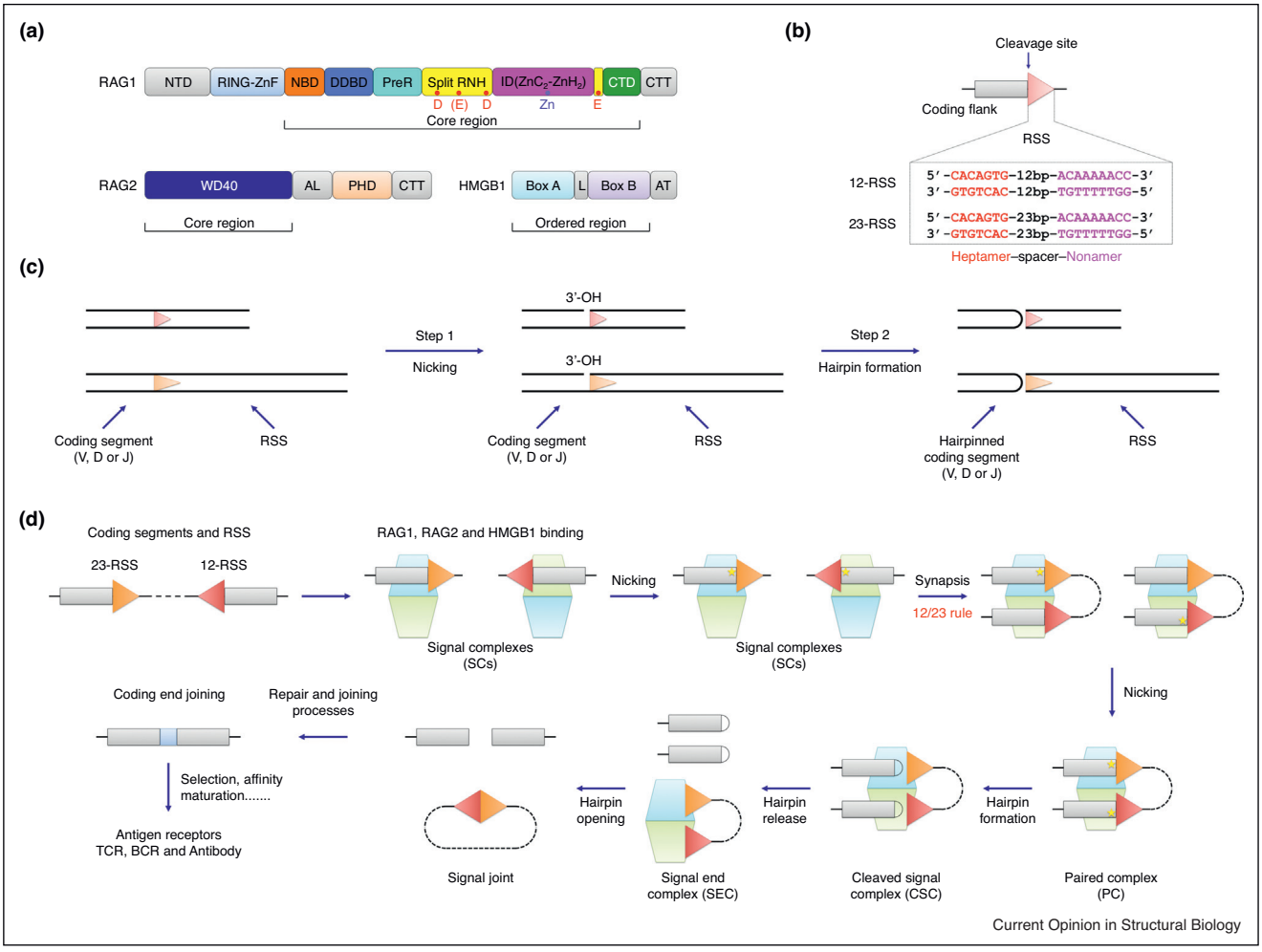
## Introduction

For optimal host defense, jawed vertebrates have evolved an elegant combinatorial mechanism to produce a large repertoire of immunoglobulin and T cell receptor genes. The coding regions of these genes can be divided into a constant region and a variable region. The variable region of a heavy chain is combinatorially assembled from variable (V), diversity (D) and joining (J) segments, and the variable region of a light chain from V and J segments [1–3]. V(D)J recombination occurs during lymphocyte development, and in addition to the diversity from the random selection of one copy of each kind of the segments, incorporation of palindromic and non-palindromic nucleotides in the downstream non-homologous end joining (NHEJ) pathway further increases the repertoire [4–6].

V, D and J gene coding segments are immediately flanked by recombination signal sequences (RSSs) required for the recombination, which is launched by recombination activating gene (RAG) proteins, RAG1 and RAG2 [7•] (Figure 1a). The $(RAG1\text{-}RAG2)_2$ complex acts as an endonuclease, which cleaves precisely at the junction between the coding flank and the RSS to induce DNA double-strand breaks [8•] (Figure 1b). An RSS has conserved heptamer (consensus of CACATGT) and nonamer (consensus of ACAAAAACC) sequences and a non-conserved 12 bp or 23 bp spacer DNA in between, leading to two kinds of RSSs, 12-RSS and 23-RSS [9,10] (Figure 1b). Each kind of gene segments is flanked by the same RSS, while different kinds of gene segments are distinguished by different RSSs. In this scenario, a recombination step strictly requires one 12-RSS and one 23-RSS, forming the 12/23 rule that assembles gene segments between different kinds but not within the same kinds to ensure recombination fidelity [11,12•]. Individuals with RAG1 or RAG2 mutations that cause deficiency in V(D)J recombination are limited in their T or B cell repertoire [13], and suffer from a spectrum of genetic disorders [14]. Aberrant V(D)J recombination is often linked to autoimmune states and cancers [15–22].

