Structural biology

Enemies join forces for gene insertion

Orsolya Barabas & Phoebe A. Rice

CRISPR–Cas is a bacterial defence system that can attack invading DNA to protect host cells, or help to insert DNA safely into the genome. Structures of this latter type of CRISPR–Cas system have now been visualized. **See p.775**

Cooperation between rival forces changed the course of history to secure an Allied victory in the Second World War. Now two studies, one by Park *et al.*[1] on page 775 and one by Schmitz *et al.* in *Cell*[2], reveal the essence of enemy cooperation in a different setting: the battle between bacterial defence systems called CRISPR–Cas and DNA from invaders. The authors visualize structures of a CRISPR–Cas complex that steers invading DNA to sites in the bacterial genome that suit both the invader and the host. The ability to add DNA sequences to specific genomic sites is valuable for modifying cells and organisms according to scientists' requirements. As such, the new blueprints are great news for research.

CRISPR–Cas systems eliminate invading 'parasitic' DNA, including viral DNA or transposons (DNA sequences that replicate themselves and move around genomes)[3]. These systems store short pieces of invading DNA, and use an RNA copy of the stored sequence as a guide to help a Cas protein identify and cut up the invading DNA, thereby preventing it from damaging the cell.

Although CRISPR–Cas has been widely repurposed as a tool for introducing modifications into genomes[4], it is generally not well suited for gene insertion. In fact, transposons themselves are more-efficient tools for gene addition[5], because they encode a transposase enzyme that catalyses movement of transposon DNA to new genomic sites[6]. However, most transposases insert their payloads at arbitrary locations in the genome, which can harm the host cell.

In 2017, a class of CRISPR-associated transposon (CAST) was discovered that acts with CRISPR in an unusual way[7]. Rather than cleaving the invading DNA, special Cas–RNA complexes, co-opted by a CAST, guide that CAST's transposition. The DNA sequence into which CASTs will be inserted can be predicted (to within 5–10 base pairs) by the guide RNA sequence, and can be reprogrammed by changing that sequence[8,9]. This specialized CRISPR–Cas action steers CASTs into 'safe' locations in the genome and thereby avoids killing the host, while enabling propagation of the CAST. The host cell can also benefit from other cargo carried on the transposon, such as antibiotic-resistance genes. The potential of these systems for programmable DNA insertion was immediately exploited in bacteria[10,11], but so far there is no evidence for CAST activity in nucleus-bearing eukaryotic cells (human cells, for example).

Studies have identified and characterized several CAST components, and have described structures for parts of the CAST machinery. These structures include: the complex of Cas with guide RNA and target DNA[12–14] (into which the transposon integrates); a TniQ, 'adapter' protein[12,14,15], which binds to the Cas–RNA–DNA complex and recruits other elements of the transposition machinery; TnsC proteins[14,15], which polymerize into spiral-like filaments along the target DNA and recruit the transposase; and TnsB (the transposase)[16,17].

However, little has been known about how all these components cooperate. In the current studies, Park *et al.*[1] and Schmitz *et al.*[2] used cryo-electron microscopy to describe the atomic structures of complete CRISPR–Cas-targeted CAST complexes, revealing how all the pieces come together (Fig. 1).

Both groups scrutinized a small CAST, dubbed *Sh*CAST, isolated from the blue-green aquatic bacterium *Scytonema hofmannii*. Schmitz and colleagues reconstructed a complex comprising the 'targeting module' of the complex (the Cas protein Cas12k and a guide RNA that was anchored to its target DNA), along with TniQ and TnsC. The structure reveals that, as expected, TniQ forms a bridge between the guide RNA and the start of a TnsC minifilament. Unexpectedly, the complex also contains a host protein, S15. This protein is known to be part of the bacterial ribosome (the molecular machinery that produces proteins). The authors used careful biochemistry to show that it is integral to helping TniQ bind to the Cas–RNA complex and recruit TnsC.

Park and colleagues visualized the final product of RNA-guided transposition, which involves an 'insertion module' comprising four subunits of TnsB bound to both ends of the transposon DNA, in addition to the other complex components. Their structures show how all the components of *Sh*CAST come together to combine the RNA-guided targeting of CRISPR–Cas systems with the insertion capability of transposases.

