Supplemental text for

Real-time Observation of CRISPR spacer acquisition by Cas1–Cas2 integrase

Jagat B. Budhathoki\textsuperscript{1}, Yibei Xiao\textsuperscript{1,#}, Gabriel Schuler\textsuperscript{1,#}, Chunyi Hu\textsuperscript{1}, Alexander Cheng\textsuperscript{1}, Fran Ding\textsuperscript{1}, Ailong Ke\textsuperscript{*}

\textsuperscript{1} Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14853, U.S.A.

\textsuperscript{#} These authors contributed equally.

\textsuperscript{*} Correspondence: ailong.ke@cornell.edu
Supplemental Information

Mapping FRET peaks under SDS condition to integration configurations

Two native FRET peaks were observed when *Efa*Cas1–Cas2 catalyzed the integration of PS(4,4) to leader-repeat-containing integration target. These two peaks partitioned to five different FRET peaks upon washing with 2% SDS, which denatured and removed *Efa*Cas1–Cas2 (Fig. 1d). These SDS-denatured peaks could not be unambiguously interpreted by cross-referencing to crystal structures. We therefore carried out a series of control measurements to define which denaturing peak corresponded to which integration state. First, the baseline was established by annealing oligonucleotides to assemble the leader-side half-integration state, with Cy3 incorporated into the identical location as in the integration target in Fig. 1, and Cy5 to equivalent locations representing the prespacer undergoing leader-side integration in two orientations (Extended Data Fig. 2e, f). Each configuration was immobilized on the surface and 25 short movies were recorded in the post-SDS-wash condition as in the integration experiments. The smFRET histogram revealed a single FRET peak ($E_{\text{FRET}}$) at 0.35±0.04 and 0.15±0.04 for configuration 1 and 2, respectively (Extended Data Fig. 2g, h). The former value agreed well with that of Peak 4 and the latter with Peak 2 in Fig. 1. We therefore concluded that Peak 2 and Peak4 represent leader-side half-integration in two prespacer orientations.

In addition, an independent set of reference points were generated by denaturing the half-integration-only reactions (Fig. 3a); full-integration was eliminated by the use of a single-deoxy-containing prespacer, PS(4,4ddC). Half integration to the leader-side corresponds to Peak 4 in Fig. 1d with this prespacer. Reaction performed for up to 2 minutes explicitly produced leader-side half-integration; space-side half-integration product became evident when we extended the reaction to 10 minutes with 2 nM of Cas1–Cas2–PS(4,4ddC) (Extended Data Fig. 2i). Two types of $E_{\text{FRET}}$ peaks at 0.08±0.05 and 0.39±0.04 were evident when 25 short movies were analyzed to generate FRET efficiency histogram (Extended Data Fig. 2i). These values are
consistent with $E_{\text{FRET}}$ values for Peak 1 and Peak 4 in Fig. 1d, respectively. Even though spacer-side reaction can occur in two prespacer orientation for PS(4,4) just like leader-side reaction (Fig. 1d, e), only one FRET peak appeared for spacer side reaction. This is because spacer reaction site is 26 bp (~8.84 nm) apart from Cy3 attached on the target. Considering the position of Cy5 on the prespacer, the distance between Cy3-Cy5 becomes more than 8.84 nm. This distance falls in an insensitive FRET range; as a result, spacer side configurations appeared as a single peak, Peak 1, in Fig. 1d, which represents both spacer side configurations.

Having assigned three FRET peaks - Peak 1, Peak 2, and Peak 4 in Fig. 1d - obtained under the SDS condition, we inferred that the remaining two peaks (i.e., Peak 3 and Peak 5) corresponds to the full-integration scenario, with prespacer in two orientations. Considering the full-integration scenario is more compact, due to the tethering of both prespacer overhangs to the target DNA, their FRET peaks were expected to be higher than their corresponding half-integration states. Indeed, the FRET values for Peak 3 and Peak 5 are higher than those for Peak 2 and Peak 4. Furthermore, these FRET values were absent in half-integration-only reactions. They were only observed when full integration is possible (Fig.1d).

