Expanded View Figures

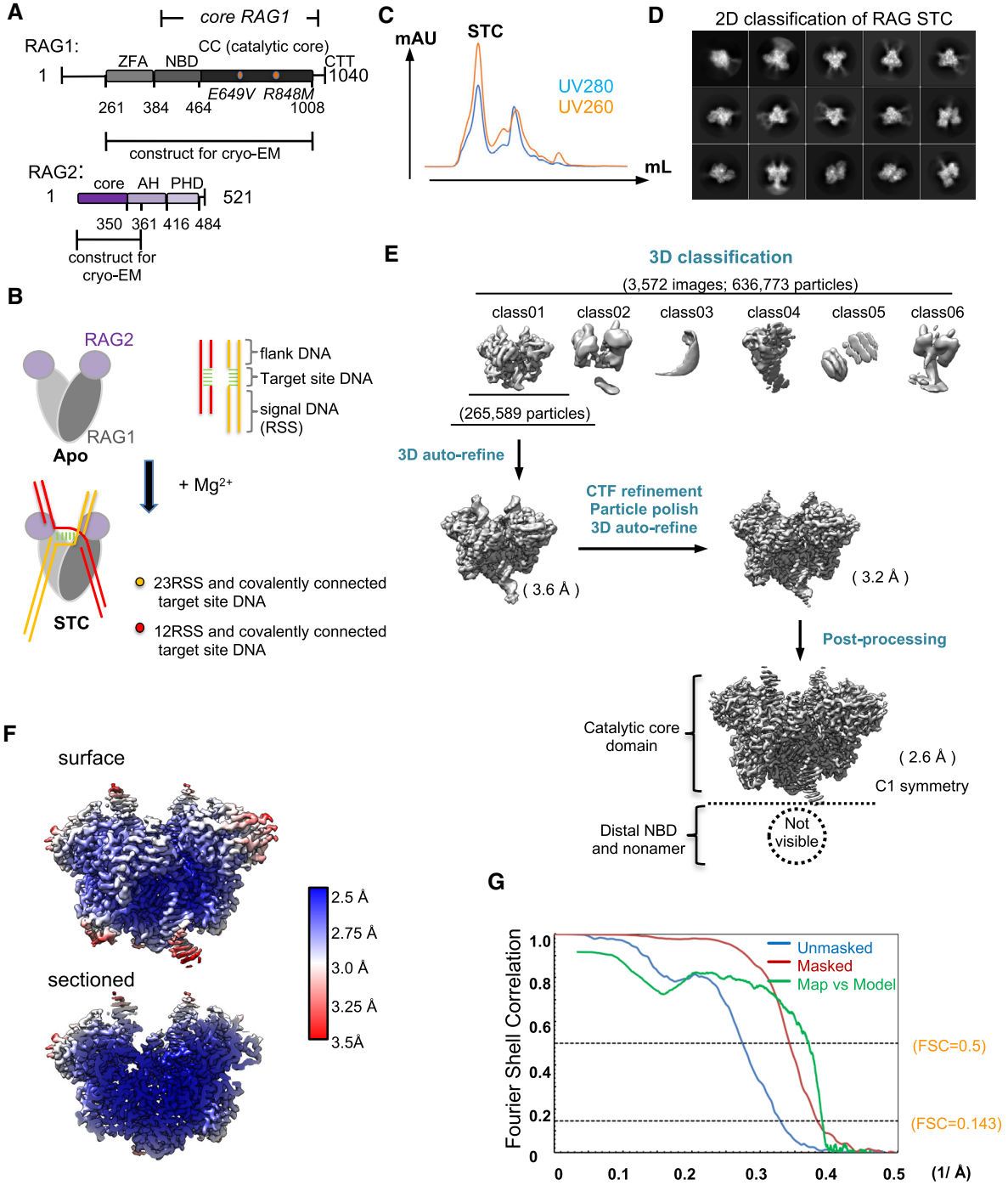

Figure EV1.

Figure EV1. Assembling the RAG STC and cryo-EM structure determination.

