Target DNA capture, deformation and target-site preference during RAG-mediated cut-and-paste transposition

Jingjing Pang[1], Siyao Jia[1], Eliza C. Martin[2], Xueming Zheng[3], Qingyi Lu[4], David G. Schatz[*2], Yuhang Zhang[*1,5]

https://doi.org/10.15302/vita.2026.10.0082

Dear Editor,

Cut-and-paste DNA transposons drive genome evolution by reshaping genome architecture through transposition and providing molecular components for the emergence of new host functions[1,2]. Transposition begins with cleavage of transposon ends to form a transposon end complex (TEC), followed by target DNA capture to form a target capture complex (TCC), and strand transfer to generate a strand-transfer complex (STC)[3]. The transposon-end 3′-OH groups attack target DNA at two symmetry-related active sites, forming staggered junctions that give rise to a target-site duplication (TSD) (Fig. 1a). However, the transient and heterogeneous nature of target-bound intermediates has limited our understanding of how target DNA is captured, deformed and positioned near the active sites, and how these processes influence target-site preference[4].

Derived from an ancestral RAGL transposase, mouse RAG retains the ancestral catalytic architecture and measurable transposition activity, which is enhanced by RAG1 R848M and deletion of two transposition-inhibitory RAG2 regions[5] (Supplementary Fig. S1a–f). Unless otherwise indicated, all target-capture and strand-transfer experiments used this activated complex. Cryo-EM resolved an unoccupied TEC and, by target-DNA proximity tethering to stabilize transient complexes, two linear TCCs (TCC-L_A and TCC-L_S), two U-shaped TCCs (TCC-U_D and TCC-U_{DX}), and two STCs (STC-1 and STC-2). These structures span the unoccupied TEC, initial linear target-DNA engagement, U-shaped accommodation within the pocket, and post-transfer DNA reordering. DNA-shape machine learning and V_{step} analysis linked target DNA deformability to target-site preference, while evolutionary comparisons and mutagenesis implicated changes in target-DNA engagement and reaction-center interactions in the suppression of transposition during RAG evolution.

Previously reported TEC-like structures contain nonspecific DNA or a RAG2 C-terminal segment in the target-binding pocket[6,7] (Supplementary Fig. S2a, b). We neutralized selected DNA-contacting surface residues and determined the structure of an unoccupied TEC at 3.0 Å resolution. Its catalytic core resembles reported hairpin-forming and strand-transfer complexes; the ~25–35 Å-deep pocket is narrow and enclosed, with RSS ends and catalytic residues at its base (Fig. 1b; Supplementary Figs. S2c–e, S3 and Tables S1, S2). Three-dimensional variability analysis revealed a relatively rigid catalytic core but mobile nonamer-binding domains (NBDs) and distal DNA (Fig. 1c; Supplementary Fig. S4a–d and Video S1). Thus, target DNA likely requires substantial deformation to reach the reaction center without large-scale opening of the catalytic framework.

To examine how target DNA is captured by the RAG TEC and subsequently enters the catalytic pocket, we sought to visualize target-capture states. However, RAG binds free target DNA only weakly, limiting efficient TCC formation (Supplementary Fig. S5a, b). Biochemical experiments showed that RAG catalyzes intramolecular transposition more efficiently than intermolecular transposition using free target DNA, suggesting that substrate proximity to the RAG complex promotes target DNA capture and strand transfer (Supplementary Fig. S5c, d). We therefore developed a target-DNA proximity-tethering system by fusing the sequence-specific zinc-finger protein ZIF268 (ZF) to the N-terminus of truncated RAG2 and introducing a cognate ZF-binding site (ZBS) into the target DNA, thereby positioning target DNA near the transposition-active RAG complex (Fig. 1d; Supplementary Fig. S6a, b and Table S3). By varying the length of the flexible linker between RAG and the ZF, the ZF-RAG system constrained integration to a more defined position and substantially reduced target-site heterogeneity while preserving RAG's 5-bp TSD and GC-rich target preference (Fig. 1e, f; Supplementary Fig. S6c–h). Neither the zinc-finger module nor the linker was resolved in the cryo-EM maps, consistent with substantial conformational flexibility and arguing against the tether constraining target DNA to a single rigid conformation (Supplementary Fig. S7a, b). Together, these results indicate that proximity tethering enhances target DNA capture and reaction efficiency while preserving the major catalytic features of RAG-mediated transposition, thereby providing an efficient strategy for stabilizing RAG–target DNA complexes and reducing variability in target-site usage. We note that the TCC structures reported here show only that the observed conformations are accessible under the stabilizing conditions used, not their relative occupancies or interconversion rates in untethered RAG reactions.