RAG1 belongs to the DDE enzyme family and the RAG complex catalyzes two consecutive reactions, strand cleavage (nicking) and strand transfer (hairpin formation), to generate a pair of hairpinned coding ends and blunt-ended signal ends (Figure 1c,d) [8•]. Briefly, RAG complex binds to a single intact RSS to form a signal complex (SC), either 12-SC or 23-SC, which can introduce a nick at the coding flank-signal end junction in the presence of $Mg^{2+}$, producing a 3′−OH on the coding flank segment and a 5′-phosphate on the signal end. Synapsis of one 12-RSS and one 23-RSS in the same RAG dimer forms the paired complex (PC), in which both 12-RSS and 23-RSS can be nicked. Then the 3′−OH of the nucleotide in the coding end attacks the scissile phosphate on the reverse strand to form a pair of hairpin DNA and a pair of signal

**Figure 1**




Overview of RAG-mediated V(D)J recombination pathway. **(a)** Domain organization of RAG1, RAG2 and HMGB1 proteins. NTD: N-terminal domain; RING-ZnF: RING domain and zinc finger domain; NBD: nonamer binding domain; DDBD: dimerization and DNA binding domain; PreR: pre-RNase H domain; RNH: RNase H-like domain; ID: insertion domain, which can be further divided into two parts, ZnC2 and ZnH2; CTD: C-terminal domain; CTT: C-terminal tail; WD40: tryptophan-aspartic acid repeat domain; AL: acidic linker; PHD: plant homeodomain; L: linker; AT: acidic tail. Potential catalytic residues are indicated as red dots in RNH and the zinc ion is indicated as a slate dot in ID. Core or ordered regions in RAG1, RAG2 and HMGB1 are indicated. **(b)** Schematic representation of RSSs. The consensus sequences of heptamer and nonamer are shown in red and magenta respectively. **(c)** Schematic representation of RAG mediated catalysis that involves two consecutive reactions, nicking and hairpin formation. **(d)** Overview of the V(D)J recombination process. The coding regions and RSSs are shown as rectangles and triangles respectively. The dimeric (RAG1-RAG2)2 complex is shown as stacked cyan and light green trapezoids. Briefly, RAG binds a single RSS in the presence of HMGB1 to form signal complexes (SCs), either 12-SC or 23-SC, which can undergo nicking (shown by a yellow pentagram) at the coding flank–RSS junction in the presence of Mg2+. Synapsis of one 12-RSS and one 23-RSS in the same RAG dimer forms the paired complex (PC), followed by generation of the cleaved signal complex (CSC) with hairpin coding end and cleaved signal end. Hairpin release produces the signal end complex (SEC). Further processing by enzymes in the non-homologous end-joining (NHEJ) DNA repair pathway results in ligation of the coding ends and circularization of the signal ends. This panel is adapted from Figure 1a of Ru *et al.* [39••].

end DNA in the cleaved signal complex (CSC). Next, NHEJ factors are recruited and hairpin DNAs are released from CSC, forming the signal end complex (SEC). *In vitro*, high-mobility group box proteins like HMGB1 have been shown to enhance the RAG catalytic activity [23–25], mainly through stabilization of the highly bent RSS conformations during catalysis (Figure 1a). In this review, we will summarize the current structural and mechanistic insights on the coupling among RSS binding, conformational changes and catalysis in the RAG complex.

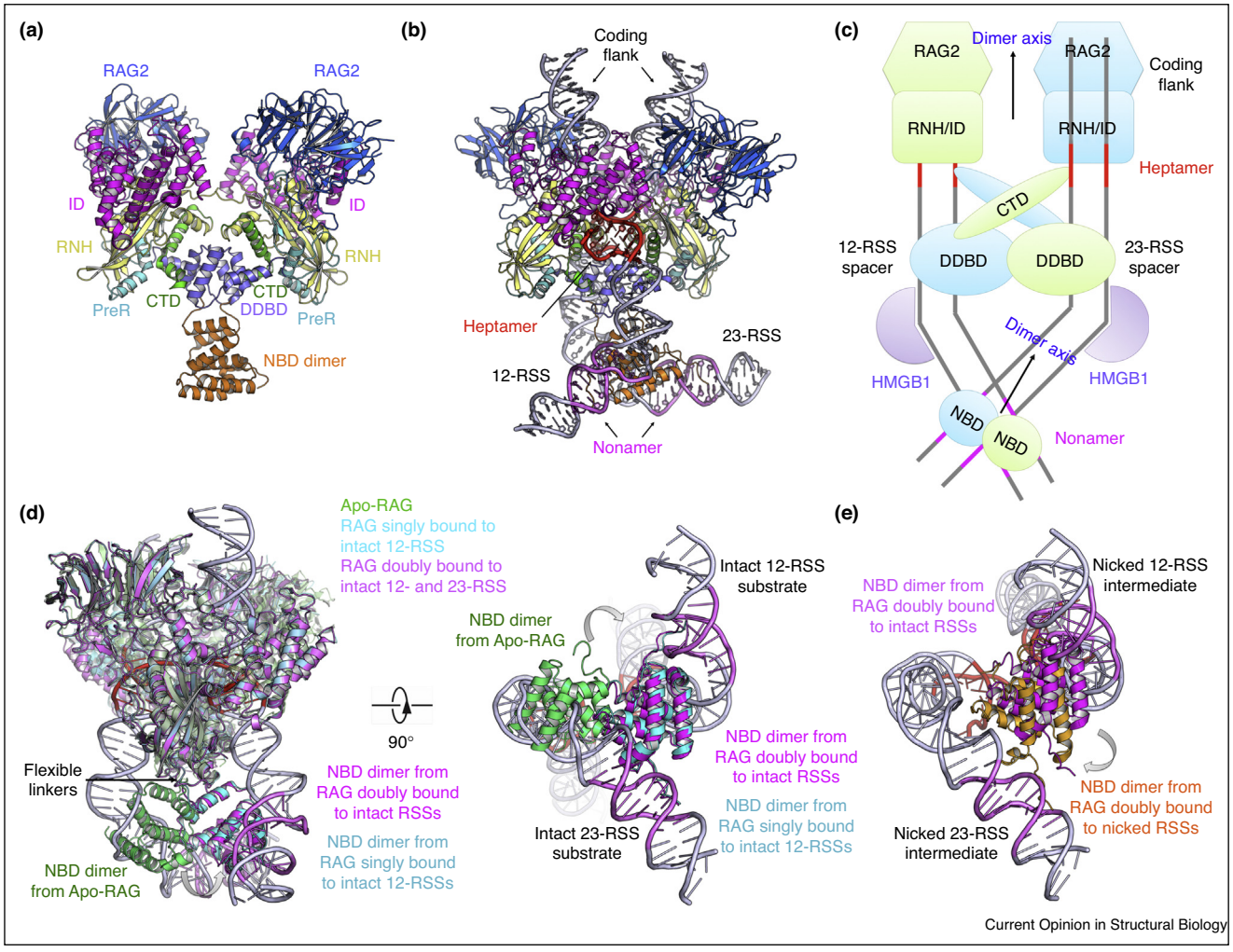
## Overall structural architecture of the $(RAG1\text{-}RAG2)_2$ complex

RAG1 and RAG2 are multi-domain proteins and both comprise a core region that is required for RAG's activity

of RSS binding and cleavage *in vitro*, and a non-core region that regulates RAG's function in cells (Figure 1a). For RAG1, the core region contains a nonamer-binding domain (NBD), a dimerization and DNA binding domain (DDBD), a pre-RNase H domain (Pre-RNH), a split RNase-H like domain (RNH) disrupted by an insertion domain (ID), and a C-terminal domain (CTD). The insertion domain was defined from the analogous domain in the *h*AT family transposase *Hermes* [26,27•], and the proposed catalytic residues are exclusively located in the RNH of RAG1. For RAG2, the core region only contains an N-terminal six-bladed WD40 domain. Crystal structures of the mouse RAG1 NBD in complex with nonamer DNA and of the mouse RAG1-RAG2 core complex (Apo-RAG) provided a first view of the Y-shaped dimeric core [28••,29•] (Figure 2a, in which RAG1 and RAG2 associate with each other and two RAG1-RAG2 protomers form a homodimer through contacts on RAG1.