Both reports show that the whole is

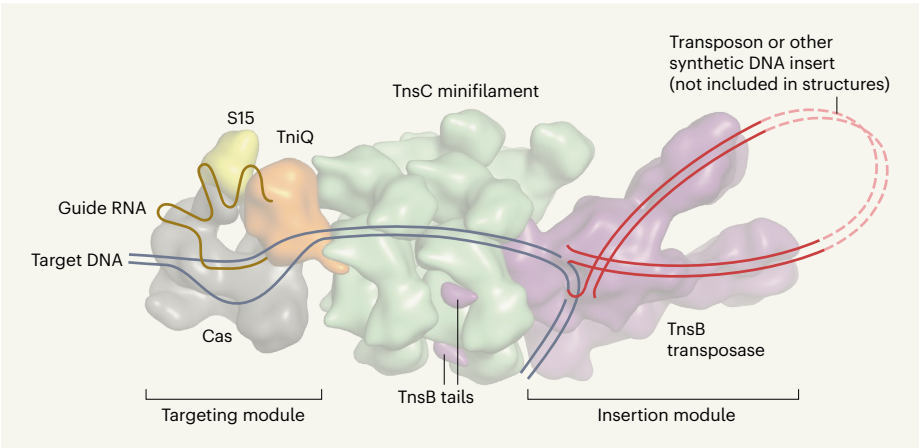

Figure 1 | CRISPR–Cas guides insertion of transposon DNA. Mobile DNA sequences called transposons can invade bacteria and propagate within their genomes. Although many transposons are inserted at random locations (not shown), one type of transposon, called CAST, uses its own CRISPR–Cas complex to steer insertion. Two research groups[1,2] have now resolved high-resolution structures of this complex. A 'targeting module' involves a Cas protein, which recognizes a particular DNA sequence that matches a guide RNA. Insertion of the transposon is mediated by an 'insertion module' involving the transposase enzyme (TnsB), which binds to the ends of the transposon. The targeting module and TnsB are brought together by TniQ and S15 proteins, and a minifilament made of several TnsC proteins, which anchors TnsB through the latter protein's tail domains. (Proteins are shown as smoothed transparent surfaces based on the structures from refs 1 and 2, and nucleic acids are overlaid in a simplified schematic form intended to indicate their general positions.)

more than the sum of the parts. In previous Cas12k–RNA–DNA structures[13,14], the DNA was only partially paired with the RNA — but the new work shows that pairing is complete when TniQ, TnsC and S15 are present. The latest structures also show that the strand of the DNA duplex in closest contact with TnsC in the full assembly is different from what was seen in previous structures of isolated TnsC–DNA complexes[14,15].

Mutual interactions between TnsC and TnsB are a key aspect of transposon targeting[18]. The TnsC filament is trimmed down to size by TnsB, to bring the transposase into close proximity with the target DNA sequence specified by the guide RNA. In turn, the catalytic activity of TnsB (unlike many transposases that do their job alone) depends on interactions with TnsC.

A surprising observation from Park and colleagues' structures was that the number of TnsC molecules in the minifilament was variable (12 or 13). This flexibility suggests how a system that can so accurately target a particular DNA sequence can also show 5–10 base pairs of variability in the insertion site selected. Such a feature might reflect an adaptation to the natural battlefield, enabling small variations in CAST insertion that minimize the harm of its addition to the host. Efforts are now needed to describe the dynamics of TnsC assembly and disassembly, and to elucidate TnsB's exact role in it.

How is transposase activity controlled? One might have expected that TnsC would alter the conformation of TnsB to activate its catalytic core. Therefore, it is interesting that the TnsB structure Park *et al.* observed in the full complex is similar to that of TnsB without TnsC[16,17]. Moreover, the authors found no contacts between TnsC and the catalytic core of TnsB that could activate TnsB.

Clues to how TnsC activates TnsB might be found by considering a related transposase called MuA. For both TnsB and MuA, the target DNA must strongly bend to fit into the active pocket of the transposase. Making such a strong bend in DNA is energetically difficult, and factors that stabilize bent target DNA enhance MuA transposition[19]. Perhaps TnsC activates transposition, at least in part, by bringing bent target DNA to TnsB. In fact, TnsB might not be able to bind to natural target DNA in the absence of TnsC, which tethers it to DNA. Moreover, in the structure reported by Park *et al.*, the TnsC minifilament is positioned to contact target DNA on both sides of TnsB, perhaps to stabilize its bent form. Finally, the authors observed that interactions between TnsB and TnsC caused a disordered segment of TnsB to fold and dock in a way that might help to stabilize bent target DNA.

Together, these two papers highlight how CRISPR and transposons — which conventionally oppose one another — can join forces to strategically choose target locations for inserting large DNA payloads. The molecular views provided will help genome engineers to develop CAST-based gene-insertion systems with more-tailored target locations, and could help researchers to design variants of the system that exhibit less off-target activity.