A Schematic diagram of RAG1 and RAG2 proteins with domains indicated with shading and colors. The portions of the proteins used for STC assembly and cryo-EM data collection are indicated; the RAG1 protein contained R848M and E649V mutations (pink circles), which activate transposition. ZFA, zinc finger A domain; NBD, nonamer binding domain; CC catalytic core; CTT, C-terminal tail; AH, acidic hinge; PHD, plant homeodomain.

B Schematic of the RAG STC and the process of complex assembly. The names of the different portions of the DNA as used throughout are indicated. RAG1, RAG2, and HMGB1 (not shown) were incubated together in the presence of Mg^{2+} with a 12RSS-containing DNA substrate (red) and a 23RSS-containing DNA substrate (yellow) to assemble the STC.

C Gel filtration profile shows the purification of the STC.

D 2D class averages from cryo-EM particles of RAG STC.

E Flowchart of cryo-EM map reconstruction. Six 3D classes were generated after 3D classification. After multiple steps of data processing as indicated, a 2.6 Å map with C1 symmetry was obtained. The expected position of the NBD/nonamer, which could not be visualized, is indicated with a dashed circle in the final 2.6 Å map.

F Color-code resolution estimation of RAG STC reconstruction map viewed on the surface and sectioned.

G Unmasked and masked Fourier shell correlation curves, as well as the model versus final map, as indicated. Resolution is read by the cutoff value at FSC = 0.143.