The non-core region of RAG1 includes an N-terminal domain (NTD) adjacent to the RING-zinc finger (RING-ZnF) domain and a C-terminal tail (CTT). The RING-

**Figure 2**




Overview of the RAG complex structures and the structural basis of the 12/23 rule. **(a)** Overall structure of the Apo-RAG dimer (PDB: 4WWX). The domains are colored as in Figure 1(a). **(b)** Overall structure of the RAG dimer in complex with nicked 12- and 23-RSS intermediates (PDB: 6DBI). The domains are colored as in Figure 1(a). The 12-RSS and 23-RSS intermediates are shown in light slate with heptamer and nonamer are highlighted in red and magenta respectively. **(c)** Schematic representation of the interactions between RAG dimer and RSSs based on the synaptic RAG complex with nicked RSS intermediates. The dimer axes of the catalytic region and the NBD dimer do not coincide. **(d)** Orthogonal views of the superposition between Apo-RAG (green), RAG dimer singly bound to intact 12-RSS (cyan), and RAG dimer doubly bound to intact 12- and 23-RSS substrates (magenta) by aligning one RAG1-RAG2 monomer (PDB: 4WWX, 6DBX, and 6DBT). The NBDs are highlighted in darker green, cyan and magenta respectively. The intact 12-RSS from the singly bound RAG complex is omitted. Tilt of NBD dimer upon substrate RSS engagement is obvious. **(e)** Bottom view of the NBD dimer movement upon nicking of the intact RSS substrates. NBD dimers from RAG bound to intact RSSs (PDB: 6DBT) and nicked RSSs (PDB: 6DBI) are colored in magenta and orange respectively.

ZnF domain is dimeric in structure [30,31]; it acts as an E3 to promote ubiquitination on histone H3 [32,33] and on itself [34] to regulate V(D)J recombination. The non-core region of RAG2 comprises an acidic linker, a non-canonical plant homeodomain (PHD), and a C-terminal tail (CTT). The PHD binds hypermethylated lysine 4 of histone H3 (H3K4Me3), an active chromatin marker, and is necessary for efficient V(D)J recombination *in vivo* [35–38].

## Bipartite RSS interaction and 12/23 rule of V(D)J recombination

Cryo-EM structures of zebrafish RAG-DNA complexes at up to 3.4 Å resolution offered a first glimpse of the core RAG synapsed with nicked or cleaved 12-RSS and 23-RSS [39••] (Figure 2b), which was further complemented by recent crystal structures of mouse RAG in complex with similar RSSs at up to 2.75 Å resolution [40••]. These structures revealed the mode of specific recognition of the conserved heptamer and nonamer in 12-RSS and 23-RSS, as well as the largely non-specific interactions at the spacer regions adjacent to heptamer and nonamer, and at the coding flank (Figure 2c). An RSS is cooperatively recognized by both RAG1-RAG2 protomers in the dimer (first RAG and second RAG), with each protomer playing different roles for a given RSS. Briefly, the coding flank is bound by the first RAG1-RAG2 protomer exclusively. The beginning part of the heptamer interacts with the ID and RNH of the first RAG1 while the more distal part of the heptamer is recognized by the RNH, DDBD and CTD of the second RAG1, with complementary electrostatic and hydrogen bonding interactions at both the phosphate backbone and the bases. The spacer region adjacent to the heptamer contacts the DDBD of the second RAG1. The nonamer is mainly bound by the second RAG1 in the intertwined NBD dimer, while the spacer region adjacent to the nonamer contacts mainly the first RAG1 of the NBD dimer. Overall, residues that are involved in the RSS interaction are highly conserved in zebrafish and mouse, suggesting their important role in DNA binding and catalysis.

