nature structural & molecular biology

ARTICLES

https://doi.org/10.1038/s41594-019-0363-2

Cutting antiparallel DNA strands in a single active site

Xuemin Chen[1], Yanxiang Cui[2], Robert B. Best[3], Huaibin Wang[1], Z. Hong Zhou[2,4], Wei Yang[1*] and Martin Gellert[1*]

A single enzyme active site that catalyzes multiple reactions is a well-established biochemical theme, but how one nuclease site cleaves both DNA strands of a double helix has not been well understood. In analyzing site-specific DNA cleavage by the mammalian RAG1–RAG2 recombinase, which initiates V(D)J recombination, we find that the active site is reconfigured for the two consecutive reactions and the DNA double helix adopts drastically different structures. For initial nicking of the DNA, a locally unwound and unpaired DNA duplex forms a zipper via alternating interstrand base stacking, rather than melting as generally thought. The second strand cleavage and formation of a hairpin–DNA product requires a global scissor-like movement of protein and DNA, delivering the scissile phosphate into the rearranged active site.

Many bacterial and eukaryotic transposases contain an RNase H-like (RNH) catalytic core and use a single active site to cleave both DNA strands at the boundary of recognition sequences[1,2]. Among them, RAG1–RAG2 (products of RAG, recombination activating genes) cleaves DNA in the immunoglobulin and T-cell receptor loci to initiate the process of V(D)J recombination and generate immune-system diversification in jawed vertebrates[2–4]. Each of the two DNA recombination signal sequences (RSSs)—12RSS and 23RSS—which are composed of a conserved heptamer and nonamer separated by a 12 or 23 bp non-conserved spacer, marks the borders of the antigen receptor V, D or J coding segments (the coding flanks). On binding a pair of 12 and 23RSS DNAs, RAG first nicks one strand of each RSS and then cleaves the second strands by forming DNA hairpins (Fig. 1a,b)[5–7].

Although resistant to structural study for two decades, zebrafish RAG (zRAG) and mouse RAG (mRAG) have in recent years yielded crystal and cryo-EM structures of an apo form, and the DNA-bound pre-reaction (PRC) and hairpin-forming complexes (HFC) (Fig. 1a)[4,8–10]. These structures reveal how a Y-shaped dimer of RAG1–RAG2 heterodimers pairs asymmetrical 12 and 23RSS DNAs and can undergo large conformational changes. The conserved catalytic core of RAG binds the recombination signal DNA in the same fashion as all RNH-type transposases[11–16]. However, RAG nicks the top strand at the 5′ boundary of the RSS first, while all bacterial and many eukaryotic transposases first cut the equivalent of the bottom strand at the 3′ boundary[1,17–19] (Fig. 1b). In both mouse and zebrafish PRC structures[9,10], the bottom strand of the B-form DNA substrate, which is the strand opposite to the one that will be nicked, is juxtaposed to the RAG active site.

Recently reported cryo-EM structures of three different zebrafish nick-forming complexes (NFCs) of zRAG reveal that the top strand is placed in the active site for nicking when RSS DNA is untwisted by 180°[10]. However, these structures with one RSS DNA (either 12 or 23RSS) or both untwisted were determined at moderate resolutions (4–5 Å) from a mixture of three PRC complexes (with one or both RSS DNAs bound) in the same cryo-EM sample. In all six zRAG structures, the active sites are reported to be fully formed before the DNA is unwound, with the catalytic DDE motif binding two metal ions and situated adjacent to either the top (NFC) or bottom strand (PRC) for DNA cleavage[10]. In these PRC structures, it is not clear why zRAG does not nick the bottom strand first.

After obtaining a pure mRAG NFC with modified DNA substrates, we have determined the structure up to 3.2-Å resolution and observed a previously unknown DNA zipper structure. The RAG active site, which is only partially formed and non-reactive in PRC, becomes fully assembled and adopts two different configurations for nicking and hairpin forming reactions in NFC and HFC, respectively. This structure completes the jigsaw puzzle of the mechanism determining how the two DNA cleavage reactions take place in a single active site.

Results

Structural characterization of a stable NFC by cryo-electron microscopy. To avoid a severe preferred orientation problem in cryo-EM, we used mRAG1 (amino acids (aa) 265–1040) and RAG2 (aa 1–520), which are longer than the catalytic core (mRAG1 of 384–1008 and mRAG2 of 1–359) used in crystallographic studies[9]. Both forms of mRAG are active in DNA cleavage assays[9,20]. To eliminate the enzymatic activities of mRAG, mutation of the third catalytic residue in the DDE motif, E962Q, was generated for structural analyses[9]. Because the previously determined crystal and cryo-EM structures of HFC composed of wild type (WT) core mRAG or the longer mRAG with E962Q mutation are superimposable, we use the longer mRAG proteins in this report, WT in cleavage assays and both WT and E962Q in cryo-EM analyses.

A normal substrate pair of 12 and 23RSS (DNA0, Fig. 1c) can be cleaved by mRAG at 37 °C but not 22 °C (Extended Data Fig. 1a,b). However, E962Q mutant mRAG–DNA0 complexes were solely in the PRC state even at 37 °C as determined by cryo-EM (Extended Data Fig. 1c and Methods). The discrepancy between the effective DNA nicking in solution and the absence of a corresponding NFC complex captured by cryo-EM was due to the E962Q mutation, which biased the mRAG–DNA complex toward inactive PRC. When WT mRAG and DNA0 were mixed in the presence of

[1]Laboratory of Molecular Biology, NIDDK, National Institutes of Health, Bethesda, MD, USA. [2]California NanoSystems Institute, University of California, Los Angeles, Los Angeles, CA, USA. [3]Laboratory of Chemical Physics, NIDDK, National Institutes of Health, Bethesda, MD, USA. [4]Department of Microbiology, Immunology and Molecular Genetics, University of California, Los Angeles, Los Angeles, CA, USA. *e-mail: weiy@niddk.nih.gov; martinge@niddk.nih.gov

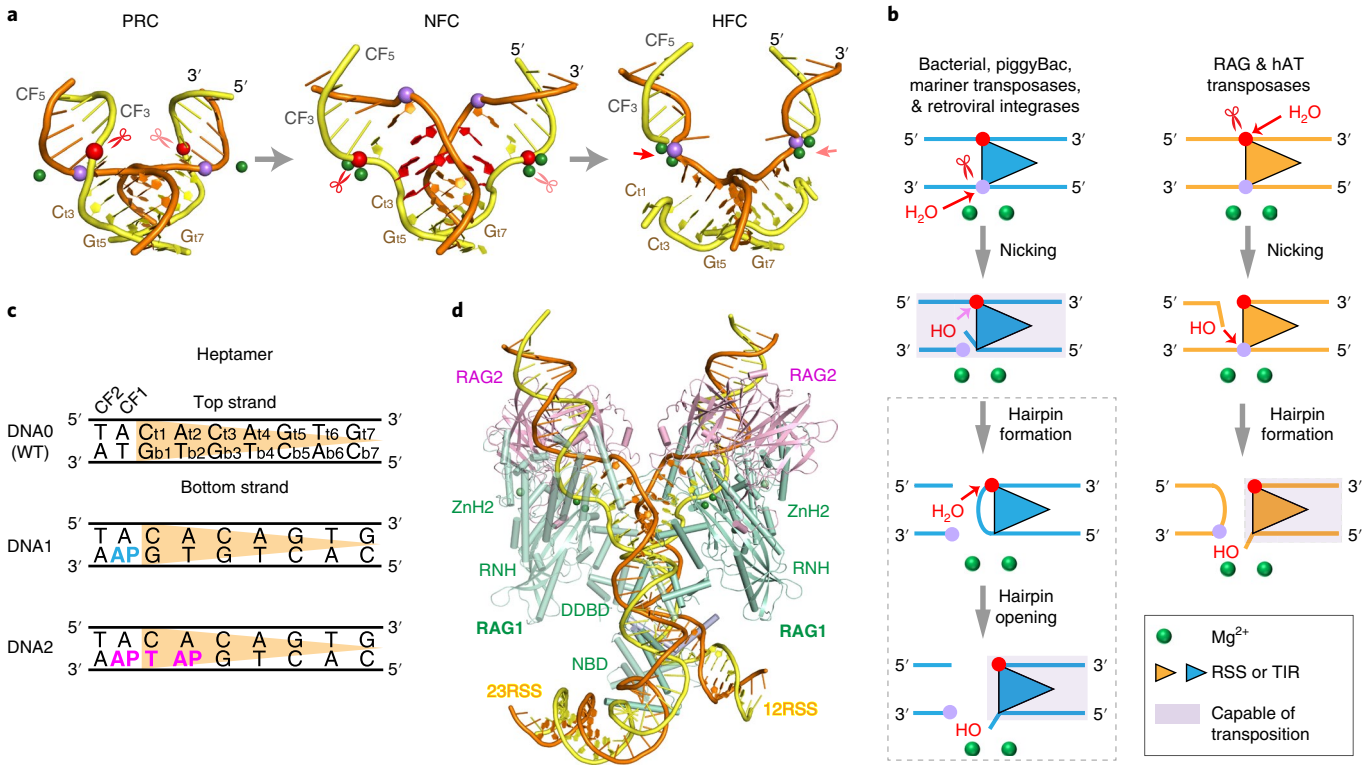


Fig. 1 | Reactions catalyzed by RAG recombinase. a, Three DNA bound states of mRAG. The top- (nicking) and bottom-strand (hairpinning) scissile phosphates are depicted as red and lilac spheres, respectively. The divalent metal ions in the active site are shown as green spheres. In the NFC, bases forming the DNA 'zipper' are shown as red sticks. **b**, Two types of DNA cleavage mechanism exhibited by RNH transposases. The site of the first DNA nick is marked with red scissors, and the nucleophilic attack is indicated with red arrows. RSS, recombination signal sequence; TIR, terminal inverted repeat of transposable elements. The dashed gray box indicates that only a subset of this type of transposase undergoes hairpin formation. **c**, DNA designs for generating NFCs. Mutations are highlighted in blue (DNA1) and magenta (DNA2). AP, abasic analogue (tetrahydrofuran); CF, coding flank. Subscripts label positions of nucleotides in the top (t) and bottom (b) strands of the heptamer. **d**, Overall structure of the NFC (DNA1). Protein domains and 12/23RSS DNAs are labeled.

Ca^{2+} at 37 °C, ~15% of the resulting complexes had one RSS DNA in the NFC conformation on cryo-EM grids, and a few were pure NFC (see Methods).

As DNA cleavage is improved by strategically placed abasic sites[21], we constructed DNA substrates with both 12 and 23RSS DNA containing an abasic site in the coding flank (CF_{b1}; designated as DNA1) or with two additional mutations (G_{b1} to T and T_{b2} to abasic) in the heptamer (DNA2) (Fig. 1c). DNA1, which was a better substrate than WT DNA0 both for nicking and hairpinning at 22 °C or 37 °C, produced more NFC at 37 °C than 22 °C when mixed with the E962Q mutant (Extended Data Fig. 1). DNA2 is nicked as efficiently as DNA1 at 22 °C, but cannot form hairpins (Extended Data Fig. 1a,b) due to the loss of mRAG1-DNA interactions necessary for the second strand cleavage[9]. To the advantage of cryo-EM analysis, DNA2 complexed with the E962Q mutant mRAG produced a pure NFC state at 22 °C (Extended Data Fig. 1c).

Cryo-EM structures with both 12 and 23RSS DNAs untwisted by 180° and the top strand in the active site of mRAG for nicking were reconstructed at 3.3-Å resolution with DNA2 and 3.7-Å resolution with DNA1 (Fig. 1d and Table 1), and improved to 3.2- and 3.4-Å resolution, respectively, by applying a two-fold symmetry to the Y-shaped RAG–DNA complex (Extended Data Figs. 2 and 3). The NFC structure of WT mRAG–DNA0 was also determined at 3.6-Å resolution (see Methods). Despite different DNA sequences adjacent to the cleavage site, the two NFC structures of DNA1 and DNA2 are superimposable with a r.m.s. deviation (r.m.s.d.) of 1.1 Å over 1,891 pairs of Cα atoms (Extended Data Fig. 4a), thus cross-validating each other. The NFC structure of WT mRAG with DNA0 has the same structural features as those of E962Q mutant RAG with DNA1 or DNA2 (Extended Data Fig. 4 and Methods). The r.m.s.d. of 878 pairs of Cα atoms (one mRAG heterodimer, excluding the nonamer-binding domain (NBD)) between the WT–DNA0 and E962Q–DNA1 NFC complexes is 0.6 Å, which is similar to the r.m.s.d. between the NFC states of DNA1 and DNA2 bound to E962Q mutant RAG (912 pairs of Cα atoms in one mRAG heterodimer and 0.7 Å).