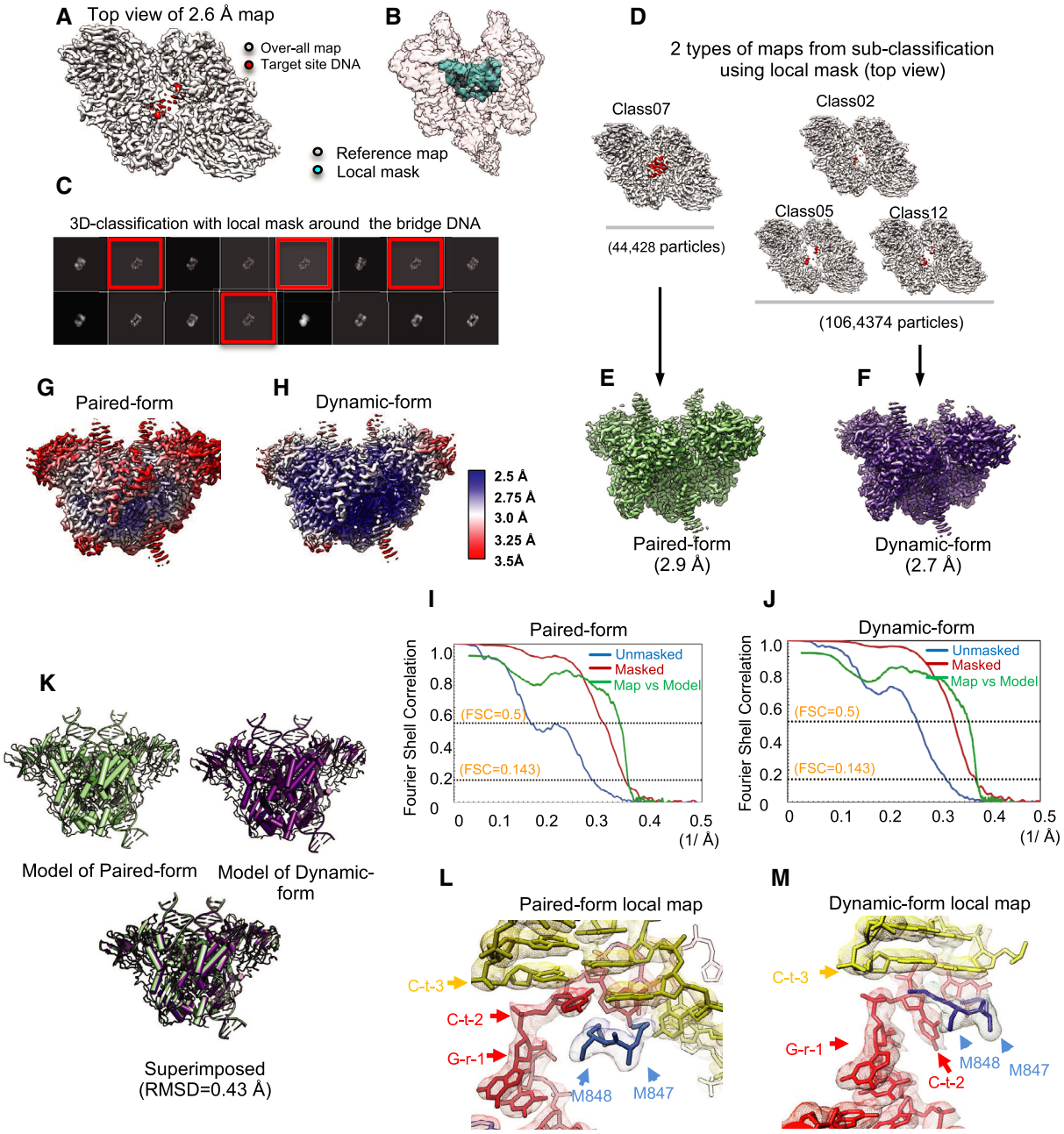

Figure EV2. Cryo-EM map of paired-form and dynamic-form RAG STC.

A Cryo-EM map of RAG STC (top view) with target site DNA density indicated in red. Target site DNA exhibits weaker density than other portions of the map.

B The mask identifying the region surrounding target site DNA (cyan) is indicated together with the RAG STC map (transparent).

C Sixteen classes from 3D classification were generated with the local mask. Four classes with better features were selected for further data processing.

D–F Cryo-EM map reconstruction of paired-form and dynamic-form RAG STC. Four 3D cryo-EM maps (top view), with target site DNA density indicated in red as in (A), were separated into two groups based on the quality of target site DNA density. The first group, with strong target site DNA density, consisted of class 07. The other three classes with weaker target site DNA density were combined for further processing. This yielded a 2.9 Å paired RAG STC map (E) and a 2.7 Å dynamic RAG STC map (F), which were used for data analysis and model building.

G–J Color-code resolution estimation maps (G and H) and unmasked and masked Fourier shell correlation curves, as well as the model versus final map (I and J) for the paired RAG STC and the dynamic RAG STC depicted as in Fig EV1E and F.

K Atomic models of paired and dynamic RAG STC structures, as well as superposition of the two structures. The two structures are extremely similar (RMSD of 0.43 Å).

L, M Cryo-EM map density in the vicinity of the integration site for the paired (L) and dynamic (M) RAG STCs, with models depicted within the maps. DNA nomenclature defined in Fig EV3E.