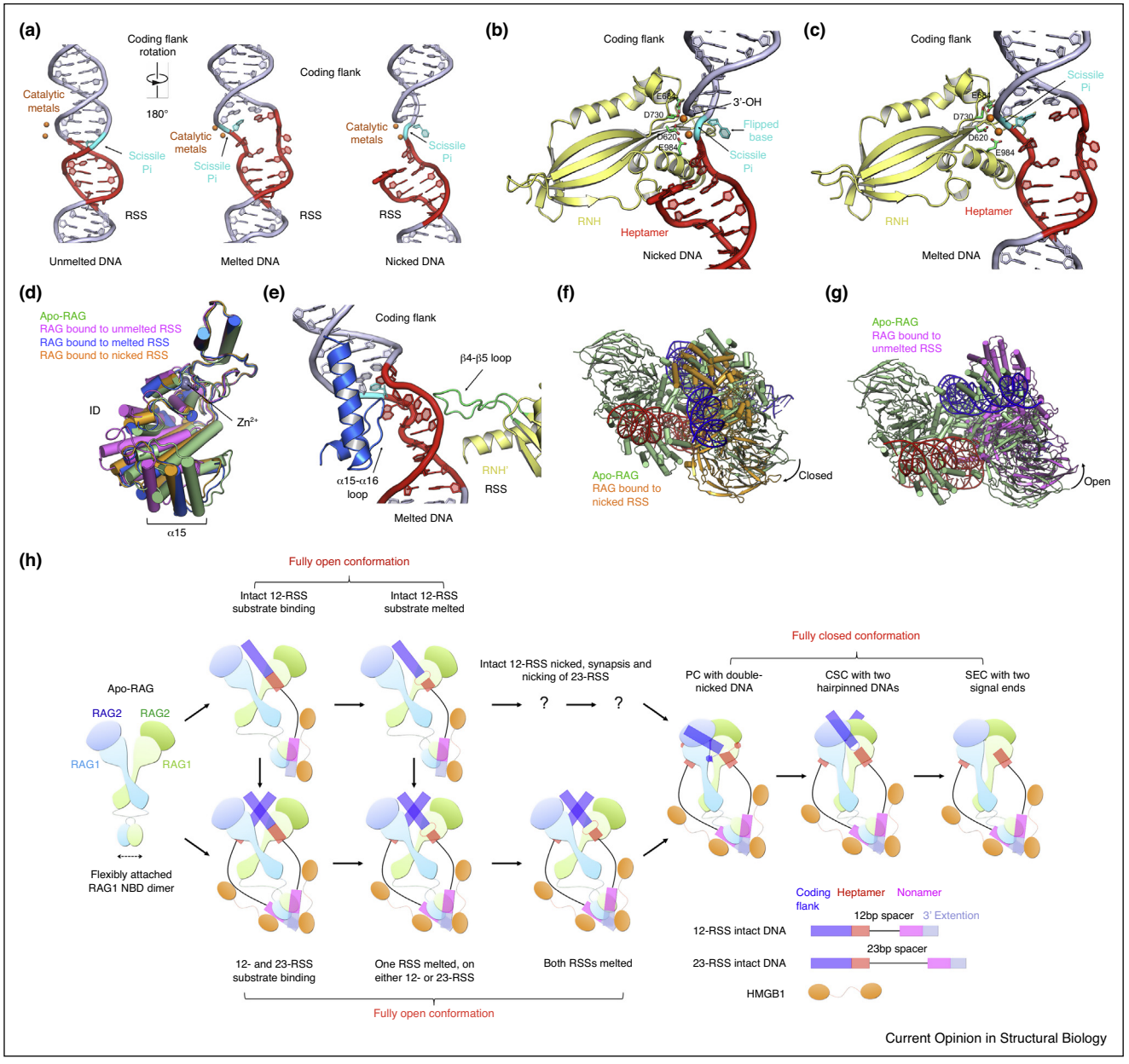
We and others have proposed previously that an induced asymmetry upon synapsis of 12-RSS and 23-RSS explains the molecular basis for the 12/23 rule [39••,40••] (Figure 2d). In this model, while both the NBD and the remaining RAG1-RAG2 complex stay as symmetric dimers, synaptic RSS binding tilts the flexibly tethered NBD dimer relative to the remaining RAG1-RAG2 complex. However, our recent structures indicate that the 12/23 rule is already set upon single 12-RSS binding, because the tilt of the NBD dimer in singly bound complexes matches that in the synaptic complexes [41••] and precisely presents for binding of the longer 23-RSS (Figure 2d). Therefore, the homodimer of the RAG1-RAG2 heterodimer can only synapse a pair of different RSSs under physiological conditions, and despite the different lengths, 12-RSS and 23-RSS essentially contact the same residues in RAG. The induced asymmetry is facilitated by the plasticity of the RSSs, which are highly bent by the bound HMGB1. In the cryo-EM structures, low resolution densities for HMGB1 exist at both the spacer regions of 12-RSS and 23-RSS [39••], while in the more recent crystal structures, stronger HMGB1 density was observed at 23-RSS than at 12-RSS [40••]. Consistently, HMGB1-binding sites have previously been mapped in solution on both 12-RSS and 23-RSS [42]. The flexibly linked NBD dimer modifies its tilt as the intact RSSs become nicked (Figure 2e). These data support that in order to synapse different coding segments during V(D)J recombination, RAG1 was evolved to have two flexibly linked DNA-binding regions for bipartite DNA recognition, and to co-opt DNA-bending proteins to accomplish the 12/23 rule.

## DNA distortion in consecutive catalytic steps of nicking and hairpin formation

RAG executes its cleavage reaction through two consecutive catalytic steps without RSS dissociation. These two steps are nicking at the coding flank-RSS junction, and attack of the opposing strand by the newly formed 3′−OH to generate a coding flank hairpin and a blunt-ended RSS. The molecular mechanism for the second catalytic step was revealed first by the cryo-EM structures of zebrafish RAG1-RAG2 in complex with nicked DNA, which indicated distortion at the coding flank-RSS junction as the basis for positioning the scissile phosphate bond and the attacking 3′−OH into the RNH active site [39••] (Figure 3a,b). The distortion includes base flipping of the first nucleotide of the RSS after nicking and of the last nucleotide of the coding flank that will be cleaved from the RSS, and twisting of the DNA duplex at the coding flank-RSS junction [39••,40••]. The flipped RSS base is extensively recognized by residues in the ID of RAG1, an observation confirmed by studies in both zebrafish and mouse RAG1-RAG2, while the flipped coding flank base lacks-specific interactions, making it possible for RAG1-RAG2 to accommodate any nucleotide at this position [39••,40••]. This hairpin forming reaction is catalyzed by four acidic residues in the RNH active site – the D(E)DE motif (D620, E684, D730 and E984 in zebrafish RAG1) (Figure 3b). The motif coordinates two metal ions for the phosphotransfer reaction, which is a mechanism extensively exploited by the DDE-family enzymes [39••,43].

However, how RAG1 performs the first nicking step remained elusive until recently because modeling of intact DNA into cryo-EM structures of zebrafish RAG1-RAG2 in complex with nicked DNA showed steric clash [39••] and the scissile phosphate for nicking in crystal structures of mouse RAG1-RAG2 in complex with intact RSSs locates ~20 Å away from the RNH active site [40••]. Cryo-EM structures in complex with intact RSSs

**Figure 3**




DNA distortions and RAG movements in the consecutive catalytic steps. **(a)** Cartoon representation of the DNAs that are bound to the RAG complex at different states. Left: unmelted RSS (PDB: 6DBU); middle: melted RSS (PDB: 6DBR); right: nicked RSS (PDB: 6DBJ). The heptamer of the RSS is shown in red and scissile phosphate in the nucleotide is highlighted in cyan. The catalytic metal ions shown as orange spheres indicate the location of the active site. **(b)** DNA distortion and base flipping in the nicked RSS intermediate facilitate hairpin formation. The heptamer in the RSS is shown in red and the flipped base and scissile phosphate are highlighted in cyan. The location of the nucleophile 3′−OH is indicated. The RNH domain is shown in yellow ribbon, and the catalytic residues and coordinated metal ions are shown in green sticks and orange spheres respectively. **(c)** DNA melting in the intact RSS substrate facilitates the nicking step. The color scheme is the same as in (b). (**d**) Different location of the α15 helices in the IDs of RAG1 from Apo-RAG (light green, PDB: 4WWX), RAG bound to unmelted RSS (light magenta, PDB: 6DBT), RAG bound to melted RSS (blue, PDB: 6DBV) and RAG bound to nicked RSS (orange, PDB: 6DBI) when RAG1-RAG2 monomers are aligned. Only the IDs are shown and the location of Zn2+ is indicated. **(e)** α15–α16 loop from the first RAG monomer and the ordered β4–β5 loop from the second RAG monomer stabilizes the melted RSS (PDB: 6DBR). The α15–α16 region is shown as blue ribbon and the β4–β5 loop from the second RAG (RNH' in yellow) is highlighted in green. The heptamer in the RSS is shown in red and the nucleotide to be nicked is highlighted in cyan. **(f)** Closure of the RAG dimer upon binding of nicked RSS intermediates. The Apo-RAG (PDB: 4WWX) and synaptic RAG (PDB: 6DBI) bound to nicked 12- and 23-RSS (red and blue) are shown as green and orange ribbons. **(g)** Opening of the RAG dimer upon intact RSS substrates binding. The Apo- RAG (PDB: 4WWX) and synaptic RAG (PDB: 6DBI) that bound to intact 12- and 23-RSS (red and blue) are shown as green and magenta ribbons. **(h)** Structure-derived insights on RAGmediated cleavage pathway in V(D)J recombination. RAG1 (light green and light cyan), RAG2 (green and cyan), and HMGB1 (orange) are represented as cartoons. Coding segments, heptamers, nonamers, and 3′ extension DNAs are shown by

