Site-Programmable Transposition:
Shifting the Paradigm for CRISPR-Cas Systems

Michael Chavez1 and Lei S. Qi1,2,3,*

1Department of Bioengineering, Stanford University, Stanford, CA, USA
2Department of Chemical and Systems Biology, Stanford University, Stanford, CA, USA
3ChEM-H, Stanford University, Stanford, CA, USA
*Correspondence: stanley.qi@stanford.edu
https://doi.org/10.1016/j.molcel.2019.07.004

Discoveries by Klompe et al. (2019) and Strecker et al. (2019) eludicate distinct CRISPR-Cas mechanisms for site-specific programmable transposition in prokaryotic organisms.

CRISPR-Cas systems are deployed by prokaryotes as a mechanism to defend against mobile genetic elements (MGEs) such as bacteriophages and plasmids. In these systems, short RNAs guide CRISPR-associated (Cas) nucleases to MGEs for sequence-specific cutting and degradation (Knot and Doudna, 2018). This mechanism has been demonstrated in diverse CRISPR-Cas systems and inspired its repurposing as a genome engineering tool in eukaryotes (Knot and Doudna, 2018). However, bioinformatic analysis of CRISPR loci from a wide variety of prokaryotes suggests that these systems may not be limited to site-specific endonuclease activity (Faure et al., 2019a; Shmakov et al., 2016). Indeed, recent studies by Strecker et al. (Strecker et al., 2019) and Klompe et al. (Klompe et al., 2019) describe two distinct CRISPR subtypes that mediate site-directed transposition, significantly broadening the space of natural functions and potential biotechnological applications for CRISPR-Cas systems.

The link between transposons and CRISPR was identified through the unique features of particular CRISPR loci. Bioinformatic analysis of a subtype of class I systems (where cleavage occurs through multiple effectors) known as type I-F revealed that some loci had features different from most CRISPR-Cas systems. These loci contained Tn7-like transposon genes and lacked the endonuclease cas3, suggesting the system was incapable of DNA cleavage (Figure 1A) (Peters et al., 2017).

An astonishing parallel was discovered with type V-K, a class II system (where cleavage occurs through a single effector), which also contains Tn7-like transposon genes and an effector with a naturally inactivated RuvC-like nuclease domain (Faure et al., 2019b; Shmakov et al., 2017). Interestingly, the two subtypes also lacked cas1 and cas2, the genes essential to acquire new sequences from mobile genetic elements into the CRISPR array. These analyses led to the work by Klompe et al. and Strecker et al., which demonstrates that type I-F and type V-K utilize this unique framework to perform site-directed transposition through their nearby Tn7-like transposase (Figure 1).

Tn7-like transposition machinery consists of TnsA, TnsB, and TnsC, which together bind and cleave donor DNA through recognition of specific sequences (Figure 1B) (Peters et al., 2017). The targeting factor TnsD, a protein that binds the Tn7 attachment site (attTn7) sequence, or TnsE, a protein that preferentially binds plasmid DNA, then recruits this TnsABC complex bound to the donor DNA strand and integrates the donor at the target site. The type I-F and type V-K CRISPR-Cas systems contain most factors of this transposition machinery, while lacking TnsD/TnsE and possessing TniQ, a homolog of TnsD.

Further, these systems are flanked by the conserved binding sites that allow for recognition and cleaving by the transposition machinery.

Kломпе et al. used a plasmid-based approach in Escherichia coli to study the transposition mechanism of the type I-F CRISPR-Cas system from Vibrio cholerae. The system was expressed by three separate plasmids containing (1) the transposition machinery, tnsA-tnsB-tnsC; (2) the CRISPR machinery, tniQ-cas8-cas7-cas6, along with a synthetic CRISPR array; and (3) a donor DNA that can be recognized by the transposition machinery. Utilizing a CRISPR array with CRISPR RNAs (crRNAs) that bind downstream of the system’s known CC protospacer-adjacent motif (PAM) and near the natural attTn7 site, the authors first show through PCR that this CRISPR-Cas system can perform targeted insertions similar to natural Tn7-like transposition. They systematically altered the construct designs and observed that transposition required all proteins (Cas6, Cas7, Cas8, TnsA, TnsB, TnsC, TniQ), the transposition machinery binding sites flanking the donor DNA, and catalytic activities of TnsB/TnsC/Cas6 proteins. To provide biochemical evidence for the essentiality of these proteins, the authors pulled down His-tagged TniQ, which copurified with Cas6, Cas7, and Cas8 and revealed that TniQ associates with other proteins that are part of the canonical Cascade complex (Peters et al., 2017).