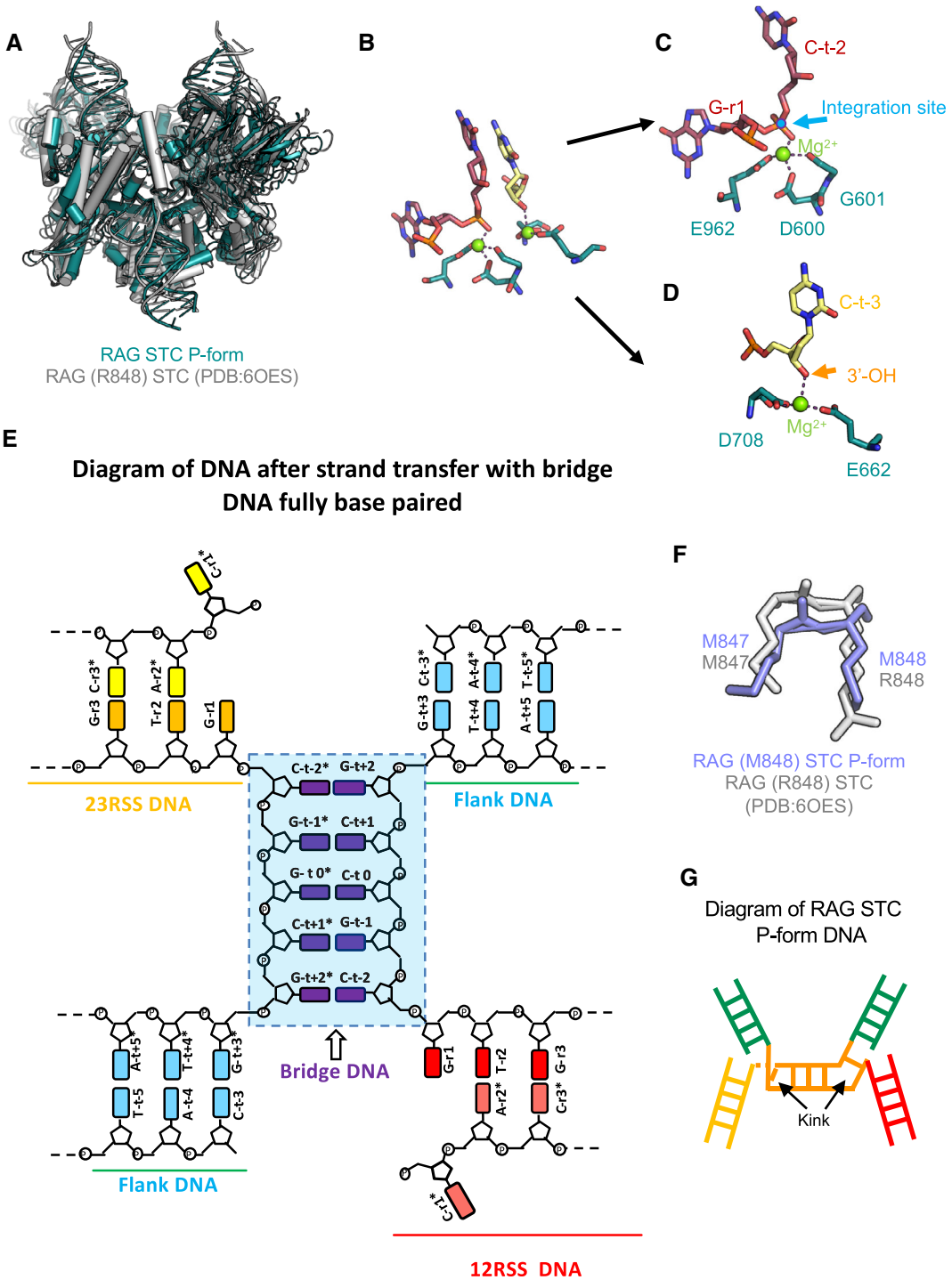

Figure EV3.

◀ **Figure EV3. Structural details of paired RAG STC.**

A Superimposition of the paired RAG STC structure determined in this study and the RAG STC structure determined with RAG1 containing R848 (Chen *et al*, 2020b).

B–D Structure of the reaction center of the paired RAG STC, depicted in intact form in (B) and as individual regions surrounding each Mg^{2+} ion (C and D) to better depict metal ion coordination. Cyan sphere, integration site phosphate; green spheres, Mg^{2+} ions. In (C), the Mg^{2+} ion is coordinated by the phosphate of C-t−2, D600, G601, and E962. In (D), the second Mg^{2+} is coordinated by the 3′-OH from flank DNA, D708, and D662. DNA nomenclature defined in (E).

E Schematic diagram of DNA after strand transfer with target site DNA fully base-paired. Yellow and red, 23RSS and 12RSS, respectively; purple, target site DNA; light blue, flank DNA. The nomenclature used for each base is indicated.

F Superimposition of M847 and M848 from the paired RAG STC and M847 and R848 from the RAG STC structure determined with RAG1 containing R848 (Chen *et al*, 2020b).

G Schematic diagram of strand transfer DNA in the paired RAG STC. The major kinks in target DNA exist one base pair in from the edges of the 5 bp target site DNA.