An untwisted but base-stacked DNA zipper in mouse NFC. In all NFC structures (DNA0, DNA1 or DNA2), DNA untwisting occurs locally in the second and third base pairs of each heptamer (CACAGTG), and the normally cylindrical DNA helix becomes a flat ribbon (Fig. 2a and Supplementary Video 1). Surrounding the 180° untwisting, the first (C/G) and fourth (A/T) base pairs of the heptamer remain hydrogen bonded, and the rest of the RSS is as in the PRC. The coding flank DNA beyond each RSS is rotated by 180° and presents its major groove rather than the minor groove (as in PRC) to contact RAG2 in the NFC (Extended Data Fig. 4b,f). The four nucleotides of the unpaired $A_{t2}C_{t3}/(T_{b2}G_{b3})$ (the parenthesis indicating the complementary strand) form an interstrand base-stacked zipper in the order of $(T_{b2})A_{t2}(G_{b3})C_{t3}$, by untwisting the strands, stretching lengthwise, breaking base pairing and sliding the two strands toward each other with interdigitated base stacking (Figs. 2b and 3a,b). The rise between these cross-stacked single bases is 3.2–3.4 Å, comparable to the base pair separation in A and B forms, and the heptamer DNA is elongated by >6 Å compared to the PRC. The interstrand base stacking brings the two nearly

Table 1 | Cryo-EM data collection, refinement and validation statistics

	PRC (DNA0-E962Q) (EMD-20030, PDB 6OEM)	PRC (DNA1-E962Q) (EMD-20031, PDB 6OEN)	NFC (DNA1-E962Q) (EMD-20032, PDB 6OEO)	NFC (DNA2-E962Q) (EMD-20033, PDB 6OER)	12RSS-NFC / 23RSS-PRC (DNA1-E962Q) (EMD-20034, PDB 6OEP)	12RSS-PRC / 23RSS-NFC (DNA1-E962Q) (EMD-20035, PDB 6OEQ)	NFC_C2 (DNA1-E962Q) (EMD-20039)	NFC_C2 (DNA2-E962Q) (EMD-20038)	NFC (DNA0-WT) (EMD-21003, PDB 6V0V)
Data collection and processing									
Magnification	130,000	130,000	130,000	130,000	130,000	130,000	130,000	130,000	130,000
Voltage (kV)	300	300	300	300	300	300	300	300	300
Electron exposure ($e^-/Å^2$)	57	42	42	50–60	42	42	42	50–60	45
Defocus range (μm)	−1.4 to −3.0	−1.4 to −3.0	−1.4 to −3.0	−1.4 to −3.0	−1.4 to −3.0	−1.4 to −3.0	−1.4 to −3.0	−1.4 to −3.0	−1.2 to −3.0
Pixel size (Å)	1.07	1.07	1.07	1.07	1.07	1.07	1.07	1.07	1.06
Symmetry imposed	*C1*	*C1*	*C1*	*C1*	*C1*	*C1*	*C2*	*C2*	*C1*
Initial particle images (no.)	590,590	2,619,084	2,619,084	1,689,209	2,619,084	2,619,084	2,619,084	1,689,209	1,282,896
Final particle images (no.)	109,865	29,224	109,388	333,280	107,398	27,374	109,388	333,280	111,362
Map resolution (Å)	3.6	4.3	3.7	3.3	3.7	4.3	3.44	3.15	3.6
FSC threshold	0.143	0.143	0.143	0.143	0.143	0.143	0.143	0.143	0.143
Map resolution range (Å)	3–6	4–7	3–6	3–6	3–7	4–7	3–5	2.5–4.5	3–7
Refinement									
Initial model used (PDB code)	6CIK	6OEM	6OER	5ZE0	6OEO, 6OEN	6OEO, 6OEN			6OEO
Model resolution (Å)	4.1	4.4	3.8	3.4	3.8	4.4			3.9
FSC threshold	0.5	0.5	0.5	0.5	0.5	0.5			0.5
Map sharpening *B* factor ($Å^2$)	−50	−100	−100	−80	−100	−100	−100	−129	−60
Model composition									
Nonhydrogen atoms	19,704	19,662	19,518	19,484	19,205	19,147			8,193
Protein residues	2,000	1,998	1,970	1,952	1,909	1,910			887
Ligands	4	3	6	6	5	4			3
B factors ($Å^2$)									
Protein	114.86	131.29	89.17	85.32	96.36	120.58			98.93
Ligand	96.17	100.76	102.68	97.79	114.42	116.88			115.51
r.m.s. deviations									
Bond lengths (Å)	0.002	0.006	0.005	0.007	0.010	0.009			0.008
Bond angles (°)	0.473	0.980	0.716	0.799	1.019	1.053			1.000
Validation									
MolProbity score	2.64	2.31	2.15	2.18	1.87	1.98			2.01
Clashscore	7.52	8.05	13.82	13.51	6.20	8.58			10.71
Poor rotamers (%)	2.01	3.72	0.06	0.12	0.37	0.62			0.40
Ramachandran plot									
Favored (%)	95.4	93.17	91.71	90.34	90.70	91.24			92.61
Allowed (%)	4.6	6.83	8.19	9.45	9.24	8.71			7.28
Disallowed (%)	0.00	0.00	0.10	0.21	0.05	0.05			0.11

parallel phosphosugar backbones 3 Å closer than in B-DNA, forming a 'slender waist' in the DNA substrate (Fig. 2a).

In the three best resolved zebrafish NFC structures, which are of either RSS DNA (12 or 23) untwisted and an average of the two[10], the outline of untwisted DNA backbones is similar to that in the mouse NFC, but the first 4 bp in each heptamer (CACAGTG) are modeled as melted and unpaired, and two bases (T_{b2} and G_{b3}) are flipped out of the duplex with no contact by zRAG (Fig. 3c). The cryo-EM

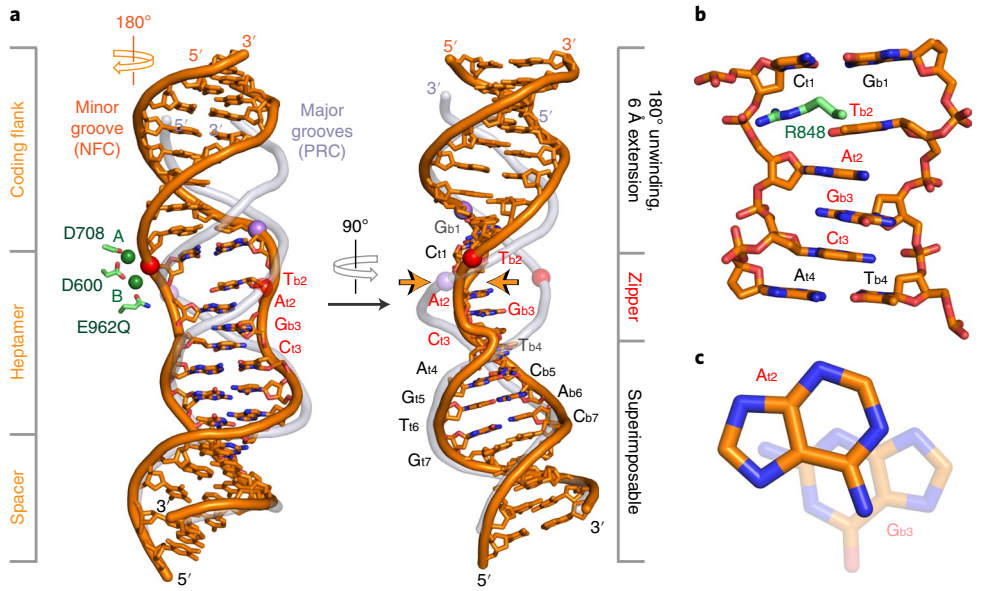


Fig. 2 | Structure of the DNA zipper in NFC. a, Two orthogonal views of the untwisted and extended DNA in NFC (orange), with the DNA in PRC (semitransparent gray) superimposed. The $(T_{b2})A_{t2}(G_{b3})C_{t3}$ zipper stabilized by the interstrand base stacking is labeled in red. Arrows highlight the flattened CACA (heptamer) in NFC. The scissile phosphates are shown as red and lilac spheres. The direction of DNA unwinding is marked in orange. The active site is marked by the catalytic DDE motif and two divalent cations. **b**, An enlarged view of the $(T_{b2})A_{t2}(G_{b3})C_{t3}$ zipper in the NFC structure of DNA1. **c**, Top view of the stacked $A_{t2}G_{b3}$ bases in the NFC structure of DNA2.

density maps for these three related structures differ somewhat. Although they were interpreted as supporting the base-flipped-out model[10], they can also support the zipper DNA (Fig. 3c). In the proposed zebrafish NFC structure with base pairs melted, how the DNA becomes lengthened by 6 Å rather than shortened and what stabilizes the flipped-out extrahelical bases remain unexplained.

In the mouse NFC structures, the 6 Å extension of DNA is a result of the zipper formation and is as essential as the 180° unwinding for placing the scissile phosphate in the RAG active site. In the DNA zipper, the bases are stabilized by stacking on each other. The conserved A_{t2} and G_{b3} are crucial for zipper formation, and each contributes a positively charged amine group (N^6 of A_{t2} and N^2 of G_{b3}) to form cation-π interactions with its stacked neighbors (Fig. 2b,c). The same zipper also forms with native DNA (DNA0) in complex with WT mRAG (Fig. 3d). As the DNA zipper is intact in DNA2 that lacks T_{b2} (Fig. 3b), the remaining $A_{t2}(G_{b3})C_{t3}$ (or A(G)C) must be sufficient to maintain the untwisted DNA conformation. Base stacking is well known to be the principal force that stabilizes the DNA double helix[22,23]. Here the interstrand base stacking stabilizes the two untwisted strands and prevents DNA 'melting'.

To assess the intrinsic propensity for DNA to form a zipper, as in the mouse NFC structures, versus a melted and base-flipped-out structure reported for zebrafish, we first carried out unbiased molecular dynamics simulations of these two DNA structures (see Methods). We found that simulations started from the zipper structure were stable, remaining ~2 Å all-atom r.m.s.d. from the initial coordinates over 120 ns simulation time, regardless of whether the Amber14[24] or CHARMM 36 force fields for DNA[25] were used. On the other hand, simulations started from the base-flipped structure quickly diverged from that state, with increasing r.m.s.d. from the initial coordinates, and became more zipper-like as evident from the decreasing r.m.s.d. to the zipper coordinates (Extended Data Fig. 6). As an additional test, we ran simulations from a canonical B-form DNA and applied a moving bias force to the terminal nucleotides to mimic the 6 Å stretching and unwinding in the NFC. Even with this simple bias, the DNA moves to within 2–3 Å r.m.s.d. from the zipper structure, while remaining ~5 Å r.m.s.d. from the base-flipped structure.

A related interdigitated base-stacking structure has also been observed in a DNA unwound by ~60° in complex with a tyrosine recombinase (Fig. 3e)[26]. There, the DNA zipper is formed by A(A)T instead of A(G)C. We suspect that R(R)Y (R for purine and Y for pyrimidine) is favorable for interstrand base stacking within a DNA zipper structure and can form at RY or YR dinucleotide sequences when DNA is unwound. Indeed, hAT transposases, which are homologous to RAG (Fig. 1b), may rely on the conserved CA or TA sequence at the transposon end[27] to form a DNA zipper for the first cleavage.

Stabilization of the DNA zipper by mRAG. To understand how the RAG protein may influence DNA unwinding, we redetermined the mouse PRC structure by cryo-EM at 3.6-Å resolution (see Methods and Extended Data Fig. 7). This cryo-EM structure differs from the existing PRC crystal structure by a 20° rotation of the Y stem (consisting of NBD and nonamer regions of DNA)[9] (Extended Data Fig. 4c). The large difference is most likely due to crystallization because the Y stems in the cryo-EM structures of PRC and NFC differ by less than 5° (Extended Data Fig. 4b). Apart from the Y stem, the PRC structures are similar (Extended Data Fig. 4c). Free from the effects of crystallization, the cryo-EM PRC structure is slightly expanded, as observed previously for the HFC[9]. As a bonus, the cryo-EM density map of mouse PRC reveals the trace of loop L_{12} (residues 606–617), which connects the first two β strands of the RNH domain (Supplementary Fig. 1 and Extended Data Fig. 7h) and extends from one RAG1 subunit towards the other, forming *trans* interactions with the RSS DNA on the opposite Y arm (Figs. 4a and 5a).

When cryo-EM structures of mouse PRC and NFC are superimposed, they are indistinguishable overall from RAG2 at the tip of the Y structure to NBD at the Y stem (Fig. 4a and Extended Data Fig. 4b). The main difference occurs in the ZnH2 domains (aa 793–951) of the RAG1 subunits (Fig. 4a and Supplementary Video 2). ZnH2, which is an insertion in the RNH domain and