Extending the system to contain crRNAs that bind lacZ revealed the programmability of the transposition. The authors further designed crRNA that could bind to 16 arbitrary intergenic sites on the E. coli genome and found that every site created a downstream (46–55 base pairs), RNA-guided DNA transposition. This transposition occurred in either a forward or reverse direction depending on the crRNA used, indicating that more characterization is needed to precisely control integration.

Strecker et al. demonstrated site-directed transposition with a single Cas effector using the type V-K system, which they named CAST (CRISPR-associated transposase). Often, class II systems require a transacting CRISPR RNA (tracrRNA) that binds crRNAs to function. In light of this, the authors sequenced the small RNAs of the cyanobacteria Syctonema hoffmani and Anabeana cylindrica,
host cells that express the CAST systems, and deduced the sequences of the tracrRNA and crRNAs. They engineered a three-plasmid system in E. coli: (1) a helper plasmid containing TnsB, TnsC, TniQ, the CAST effector (Cas12k), the tracrRNA, and a synthetic crRNA array; (2) a donor plasmid containing DNA flanked by transposition machinery binding sites; and (3) a target plasmid library containing a synthetic protospacer sequence flanked by a short random motif upstream of the protospacer. With this system they defined an optimal NGTN (N is any nucleotide) PAM for both S. hoffmani and A. cylindrica Cas12k effectors. They modified the S. hoffmani CAST (shCAST) system and demonstrated that TnsB, TnsC, TniQ, Cas12k, and the tracrRNA are all needed for transposition.

Engineered class 2 systems such as Cas9 use a single guide RNA (sgRNA) by fusing the tracrRNA and crRNA via a short hairpin-loop structure (Knott and Doudna, 2018). Strecker et al. applied this principle to shCAST and demonstrated comparable insertion efficiencies to the natural system using an sgRNA to the tracrRNA/crRNA complex. They purified the four shCAST proteins, reconstituted the system in vitro with an sgRNA or tracrRNA/crRNA, and showed that transposition could occur with same plasmid system and that Cas12k was not capable of DNA cleavage in any conditions tested.

By designing and testing sgRNAs that target various sites on the E. coli genome, the authors confirmed that shCAST is capable of programmable transposition. Of the 48 sites tested, 29 of them created transpositions detectable with PCR. The integrations occurred 60–66 bp downstream of the sgRNA binding site while leaving the transposition binding sites intact. Interestingly, these integrations occur unidirectionally, unlike those of the type I-F system.

Both studies observed no induced transposition when simply co-expressing catalytically dead class II (dCas9) or nuclease-free class I CRISPR-Cas (Cascade without Cas2 or Cas3) systems with the transposase proteins. This validates that the transposition mechanism is dependent on the formation of the type I-F Cascade-TniQ complex or association of Cas12k, TniQ, TnsB, and TnsC and not simply the formation of an R-loop in the presence of transposon machinery. Future studies defining the biochemical and structural mechanisms used by the systems will better elucidate the molecular mechanism for transposition. Importantly, the observation that CRISPR-Cas systems can perform natural transposon mobilization and not nuclease DNA cutting implies that our understanding of the CRISPR-Cas mechanisms is far from complete.

Applying these systems to achieve efficient targeted DNA insertions in eukaryotic cells holds tremendous potential for genome engineering. Current methods to perform targeted gene knockins largely rely on homology-directed repair (HDR), which requires introducing site-specific double-stranded breaks in the presence of donor DNA, often providing limited efficacy. Transposon-linked CRISPR-Cas systems that do not rely on HDR could dramatically improve the efficiency and insertion size of these knockins, significantly improving our ability to perform gene therapy. For example, diseases such as Duchene’s Muscular Dystrophy are often caused by partial gene deletion. Transposon-based CRISPR-Cas systems could allow efficient and targeted DNA insertion for gene repair. Furthermore, cell therapies such as CAR T cells often require stable insertion of gene payloads in the genome to create therapeutic functionality. Transposon-based CRISPR systems could insert large payloads such as the CAR cassette into a safe harbor locus.
in the genome. These tasks have been limited using nuclease CRISPR systems but could potentially benefit from the transposon CRISPR systems. Overall, discoveries of targeted transposition in type I-F and type V-K subtypes provide a new mechanism for CRISPR-Cas biology and shed light for overcoming existing hurdles faced by genome engineering.

REFERENCES