Initial three-dimensional classification of the assembled TCCs separated the particles into two broad categories: one containing extended target DNA outside the pocket and

1. School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China. 2. Department of Immunobiology and Department of Molecular Biophysics and Biochemistry, Yale School of Medicine, New Haven, CT, USA. 3. Department of Biochemistry and Molecular Biology, School of Medicine, Jiangsu University, Zhenjiang, Jiangsu, China. 4. School of Life Sciences, Beijing University of Chinese Medicine, Beijing, China. 5. Sheng Yushou Center of Cell Biology and Immunology, Shanghai Jiao Tong University, Shanghai, China. *Correspondence: David Schatz (david.schatz@yale.edu), Yuhang Zhang (yuhang.zhang@sjtu.edu.cn)

Received: August 6, 2026; Accepted: October 19, 2026; Published: November 2, 2026

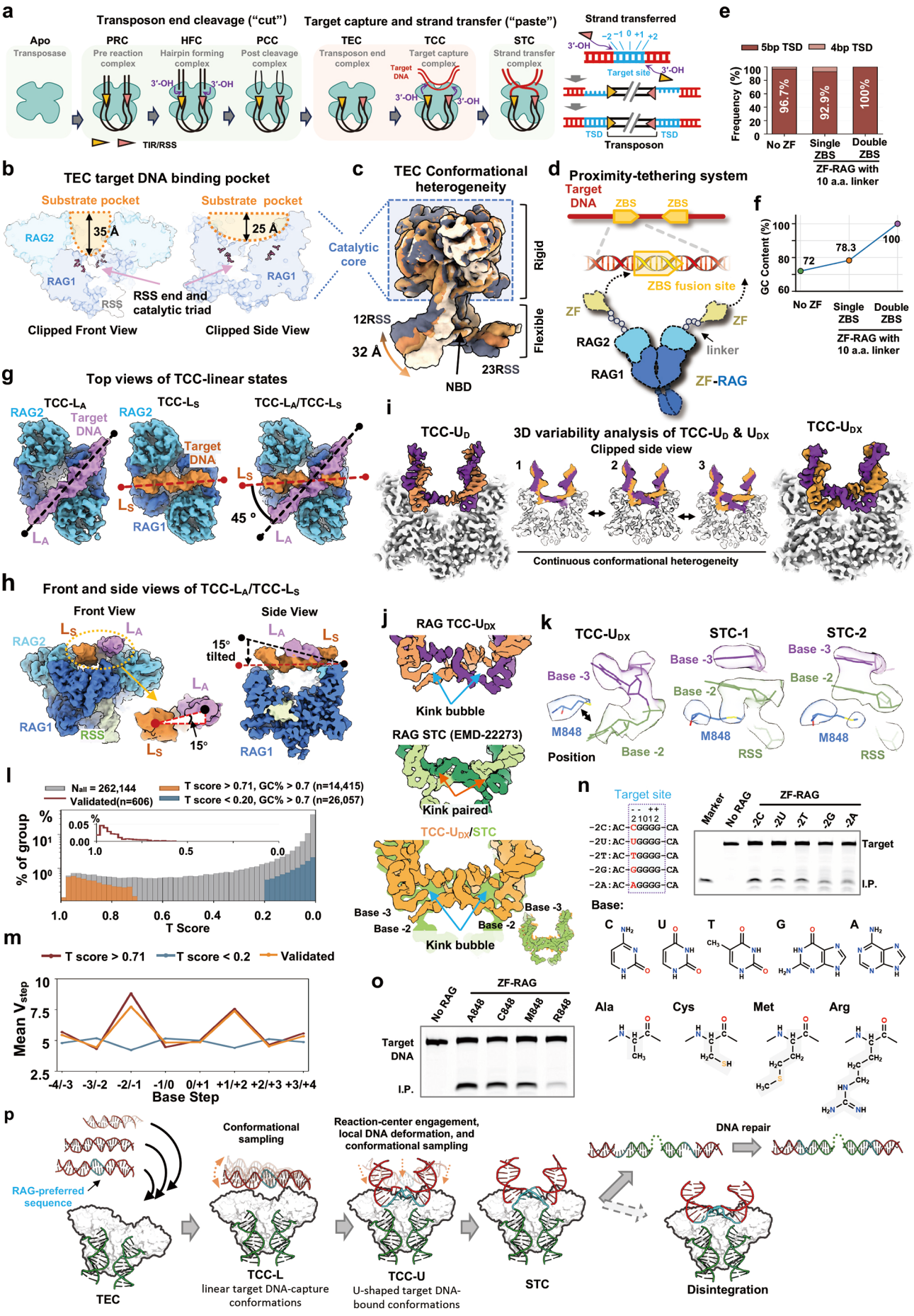

a
Transposon end cleavage ("cut")
Target capture and strand transfer ("paste")
Apo
Transposase
PRC
Pre reaction complex
HFC
Hairpin forming complex
PCC
Post cleavage complex
TEC
Transposon end complex
TCC
Target capture complex
STC
Strand transfer complex
Target DNA
3'-OH
TIR/RSS
Strand transferred
Target site
TSD
Transposon
e
5bp TSD
4bp TSD
Frequency (%)
96.7%
92.9%
100%
No ZF
Single ZBS
Double ZBS
ZF-RAG with 10 a.a. linker
b
TEC target DNA binding pocket
Substrate pocket
35 Å
25 Å
RAG2
RAG1
RSS