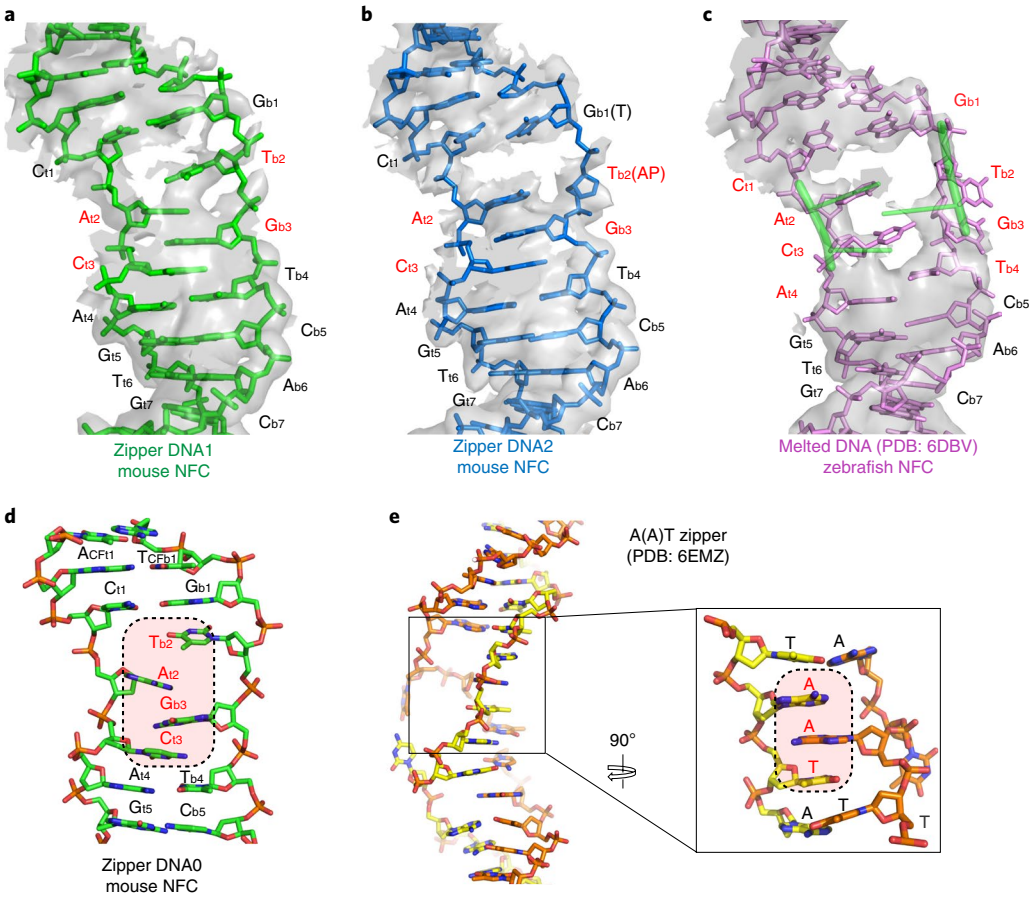


Fig. 3 | A DNA zipper is associated with untwisted DNA. a,**b**, The DNA zipper in mouse NFC structures with DNA1 or DNA2. The bases forming the zipper are labeled in red letters. **c**, The melted DNA and flipped-out bases (labeled in red) in zebrafish NFC (PDB 6DBV). The three bases $A_{t2}(G_{b3})C_{t3}$ forming the zipper in mouse NFC are superimposed and shown as a green cartoon. In **a**–**c**, cryo-EM density maps (contoured at 5σ) are superimposed onto the corresponding structures in semitransparent gray. **d**, The (T)A(G)C zipper in the wild-type mouse NFC structures. The zipper is highlighted. **e**, The A(A)T zipper in the Tn1549 transposon structure (PDB 6EMZ) is shown in orthogonal views.

forms an appendage on each Y arm, opens outward upon DNA binding (25° rotation and 13 Å translation)[9,10] and then closes inward in transitioning from PRC to NFC (12° rotation and 6 Å translation). In the PRC before DNA is unwound, R848 in ZnH2 is buried deep in the minor groove of the coding flank abutting the heptamer, and the nearby M849, N850, G851 and N852 track along the first three bases (CAC) of the RSS (Fig. 4b). In passing from PRC to NFC, each ZnH2 domain must traverse a DNA phosphosugar backbone and transition into the major groove. In the NFC structures, side chains of M849 to N852 still track the backbone of CAC but now insert into the major groove, forming both hydrophobic and charge interactions with the DNA (Fig. 4c). Instead of serving as a 'piston' to actively drive DNA untwisting as proposed[10,28], the movement of each ZnH2 domain is in the direction opposite to that of DNA unwinding (Fig. 4a,d).

If ZnH2 were to drive DNA untwisting, R848, which is inserted in the minor groove in PRC (Fig. 4b,c), would play a positive role in untwisting DNA and stabilizing the DNA zipper afterwards (Fig. 2b). As RAG is not an ATPase and cannot move any domain in a directional manner, R848 may instead limit Brownian motions of ZnH2 and prevent it from moving between DNA grooves. To discern how DNA untwisting is initiated and what role the ZnH2 domain plays, we substituted Ala for R848 and found that, without the R848 side chain, the mutant mRAG is more rather than less active in DNA nicking and hairpinning (Fig. 4e). R848 thus appears to act as a barrier to ZnH2 movement and DNA untwisting, and its inhibitory effect appears to outweigh its role in stabilizing the zipper.

Opposite ZnH2, loop L_{12}, which is present but not included in the model of the zebrafish structures due to the limited resolution of the cryo-EM map[10], contacts the heptamer DNA and also changes conformation from PRC to NFC (Fig. 4a). In PRC, loop L_{12} contacts the bottom strand of the heptamer mostly on the major groove side (Fig. 5a). Accompanying DNA unwinding, L_{12} transforms from a stubby hairpin in PRC to an elongated one in NFC (Fig. 5b), and the tip of the β-turn (G610 and S611) moves >6.5 Å, again in a direction opposite to DNA unwinding (Supplementary Video 2). In NFC, L_{12} is entirely in the minor groove, forming close contacts with the top of the DNA zipper ($T_{b2}A_{t2}$) and stabilizing the phosphate group immediately downstream of the scissile phosphate. The unwound and flattened DNA zipper ($(T_{b2})A_{t2}(G_{b3})C_{t3}$) is sandwiched between ZnH2 on the major groove side and L_{12} on the minor groove side (Fig. 4c). Each RAG-heptamer DNA interface is increased by 370 Å^2 from PRC to NFC.

The RAG active site and strand-specific nicking. The RAG1 active site is not fully assembled in PRC, as the third catalytic residue, E962, is 10 Å away from the first two, D600 and D708 (Extended Data Fig. 4e)[9]. This is because helix αX (Supplementary Fig. 1), on which E962 is situated, is not properly oriented in the apo form or in PRC, where the scissile phosphate is far away from the active site. In NFC, when the scissile phosphate is placed in the active site upon formation

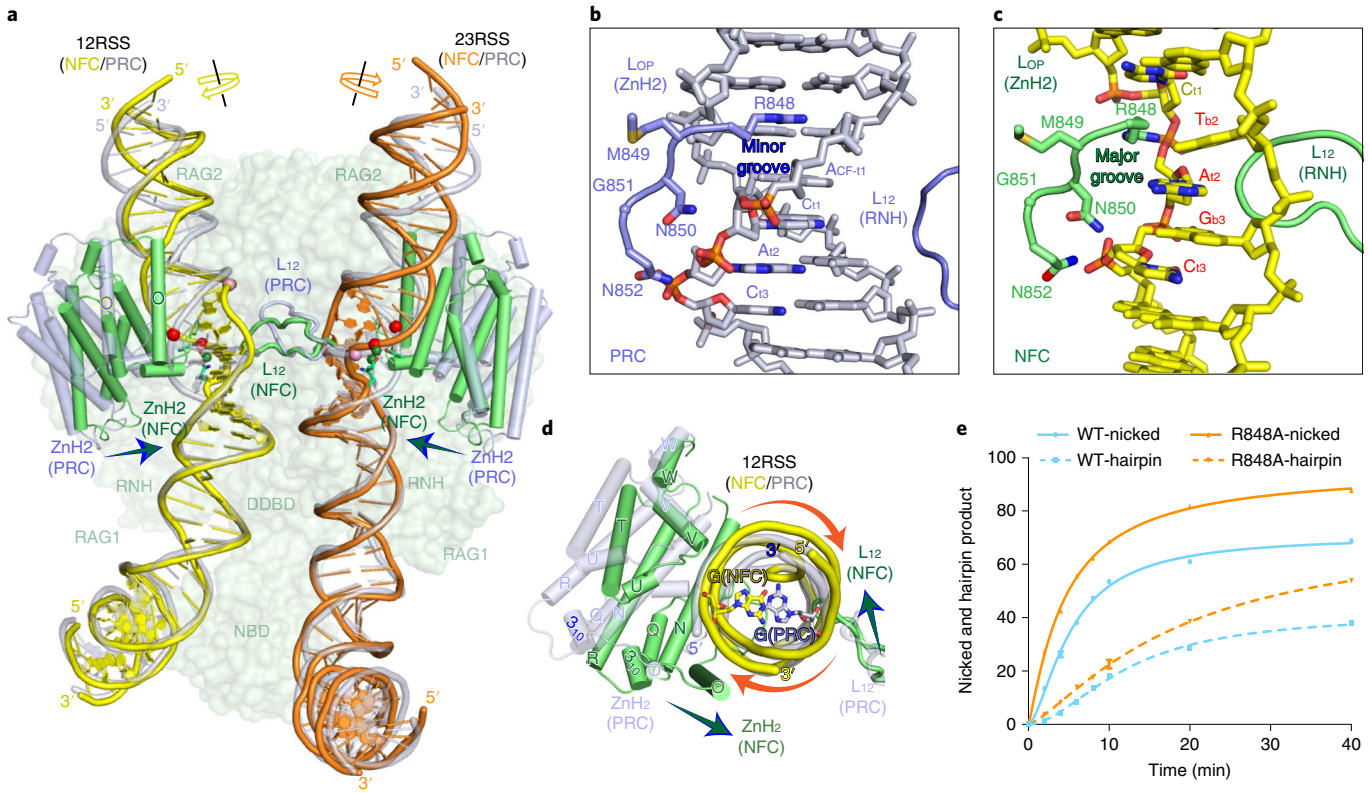


Fig. 4 | Protein-DNA interactions in NFC. a, Superimposition of cryo-EM structures of PRC and NFC. The RAG protein in NFC is shown in a semitransparent green surface representation. ZnH2 domains and one L_{12} loop are shown in green (NFC) and light blue (PRC) cartoons. DNAs are shown as yellow and orange (NFC) or gray (PRC) cartoons. Directions of ZnH2 domain and loop L_{12} movement from PRC to NFC are indicated by arrowheads, and those of 12 and 23RSS DNA by curved arrows. **b**,**c**, The ZnH2 domain contacts the coding flank and heptamer in the minor groove in PRC (**b**), but in the major groove in NFC (**c**). **d**, The ZnH2 domain and loop L_{12} (blue-green arrow) of RAG1 and the RSS DNA (red arrow) move in opposite directions during the PRC to NFC transition. One CF base (Gua, G) is shown to mark the DNA rotation. **e**, DNA cleavage activities of WT and R848A mutant mRAG on the normal substrate (DNA0). Mean values and s.d. were obtained from three independent samples.

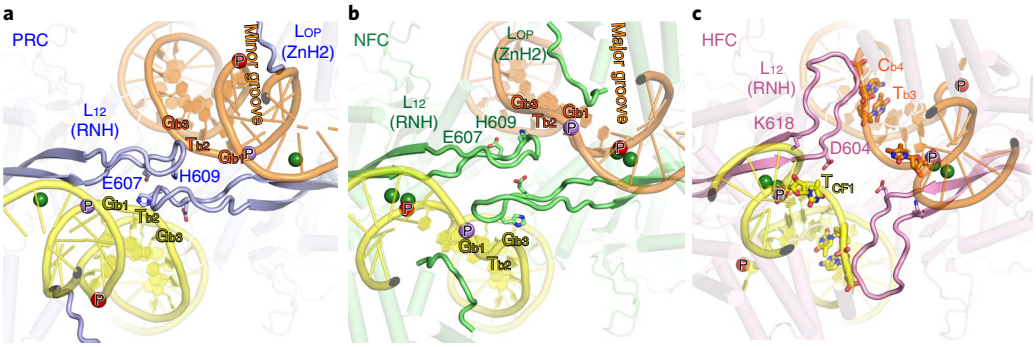


Fig. 5 | Different configurations of L_{12} loops in PRC, NFC and HFC. a–**c**, The L_{12} loops bridge the dimer interface between RNH domains in PRC (**a**), NFC (**b**) and HFC (**c**). L_{12} tracks $C_{t1}A_{t2}C_{t3}$ of the heptamer in the major groove in PRC but in the minor groove in NFC. In HFC, each L_{12} forms both *cis* and *trans* interactions with the two RSS DNA substrates. Only one Me^{2+} ion (green sphere) is observed in each active site in PRC, but two are present in NFC and HFC, when the scissile phosphate (red and lilac spheres, respectively) is brought into the active site. DNA bases are shown as ladders except for those in the heptamer, where bases and sugars are fully shown.

of the DNA zipper, the ZnH2 domain becomes properly closed and the connected helix αX is re-oriented to bring E962(Q) into the active site (Extended Data Fig. 8a and Supplementary Video 1).

When E962 is far from D600 and D708, only one divalent cation (Mg^{2+} or Mn^{2+}) is observed in the active site of mouse PRC, as revealed by both cryo-EM and crystallography. As two divalent cations were modeled into the active site of zebrafish PRC[10], we soaked mouse PRC crystals in a buffer containing 5 mM Mn^{2+} and collected diffraction data to 3.2-Å resolution. Based on the anomalous signal of Mn^{2+}, we confirmed that only one Mn^{2+} is in each active site (Extended Data Fig. 8b). The incomplete assembly of the mRAG active site, with its dislocated E962 and absence of the second divalent cation, ensures that the wrong (bottom) strand is not cleaved even when juxtaposed with the active site in PRC.