**Photostability test of Cy3 and Cy5**

Fluorophore blinking interrupts a continuous smFRET trace. Fluorophore photobleaching prematurely terminates smFRET recording. Neither events are desirable. Because long movie recordings were needed to capture full cycles of integration/disintegration, we lowered excitation laser power to extend fluorophore lifetime and used longer exposure time to reach an acceptable signal-to-noise level. We evaluated the photostability of Cy3 and Cy5 fluorophores under the experimental condition. Cy3 in the assembled construct described in Extended Data Fig. 12a was excited by the green laser at 3-5 mW, FRET-excited Cy5 signal was collected at
3.3 Hz frame-refresh-rate for a minimum of twenty minutes. Only 2.7% (18 of 679) of the Cy3 and 0.45% (2 of 564) of the Cy5 molecules showed blinking behavior, which converted to data reliability of 97.8% and 99.5% for Cy3 and Cy5 respectively. The survival time distribution of Cy3 and Cy5 were also plotted against the continuous laser exposure (Extended Data Fig. 7c), which shows excellent durability of Cy3 and Cy5 (86.7% of the Cy3 and 97.0% of the Cy5).
# Supplementary Table 1: DNA oligonucleotides used in smFRET and ensemble experiments

| Name*                  | Sequence (5'→3')                                                                 |
|------------------------|----------------------------------------------------------------------------------|
| Wildtype target, Top   | /5Bio/CTTTGGAAAAATAATTCTCCGAGGGTTTTTAGAGTCATGTTGTTTAGAGATGGGTACCAAAAAACGCACACACACGGTTCA |
| Wildtype target, Bottom| TG/iCy3/ACTCTAAAACCTCGAGAATTATTGAGATGGGTACCAAAAAACGCACACACACGGTTCA               |
| **Leader-mutant**, Top | /5Bio/CTTTGGAAAAATAATTCTCTCTACTATATTATATTAGAGAGATGGGTACCAAAAAACGCACACACACGGTTCA |
| **Leader-mutant**, Bottom| TGAACGTGTGTGTGCGTTTTGGTACCATTCTAAACAACAAG/iCy3/ACTCTAAAACCTCGAGAATTATTGAGATGGGTACCAAAAAACGCACACACACGGTTCA |
| Inverted Repeat-mutant, Top | /5Bio/CTTTGGAAAAATAATTCTCTCTACTATATTATATTAGAGAGATGGGTACCAAAAAACGCACACACACGGTTCA |
| Inverted Repeat-mutant, Bottom| TGAACGTGTGTGTGCGTTTTGGTACCATTCTAAACAACAAG/iCy3/ACTCTAAAACCTCGAGAATTATTGAGATGGGTACCAAAAAACGCACACACACGGTTCA |
| PS(4, 4), Top          | AG CTACTCCGATGGCCATATGCGGACT                                                    |
| PS(4, 4), Bottom       | TACG CGCATATGGCCCATATGCGGACT                                                    |
| PS(4, 4ddC), Top       | CTACTCCGATGGCCCATATGCGGACT                                                     |
| PS(4, 4ddC), Bottom    | CGCATATGGCCCATATGCGGACT/3ddC/                                                  |
| PS(4, 5), Top          | AGCTACTCCGATGGCCCATATGCGGACT                                                    |