RSS end and catalytic triad
Clipped Front View
Clipped Side View
c
TEC Conformational heterogeneity
Catalytic core
Rigid
Flexible
12RSS
32 Å
23RSS
NBD
d
Proximity-tethering system
Target DNA
ZBS
ZBS fusion site
ZF
linker
RAG2
RAG1
ZF-RAG
f
GC Content (%)
72
78.3
100
No ZF
Single ZBS
Double ZBS
ZF-RAG with 10 a.a. linker
g
Top views of TCC-linear states
TCC-LA
TCC-LS
TCC-LA/TCC-LS
RAG2
Target DNA
RAG1
LA
LS
45 °
i
TCC-UD
3D variability analysis of TCC-UD & UDX
Clipped side view
1
2
3
Continuous conformational heterogeneity
TCC-UDX
h
Front and side views of TCC-LA/TCC-LS
Front View
Side View
15° tilted
15°
RAG2
RAG1
RSS
j
RAG TCC-UDX
Kink bubble
RAG STC (EMD-22273)
Kink paired
TCC-UDX/STC
Base -3
Base -2
Kink bubble
k
TCC-UDX
STC-1
STC-2
Base -3
Base -2
M848
RSS
Position
l
Nall = 262,144
Validated(n=606)
T score > 0.71, GC% > 0.7 (n=14,415)
T score < 0.20, GC% > 0.7 (n=26,057)
% of group
T Score
m
T score > 0.71
T score < 0.2
Validated
Mean Vstep
Base Step
-4/-3
-3/-2
-2/-1
-1/0
0/+1
+1/+2
+2/+3
+3/+4
n
Target site
Marker
No RAG
ZF-RAG
-2C
-2U
-2T
-2G
-2A
Target
I.P.
Base:
C
U
T
G
A
Ala
Cys
Met
Arg
o
No RAG
ZF-RAG
A848
C848
M848
R848
Target DNA
I.P.
p
RAG-preferred sequence
TEC
Conformational sampling
TCC-L
linear target DNA-capture conformations
Reaction-center engagement, local DNA deformation, and conformational sampling
TCC-U
U-shaped target DNA-bound conformations
STC
DNA repair
Disintegration