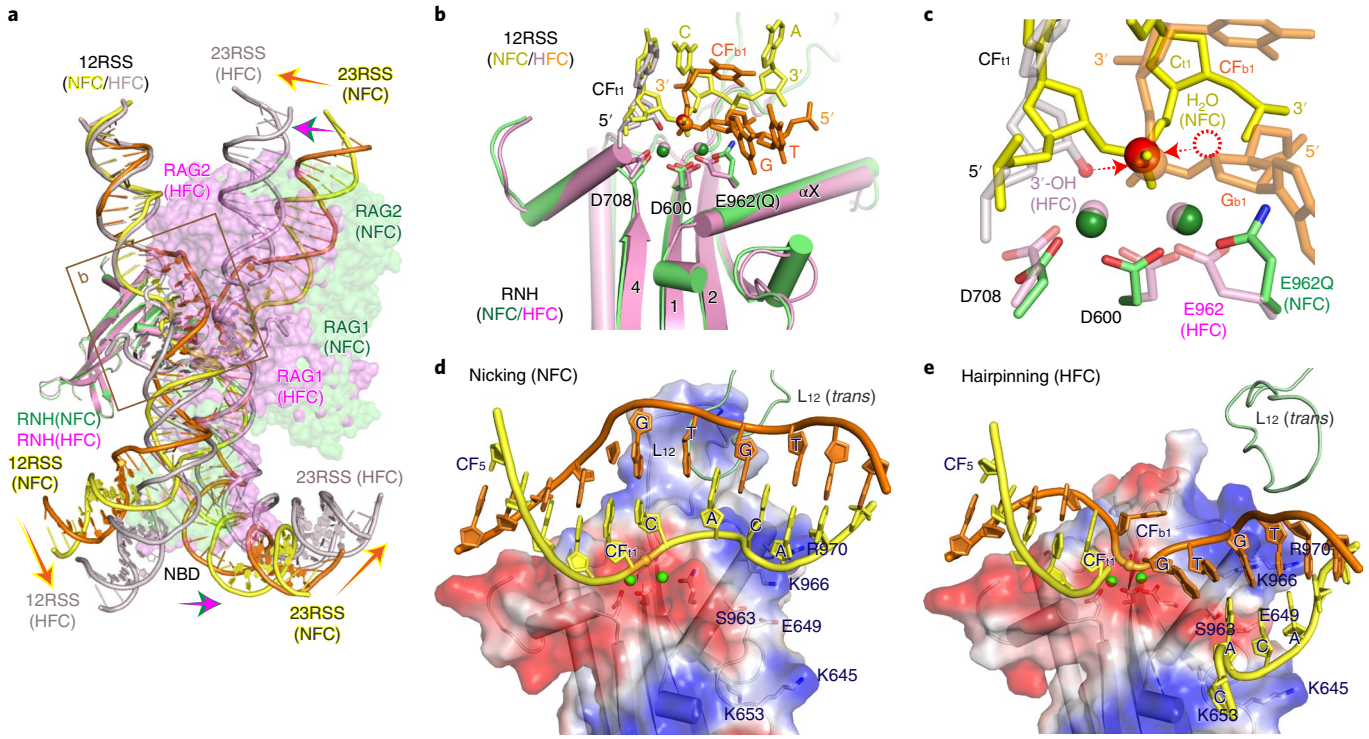


Fig. 6 | Cutting two DNA strands in a single active site. a, After superimposition of one RNH domain in NFC and HFC, RAG on the other Y arm is shown as a green (NFC) and pink (HFC) molecular surface. RSS DNAs are shown as yellow (top strand) and orange (bottom strand) (NFC) or gray (HFC) cartoons. Movements of the protein and RSS DNA are indicated by color-coded arrows. **b**, The RNH domains of NFC (green) and HFC (pink) are superimposable, as are the scissile phosphates and CF_{t1} carrying the 3′-OH product of nicking, which is the nucleophile for hairpin formation. The cleavage strands in NFC (yellow) and HFC (orange) have opposite polarity and bind differently to RNH. CF_{t1} in HFC is shown as light pink sticks. **c**, An enlarged view of the catalytic center from **b**. For the nicking and hairpinning reactions, the nucleophiles are on opposite sides of the scissile phosphates, and E962 assumes different rotamer conformations. The directions of the two reactions are marked by dashed red arrows. **d**,**e**, Different protein and DNA structures of NFC (**d**) and HFC (**e**). The RNH domain is shown as an electrostatic surface (blue and red represent the positive and negative charge potentials, respectively), and DNAs are shown as yellow (top strand) and orange (bottom strand) cartoons. The protein surface is changed due to the large movement of loop L_{12} and reorientation of a few polar side chains.

After hydrolysis of the top strand, the 3′-OH product remains in the active site and serves as the nucleophile for cleaving the bottom strand and forming a hairpin product[9]. When transitioning from NFC to HFC (2.75 Å)[9], the two Y arms of RAG and the bound DNA pivot around the dimerization and DNA binding domain (DDBD) in unison by 12–14° toward each other (Fig. 6a,b and Supplementary Videos 3 and 4). Balancing the Y arm rotation, the NBD domains and the nonamers on the other side of DDBD (the pivoting point) undergo twice as much rotation (24°) and translation (12 Å) (Fig. 6a). In HFC, the heptamer resumes base pairing and moves sideways as a duplex by ~18 Å, thus moving the bottom strand into the active site. Accompanying the global scissor-like movement, the bottom strand is bent 90° immediately beyond the scissile phosphate for hairpin formation (Fig. 6b–e). A near 90° bend of the bottom strand has also been observed in the HFC structure of Hermes transposase[29]. Meanwhile the two L_{12} loops dissociate from each other, and each moves as much as 17 Å. The β-turn of loop L_{12} maintains contact with the heptamer in *trans*, while D604 and K618 at the base of L_{12} help to flip out the first coding-flank base (CF_{b1}) in *cis* to orient the scissile phosphate for hairpin formation (Figs. 5c and 6d,e).

Discussion

DNA untwisting and zipper formation. A remaining question is what causes DNA to untwist by 180° and form a zipper in the absence of an ATPase or other external energy source. It has been proposed that RAG recombinase functions as a piston and its domain movement forcefully drives the DNA to unwind[10,28], but during the transition from the PRC to the NFC state in both mRAG and zRAG, the ZnH2 domain and loop L_{12} move in the opposite direction to DNA unwinding (Fig. 4d), and thus it is unlikely that these protein movements are the cause of DNA unwinding. Moreover, the rest of the Y-shaped RAG dimer remains unchanged between PRC and NFC in both mouse (Extended Data Fig. 4b,e,f) and zebrafish[10]. Even though the coding flank DNAs are untwisted by 180° and present different DNA grooves to the protein, both mRAG and zRAG accommodate DNA changes with the same interface[9,10]. A 'spring-loaded' motion of the protein might occur, but the cryo-EM structures of mouse and zebrafish NFC offer no indication that it takes place. The energy source for DNA untwisting and extension is unclear. We suspect that the CAC sequence in the heptamer plays a key role in DNA untwisting and DNA zipper formation, while domain movement in RAG and engagement of the active site probably stabilize the unwound DNA. In agreement with this DNA-centric view, CAC in the heptamer sequence is conserved in all species undergoing V(D)J recombination, and any mutation in CAC diminishes DNA cleavage and V(D)J recombination[30–32]. As DNA nicking and NFC formation are both temperature-dependent for mRAG and zRAG[10] (Extended Data Fig. 1), thermal energy probably supports DNA untwisting. Analogous to higher temperature, the abasic sites in DNA1 and DNA2 destabilize the double-helix structure by disrupting base stacking and thereby favor DNA distortion and nicking by RAG.

Two consecutive reactions in one active site. The catalytic RNH domain remains unchanged for DNA nicking by hydrolysis (NFC)

and hairpin formation by transesterification (HFC), and the scissile phosphates for the two consecutive reactions are superimposable (Fig. 6b and Extended Data Fig. 8c). However, E962 adopts rotamer conformations that differ by 120° in mouse HFC and NFC (Fig. 6c). The rearrangement is essential to avoid clashes with the DNA in HFC and accommodate the opposite polarity of the DNA substrate. Perhaps due to limited resolution, in the zebrafish PRC and NFC structures, the E962 equivalent (E984) and αX were modeled identically to the HFC structure and the active site reconfigurations were not noted[10]. Nevertheless, the nucleophilic water for nicking and the 3′-OH for hairpin formation are situated on opposite sides of the scissile phosphate in both mRAG and zRAG.

The flexibility of the RAG active site is largely due to two evolutionary changes in the RNase H domain. First, enzymes in the RNase H superfamily usually contain an Asp as the last carboxylate in the catalytic triad[33], but among most RNH-type transposases the Asp is replaced by Glu (E962 equivalent) to form the signature DDE motif[34]. We hypothesize that, because of the longer Glu side chain and more rotamer possibilities, this substitution allows a single active site in the transposases to change configurations and cleave two antiparallel DNA strands as observed in the NFC and HFC of RAG. Second, the RNH transposases often acquire an insertion between the second and third catalytic carboxylates immediately before the last helix (αX in RAG1) in the RNH domain (equivalent to ZnH2 in RAG1)[11,15]. The inserted domain helps to bind DNA strands of opposite polarity and enables the single active site to catalyze multiple reactions.

Concluding remarks. RAG recombinase is a specialized RNH transposase, whose primary function is to generate hairpin ends on coding flank DNA (Fig. 1a,b) for subsequent processing to generate diverse antigen receptors[2,3,35,36], without transposing the cleaved DNA to a new genomic site. Nevertheless, catalysis of multiple hydrolytic and transesterification reactions on double helical DNA is a general feature for all RNH-type transposases. Unlike RAG, the majority of these transposases cleave the bottom strand first (Fig. 1b) and thus probably do not require DNA untwisting by 180° or zipper formation. However, like RAG, cleavage of two antiparallel DNA strands in a single active site will depend on structural rearrangements of both the DNA substrate and the catalytic residues during each reaction cycle.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41594-019-0363-2.

Received: 24 June 2019; Accepted: 13 December 2019;
Published online: 3 February 2020

References

1. Mizuuchi, K. Transpositional recombination: mechanistic insights from studies of Mu and other elements. *Annu. Rev. Biochem.* **61**, 1011–1051 (1992).
2. Gellert, M. V(D)J recombination: RAG proteins, repair factors, and regulation. *Annu. Rev. Biochem.* **71**, 101–132 (2002).
3. Schatz, D. G. & Swanson, P. C. V(D)J recombination: mechanisms of initiation. *Annu. Rev. Genet.* **45**, 167–202 (2011).
4. Kim, M. S., Lapkouski, M., Yang, W. & Gellert, M. Crystal structure of the V(D)J recombinase RAG1–RAG2. *Nature* **518**, 507–511 (2015).
5. Sakano, H., Huppi, K., Heinrich, G. & Tonegawa, S. Sequences at the somatic recombination sites of immunoglobulin light-chain genes. *Nature* **280**, 288–294 (1979).
6. Lewis, S. M. The mechanism of V(D)J joining: lessons from molecular, immunological, and comparative analyses. *Adv. Immunol.* **56**, 27–150 (1994).
7. Lapkouski, M., Chuenchor, W., Kim, M. S., Gellert, M. & Yang, W. Assembly pathway and characterization of the RAG1/2–DNA paired and signal-end complexes. *J. Biol. Chem.* **290**, 14618–14625 (2015).
8. Ru, H. et al. Molecular mechanism of V(D)J recombination from synaptic RAG1–RAG2 complex structures. *Cell* **163**, 1138–1152 (2015).
9. Kim, M. S. et al. Cracking the DNA code for V(D)J recombination. *Mol. Cell* **70**, 358–370 (2018).
10. Ru, H. et al. DNA melting initiates the RAG catalytic pathway. *Nat. Struct. Mol. Biol.* **25**, 732–742 (2018).
11. Davies, D. R., Goryshin, I. Y., Reznikoff, W. S. & Rayment, I. Three-dimensional structure of the Tn5 synaptic complex transposition intermediate. *Science* **289**, 77–85 (2000).
12. Richardson, J. M., Colloms, S. D., Finnegan, D. J. & Walkinshaw, M. D. Molecular architecture of the Mos1 paired-end complex: the structural basis of DNA transposition in a eukaryote. *Cell* **138**, 1096–1108 (2009).
13. Hare, S., Gupta, S. S., Valkov, E., Engelman, A. & Cherepanov, P. Retroviral intasome assembly and inhibition of DNA strand transfer. *Nature* **464**, 232–236 (2010).
14. Montano, S. P., Pigli, Y. Z. & Rice, P. A. The Mu transpososome structure sheds light on DDE recombinase evolution. *Nature* **491**, 413–417 (2012).
15. Hickman, A. B. et al. Structural basis of hAT transposon end recognition by Hermes, an octameric DNA transposase from *Musca domestica*. *Cell* **158**, 353–367 (2014).
16. Passos, D. O. et al. Cryo-EM structures and atomic model of the HIV-1 strand transfer complex intasome. *Science* **355**, 89–92 (2017).
17. Yusa, K. piggyBac transposon. *Microbiol. Spectr.* **3**, MDNA3-0028-2014 (2015).
18. Lesbats, P., Engelman, A. N. & Cherepanov, P. Retroviral DNA integration. *Chem. Rev.* **116**, 12730–12757 (2016).
19. Nowotny, M., Gaidamakov, S. A., Crouch, R. J. & Yang, W. Crystal structures of RNase H bound to an RNA/DNA hybrid: substrate specificity and metal-dependent catalysis. *Cell* **121**, 1005–1016 (2005).
20. Grundy, G. J., Yang, W. & Gellert, M. Autoinhibition of DNA cleavage mediated by RAG1 and RAG2 is overcome by an epigenetic signal in V(D)J recombination. *Proc. Natl Acad. Sci. USA* **107**, 22487–22492 (2010).
21. Grundy, G. J., Hesse, J. E. & Gellert, M. Requirements for DNA hairpin formation by RAG1/2. *Proc. Natl Acad. Sci. USA* **104**, 3078–3083 (2007).
22. Mills, J. B. & Hagerman, P. J. Origin of the intrinsic rigidity of DNA. *Nucleic Acids Res.* **32**, 4055–4059 (2004).
23. Yakovchuk, P., Protozanova, E. & Frank-Kamenetskii, M. D. Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic Acids Res.* **34**, 564–574 (2006).
24. Zgarbova, M. et al. Refinement of the Cornell et al. nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. *J. Chem. Theory Comput.* **7**, 2886–2902 (2011).
25. Hart, K. et al. Optimization of the CHARMM additive force field for DNA: improved treatment of the BI/BII conformational equilibrium. *J. Chem. Theory Comput.* **8**, 348–362 (2012).
26. Rubio-Cosials, A. et al. Transposase–DNA complex structures reveal mechanisms for conjugative transposition of antibiotic resistance. *Cell* **173**, 208–220 (2018).
27. Atkinson, P. W. hAT transposable elements. *Microbiol. Spectr.* **3**, MDNA3-0054-2014 (2015).
28. Ru, H., Zhang, P. & Wu, H. Structural gymnastics of RAG-mediated DNA cleavage in V(D)J recombination. *Curr. Opin. Struct. Biol.* **53**, 178–186 (2018).
29. Hickman, A. B. et al. Structural insights into the mechanism of double strand break formation by Hermes, a hAT family eukaryotic DNA transposase. *Nucleic Acids Res.* **46**, 10286–10301 (2018).
30. Ramsden, D. A., McBlane, J. F., van Gent, D. C. & Gellert, M. Distinct DNA sequence and structure requirements for the two steps of V(D)J recombination signal cleavage. *EMBO J.* **15**, 3197–3206 (1996).
31. Hesse, J. E., Lieber, M. R., Mizuuchi, K. & Gellert, M. V(D)J recombination: a functional definition of the joining signals. *Genes Dev.* **3**, 1053–1061 (1989).
32. Hu, J. et al. Chromosomal loop domains direct the recombination of antigen receptor genes. *Cell* **163**, 947–959 (2015).
33. Nowotny, M. Retroviral integrase superfamily: the structural perspective. *EMBO Rep.* **10**, 144–151 (2009).
34. Yuan, Y. W. & Wessler, S. R. The catalytic domain of all eukaryotic cut-and-paste transposase superfamilies. *Proc. Natl Acad. Sci. USA* **108**, 7884–7889 (2011).
35. Boboila, C., Alt, F. W. & Schwer, B. in *Adv. Immunol* Vol 116 (ed. Alt, F. A.) Ch. 1 (Elsevier, 2012).
36. Deriano, L. & Roth, D. B. Modernizing the nonhomologous end-joining repertoire: alternative and classical NHEJ share the stage. *Annu. Rev. Genet.* **47**, 433–455 (2013).