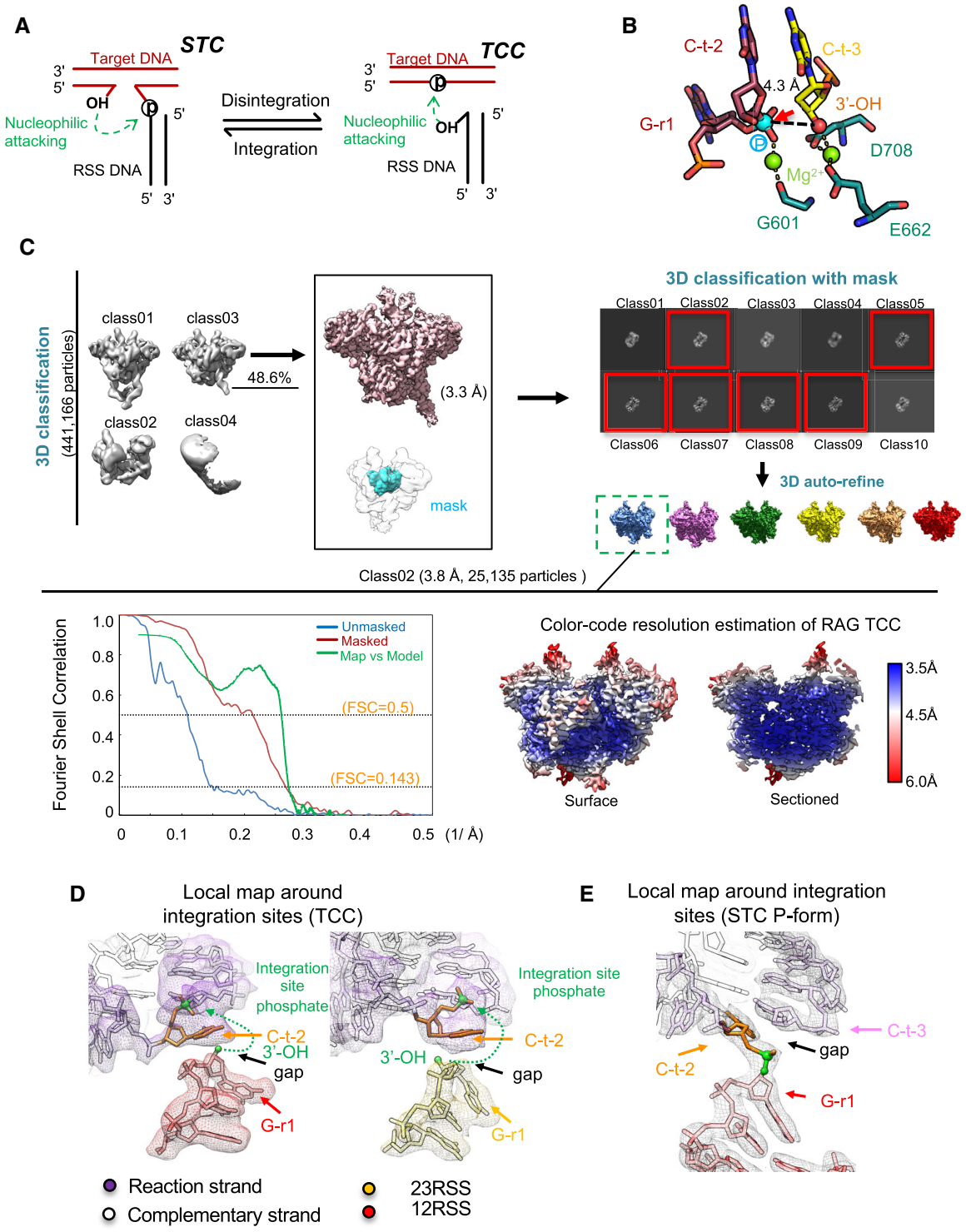

Figure EV4.

Figure EV4. The RAG target capture complex (TCC).