Fig. 1 RAG-mediated target DNA capture, deformation and target-site preference. a Schematic of cut-and-paste transposition, showing transposon-end cleavage, target DNA capture and strand transfer. **b** Clipped TEC cryo-EM map showing the target-DNA pocket (orange dashed boxes) and catalytic residues (sticks). **c** Overlay of four cryo-EM TEC conformations. **d** Schematic of the proximity-tethering system showing target DNA (top) and ZIF268-fused RAG (bottom). **e** TSD-length distributions under different ZBS conditions. **f** Integration-site GC enrichment under different ZBS conditions. **g** Unsharpened TCC-L_A (left) and TCC-L_S (middle) cryo-EM maps; target-DNA overlay (right). DNA: pink (L_A), orange (L_S); dashed lines, DNA axes. **h** Front and clipped side views comparing target-DNA orientations in TCC-L_A and TCC-L_S. **i** TCC-U_D and TCC-U_{DX} cryo-EM maps with three representative 3D-variability maps from TCC-U_D. Protein density is white; TCC-U_D and TCC-U_{DX} are displayed at the same RMS contour level. **j** Target-site DNA densities in TCC-U_{DX} (top) and STC (middle; PDB 6XNY), with an overlay (bottom). **k** Densities of RAG1 Met848 and target base −2 in TCC-U_{DX}, STC-1 and STC-2. **l** T score distribution of all possible 9-mers. Orange and blue bars indicate high-scoring (T > 0.71) and low-scoring (T < 0.2) GC-rich sequences (GC fraction > 0.7 across positions −3 to +3). Inset, T score distribution of validated transposition sites, with bin counts normalized to all 9-mers. **m** Mean V_{step} profiles for high-scoring (red, *n* = 21,822), low-scoring (blue, *n* = 193,077) and validated (orange, *n* = 606) sequences. **n** Integration assay using target substrates with different bases at position −2. I.P., integration product. **o** Integration assay for RAG1 residue-848 variants. **p** Model of RAG-mediated target DNA capture and strand transfer. TCC-L, TCC-U and STC represent conformational ensembles rather than obligatory discrete intermediates.

another containing U-shaped target DNA within it (Supplementary Fig. S7a). The extended particles represented only a small fraction of the dataset, consistent with a weak and dynamic interaction before pocket entry. To enrich these conformations, we introduced a glutamate substitution at RAG1 residue 848 in the transposition-activated RAG complex, aiming to reduce target DNA insertion through increased electrostatic repulsion at the pocket base while preserving RSS cleavage activity. This yielded two structures with linear target DNA positioned above the RAG tetramer (Supplementary Figs. S8, S9): the asymmetric TCC-L_A and the symmetric TCC-L_S (Fig. 1g). In TCC-L_A, target DNA is tilted ~15° and shifted toward one RAG1–RAG2 heterodimer, producing a predominantly one-sided anchor (Fig. 1h). In TCC-L_S, the duplex lying more symmetrically in the RAG1 groove, is positioned deeper above the pocket and approaches both catalytic centers (Supplementary Fig. S10a). Compared with TCC-L_A, TCC-L_S appears to form more extensive contacts with target DNA, and its target DNA aligns more closely with that in the product STC (Supplementary Fig. S10b–d). Target DNA density in both maps is weaker than that of the protein core, permitting approximate helix placement and qualitative trajectory comparisons but not base-specific contact assignments (Supplementary Fig. S9b). Given the dynamic nature of the linear TCC states, TCC-L_A and TCC-L_S likely represent two relatively stable conformations within a broader ensemble of early linear target-DNA engagement states sampled by the RAG1 R848E complex, and these structures do not establish an obligatory temporal order between them.

RAG2 also contacts target DNA in linear TCCs. Its target-facing D6–A1 loop differs among RAG2/RAG2L proteins: family B and D RAG2L loops are longer and more basic than those of family A RAG2L and vertebrate RAG2 (Supplementary Fig. S11a–c). The amphioxus BbRAGL TEC bound target DNA more strongly than the mouse RAG TEC. Introducing H10KR or a BbRAG2L-like basic loop into mouse RAG2 enhanced binding; H10KR also increased cellular transposition without altering RSS cleavage, whereas H10E reduced binding (Supplementary Fig. S11d–h). Thus, this loop modulates target engagement independently of RSS cleavage, suggesting that evolutionary attenuation of its target-binding function might have contributed to reduced vertebrate RAG transposition.