Methods

Cell lines. HEK293T cells were originally obtained from Thermo Fisher Scientific and maintained as stock in the Yang laboratory. None of the cell lines used were authenticated or tested for mycoplasma contamination.

Protein and DNA preparation. The mRAG proteins, which comprise active (WT or R848A) or catalytic inactive mutant (E962Q) RAG1 (aa 265–1040) and degradation-resistant T490A mutant RAG2 (aa 1–520), were expressed as N-terminal His6-MBP fusions (on both RAG1 and RAG2) in HEK293T cells and purified as previously described[4,9]. The extended domains of RAG beyond the catalytic core regions help to reduce the preferred orientation problem on cryo-EM grids. In addition to amylose affinity purification, a step of Mono Q anion exchange chromatography improved protein purity and eliminated a trace amount of DNA contamination. The buffer used in amylose affinity purification comprised 20 mM HEPES (pH 7.4), 500 mM KCl, 5% glycerol, 2 mM DTT, 0.5 mM EDTA. The salt concentration of protein samples coming off the amylose column was lowered to 100 mM before loading onto a Mono Q column (GE Healthcare), which was pre-equilibrated with 20 mM HEPES (pH 7.4), 100 mM KCl, 5% glycerol, 2 mM DTT, 0.5 mM EDTA. mRAG protein was eluted by a linear gradient of 100–500 mM KCl. The purified mRAG protein was buffer-exchanged into a storage buffer containing 20 mM HEPES (pH 7.4), 500 mM KCl, 20% glycerol, 0.1 mM EDTA, 2 mM DTT, concentrated to 6–8 mg ml^{-1}, and stored at −80 °C. Human HMGB1 (aa 1–163) was prepared as reported previously[37].

DNA of 12 and 23RSS used for structural analyses and biochemical assays (Supplementary Table 1) was synthesized as ssDNA (Integrated DNA Technologies). Long oligonucleotides (>20 nucleotides) were purified by 8–15% TBE-urea PAGE in a small gel cassette (Life Technologies). Gel-purified oligonucleotides were then loaded onto a Glen Gel-Pak column (Glen Research) and eluted in deionized H_2O. dsDNA was annealed in a Thermocycler in annealing buffer containing 20 mM Tris-HCl, pH 8.0, 0.5 mM EDTA, 50 mM NaCl.

DNA cleavage assays. All assays were performed in a reaction buffer containing 25 mM HEPES (pH 7.4), 100 mM KCl, 1 mM DTT, 0.1 mg ml^{-1} BSA and 5 mM $MgCl_2$. A histone H3K4me3 peptide was included to bind and activate the PHD domain in the extended RAG2[20]. 50 nM each of 12 and 23RSS DNA with a Cy5 or FAM label on the 20 bp coding flank (DNA0, DNA1 and DNA2 for cleavage assay, and pre-nicked WT substrates for hairpin formation assay) (Supplementary Table 1) were incubated with 50 nM of heterotetrameric WT or mutant (R848A) mRAG (tetramer), 100 nM HMGB1 and 200 nM H3K4Me3 peptide (Epicypher) at 22 °C or 37 °C for 0–40 min. Reactions were stopped by adding an equal volume of formamide buffer (95% (vol/vol) formamide, 12 mM EDTA and 0.3% bromophenol blue) and heating at 95 °C for 10 min. Cleavage products were separated by 15% TBE-urea PAGE, visualized and quantified using a Typhoon PhosphorImager (GE Healthcare). Plots of biochemical data show the mean ± s.d. from three independent experiments using Graphpad Prism software (version 7.0).

Cryo-electron microscopy sample preparation and data collection. The procedure for assembly and purification of mRAG complexed with 12 and 23RSSs was similar to that described previously[4,9]. The purified mRAG (WT or E962Q) contained MBP tags on both RAG1 and RAG2 subunits. These MBP tags further helped to even out orientation distributions on QUANTIFOIL R 1.2/1.3 (Cu, 300 mesh) grids. MBP-mRAG protein, 12 and 23RSS DNAs (DNA0, DNA1 or DNA2) (Supplementary Table 1), HMGB1 (aa 1–163) and H3K4Me3 peptide were mixed at a 1:1.2:1.2:2.4:4 molar ratio in buffer containing 20 mM HEPES (pH 7.4), 100 mM KCl, 5 μM $ZnCl_2$, 1 mM DTT, 5% glycerol and 5 mM divalent cation (Mg^{2+} in the E962Q–DNA0 complex, and Ca^{2+} in the E962Q–DNA1, E962Q–DNA2 and WT–DNA0 complexes) and incubated at 37 °C for 15 min. Each mixture of mRAG–DNA was then further purified at 4 °C by size exclusion chromatography on a Superdex 200 Increase 10/300 GL column (GE Healthcare) in buffer containing 20 mM HEPES (pH 7.3), 100 mM KCl, 1% glycerol, 1 mM DTT and 5 mM divalent cation (Ca^{2+} or Mg^{2+}). Only samples in the elution peak fractions were pooled and used for cryo-EM grid preparation.

To capture DNA in the reactive but not yet cleaved NFC state, we incubated a catalysis-deficient (E962Q) mutant mRAG with WT DNA0, mutant DNA1 or DNA2 as substrate (Fig. 1c) at 22 °C or 37 °C for 19 s to 5 min and prepared cryo-EM grids at the selected temperature for data acquisition. For grids of PRC with DNA0, the purified sample (0.4 mg ml^{-1}) was loaded on C-flat CF-1.2/1.3-4C holey carbon grids (3 μl sample on each grid) at 22 °C or 37 °C at 100% humidity, blotted for 4 s, and flash-frozen in liquid ethane in a Vitrobot. To prepare grids of NFC with DNA1 or DNA2, the samples (0.4 mg ml^{-1}) were first incubated at 22 °C or 37 °C from 15 s to 5 min and then loaded, blotted and frozen on QUANTIFOIL R 1.2/1.3 (Cu, 300 mesh) grids. The frozen grids were stored in liquid nitrogen before use.

For structure determination, the frozen grids of the E962Q samples were loaded into a Titan Krios electron microscope operated at 300 kV for automated image acquisition with Leginon 3.1[38] at University of California, Los Angeles (UCLA). Movies were recorded on a Gatan K2 Summit direct electron detector using the super-resolution mode at 130k nominal magnification (calibrated pixel size of 1.07 Å at the sample level, corresponding to 0.535 Å in super-resolution mode) and defocus values ranging from −1.4 to −3.0 μm. During data collection, the total dose was 57 e^-/A^2 on E962Q mutant PRC with DNA0, 42 e^-/A^2 on E962Q mutant NFC with DNA1 and 50–60 e^-/A^2 on E962Q mutant NFC with DNA2. Details of the collection statistics are shown in Table 1.

To determine if the DNA zipper structure exists in WT NFC, the WT mRAG–DNA0 complex was prepared by mixing purified WT mRAG and DNA0 and incubating it at 37 °C for 5 or 30 min before freezing on QUANTIFOIL R 1.2/1.3 (Cu, 300 mesh) grids. Cryo-EM data were collected on a Titan Krios electron microscope operated at 300 kV using the SerialEM program at the Multi-Institute Cryo-EM Facility (MICEF) of the National Institutes of Health. Videos were recorded on a Gatan K2 Summit direct electron detector using the super-resolution mode at 130k nominal magnification (calibrated pixel size of 1.06 Å at the sample level, corresponding to 0.53 Å in super-resolution mode), with a total dose of 45 e^-/A^2 and defocus values ranging from −1.2 to −3.0 μm.

Structure determination and model refinement. All frames in each collected video were aligned and summed to generate both dose-weighted and dose-unweighted micrographs using Motioncorr2[39]. The latter were only used for defocus determination. Particles on dose-weighted micrographs were picked using Gautomatch (developed by K. Zhang; https://www.mrc-lmb.cam.ac.uk/kzhang/Gautomatch) and extracted in RELION-2.1 using a box size of 280 × 280 pixels[40]. Using the extracted particles, initial maps were obtained with cryoSPARC[41], and then served as the reference for template-based particle picking in Gautomatch and three-dimensional (3D) classification in RELION[42]. Two-dimensional classification and 3D classification were used to remove contamination and screen for the most homogeneous particles used for in-depth 3D structural analyses. The criteria for selection are integrity or completeness of the protein–DNA complexes and well-resolved protein secondary structures and DNA helices. In the mixed PRC and NFC structures of E962Q–DNA1 complexes (PDB 6OEP and 6OEQ, Table 1), the half with the untwisted RSS DNA is superimposable with the pure NFC structures made of DNA1 or DNA2 (PDB 6OEO and 6OER), and the other half is superimposable with the pure PRC structures made of DNA0 or DNA1 (PDB 6OEM and 6OEN). In all PRC and NFC structures, tracings of HMGB1 are similar to but less complete than those in the HFC structures, probably due to the reduced resolution.

For WT mRAG complexed with DNA0 and E962Q with DNA1, a masked 3D classification without alignment was applied on either the12RSS-side or 23RSS-side to classify different conformations from the datasets[43] (Extended Data Figs. 2 and 5). Among the WT RAG–DNA0 complexes, only ~10% of particles contained one untwisted RSS DNA (Extended Data Fig. 5). From the merged NFC particles after initial 3D classification, however, a pure NFC state was isolated by further classification of the 12RSS- and 23RSS-half separately, and averaging of the two NFC halves. We obtained a 3.7-Å resolution averaged map containing a pure NFC state on one half and a mixed state of NFC and PRC on the other half of RAG complexes. The 3.6-Å resolution map for model building of the NFC half was generated by using a soft mask covering the target region during postprocessing in RELION. The trimmed mutant NFC_DNA1 structure (6OEO) was used as the initial model for the WT NFC structure, containing only one RAG1–RAG2 heterodimer and one RSS DNA.

All reported resolutions are based on the 'gold standard' refinement procedure and the 0.143 Fourier shell correlation (FSC) criterion[44]. Local resolution was estimated using Resmap[45]. For model building, we used the reported 3.15-Å resolution PRC and 2.75-Å resolution HFC crystal structures as initial models to build cryo-EM structures of PRC and NFC, respectively. We first fit the coordinates into the cryo-EM map using Chimera, and then manually adjusted and rebuilt the model according to the cryo-EM density in COOT[46]. Phenix real-space refinement was used to refine the model. MolProbity and EMRinger[47] were used to validate the final model. The refinement statistics are shown in Table 1. The detailed classifications and map qualities of mRAG complexed with DNA1, DNA2 and DNA0 are provided in the Supplementary Information (Extended Data Figs. 2, 3, 5 and 7, respectively). The r.m.s.d. of 878 pairs of Cα atoms (one mRAG heterodimer, excluding the NBD domain) between the WT–DNA0 and E962Q–DNA1 NFC complexes is 0.6 Å, which is similar to the r.m.s.d. between the NFC states of DNA1 and DNA2 bound to E962Q mutant RAG (912 pairs of Cα atoms in one mRAG heterodimer and 0.7 Å).

To determine the NFC and PRC populations of mRAG complexed with DNA0, DNA1 or DNA2 at 22 °C or 37 °C, different cryo-EM datasets were collected on either a 300 kV Titan Krios electron microscope or a 200 kV FEI Tecnai F20 electron microscope equipped with a Gatan K2 Summit direct electron detector. Motion correction, contrast transfer function estimation and particle picking were done as described above. 2D classifications were done first to remove obvious contaminants. The selected particles from 2D classification were used to refine an initial model generated from cryoSPARC. Because of a substantial positional change of the ZnH2 domain between the pre-reaction and nick-forming state and the 180° rotation of DNA, which results in a switch between major and minor groove, it is easy to distinguish NFC from PRC even in moderate to low-resolution cryo-EM reconstructions. Masked 3D classifications as described above were used to classify PRC or NFC conformations on both the12 and 23RSS sides. The percentages of PRC or NFC were then counted, with the sum of the two being 100%.