| PS(4, 6), Top          | AGCTACTCCGATGGCCCATATGCGGACT                                                    |
| PS(4, 26), Top         | CTACTCCGATGGCCCATATGCGGACT                                                     |
| PS(4, 4 bp duplex), Top| AG CTACTCCGATGGCCCATATGCGGACT                                                   |
| PS(4, 4 bp duplex), Bottom| TACG CGCATATGGCCCATATGCGGACT                                                     |
| PS(4 bp duplex, 4 ), Top| AGCTACTCCGATGGCCCATATGCGGACT                                                    |
| PS(4 bp duplex, 4), Bottom| TACG CGCATATGGCCCATATGCGGACT                                                     |
| PS(4, 8bp duplex), Top | AGCTACTCCGATGGCCCATATGCGGACT                                                    |
| PS(4, 8bp duplex), Bottom| GATCGATGCTGCGTTTCTCTACGCGCATATGGCCCATATGCGGACT/iCy5/GAGTAGGACT                  |
| PS(4, 20 bp duplex), Top| AGCTACTCCGATGGCCCATATGCGGACT                                                    |
| PS(4, 20 bp duplex), Bottom| GATCGATGCTGCGTTTCTCTACGCGCATATGGCCCATATGCGGACT/iCy5/GAGTAGGACT                  |
| Leader-integration mimic, Biotin Top (common strand for both mimics) | /5Bio/CTTTGGAAAAATAATTCTCCGAG |
| Leader-integration mimic1, Top | TACGCGCATATGGCCATCG/iCy5/GAGTAGGACTGTTT TAGAGTCATGTTGTTTAGAATGTGTAACAAAACGCACACA CACGTTCA |
|--------------------------------|------------------------------------------------------------------|
| Leader-integration mimic2, Top | AGCTACTCCGATGCCCATATGCGAECTGTTTATTAGAGT CATGTTGTTTAGAATGTGTAACAAAACGCACACA CACGTTCA |
| 256bp Promoter-Leader-Repeat-Spacer target, Top | Cy3/GCAGCAGGAGAGACAATTTAAAGAGACTTTAAGAAGATACTTATAAAAGAAGATTTGACTTTAAAGTC TAACCTATAGACTTACATCACGAGGCAGAGGGAGACG GGGAGACAGCAACACTCCGGAAGTTGTGACTGACGA ACATGCTTTGATTTAATCTCCACTCGAG ATGTTGTTTGGATACATGGAATCACAACACACACGACGTTCA |
| 256bp Promoter-Leader-Repeat-Spacer target, Bottom | Cy5/CTGGTTTGAATTTGAAGCTGCCAGACCAGACTGC AcgggGTTTTGGTACCATTCTAACAACATGACTCTAAA CCTCGGAGAATTTTTTTTTAAGAAGATTTGACTTTCTGCAC ATAGACCTTCCGATGTTGCTGTCCTCCCCCTGTGCTCT CTCGATGGGATGATGTCAGCTTACAAATTCAAACCA GTGCTTTGTTTAGAATGACCTTTTAAGTACCAACACACACCG TGCAGTCTGGCACCCGACGTTCA |
| 200 bp Promoter-Leader-Repeat-Spacer target, Top, smFRET | /5'Bio/GAGACAACCTAAGAGACCTTTAAGAGATTATTTTAA AATTTCATAAAAAGAGATTTAGACTTTAAGCTCTAAACCT CACGAGACGGAGAGGCAACACGACGGGGAGATCC AGGAACTTTTGGAAAAATAATTCTCGAGTCGAGGATTTAGC TACGTGTTTGGATACATGGAATCACAACACACACGACGTTCA |
| 200 bp Promoter-Leader-Repeat-Spacer target, Bottom, smFRET | GCTCGCAAGACGACTCGACgGTTTTGGTACCATTCT AACACATGACTCCTAAACACTCCGGAAGATTATTTTTT TCAAAGTTCCGGACCCCGTTCCTGACGTGACGTC ATCTGTCCATAGGCTAGACTTTAAGTCAATAACTCTTTTC TATGAAAAATAATTAATCCTTTCAGTTAGTTAGTTCGCTC |
