



STAGE: A compact and versatile TnpB-based genome editing toolkit for *Streptomyces*

Jing Luo^{a,1}, Natalie Chia^{a,1} , Yuxi Qin^{a,1}, Pan Tan^{b,c,1}, Lingwen Zhang^a, Sihan Yang^a, Zihan Yuan^a, Liang Hong^{b,c,d}, Sang Yup Lee^{e,f,g,2} , and Yaojun Tong^{a,2}

Affiliations are included on p. 12.

Edited by George Chen, Tsinghua University, Beijing, China; received April 15, 2025; accepted July 27, 2025 by Editorial Board Member Jens Nielsen

Streptomyces are naturally endowed with the capacity to produce a wide array of natural products with biomedical and biotechnological value. They have garnered great interest in synthetic biology applications given the abundance of uncharacterized biosynthetic gene clusters (BGCs). However, progress has been hindered by the limited availability of genetic tools for manipulating these bacteria. Several representative CRISPR-Cas systems have been established in *Streptomyces* to streamline experimental workflows and improve editing efficiency. Nevertheless, their broader applicability has been constrained by issues such as nuclease activity-related cytotoxicity and the large size of effector proteins. To address these challenges, we present *Streptomyces*-compatible TnpB-assisted genome editing (STAGE), a genetic toolkit based on ISDra2 TnpB, which is approximately one-third the size of Cas9 and enables precise, site-specific gene editing. We demonstrated that STAGE introduces genetic mutations with high efficiency and minimal off-target effects in two industrially important *Streptomyces* strains. Building on this platform, we developed STAGE-cBEST and STAGE-McBEST, enabling single and multiplexed C-G-to-T-A base editing, respectively, with editing efficiencies exceeding 75%. To further enhance performance, we engineered the ISDra2 TnpB system using an AI-assisted protein engineering framework, resulting in two variants that achieve nearly 100% genome editing efficiency. Additionally, through sequence homology analysis, we identified a TnpB ortholog from the same biological origin of ISDra2 TnpB, which also functions effectively as a gene editing tool. Our study establishes STAGE as a highly precise, programmable, and versatile genome editing platform for *Streptomyces*, paving the way for advanced genetic manipulation and synthetic biology applications in these industrially important bacteria.

TnpB | *Streptomyces* | base editing | genome editing | multiplexed genome editing

Streptomyces are gram-positive bacteria with high gene cluster (GC)-content genomes, recognized as key producers of natural products with significant biological activities across medicine, agriculture, food, and environment industries (1). Advances in whole-genome sequencing (WGS) have revealed that *Streptomyces* still holds immense potential for discovering new natural products, as it harbors numerous uncharacterized biosynthetic GCs (BGCs) (1, 2). Mining these unexplored BGCs could greatly enhance the discovery, characterization, engineering, and production of novel bioactive natural products. Consequently, efficient genetic manipulation tools are essential for unlocking these uncharacterized BGCs. In the past decade, RNA-guided CRISPR, CRISPR-Cas/CRISPR-associated protein (Cas) systems, including Cas9- and Cas12-based systems, have become the primary genome editing tools in *Streptomyces*, greatly improving the efficiency and convenience of genome editing compared to traditional recombination-based approaches (3–8). However, several challenges remain. First, the efficient DNA cleavage by Cas introduces DNA double-stranded breaks (DSBs) that can lead to chromosomal rearrangements, genomic instability, and even cell death (cytotoxicity). Second, Cas proteins are typically large, exceeding 1,300 amino acids (aa), posing difficulties in plasmid construction and reducing transformation efficiency in *Streptomyces*. Therefore, there is a critical need for compact, highly efficient, low off-target effects, and versatile genome editing systems tailored to *Streptomyces* to support the next generation of synthetic biology-driven discovery of bioactive natural products.