solved this conundrum because of the two observed conformations, an unmelted RSS state similar to the crystal structures, and a melted RSS state in which unwinding at the second and third positions of the heptamer leads to the positioning of the scissile phosphate into the active site for nicking [41••] (Figure 3a and c). RSS melting leads to a dramatic, ~180° corkscrew-like rotation of the coding flank (Figure 3a), and the rotation may be tolerated by the non-specific nature of the coding flank interaction with RAG1-RAG2 [41••]. RSS melting is promoted by higher temperature, and either 12-RSS or 23-RSS may melt first to be nicked [41••]. Unlike hairpin formation, the nicking step does not require RSS synapsis, and combinations of 12-RSS unmelted/23-RSS unmelted, 12-RSS melted/23-RSS unmelted, 12-RSS unmelted/23-RSS melted and 12-RSS melted/23-RSS melted all exist within the same cryo-EM sample [41••]. In Apo-RAG, the RNH active site residues are not placed correctly for catalysis [28••]; upon DNA binding, the active site is fully formed, but undergoes limited conformational changes between the two catalytic steps [39••,41••]. Conceptually, it is the two different RSS deformations that enable the use of the same active site for both the nicking and hairpin formation steps, rather than alterations in the active site to accommodate the different substrates.

## Major conformational gymnastics of RAG1-RAG2 that distorts the DNA

Structures of RAG1-RAG2 bound to RSSs in different states have allowed comparison of their detailed conformations by superposition (Figure 3d). One surprising major difference resides at the ID of RAG1: in Apo-RAG, it is most inward and closest to DNA if an RSS is bound [28••]; in complexes with intact RSSs in the unmelted state, ID is most outward to accommodate the DNA [40••,41••]; in complexes with intact RSSs in the melted state, ID wedges melted DNA through its α15–α16 loop region [41••]. We propose that because the ID of Apo-RAG is inwardly located, the outward movement upon intact RSS binding may have created strain in the protein. Thus, like a loaded spring, the ID nudges in, likely facilitating and stabilizing DNA melting. The ID stays inward during the rest of the catalytic cycle. These conformational changes of the ID create a piston-like movement in the RAG1-RAG2 molecular machine. The inward movement of the ID upon melting of an intact RSS is also accompanied by conformational changes in the β4–β5 loop (Figure 3e) from the RNH domain of the second RAG1 to precisely position the catalytic residues into the active site [41••].

In comparison with Apo-RAG, we initially observed dramatic RAG1-RAG2 dimer closure upon binding to nicked, highly distorted RSS intermediates [39••] (Figure 3f). In this closed conformation, a new interaction interface is created between RAG2 of one protomer and the ID of RAG1 from the symmetric protomer, and each RSS is recognized intimately by both subunits of RAG1. Surprisingly however, the recent cryo-EM and crystal structures of RAG1-RAG2 in complex with intact RSS substrates revealed dimer conformations that are much more open, instead of more closed, than Apo-RAG [40••,41••] (Figure 3g). This dimer opening upon intact RSS substrate binding cooperates with the outward piston-like movement of the ID of RAG1, and may also be necessary to avoid clash of the RAG1-RAG2 complex with the substrates [41••]. Unlike large ID movement upon RSS melting, only minor changes in dimer closure are associated with RSS melting; instead, dimer closure is most important for distortion of nicked DNA to position it into the active site for hairpin formation.

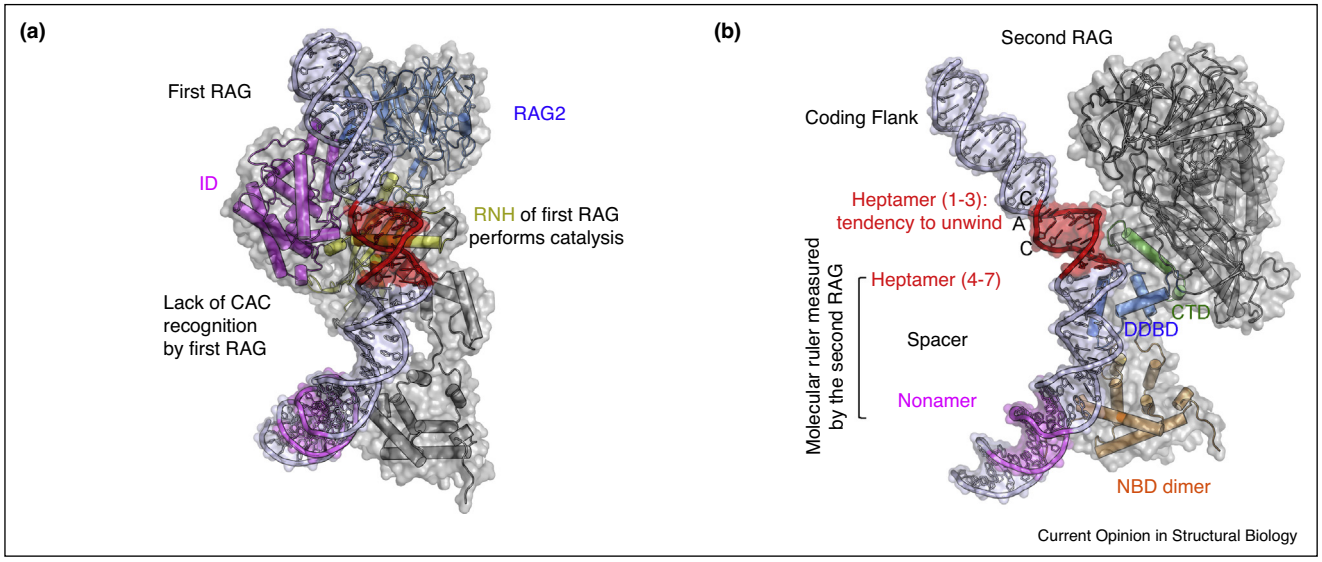
Therefore, RAG undergoes dramatic conformational changes during its catalytic cycle, both within the subunit and between the protomers in the dimer, to deform the DNA. While the ID movement is the main driving force for RSS melting and substrate nicking, dimer closure offers the major power stroke for base flipping and other distortions important for placing the nicked DNA into the active site for hairpin formation [41••] (Figure 3h).