A Schematic depicting the integration and disintegration reactions. The free 3′-OH from the RSS DNA or flank DNA serves as the nucleophile that attacks the phosphate in either target DNA or strand transfer DNA.

B The reaction center of the paired RAG STC showing the distance between the 3′-OH from C-t−3 and the disintegration scissile phosphate of C-t−2. The 3′-OH and phosphate are constrained by coordinating Mg^{2+} ions. The red arrow indicates the direction for in-line nucleophilic attack for the disintegration reaction.

C Flowchart of cryo-EM map reconstruction of RAG TCC. *De novo* data processing beginning with particle picking. Four 3D classes were generated after 3D classification as indicated. Class 03, with the best resolution and most particles, was selected for further processing. After multiple steps of data processing, 3.3 Å maps were obtained. One sub-classification of cryo-EM map was identified to obtain the atomic model of the target capture complex (TCC) by applying the mask surrounding the target site DNA and the integration reaction center. Fourier shell correlation curves and color-code resolution estimation map are at the bottom.

D Cryo-EM map density in the vicinity of the 3′-OH of 23RSS (right) and 12RSS (left) signal DNA of RAG TCC, with models depicted within the maps. The oxygen of 3′-OH and the phosphate at integration site are colored green and are connected by a green dashed arrow. A gap between the density of RSS signal DNA and target DNA is indicated by the black arrow.

E Cryo-EM map density of paired RAG STC at the equivalent position in (D) is shown in the same view. The phosphodiester bond formed between the RSS signal 3′-OH and integration site phosphate is colored green. Black arrow indicates the gap generated in the reaction strand of target DNA after the integration reaction. DNA nomenclature defined in Fig EV3E.