Molecular simulation of the untwisted DNA. Molecular simulations of the DNA duplex with sequence 5′-ACACAG-3′ were carried out using the GROMACS 5.1.4 simulation code[48] in combination with the PLUMED 2.4.3 plug-in[49]. Simulations were run with either the Amber 14 DNA force field with explicit TIP3P water[24] or with the CHARMM 36 DNA force field with the modified CHARMM TIP3P water[25]. Sodium and chloride ions were added to a total ionic strength of ~100 mM and so that the net charge of the system was zero. Simulations were run using periodic boundary conditions, with a 6.5 nm truncated octahedron cell. Lennard–Jones interactions were treated with a twin-range cutoff with inner and outer radii of 0.9 and 1.4 nm, respectively, while the electrostatic energy and forces were calculated via the particle mesh Ewald method with a grid spacing of 0.12 nm. All bonds were fixed in length using the LINCS constraint algorithm, and the equations of motion were integrated via a leapfrog algorithm with a 2 fs time step. The temperature was kept constant with a velocity rescaling thermostat[50], while a Parrinello–Rahman barostat[51] was used to maintain the average pressure at 1 bar. Before starting each simulation, a short steepest descent energy minimization was run to relieve any close contacts introduced during the set-up process.

Unbiased simulations were run starting from either the published structure with bases flipped out (PDB 6DBR) or from the DNA zipper structure determined in the present work, and were run at constant temperature and pressure for ~120 ns with both the Amber 14 and CHARMM 36 force fields. To maintain the DNA distortion, position restraints were applied to the terminal residues on each strand with a force constant of 1,000 kJ mol^{-1} nm^{-2} in each dimension. Otherwise, the dynamics of the interior residues was completely unrestrained.

A second set of simulations was run by applying a twisting and stretching force to the same DNA duplex in an initially canonical B-DNA structure. The bias was applied only to the terminal residues of each strand, by defining a distance matrix r.m.s.d.[52] coordinate comprising all the heavy atoms of the terminal residues, relative to the experimental structure. This coordinate has a minimum when the relative positions of these atoms are the same as in the experimental structure. A time-dependent umbrella bias of the form $V_{\mathrm{umb}}(t) = 1/2k_{\mathrm{umb}}(d_{\mathrm{r.m.s.d.}}(t) - d_0(t))^2$ was employed, with the target value of the coordinate $d_0(t)$ being linearly reduced from its value in the initial structure to zero over the course of the 40 ns biased simulation. A force constant k_{umb} of 10,000 kJ mol^{-1} nm^{-2} was used.

Determination of the number of divalent cations in the active site of PRC. Crystals of WT core mRAG1–RAG2 (aa 384–1008 and 1–359, respectively) complexed with a nicked 12RSS and intact 23RSS were grown as previously described[9]. The complex was assembled in purification buffer containing 1 mM Ca^{2+}. Dehydration and Mn^{2+} soaking of the crystals, X-ray diffraction data collection and processing were carried out as described previously[9]. Searching for Mn^{2+} and Zn^{2+} was perform using AUTOSOL[53,54] based on the anomalous diffraction data, RAG sequence and a structural model of RAG recombinase (PDB 6CIM) in the absence of metal ions and HMGB chains. Finally, two Zn^{2+} and two Mn^{2+} were found in each RAG molecule comprising two RAG1 and two RAG2 subunits (Extended Data Fig. 8b).

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The accession numbers for the cryo-EM structures and associated density maps of the mouse PRC and NFC complexes reported in this paper have been deposited to the PDB and EMDB under accession codes PDB 6OEM to 6OER and 6V0V and EMD-20030 to EMD-20035, EMD-20038, EMD-20039 and EMD-21003, as specified in Table 1.

References

37. Grundy, G. J. et al. Initial stages of V(D)J recombination: the organization of RAG1/2 and RSS DNA in the postcleavage complex. *Mol. Cell* **35**, 217–227 (2009).
38. Suloway, C. et al. Automated molecular microscopy: the new Leginon system. *J. Struct. Biol.* **151**, 41–60 (2005).
39. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
40. Fernandez-Leiro, R. & Scheres, S. H. W. A pipeline approach to single-particle processing in RELION. *Acta Crystallogr. D Struct. Biol.* **73**, 496–502 (2017).
41. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
42. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
43. Bai, X. C., Rajendra, E., Yang, G., Shi, Y. & Scheres, S. H. Sampling the conformational space of the catalytic subunit of human γ-secretase. *Elife* **4**, e11182 (2015).
44. Swint-Kruse, L. & Brown, C. S. Resmap: automated representation of macromolecular interfaces as two-dimensional networks. *Bioinformatics* **21**, 3327–3328 (2005).
45. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **11**, 63–65 (2014).
46. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
47. Barad, B. A. et al. EMRinger: side chain-directed model and map validation for 3D cryo-electron microscopy. *Nat. Methods* **12**, 943–946 (2015).
48. Páll, S., Abraham, M. J., Kutzner, C., Hess, B. & Lindahl, E. in *Solving Software Challenges for Exascale* (eds Markidis, S. & Laure, E.) 3–27 (Lecture Notes in Computer Science 8759, Springer, 2015).
49. Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C. & Bussi, G. PLUMED2: new feathers for an old bird. *Comput. Phys. Commun.* **185**, 604–613 (2014).
50. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101 (2007).
51. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190 (1981).
52. Domanski, J., Sansom, M. S. P., Stansfeld, P. J. & Best, R. B. Balancing force field protein–lipid interactions to capture transmembrane helix–helix association. *J. Chem. Theory Comput.* **14**, 1706–1715 (2018).
53. Grosse-Kunstleve, R. W. & Adams, P. D. Substructure search procedures for macromolecular structures. *Acta Crystallogr. D Biol. Crystallogr.* **59**, 1966–1973 (2003).
54. McCoy, A. J., Storoni, L. C. & Read, R. J. Simple algorithm for a maximum-likelihood SAD function. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 1220–1228 (2004).

Acknowledgements

This research was supported by the National Institute of Diabetes and Digestive and Kidney Diseases (M.G., DK036167; W.Y., DK036147 and DK036144; Z.H.Z., GM071940). The authors acknowledge the use of instruments at the Electron Imaging Center for NanoMachines supported by the NIH (1S10RR23057, 1S10OD018111 and U24GM116792), NSF (DBI-1338135 and DMR-1548924) and CNSI at UCLA.

Author contributions

X.C. carried out all experiments and structure determination. Y.C. collected cryo-EM micrographs on the Krios microscope at UCLA and helped with structure determination and refinement. H.W. helped with cryo-EM data collection on the TF20 and Krios at NIH. R.B.B. carried out molecular dynamics simulations. Z.H.Z., W.Y. and M.G. supervised the research project. X.C., R.B.B., W.Y. and M.G. prepared the manuscript.

Competing interests

The authors declare no competing interests.

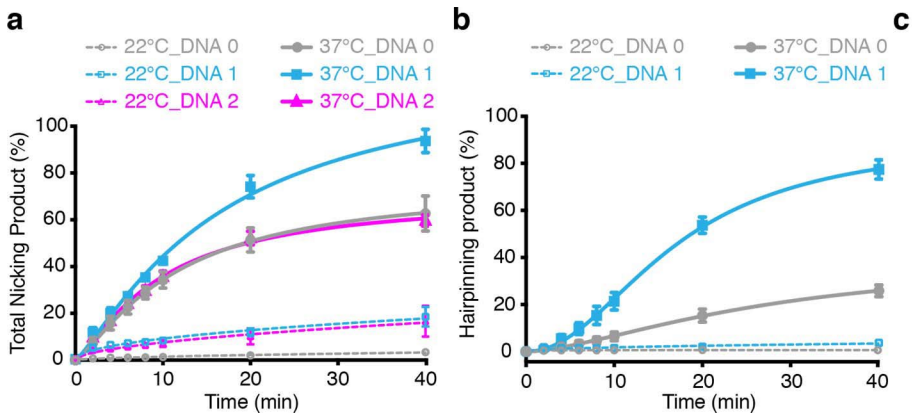
Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41594-019-0363-2.

Supplementary information is available for this paper at https://doi.org/10.1038/s41594-019-0363-2.

Correspondence and requests for materials should be addressed to W.Y. or M.G.

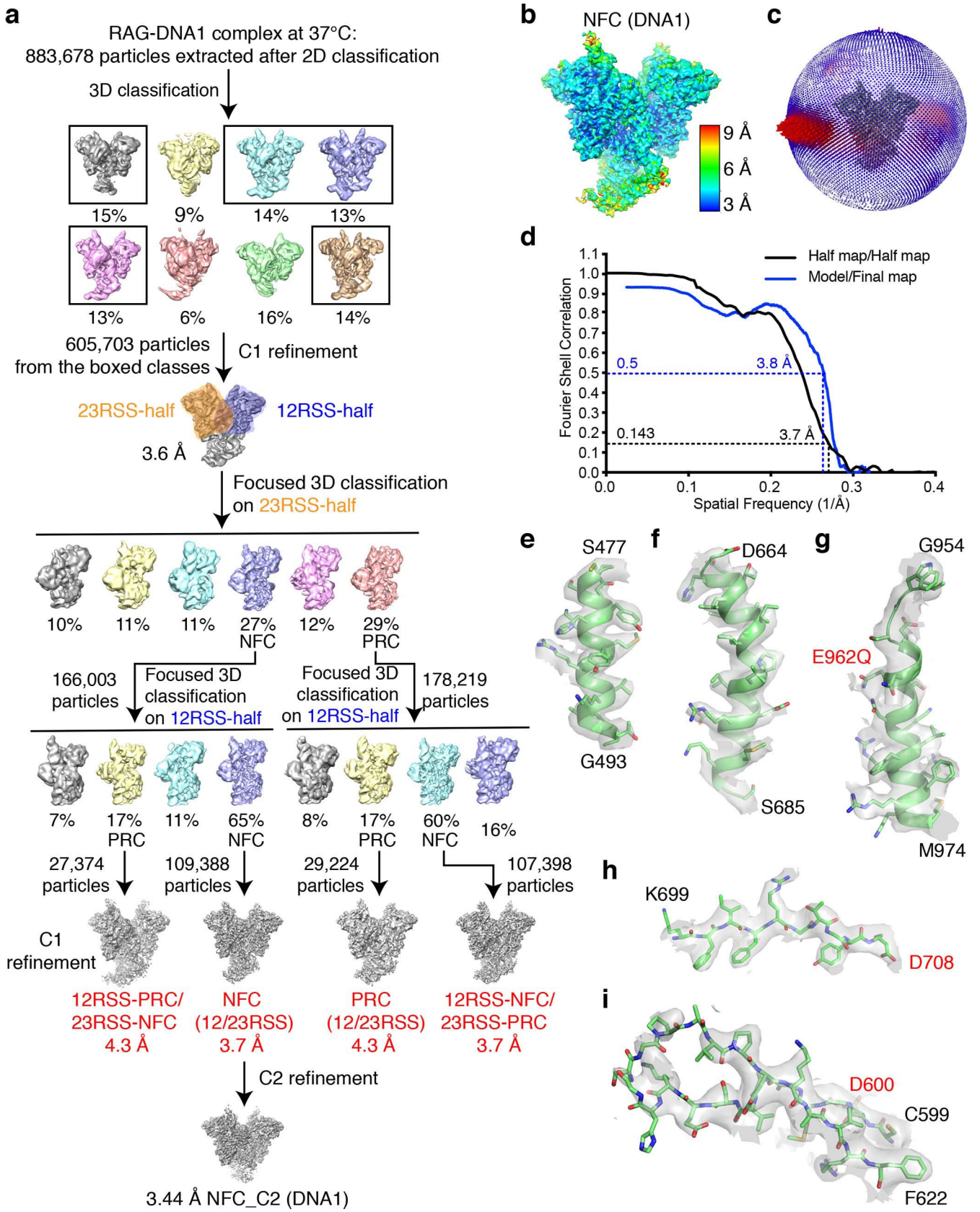
Peer review information Anke Sparmann was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.