## Unwinding tendency of the heptamer and the molecular ruler that assists its positioning

Unexpectedly, there are limited base-specific contacts between RAG1 and the terminal conserved CAC/GTG sequence of the heptamer in the intact (both unmelted and melted) RSS state [41••] (Figure 4a). In contrast, the heptamer in the nicked or cleaved states is extensively recognized. Of note, previous biochemical experiments have suggested lack of strong contacts [42], tendency of unwinding [44,45], RSS distortion [42,46] and robust

**(Figure 3 Legend Continued)** semi-transparent blue, red, magenta, and light slate rectangles, respectively. 12 and 23 bp spacers are shown as black lines. Briefly, without binding to an RSS, apo-RAG is in an open conformation with the flexibly attached NBD dimer. When bound to a singly intact RSS (e.g. 12-RSS) or simultaneously bound to 12- and 23-RSS intact DNA in the presence of HMGB1, an even open conformation of RAG dimer is induced. The NBD dimer is tilted through the interaction with nonamer(s) from the RSS(s). Then, the heptamer in the bound RSS(s) will undergo the melting step, which induces the rotation of the coding flank and the localization of the scissile phosphate into the active site to facilitate nicking. When both 12- and 23-RSS are synapsed and nicked in the same RAG dimer, the PC is formed, which assumes a fully closed conformation with flipped bases in the coding flank (blue filled circle) and the RSS (red filled circle) on both nicked RSS intermediates to facilitate the hairpin formation. The coding ends are then linked into hairpin DNA to form the CSC. NHEJ factors are recruited to dissociate the hairpin coding ends, leaving the SEC with only the bound signal ends. However, the steps of how RAG captures the other RSS after nicking on the first RSS is not known.

**Figure 4**




Molecular ruler supervises the accurate positioning and cleavage on the RSS. **(a)** Cartoon representation of the first RAG monomer that performs catalysis. The first RAG monomer is shown in gray surface. In the superimposed ribbon, ID and RNH are in magenta and yellow with the remaining RAG1 in gray, and RAG2 is in blue. The heptamer and nonamer of bound RSS is highlighted in red and magenta. Of note, the conserved CAC sequence in the heptamer lacks recognition by RAG. **(b)** Cartoon representation of the second RAG monomer that recognizes the last three positions of the heptamer, the spacer and the nonamer of the RSS. The second RAG monomer is shown in gray surface with superimposed ribbon. The heptamer and nonamer of bound RSS is highlighted in red and magenta. The NBD dimer (orange), DDBD (blue) and CTD (green) from the second RAG recognize the nonamer, spacer and last three positions of the heptamer and thus serve as a molecular ruler that positions the CAC of the heptamer into the active site of the first RAG.

affinity for single-stranded DNA [47] as characteristics of intact DNA binding at the heptamer. We thus proposed that the CAC/GTG sequence is conserved for its structural property of weak base pairing [41••] as an alternating purine-pyrimidine tract [48]. In this context, the unwound AC/TG base pairs exactly set the stage for the previously recognized 3′-flap endonuclease activity of RAG [49] to nick at the nucleotide 5′ to the first C/G base pair.

How is then the CAC/GTG sequence placed into the RAG1 active site? We posit that interactions at the nonamer and the second part of the heptamer help to position the coding flank-heptamer junction by measuring from the nonamer as a molecular ruler [41••] (Figure 4b). These interactions are relatively unchanged during the RAG catalytic cycle, clamping down the part of RSS away from the coding flank-heptamer junction but allowing the first three positions of the heptamer and the coding flank to go through the conformational gymnastics in the different catalytic steps.

## Implications on other DDE family transposases and retroviral integrases

Finally, similar to RAG, nicking of the substrate DNA is a universal first step of catalysis for DDE family transposases and retroviral integrases [50], which include, among others, the bacteriophage transposase MuA, bacteria transposases Tn5 and Tn10, Tc1/mariner family transposase Mos1, and *hAT* family transposase Hermes [27•,51–53]. However, equivalent complexes with intact substrate DNA are not yet available for these DDE enzymes [54]. For Hermes and Tn5, modeling of intact DNA into known structural complexes with cleaved DNA showed large distances of the nicking site to the active site, suggesting that DNA unwinding may also be required at this catalytic step [41••]. Remarkably, in the human genome, CA and TG are the most frequently observed terminal sequences in inverted repeats of DNA transposons and long terminal repeats of retrotransposons [55]. Collectively, these data indicate that DNA melting may represent a universal mechanism for initiation of retroviral integration and DNA transposition in the DDE enzyme family [41••].

## Conflict of interest

There are no conflicts of interest to declare.

## Acknowledgements

This work was supported by research funding from the National Institutes of Health (AI125535 to H.W.), and the Irvington Postdoctoral Fellowship from the Cancer Research Institute (to H.R.).

## References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as

- • of special interest
- •• of outstanding interest


1. Fanning L, Connor A, Baetz K, Ramsden D, Wu GE: **Mouse RSS spacer sequences affect the rate of V(D)J recombination**. *Immunogenetics* 1996, **44**:146-150.

2. Tonegawa S: **Somatic generation of antibody diversity**. *Nature* 1983, **302**:575-581.

3. Early P, Huang H, Davis M, Calame K, Hood L: **An immunoglobulin heavy chain variable region gene is generated from three segments of DNA: VH, D and JH**. *Cell* 1980, **19**:981-992.

4. Taccioli GE, Rathbun G, Oltz E, Stamato T, Jeggo PA, Alt FW: **Impairment of V(D)J recombination in double-strand break repair mutants**. *Science* 1993, **260**:207-210.

5. Sakano H, Maki R, Kurosawa Y, Roeder W, Tonegawa S: **Two types of somatic recombination are necessary for the generation of complete immunoglobulin heavy-chain genes**. *Nature* 1980, **286**:676-683.

6. Sakano H, Kurosawa Y, Weigert M, Tonegawa S: **Identification and nucleotide sequence of a diversity DNA segment (D) of immunoglobulin heavy-chain genes**. *Nature* 1981, **290**:562-565.

7. • Schatz DG, Oettinger MA, Baltimore D: **The V(D)J recombination activating gene, RAG-1**. *Cell* 1989, **59**:1035-1048.

This paper identifies RAG1 as a potential V(D)J recombinase that is responsible for activating the V(D)J recombination process*in vivo*.