Recently, Obligate Mobile Element Guided Activity (OMEGA) systems, which feature compact endonucleases (approximately 400 aa) derived from ancestral Cas proteins, have been characterized for their ability to cleave double-stranded DNA (dsDNA) and have been successfully developed into effective genome editing tools (9, 10). Among these, the transposon (IS200/IS605)-encoded endonuclease TnpB, an evolutionary precursor of

Significance

Streptomyces species are essential biofactories to produce diverse natural products with significant pharmaceutical and industrial value; however, their genetic engineering has been hampered by the limited availability and efficiency of existing tools. Here, we present *Streptomyces*-compatible TnpB-assisted genome editing (STAGE), a compact and highly efficient TnpB-based genome editing platform specifically designed for *Streptomyces*. By leveraging a compact, programmable TnpB nuclease and AI-assisted optimization, STAGE overcomes key limitations of conventional CRISPR systems, including large effector protein size and cytotoxic off-target effects. This versatile toolkit enables a wide range of genome modifications, including targeted indels, precise base editing, multiplexed gene editing, and large-fragment deletions. STAGE substantially expands the synthetic biology toolbox for *Streptomyces* and accelerates their exploitation for biotechnological innovation.

The authors declare no competing interest.

This article is a PNAS Direct Submission. G.C. is a guest editor invited by the Editorial Board.

Copyright © 2025 the Author(s). Published by PNAS. This article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](#).

¹J.L., N.C., Y.Q., and P.T. contributed equally to this work.

²To whom correspondence may be addressed. Email: leesy@kaist.ac.kr or yaojun.tong@sjtu.edu.cn.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2509146122/-/DCSupplemental>.

Published August 26, 2025.

Cas12, has emerged as a promising candidate for genome editing due to its small size and the vast potential of unexplored miniature genome editors arising from its widespread distribution across organisms (10–12). ISDra2 TnpB, derived from *Deinococcus radiodurans* and one of the relatively well-studied TnpB effectors, has demonstrated effective application in mammalian cells, *Escherichia coli*, and plants (10, 12, 13). However, the potential of TnpB systems for genome editing in *Streptomyces* remains largely unexplored and requires systematic evaluation.

In this study, we used ISDra2 TnpB (408 aa), approximately one-third the size of commonly used Cas nucleases such as SpCas9 and FnCas12a, as a scaffold to develop *Streptomyces*-compatible TnpB-assisted genome editing (STAGE), a STAGE toolkit. STAGE is designed to meet diverse gene and BGC editing needs in *Streptomyces*, including the creation of random-sized deletion libraries, induction of indel mutations, execution of in-frame knock-outs and knock-ins, base editing, and multiplexed gene engineering. We assessed the fidelity of STAGE through WGS and confirmed that TnpB-mediated genome editing in *Streptomyces* is highly efficient, with negligible off-target effects. To enable effective genome editing, we systematically optimized STAGE through AI-assisted TnpB engineering, reRNA engineering, and spacer length optimization. Furthermore, we adapted TnpB* from *D. radiodurans*, an ISDra2 TnpB ortholog, into *Streptomyces coelicolor* M145 and assessed its genome editing performance. These explorations, including reRNA sequence exchange and swapping of potential TAM recognition domains, further facilitated the development of an enhanced version of STAGE with significantly expanded targeting capabilities. Overall, this study presents a comprehensive, powerful, and precise genome editing toolkit, paving the way for the systematic exploration and engineering of valuable natural products in *Streptomyces*.

Results

Design and Construction of STAGE. TnpB proteins from the IS200/IS605 transposon family are considered ancestors of the Class 2 type V CRISPR effector Cas12 (14). Recent studies have identified TnpB (approximately 400 aa) as a compact RNA-guided DNA endonuclease that associates with a single long noncoding RNA (referred to as reRNA or ω RNA) to cleave dsDNA in a Target Adjacent Motif [TAM, originally termed as transposon-associated motif (10)]-dependent manner (Fig. 1A) (10, 15). The size of TnpB is significantly smaller than the CRISPR-Cas endonucleases currently used in *Streptomyces* (Fig. 1B). To expand the genetic manipulation toolkit for *Streptomyces*, we developed a TnpB-based genome editing toolkit, STAGE for *Streptomyces*, and it would be evaluated in *S. coelicolor* and *Streptomyces lividans*.