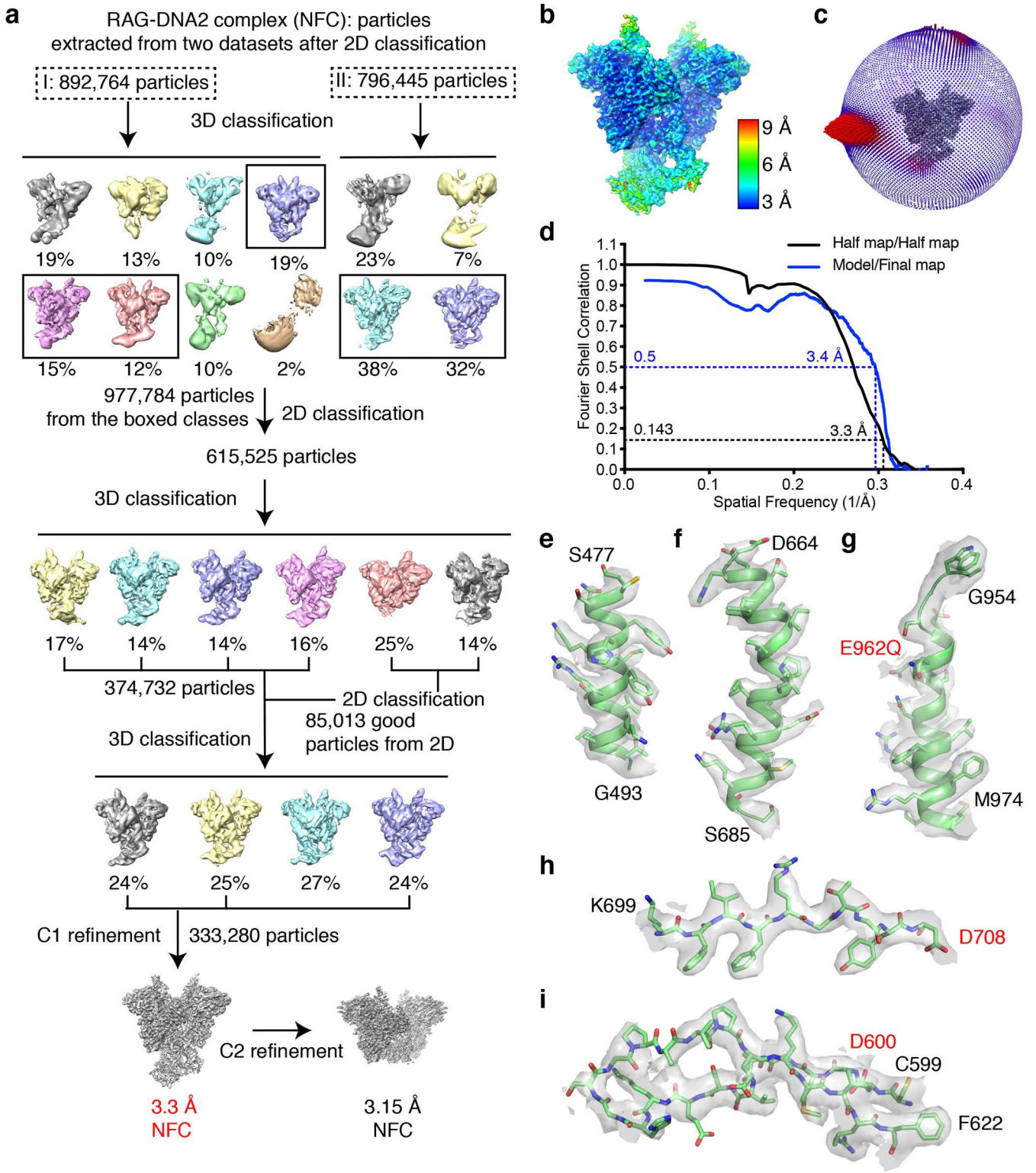
c

NFC/PRC	DNA 0	DNA 1	DNA 2
22°C (%)	0 / 100	10 / 90*	100 / 0
37°C (%)	0 / 100*	62 / 38	—

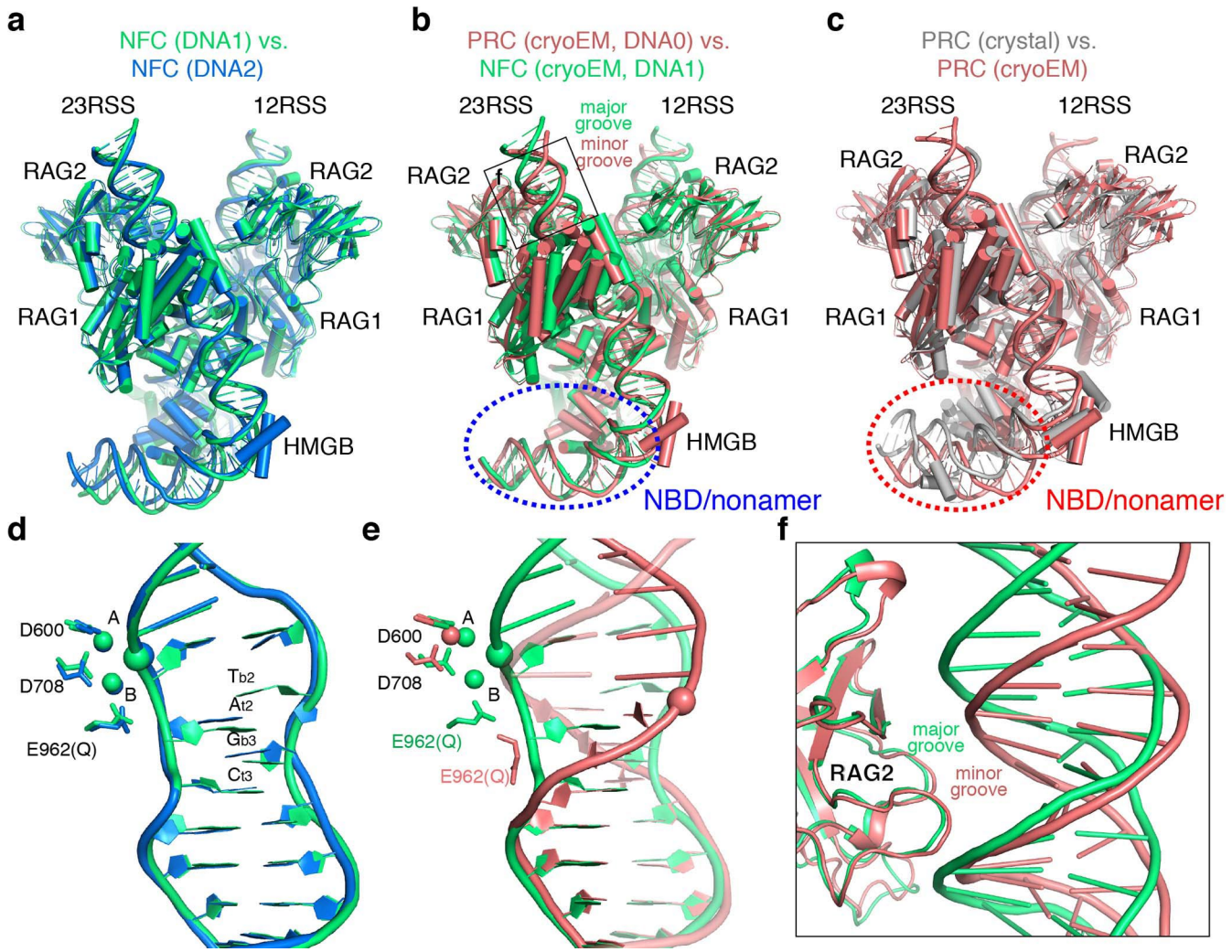
Extended Data Fig. 1 | Cleavage and cryo-EM analysis of DNA substrates in NFC. a, **b**, Cleavage efficiencies (nicking and hairpinning) of the three DNA variants by WT mRAG at 22 and 37 °C (mean and s.d., n = 3 independent samples). **c**, Percentage of NFC and PRC (NFC/PRC) in cryo-EM 3D classification from samples made of DNA0, DNA1 or DNA2 substrate with E962Q mutant mRAG at 22 and 37 °C. Asterisk (*) indicates that the dataset was collected on a Tecnai F20 electron microscope instead of Titan Krios.



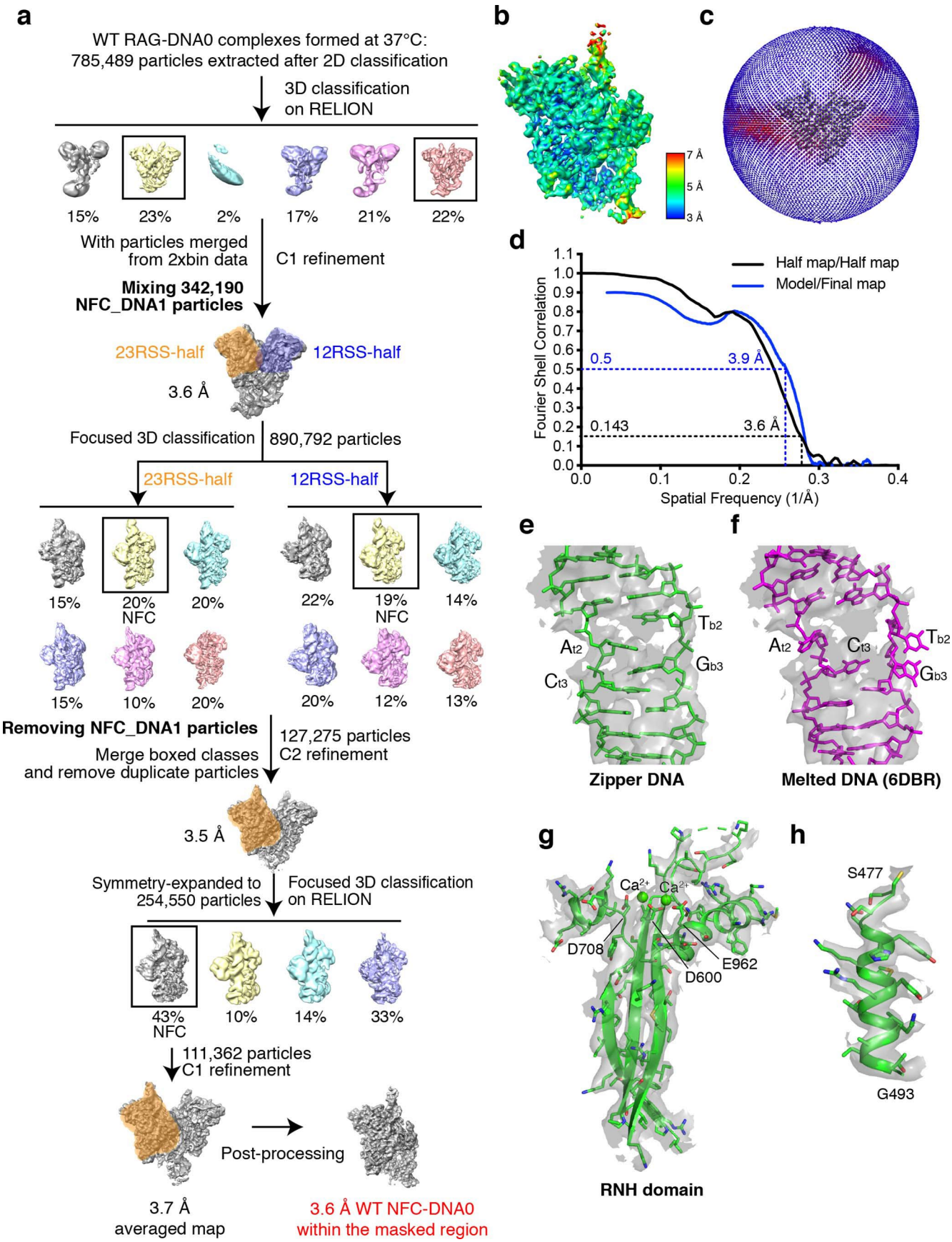
Extended Data Fig. 2 | Structure determination of mouse NFC with DNA1 by cryo-EM. a, Flow chart for cryo-EM data processing of mRAG complexed with DNA1. The maps with red bold letter are used for final model building. **b**, A surface presentation of the 3.7 Å NFC (DNA1) map (C1 symmetry). Colors are according to the local resolution estimated by ResMap, and the color scale bar is shown on its right. **c**, Angular distributions of all particles used for the final three-dimensional reconstruction shown in b. **d**, The FSC curves of the NFC (DNA1) map (C1). The "gold standard" FSC between two independent halves of the map (black line) indicates a resolution of 3.7 Å, and the blue line is the FSC between the final refined model and the final map. **e** to **i**, Representative regions of the C1 map (transparent grey surface). The maps are shown with the final structural models (cartoon or stick) superimposed.



Extended Data Fig. 3 | Structure determination of mouse NFC with DNA2 by cryo-EM. a, Flow chart for cryo-EM data processing of mRAG complexed with DNA2. The maps with red bold letter are used for final model building. **b**, A surface presentation of the 3.3 Å NFC (DNA2) map (C1 symmetry). Colors are according to the local resolution estimated by ResMap, and the color scale bar is shown on its right. **c**, Angular distributions of all particles used for the final three-dimensional reconstruction shown in b. **d**, The FSC curves of NFC (DNA2) map (C1). The "gold standard" FSC between two independent halves of the map (black line) indicates a resolution of 3.3 Å, and the blue line is the FSC between the final refined model and the final map. **e** to **i**, Representative regions of the C1 map (transparent grey surface). The maps are shown with the final structural models (cartoon or stick) superimposed.