A second class of particles revealed a pre-strand-transfer state in which U-shaped target DNA had entered the substrate-binding pocket but showed weak density across the central 5-bp target site, suggesting disorder near the active center (Supplementary Fig. S7a). To prevent strand transfer and enrich these states, we introduced the active-site mutation E962N, which abolished detectable transfer (Supplementary Figs. S5b, S12a–d). Focused classification yielded TCC-U_D, an averaged reconstruction with weak target-site density (Supplementary Fig. S12a). Local three-dimensional variability analysis revealed a continuum of target-site density states from absent to X-shaped, with the latter conformations yielding TCC-U_{DX}, an averaged reconstruction of a subset rather than a discrete obligatory intermediate (Fig. 1i; Supplementary Video S2).

Local analysis of TCC-U_{DX} showed that the map reached ~3–3.5 Å resolution around target DNA positions −2 and −1, close to the overall resolution of the complex and better than that near position 0 (Supplementary Fig. S12c, d). We therefore restricted model interpretation to regions with continuous backbone density and recognizable base features. At the kink near position −2, the density features are consistent with pronounced local deformation and partial disruption of base pairing; by contrast, target-site DNA in the STC is better ordered and more fully paired (Fig. 1j). Since the RAG substrate-binding pocket closes early during DNA cleavage and remains structurally stable, thereafter, the disorder observed in TCC-U_D and TCC-U_{DX} is unlikely to be driven by large conformational changes of the protein pocket (Supplementary Fig. S3). Rather, the disorder is more likely to arise from deformation of target DNA as it enters the deep, narrow pocket. Molecular-dynamics simulations provided complementary support: U-shaped duplex DNA displayed substantial heterogeneity at the base of the bend when simulated in water, and local unwinding near the kink remained detectable when the same U-shaped DNA was simulated in the RAG binding pocket (Supplementary Fig. S13a). Twenty independent simulations initiated from the experimental TCC-U_{DX} geometry produced similar overall conformations and fluctuation profiles (Supplementary Fig. S13b). Although these simulations do not define a transition pathway, they support the structural inference that imposing the observed U-shaped trajectory can destabilize the target-site duplex. Comparison of TCC-U_{DX} and STC further showed that they retain similar overall bent trajectories but differ markedly in local order. Indeed, site-specific introduction of a mismatch at the −2 position significantly increased transposition activity (Supplementary Fig. S14a), consistent with previous reports that several transposases, including RAG, preferentially catalyze

strand transfer at sites of local DNA unwinding[8–10]. Target DNA in TCC-U_{DX} forms a larger contact interface with RAG than in TCC-L_S, consistent with increased protein–DNA engagement upon deeper accommodation within the substrate-binding pocket (Supplementary Figs. S10b, S14b). Together with the linear TCC states, these structures support a model in which target DNA is first captured in linear conformations and subsequently samples U-shaped trajectories within the substrate-binding pocket before strand transfer, with target-site DNA becoming locally distorted and partially unpaired as it moves closer to the active center. Similar local DNA unwinding during RAG-catalyzed RSS cleavage and other DDE-family transposition reactions suggests a possible role for local DNA destabilization in positioning DNA near the active center during catalysis[10–12].

Further focused classification of post-strand-transfer particles resolved two related RAG–transposition product structures, STC-1 and STC-2. STC-2 more closely resembles the previously reported RAG STC and displays more continuous target DNA base pairing (Supplementary Figs. S7, S15). Comparison of TCC-U_{DX} with STC-1 and STC-2 revealed local density features consistent with a positional rearrangement between the target −2 base region and the RAG1 Met848 side-chain region across strand transfer (Fig. 1k); these observations indicate that target DNA in the post-strand-transfer complexes retains conformational heterogeneity and that local ordering around the reaction center differs between the pre-strand-transfer TCC-U_{DX} structure and the post-strand-transfer STC structures.