| 192 bp Promoter-Leader-Repeat, Top | Cy3/GAGACAACCTAAGAGACCTTTAAGAGATTATTTTAA AATTTCATAAAAAGAGATTTAGACTTTAAGCTCTAAACCT CACGAGACGGAGAGGCAACACGACGGGGAGATCC AGGAACTTTTGGAAAAATAATTCTCGAGTCGAGGATTTAGC TACGTGTTTGGATACATGGAATCACAACACACACGACGTTCA |
| 192 bp Promoter-Leader-Repeat, Bottom | Cy5/ACCAGAATCGACgGTTTTGGTACCATTCTAACAAC CATGACTCTAAACACTCCGGAAGATTATTTTTTTCAAG TTTCCGGATCCTCCGGCTGTCCCTGACGTGACGTC ATCTGTCCATAGGCTAGACTTTAAGTCAATAACTCTTTTC TATGAAAAATAATTAATCCTTTCAGTTAGTTAGTTCGCTC |
| 84bp Spacer side-labeled target, Top, smFRET | /5Bio/CTTTGGAAAAAAATAATTCTCCGAGGTTTTAGAGT CATGTTGGTTTAAGATGGTACCAAAAACcccgTCAAGTTCTG GTCTCGAGAC |
|---------------------------------------------|--------------------------------------------------|
| 84bp Spacer side-labeled target, Bottom, smFRET | GCTCGCAGACCGAGACTCGACgggGTTTTGGTACCATTCT AAACAAATGACTCTAATACCTCGGAGAATTATTTTTTCAAAAG |
| Cy3-IR₁ probe | Cy3/GCCAGTTTTAGAGTCA |
| F_pTarget_gRNA | ttaaacataaGTTTTAGAGCTAGAAATAGC |
| R_pTarget_gRNA | gagtacttaaACTAGTATTATACCTAGGAC |
| F_pTarget_pTrc | acttgacaatattcatccgcgtctgtattgttgaggGGGTTCGACAC |
| R_pTarget_pTrc | GACATCCGGATATAGTTCTCC |
| F_genome_check | GATTAGCCAAAAAGGATGAGC |
| R_genome_check | CGACGTAAAGCCATTAACG |
| F_Efa_array (primer set 1) | GTAAATGGGCGGGAGCAGAG |
| R_Efa_array (primer set 1) | GACATCCGGGATATAGTTCTCC |
| F_Efa_array_integration (primer set 2) | GTAAATCGCATGGGACCAAC |
| R_Efa_array_integration (primer set 2) | CATGACTCTAAAAACAGTTCGACATAG |
| Top_Efa_prespacer | CTACTCCGATGGCCCCATATGGCAACT |
| Bottom_Efa_prespacer | CGCATATGGGCCCATCGAGTAGAC |
| BL21-Al_Efa array sequence (NC_012947.1 position 1002802-1004320) | AACCCCGCCGGGGCTCTTTGGGAGGCTCGCAGCTCGCGGGGGTTTTGGGTCATAAAAAAGGGAAGCCTAGCTCAGTCAAGAAGCTGGGACCTTTGGGTCTTTGCTGATTACAAAAGCCCAAGAGACCATGCGGAGAGCAGTGAGTCAGACGAGAGCAAGAGACGAGATTTCCAACCTGAGATGAGTCAGGAGAGCAAGAGACGAGATTTCCAAGAGGAGGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCAAGACGAGATTTCCA91
* The following DNA oligos were bought from Integrated DNA Technologies, Inc., (IDT) and used for smFRET measurements. The oligos were annealed in duplex annealing buffer (100 mM Potassium acetate, 30 mM HEPES, pH=7.5) to form different constructs.

| position 1002802-1004320) | AGTTGGCTGCTGCCACCAGCTGAGCAATAACTAGCATAACCCCTTGCCCCTCTATAAAGTTGAGGTTTTTGCTGAAAGGAGGAACTATATCCGGATGTCACTTGACAAATTATCATCCGCTCGTATAAATGTGTGGAGGAGGGTTTCGACACTTCACAGATAGTGAGGGATCCGGGCAAAGGGCGATTAATTGCGGTCCAACATAGGCGTAAACTACGATGGCACAACACTCAGTCGCAGCTGCTTTGGAAAAATAATTCTCCGAGGTTTAGAGTCATGTGGTTTAGATAATGGGTACCACAAACCAGTGCTGCAGTCTGGTGCTGGCAGCTTACAAACGCCAGGTGGAAATCAGTTGAGGTACGATGGGAACAGTCTGGGTGGGATTGAGAAGAATGAAAAACCGCCGATCCTGACACCGCATTACTGCAAGGTAGTGGACAAGACCAGGCCTGACCTGAA |