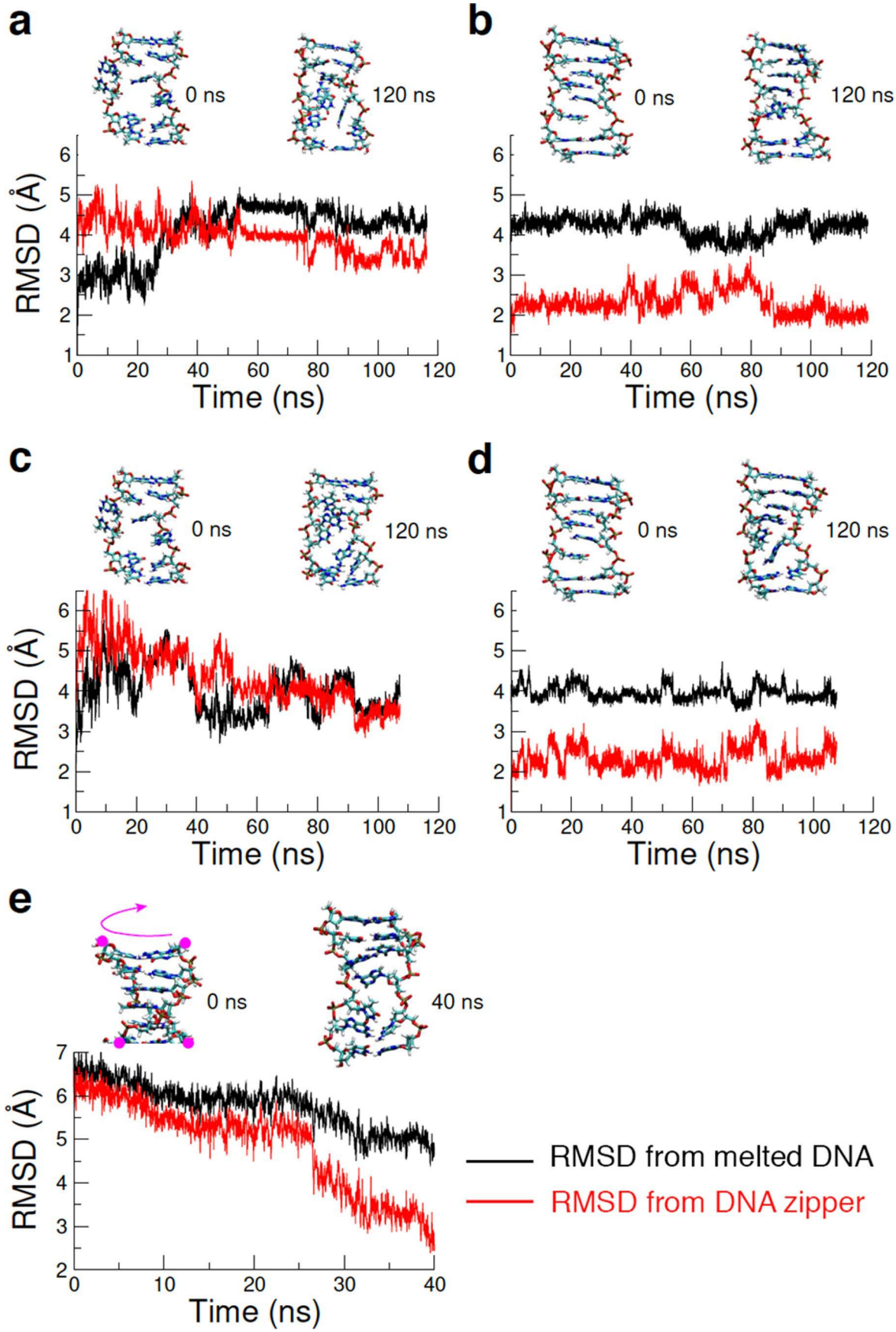


Extended Data Fig. 4 | Structural comparisons of mouse PRC and NFC. a, Cryo-EM structures of NFC with DNA1 (green) and DNA2 (blue) are superimposable. **b**, Comparison of mouse PRC and NFC structures. Superposition of cryo-EM PRC (red) and NFC (DNA1) (green) structures reveals limited NBD and nonamer movement, which is marked with blue dashed circle (right panel). **c**, Superposition of crystal (grey) and cryo-EM (red) PRC structures reveals the different NBD and nonamer region (circled in red dashes) due to crystal-lattice contacts. **d, e**, The zoom-in views of the active center and DNA distortions in the superimposed structures shown in a-b. The catalytic DDE motif and two metal ions (a and b) are labeled; the heptamer of RSS DNA is shown in detailed cartoon presentation; the scissile phosphate of top strand is marked by a large ball. In panel d, bases forming the DNA zipper are labeled. **f**, A zoom-in view of boxed area in panel b. RAG2 interacts with the minor groove in PRC or the major groove in NFC.



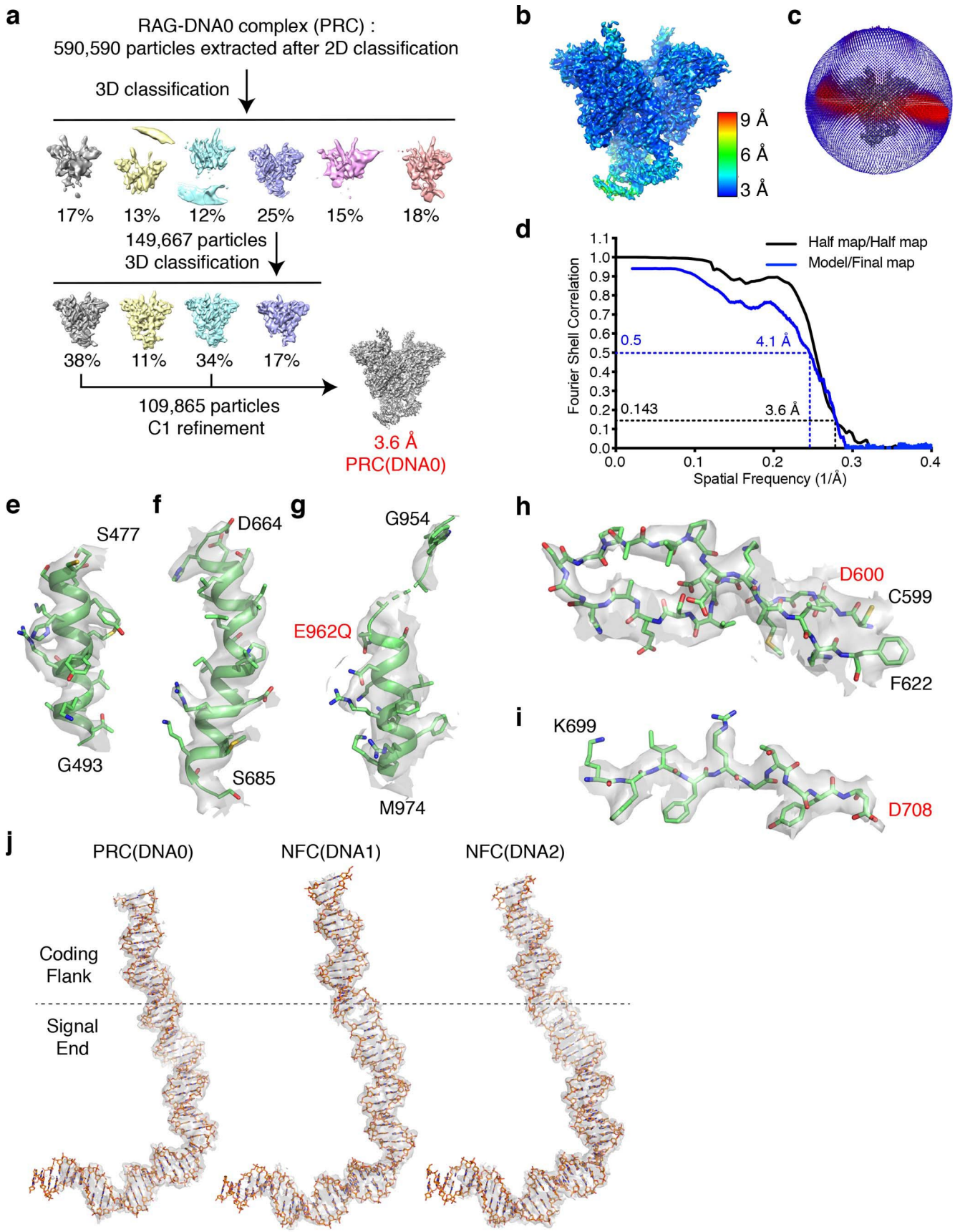
Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Structure determination of mouse WT NFC with DNA0 by cryo-EM. a, Flow chart for cryo-EM data processing of WT mRAG complexed with DNA0. Data processing was done using RELION. The 3.6Å map labeled in red was used for final model building. **b**, A surface presentation of the 3.6 Å map of WT NFC (DNA0). Colors are according to the local resolution estimated by ResMap, and the color scale bar is shown on its right. **c**, Angular distributions of all particles used for the final three-dimensional reconstruction. **d**, The FSC curves of WT NFC (DNA0) map. The "gold standard" FSC between two independent halves of the map indicates an overall resolution of 3.6 Å. **e**–**h**, Representative regions of the map (transparent grey surface). The refined zipper DNA fits the map better (**e**) than the melted DNA (PDB: 6DBR) (**f**).



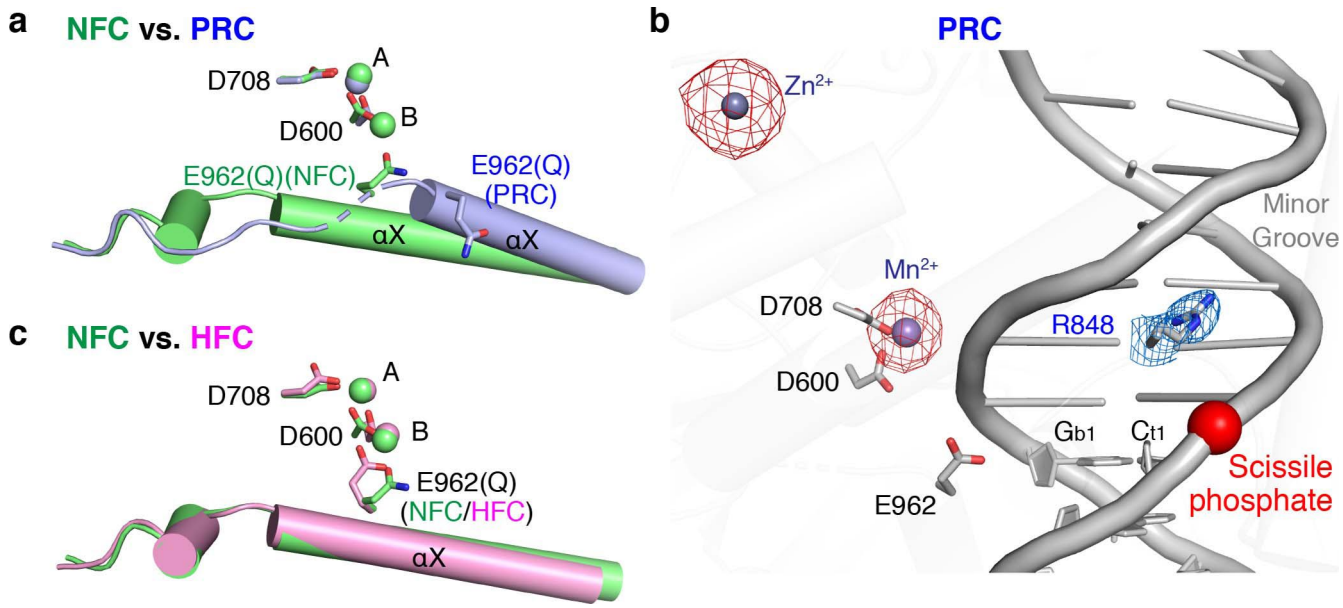
Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Molecular simulations of the untwisted region of DNA. a, **b**, Unbiased simulations were run starting from the base-flipped out structure (PDB 6DBR) (a), and the zipper structure (**b**) using the Amber 14 force field. Plotted in each case are the all-atom RMSD to the base-flipped out structure (black) and the zippered structure (red). The structures at the start and end of each run are shown above the RMSD plots. **c**, **d**, The analogous results are given for simulations with the CHARMM 36 force field. **e**, Biased simulations were run from a canonical B-DNA form, in which the terminal residues were driven to mimic the stretched and untwisted DNA observed in the mouse and zebrafish NFC structures. The RMSD and initial and final structures are shown as before.



Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Structure determination of mouse PRC with DNA0 by cryo-EM. a, Flow chart for the cryo-EM data processing of mRAG complexed with DNA0. The maps with red bold letter are used for final model building. **b**, A surface presentation of the 3.6 Å NFC map (C1 symmetry). Colors are according to the local resolution estimated by ResMap, and the color scale bar is shown on its right. **c**, Angular distributions of all particles used for the final three-dimensional reconstruction. **d**, The FSC curves of PRC (DNA0) map (C1). The "gold standard" FSC between two independent halves of the map (black line) indicates a resolution of 3.6 Å, and the blue line is the FSC between the final refined model and the final map. **e** to **i**, Representative regions of the C1 map (transparent grey surface). The maps are shown with the final structural models (cartoon or stick) superimposed. **j**, The maps of 23RSS in PRC (DNA0), NFC (DNA1) and NFC (DNA2).



Extended Data Fig. 8 | Remodeling of the active site in NFC and HFC. a, Repositioning of αX in the RNH domain during PRC to NFC transition (both are cryo-EM structures). E962 is far from the active site in PRC (light blue) but is positioned for catalysis in NFC (green). **b**, Anomalous X-ray scattering of the PRC crystals confirms that one Mn^{2+} and one Zn^{2+} are bound to each RAG1 subunit. The anomalous map is contoured at 3σ in red. The blue 2Fo-Fc map (contoured at 1σ) highlights R848, which is buried in the minor groove. **c**, The reconfigured E962 in HFC (pink) after the first DNA cleavage by nicking (green).

nature research

Corresponding author(s): Wei Yang, Martin Gellert

Last updated by author(s): Nov 20, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Legion, Serial EM, Gatan Digital Micrograph
Data analysis	Relion, Gautomatch, CTFFIND4, MotionCor2, EMAN2, Phenix, Coot, UCSF Chimera, Pymol2

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The cryo-EM maps and PDB coordinates are deposited under accession numbers:
EMD-20030, PDB-6OEM; EMD-20031, PDB-6OEN; EMD-20032, PDB-6OEO; EMD-20035, PDB-6OER; EMD-20033, PDB-6OEP; EMD-20034, PDB-6OEQ; EMD-20038; EMD-20039; EMD-21003, PDB-6V0V.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical methods were used to predetermine sample size, not applicable
Data exclusions	No data excluded
Replication	All replicates were successful
Randomization	Animals and humans were not used in the study, not applicable
Blinding	Animals and humans were not used in the study, blinding not applicable

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	HEK293T was purchased from Thermo Fisher Scientific and maintained as Yang laboratory stock.
Authentication	Not authenticated.
Mycoplasma contamination	Not tested
Commonly misidentified lines (See ICLAC register)	No misidentified lines used.