We next examined whether the previously reported GC enrichment fully accounts for RAG target-site preference or whether additional position-specific sequence features contribute. To address this question, we reanalyzed 1,302 non-redundant 13-mer sequences from cellular RAG transposition sites. This analysis confirmed broad GC enrichment but also revealed a marked pyrimidine preference at position −2 and a reciprocal purine preference at the symmetry-related +2 position, indicating that RAG target-site preference contains position-specific information beyond overall GC content (Supplementary Fig. S16a–c). To characterize these features more comprehensively and relate them to structural properties, we encoded all possible 9-mers using DeepDNAshape-predicted parameters, including minor-groove width, electrostatic potential, shift, slide, rise, tilt, roll and helical twist. Dimensionality-reduction analysis showed that experimentally validated target sites were concentrated within a subset of the DNA-shape feature space (Supplementary Fig. S17a–c). We then trained an XGBoost model against genomic background sequences and generated a target DNA association (T) score for all 262,144 possible 9-mers (Supplementary Fig. S17a and Table S4). Experimentally validated sites were enriched among sequences with high T scores, whereas GC-rich sequences were present in both the high-scoring and low-scoring groups, further indicating that GC content alone is insufficient to account for RAG target-site preference (Fig. 1l). SHAP analysis showed that the T score reflected contributions from multiple DNA-shape parameters, with roll providing the largest overall contribution (Supplementary Fig. S17d). The contribution of roll was strongly concentrated at the −2/−1 and +1/+2 steps, where roll values were elevated, suggesting that local bending geometry at these positions might be particularly important for RAG target-site selection (Supplementary Fig. S17e, f).

To examine whether these sequence-associated differences extend to local DNA deformability, we analyzed V_{step}, a scalar descriptor of the conformational-space volume accessible to a dinucleotide step[13], previously applied to Sleeping Beauty target selection[14]. Validated target sites and high-T-score sequences exhibited distinct V_{step} profiles, with pronounced peaks at the −2/−1 and +1/+2 steps compared with low-scoring sequences (Fig. 1m; Supplementary Fig. S18a). These differences persisted among GC-rich sequences, further indicating that GC content alone does not explain RAG target-site preference (Supplementary Fig. S18b, c). Notably, the −2/−1 step corresponds to the strongly deformed DNA kink observed in TCC-U_{DX}, linking the V_{step} predictions to our structural observations.

Residue 848 offers a structural link between evolution and site specificity. The relative rearrangement of the residue 848 side-chain region and the −2 target base region between TCC-U_{DX} and STC (Fig. 1k) suggests that smaller pyrimidines may better accommodate this local geometry, consistent with validated −2 preferences (Supplementary Fig. S16c). *In vitro*, −2 pyrimidines supported more efficient strand transfer than purines, with efficiency generally decreasing as base size increased; when G occupied −2 position, additional strand transfer occurred at the adjacent −1 C position (Fig. 1n; Supplementary Fig. S19a, b). Replacing vertebrate RAG1 Arg848 with shorter side chains (Ala, Cys or Met) enhanced integration. Notably, Met additionally reduced disintegration *in vitro* (Fig. 1o; Supplementary Fig. S19c, d). Invertebrate RAG1L proteins predominantly contain Met at this position, whereas vertebrate RAG1 proteins contain Arg. Thus, steric compatibility between residue 848 and the −2 target base is one factor contributing to integration and target-site selection, suggesting that the evolutionary replacement of Met with Arg contributed to the suppression of RAG transposition.

Together, these structures support a model in which RAG initially captures target DNA in an ensemble of linear conformations, followed by sampling of U-shaped trajectories within the substrate-binding pocket. During pocket entry, target DNA undergoes local deformation and possible partial unpairing near the catalytic centers, becoming better ordered after strand transfer (Fig. 1p). Sequence-encoded DNA deformability at key dinucleotide steps and steric interactions contribute to target-site preference beyond GC content. Comparison with BbRAGL implicates reduced target engagement and altered reaction-center interactions in the suppression of transposition during vertebrate RAG domestication. Related DNA destabilization occurs during RAG-mediated RSS cleavage and other DDE-family reactions[10–12], although individual transposases stabilize distinct DNA geometries. HzTransib also bends target DNA substantially (Supplementary Fig. S20), suggesting that DNA deformation may be an evolutionarily conserved feature of RAG-related transposases. Notably, our L_A/L_S and U_D/U_{DX} structures do not establish an obligatory temporal order or demonstrate that every captured conformation is a required reaction intermediate. Moreover, proximity tethering and specific mutations may alter the relative proportions of conforma-

tional populations. Their native occupancies and interconversion rates therefore remain unresolved. In summary, our findings provide mechanistic insights into RAG-mediated target DNA capture, deformation and site selection, while revealing molecular features associated with the evolutionary suppression of transposition.