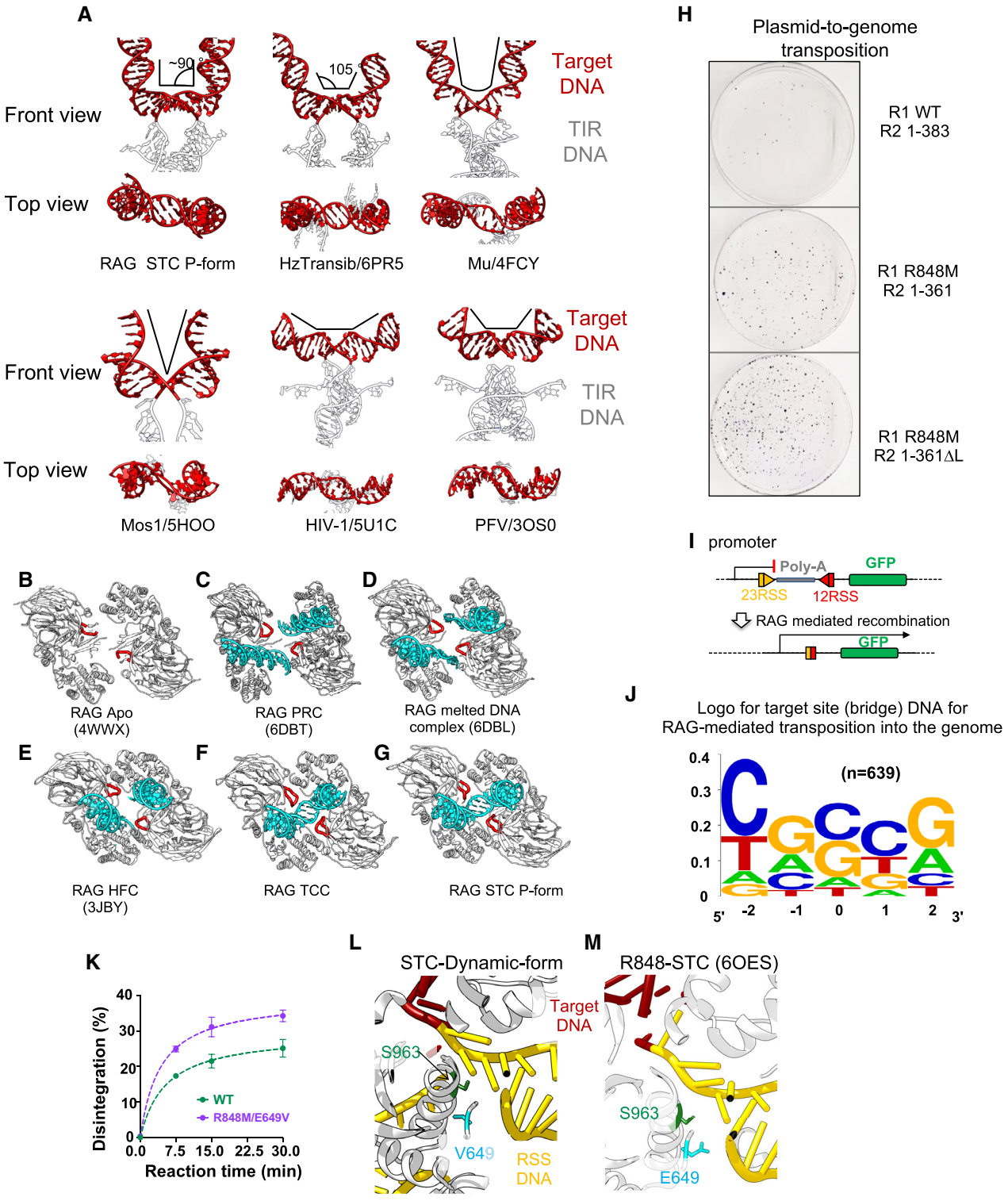

Figure EV5.

Figure EV5. Deletion of four amino acids from the tip of RAG2 loop 333–342 increases transposition efficiency.

A Diverse target DNA conformations and bending angles in transposase and retroviral integrase STC structures. Front and top views of target DNA (red) and RSS/TIR DNA (white) for the paired RAG STC and the indicated transposases or retroviral integrases. Black lines are used to approximate the DNA bends.

B–G Comparison of RAG2 loop 333–342 between different RAG complexes during the cut-and-paste transposition. (B) Free RAG complex (mouse) without the DNA substrate; (C) RAG pre-reaction complex (zebrafish) with intact 12RSS and 23RSS; (D) RAG pre-reaction complex (zebrafish) with intact 12RSS and 23RSS melted at the catalytic center; (E) RAG hairpin-forming complex (zebrafish) with nicked 12RSS and 23RSS; (F) RAG TCC (mouse) with target DNA and 12RSS and 23RSS signal DNA; (G) paired RAG STC (mouse) with 12RSS and 23RSS integrated into target DNA. RAG1 and RAG2 are colored white with RAG2 loop 333–342 highlighted in red. DNA is shown in cyan.

H Example of plates stained with crystal violet to reveal puromycin-resistant colonies from the plasmid-to-genome transposition assay. The RAG2 ΔL mutation (deletion of RAG2 aa 336–339) increases colony numbers, which is a measure of transposition efficiency.

I Schematic of GFP fluorescence recombination assay, used for testing RAG V(D)J recombination activity *in vivo*.

J DNA sequence logo obtained from 5 bp target site duplications from 639 plasmid-to-genome transposition events mediated by RAG1 E649V/R848M and RAG2 1–350 in HEK293T cells (data derived from Zhang *et al*, 2019). Target site duplications exhibit a GC preference (Zhang *et al*, 2019) and a pyrimidine/purine step at the edges.

K Comparison of disintegration activity of WT (green) and RAG1 double mutant (R848M/E649V, purple). Data were fit to Michaelis–Menten equation to draw the curve. Data points represent the mean of 3 biological replicates and are shown ± SEM.

L, M Diagram of the valine or glutamic acid at RAG1 position 649 in the structures of dynamic RAG STC (left) or R848 RAG STC (Chen *et al*, 2020b) (right). S963 and V649 or E649 are shown as stick mode and colored green and cyan, respectively. Protein, target, and signal DNA are colored white, dark-red, and yellow, respectively. V649 points into the protein core but E649 points toward signal DNA.