8. • McBlane JF, van Gent DC, Ramsden DA, Romeo C, Cuomo CA, Gellert M, Oettinger MA: **Cleavage at a V(D)J recombination signal requires only RAG1 and RAG2 proteins and occurs in two steps**. *Cell* 1995, **83**:387-395.

This paper shows that purified RAG1 and RAG2 proteins are sufficient to carry out double strand breaks*in vitro*, through nicking and hairpin formation steps.

9. Max EE, Seidman JG, Leder P: **Sequences of five potential recombination sites encoded close to an immunoglobulin kappa constant region gene**. *Proc Natl Acad Sci U S A* 1979, **76**:3450-3454.

10. Sakano H, Huppi K, Heinrich G, Tonegawa S: **Sequences at the somatic recombination sites of immunoglobulin light-chain genes**. *Nature* 1979, **280**:288-294.

11. Eastman QM, Leu TM, Schatz DG: **Initiation of V(D)J recombination in vitro obeying the 12/23 rule**. *Nature* 1996, **380**:85-88.

12. • van Gent DC, Ramsden DA, Gellert M: **The RAG1 and RAG2 proteins establish the 12/23 rule in V(D)J recombination**. *Cell* 1996, **85**:107-113.

These two papers above show that DNA double strand breaks require synapsis of two different RSSs that contain either 12 or 23 bp spacer DNA in bipartite signal sequences, which forms the 12/23 rule.

13. Lee YN, Frugoni F, Dobbs K, Tirosh I, Du L, Ververs FA, Ru H, Ott de Bruin L, Adeli M, Bleesing JH *et al.*: **Characterization of T and B cell repertoire diversity in patients with RAG deficiency**. *Sci Immunol* 2016, **1**.

14. Lee YN, Frugoni F, Dobbs K, Walter JE, Giliani S, Gennery AR, Al-Herz W, Haddad E, LeDeist F, Bleesing JH *et al.*: **A systematic analysis of recombination activity and genotype-phenotype correlation in human recombination-activating gene 1 deficiency**. *J Allergy Clin Immunol* 2014, **133**:1099-1108.

15. Brandt VL, Roth DB: **Recent insights into the formation of RAG-induced chromosomal translocations**. *Adv Exp Med Biol* 2009, **650**:32-45.

16. Hsu LY, Lauring J, Liang HE, Greenbaum S, Cado D, Zhuang Y, Schlissel MS: **A conserved transcriptional enhancer regulates RAG gene expression in developing B cells**. *Immunity* 2003, **19**:105-117.

17. Andrade D, Redecha PB, Vukelic M, Qing X, Perino G, Salmon JE, Koo GC: **Engraftment of peripheral blood mononuclear cells from systemic lupus erythematosus and antiphospholipid syndrome patient donors into BALB-RAG-2-/- IL-2Rgamma-/- mice: a promising model for studying human disease**. *Arthritis Rheum* 2011, **63**:2764-2773.

18. Girschick HJ, Grammer AC, Nanki T, Vazquez E, Lipsky PE: **Expression of recombination activating genes 1 and 2 in peripheral B cells of patients with systemic lupus erythematosus**. *Arthritis Rheum* 2002, **46**:1255-1263.

19. Morbach H, Singh SK, Faber C, Lipsky PE, Girschick HJ: **Analysis of RAG expression by peripheral blood CD5+ and CD5- B cells of patients with childhood systemic lupus erythematosus**. *Ann Rheum Dis* 2006, **65**:482-487.

20. Hillion S, Garaud S, Devauchelle V, Bordron A, Berthou C, Youinou P, Jamin C: **Interleukin-6 is responsible for aberrant B-cell receptor-mediated regulation of RAG expression in systemic lupus erythematosus**. *Immunology* 2007, **122**:371-380.

21. Niu N, Zhang J, Huang T, Sun Y, Chen Z, Yi W, Korteweg C, Wang J, Gu J: **IgG expression in human colorectal cancer and its relationship to cancer cell behaviors**. *PLoS One* 2012, **7** e47362.

22. Da Sylva TR, Fong IC, Cunningham LA, Wu GE: **RAG1/2 re-expression causes receptor revision in a model B cell line**. *Mol Immunol* 2007, **44** 889-899.

23. Bergeron S, Madathiparambil T, Swanson PC: **Both high mobility group (HMG)-boxes and the acidic tail of HMGB1 regulate recombination-activating gene (RAG)-mediated recombination signal synapsis and cleavage in vitro**. *J Biol Chem* 2005, **280**:31314-31324.

24. Zhang M, Swanson PC: **HMGB1/2 can target DNA for illegitimate cleavage by the RAG1/2 complex**. *BMC Mol Biol* 2009, **10**:24.

25. Ciubotaru M, Trexler AJ, Spiridon LN, Surleac MD, Rhoades E, Petrescu AJ, Schatz DG: **RAG and HMGB1 create a large bend in the 23RSS in the V(D)J recombination synaptic complexes**. *Nucleic Acids Res* 2013, **41**:2437-2454.

26. Hickman AB, Perez ZN, Zhou L, Musingarimi P, Ghirlando R, Hinshaw JE, Craig NL, Dyda F: **Molecular architecture of a eukaryotic DNA transposase**. *Nat Struct Mol Biol* 2005, **12**:715-721.

27. • Hickman AB, Ewis HE, Li X, Knapp JA, Laver T, Doss AL, Tolun G, Steven AC, Grishaev A, Bax A *et al.*: **Structural basis of hAT transposon end recognition by Hermes, an octameric DNA transposase from Musca domestica**. *Cell* 2014, **158**:353-367.

This paper describes the crystal structure of DNA transposase Hermes in complex with cleaved DNA, which serves as a parallel to RAG in terms of the similar catalytic steps.

28. •• Kim MS, Lapkouski M, Yang W, Gellert M: **Crystal structure of the V(D)J recombinase RAG1-RAG2**. *Nature* 2015, **518**:507-511.

This paper reports the crystal structure of mouse RAG1-RAG2 core region complex and describes its Y-shaped architecture.

29. • Yin FF, Bailey S, Innis CA, Ciubotaru M, Kamtekar S, Steitz TA, Schatz DG: **Structure of the RAG1 nonamer binding domain with DNA reveals a dimer that mediates DNA synapsis**. *Nat Struct Mol Biol* 2009, **16**:499-508.

This paper reports the crystal structure of RAG1 NBD dimer in complex with nonamer DNA and provides the structure basis for nonamer recognition.

30. Bellon SF, Rodgers KK, Schatz DG, Coleman JE, Steitz TA: **Crystal structure of the RAG1 dimerization domain reveals multiple zinc-binding motifs including a novel zinc binuclear cluster**. *Nat Struct Biol* 1997, **4**:586-591.