DATA AVAILABILITY

Atomic models and cryo-EM maps are deposited in the PDB (9LQO, 9LQP, 9LQQ, 9LQV, 9LR0) and EMDB (EMD-63299, EMD-65775, EMD-65776, EMD-63300, EMD-63301, EMD-63306, EMD-63311), respectively.

ACKNOWLEDGMENTS

We thank the Instrument Analysis Center, Shanghai Jiao Tong University, and the Electron Microscopy System, National Facility for Protein Science in Shanghai, for cryo-EM data collection. This work was supported by the National Key R&D Program of China (2022YFE0132100 to Y.Z.), National Natural Science Foundation of China (32171250 to Y.Z.), and unrestricted funds from Yale University (to D.G.S.).

AUTHOR CONTRIBUTIONS

Y.Z., J.P. and D.G.S. designed, and J.P. performed, biochemical experiments. J.P. and Y.Z. performed structural experiments, analyzed structures and built models. E.C.M. performed molecular dynamics simulations. S.J., X.Z., E.C.M. and J.P. analyzed sequences and target-site preferences; Q.L. and J.P. established the proximity-tethering system. D.G.S. and Y.Z. wrote the manuscript with input from all authors.

COMPETING INTERESTS

The authors declare no competing interests.

REFERENCES

1. Cosby, R.L., Chang, N.C. & Feschotte, C. *Genes Dev.* **33**, 1098–1116 (2019).
2. Slotkin, R.K. & Martienssen, R. *Nat. Rev. Genet.* **8**, 272–285 (2007).
3. Craig, N.L. *Mobile DNA III* https://doi.org/10.1128/9781555819217.ch1 (2015).
4. Arinkin, V., Smyshlyaev, G. & Barabas, O. *Curr. Opin. Struct. Biol.* **59**, 168–177 (2019).
5. Liu, C., Zhang, Y.H., Liu, C.C. & Schatz, D.G. *Nat. Rev. Immunol.* **22**, 353–370 (2022).
6. Ru, H. et al. *Cell* **163**, 1138–1152 (2015).
7. Chen, X.M. et al. *Proc. Natl. Acad. Sci. USA* **122**, e2512362122 (2025).
8. Yanagihara, K. & Mizuuchi, K. *Proc. Natl. Acad. Sci. USA* **99**, 11317–11321 (2002).
9. Pflieger, A., Jaillet, J., Petit, A., Augé-Gouillou, C. & Renault, S. *J. Biol. Chem.* **289**, 100–111 (2014).
10. Ru, H. et al. *Nat. Struct. Mol. Biol.* **25**, 732–742 (2018).
11. Morris, E.R., Grey, H., McKenzie, G., Jones, A.C. & Richardson, J.M. *Elife* **5**, e15537 (2016).
12. Fuller, J.R. & Rice, P.A. *Elife* **6**, e21777 (2017).
13. Olson, W.K., Gorin, A.A., Lu, X.J., Hock, L.M. & Zhurkin, V.B. *Proc. Natl. Acad. Sci. USA* **95**, 11163–11168 (1998).
14. Liu, G.Y. et al. *J. Mol. Biol.* **346**, 161–173 (2005).

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.15302/vita.2026.10.0082.

Correspondence and requests for materials should be addressed to David G. Schatz and Yuhang Zhang.

Reprints and permission information is available at https://www.vita-journal.com/.

© The Author(s) 2026. Published by Higher Education Press. This is an Open Access article distributed under the terms of the CC BY license (https://creativecommons.org/licenses/by/4.0/).