The unexpected discovery that the bacterial S15 protein is an intrinsic *Sh*CAST component could prove to be a breakthrough in achieving large targeted insertions in eukaryotic genomes. Schmitz and colleagues show that S15's human relative does not promote *Sh*CAST activity, which could explain why previous attempts to adapt the system for use in eukaryotic cells have failed. The ability of bacterial S15 to boost *Sh*CAST activity in eukaryotes will surely be tested soon.

Finally, the CAST system could potentially be used as a blueprint to establish programmable targeting of simpler transposons. This could aid CRISPR-mediated gene insertion in diverse cells and organisms — an exciting prospect for future research.

Orsolya Barabas is in the Department of Molecular and Cellular Biology, University of Geneva, 1211 Geneva, Switzerland. **Phoebe A. Rice** is in the Department of Biochemistry and Molecular Biology, The University of Chicago, Chicago, Illinois 60637, USA.
e-mails: orsolya.barabas@unige.ch; price@uchicago.edu

1. Park, J.-U. *et al. Nature* **613**, 775–782 (2023).
2. Schmitz, M., Querques, I., Oberli, S., Chanez, C. & Jinek, M. *Cell* **185**, 4999–5010 (2022).
3. Hille, F. *et al. Cell* **172**, 1239–1259 (2018).
4. Knott, G. J. & Doudna, J. A. *Science* **361**, 866–869 (2018).
5. Ivics, Z. *et al. Nature Methods* **6**, 415–422 (2009).
6. Hickman, A. B. & Dyda, F. *Microbiol. Spectr.* **3**, MDNA3-0034-2014 (2015).
7. Peters, J. E., Makarova, K. S., Shmakov, S. & Koonin, E. V. *Proc. Natl Acad. Sci. USA* **114**, E7358–E7366 (2017).
8. Klompe, S. E., Vo, P. L. H., Halpin-Healy, T. S. & Sternberg, S. H. *Nature* **571**, 219–225 (2019).
9. Strecker, J. *et al. Science* **365**, 48–53 (2019).
10. Vo, P. L. H. *et al. Nature Biotechnol.* **39**, 480–489 (2021).
11. Rubin, B. E. *et al. Nature Microbiol.* **7**, 34–47 (2022).
12. Halpin-Healy, T. S., Klompe, S. E., Sternberg, S. H. & Fernández, I. S. *Nature* **577**, 271–274 (2020).
13. Xiao, R. *et al. Mol. Cell* **81**, 4457–4466 (2021).
14. Querques, I., Schmitz, M., Oberli, S., Chanez, C. & Jinek, M. *Nature* **599**, 497–502 (2021).
15. Park, J.-U. *et al. Science* **373**, 768–774 (2021).
16. Park, J. U., Tsai, A. W.-L., Chen, T. H., Peters, J. E. & Kellogg, E. H. *Proc. Natl Acad. Sci. USA* **119**, e2202590119 (2022).
17. Tenjo-Castaño, F. *et al. Nature Commun.* **13**, 5792 (2022).
18. Skelding, Z., Sarnovsky, R. & Craig, N. L. *EMBO J.* **21**, 3494–3504 (2002).
19. Fuller, J. R. & Rice P. A. *eLife* **6**, e21777 (2017).

O.B. declares competing interests. See go.nature.com/3jqejfi for details.
This article was published online on 11 January 2023.

Catalysis

Machine learning classifies reaction mechanisms

Danilo M. Lustosa & Anat Milo

The study of how chemical reactions work is key to the design of new reactions, but relies on hard work and expert knowledge. A machine-learning tool has been developed that could change the way this challenge is approached. **See p.689**

The discovery of chemical reactions is influenced not only by how fast experimental data can be acquired, but also by how easily chemists can make sense of these data. Unravelling the mechanistic underpinnings of new catalytic reactions is a particularly intricate problem, often requiring expert knowledge of computational and physical organic chemistry. Nevertheless, it is important to study catalytic reactions because they represent the most efficient chemical processes. On page 689, Burés and Larrosa[1] report a machine-learning model that classifies the mechanisms of catalytic reactions on the basis of the time-course signatures of the reactions. This method could streamline the investigation of reaction mechanisms and requires minimal experimental effort.

The determination of catalytic-reaction mechanisms involves collecting a plethora of clues about how starting materials come together and interact with a catalyst and each other to form products. One of the most powerful techniques for drawing hypotheses from experimental data is to analyse the consumption of starting materials and the formation of products over time, a process called kinetic analysis[2]. These rates of reactant decay and product formation are best described by equations known as rate laws.

The basic premise behind these laws is that the rate of a chemical transformation depends on the concentration of reagents, the number of reagent molecules involved in each step of the reaction, and how much energy is needed to transform the reactants to transient intermediates that then form the products. Disentangling complex rate laws can lead to