31. Rodgers KK, Bu Z, Fleming KG, Schatz DG, Engelman DM, Coleman JE: **A zinc-binding domain involved in the dimerization of RAG1**. *J Mol Biol* 1996, **260**:70-84.

32. Deng Z, Liu H, Liu X: **RAG1-mediated ubiquitylation of histone H3 is required for chromosomal V(D)J recombination**. *Cell Res* 2015, **25**:181-192.

33. Grazini U, Zanardi F, Citterio E, Casola S, Goding CR, McBlane F: **The RING domain of RAG1 ubiquitylates histone H3: a novel activity in chromatin-mediated regulation of V(D)J joining**. *Mol Cell* 2010, **37**:282-293.

34. Singh SK, Gellert M: **Role of RAG1 autoubiquitination in V(D)J recombination**. *Proc Natl Acad Sci U S A* 2015, **112**:8579-8583.

35. West KL, Singha NC, De Ioannes P, Lacomis L, Erdjument-Bromage H, Tempst P, Cortes P: **A direct interaction between the RAG2 C terminus and the core histones is required for efficient V(D)J recombination**. *Immunity* 2005, **23**:203-212.

36. Liu Y, Subrahmanyam R, Chakraborty T, Sen R, Desiderio S: **A plant homeodomain in RAG-2 that binds Hypermethylated lysine 4 of histone H3 is necessary for efficient antigen-receptor-gene rearrangement**. *Immunity* 2007, **27**:561-571.

37. Matthews AG, Kuo AJ, Ramon-Maiques S, Han S, Champagne KS, Ivanov D, Gallardo M, Carney D, Cheung P, Ciccone DN *et al.*: **RAG2 PHD finger couples histone H3 lysine 4 trimethylation with V(D)J recombination**. *Nature* 2007, **450**:1106-1110.

38. Ramon-Maiques S, Kuo AJ, Carney D, Matthews AG, Oettinger MA, Gozani O, Yang W: **The plant homeodomain finger of RAG2 recognizes histone H3 methylated at both lysine-4 and arginine-2**. *Proc Natl Acad Sci U S A* 2007, **104**:18993-18998.

39. •• Ru H, Chambers MG, Fu TM, Tong AB, Liao M, Wu H: **Molecular mechanism of V(D)J recombination from synaptic RAG1-RAG2 complex structures**. *Cell* 2015, **163**:1138-1152.

This paper reports cryo-EM structures of synaptc RAG complexes with nicked and cleaved RSS intermediates and products. The study elaborates the conformational remodeling upon RSS binding, the structural basis of RSS and coding flank recognition, the catalytic mechanism of hairpin formation, and the molecular explanation for the 12/23 rule.

40. •• Kim MS, Chuenchor W, Chen X, Cui Y, Zhang X, Zhou ZH, Gellert M, Yang W: **Cracking the DNA code for V(D)J recombination**. *Mol Cell* 2018, **70**:358-370 e354..

This paper describes higher resolution crystal structures of RAG in complex with various RSSs, and reports the detailed interaction with HMGB1. The structures of RAG in complex with intact RSS substrates raise the puzzle that the scissile phosphate in RSS is 20 Å away from the active center.

41. •• Ru H, Mi W, Zhang P, Alt FW, Schatz DG, Liao M, Wu H: **DNA melting initiates the RAG catalytic pathway**. *Nat Struct Mol Biol* 2018, **25**:732-742.

This paper captures by cryo-EM the unmelted and melted states of intact RSS substrates that are bound to the RAG complex. The study unravels that DNA melting driven by the inward movement of the ID leads to the rotation of the coding flank by 180° and positions the scissile phosphate into the active site. DNA melting may be a universally adopted mechanism at the first step of catalysis by DDE family DNA transposases and retroviral integrases.

42. Swanson PC: **The bounty of RAGs: recombination signal complexes and reaction outcomes**. *Immunol Rev* 2004, **200**:90-114.

43. Hickman AB, Dyda F: **DNA transposition at work**. *Chem Rev* 2016, **116**:12758-12784.

44. Cuomo CA, Mundy CL, Oettinger MA: **DNA sequence and structure requirements for cleavage of V(D)J recombination signal sequences**. *Mol Cell Biol* 1996, **16**:5683-5690.

45. Ramsden DA, McBlane JF, van Gent DC, Gellert M: **Distinct DNA sequence and structure requirements for the two steps of V(D)J recombination signal cleavage**. *EMBO J* 1996, **15**:3197-3206.

46. Swanson PC, Desiderio S: **V(D)J recombination signal recognition: distinct, overlapping DNA-protein contacts in complexes containing RAG1 with and without RAG2**. *Immunity* 1998, **9**:115-125.

47. Peak MM, Arbuckle JL, Rodgers KK: **The central domain of core RAG1 preferentially recognizes single-stranded recombination signal sequence heptamer**. *J Biol Chem* 2003, **278**:18235-18240.

48. Patel DJ, Shapiro L, Hare D: **Nuclear magnetic resonance and distance geometry studies of DNA structures in solution**. *Annu Rev Biophys Biophys Chem* 1987, **16**:423-454.

49. Santagata S, Besmer E, Villa A, Bozzi F, Allingham JS, Sobacchi C, Haniford DB, Vezzoni P, Nussenzweig MC, Pan ZQ *et al.*: **The RAG1/RAG2 complex constitutes a 3' flap endonuclease: implications for junctional diversity in V(D)J and transpositional recombination**. *Mol Cell* 1999, **4**:935-947.

50. Montano SP, Rice PA: **Moving DNA around: DNA transposition and retroviral integration**. *Curr Opin Struct Biol* 2011, **21**:370-378.

51. Montano SP, Pigli YZ, Rice PA: **The mu transpososome structure sheds light on DDE recombinase evolution**. *Nature* 2012, **491**:413-417.

52. Davies DR, Goryshin IY, Reznikoff WS, Rayment I: **Three-dimensional structure of the Tn5 synaptic complex transposition intermediate**. *Science* 2000, **289**:77-85.

53. Richardson JM, Colloms SD, Finnegan DJ, Walkinshaw MD: **Molecular architecture of the Mos1 paired-end complex: the structural basis of DNA transposition in a eukaryote**. *Cell* 2009, **138**:1096-1108.

54. Dyda F, Rice PA: **A new twist on V(D)J recombination**. *Nat Struct Mol Biol* 2018, **25**:648-649.

55. Lee I, Harshey RM: **Patterns of sequence conservation at termini of long terminal repeat (LTR) retrotransposons and DNA transposons in the human genome: lessons from phage Mu**. *Nucleic Acids Res* 2003, **31**:4531-4540.