Based on previous studies showing that reRNA overlaps the 3' region of tnpB genes in bacteria and archaea (20, 21), we constructed three different architectures of ISDra2 TnpB-based genome editing components: one mimicking the natural composition (pArch.1), one separating TnpB and reRNA (pArch.2), as well as the one which is the same as pArch.2 except that a ribozyme is placed 3' to the spacer (pArch.3) (Fig. 1C). Specifically, the pArch.1 construct carries *S. coelicolor* M145 codon-optimized TnpB, the nonoverlapped reRNA, spacer sequence, and HDV (hepatitis delta virus ribozyme) in sequential order. In contrast, pArch.2 harbors the codon-optimized, full-length TnpB encoding gene and target-specific reRNA cassette (reRNA scaffold-spacer sequence-HDV), with each located in separate regions. pArch.3 is the same with pArch.2 except that the 3'-guide region extended continuously without HDV. In a previous study, we have demonstrated that *S. coelicolor* employs three DSB repair pathways (3).

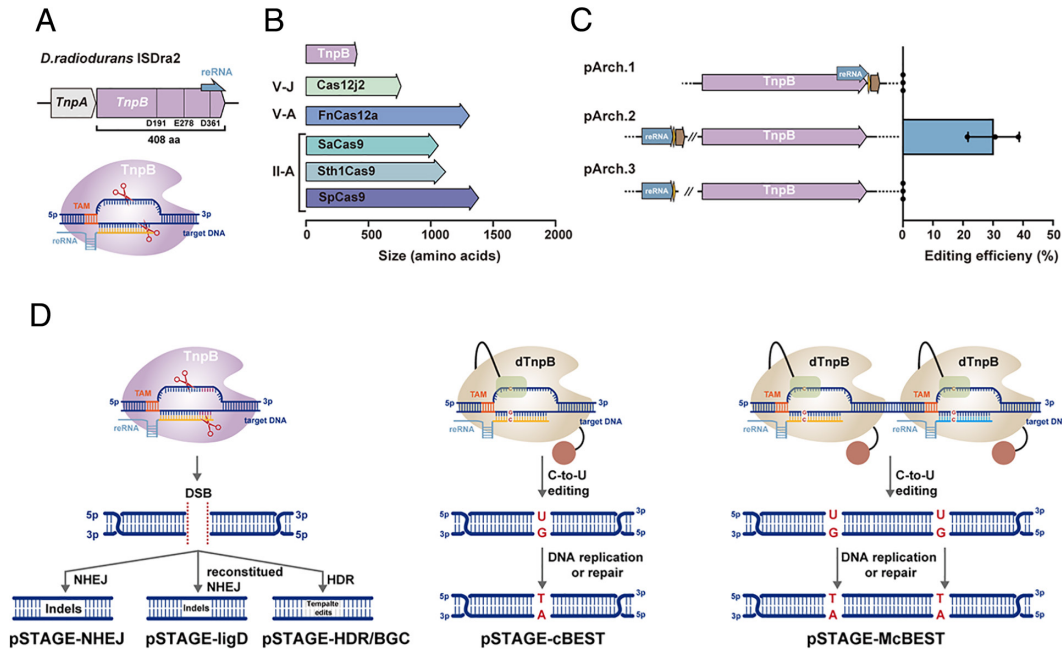


Fig. 1. TnpB as a versatile tool for *Streptomyces* genome editing. (A) Schematic representation of the *D. radiodurans* ISDra2 genomic structure and the TnpB system, illustrating its role in dsDNA cleavage. (B) Comparison of TnpB with CRISPR Class 2 RNA-guided endonucleases used for *Streptomyces* genome editing, including SpCas9 (*Streptococcus pyogenes* Cas9) (3); SaCas9 (*Staphylococcus aureus* Cas9) (16); Sth1Cas9 (*Streptococcus thermophilus* Cas9) (17); Cas12a (FnCpf1) (*Francisella tularensis* subsp. *novicida* U112 Cas12a) (18); and Cas12j2 (*Acidaminococcus* sp. Cas12j) (19). (C) Schematics and editing efficiency of TnpB constructs used for genome editing in *S. coelicolor* M145. (D) Mechanism of STAGE-mediated genome editing, highlighting targeting of single and multiple genes. Plasmids used in the DSB-dependent tools (pSTAGE-NHEJ, pSTAGE-ligD, and pSTAGE-HDR/pSTAGE-BGC, pSTAGE means the plasmid of the STAGE system) generate mutants via three distinct repair pathways: Error-prone NHEJ and reconstituted NHEJ, which induce random indels; HDR, which requires a donor repair template to introduce precise in-frame modifications. In contrast, pSTAGE-cBEST (cytosine Base Editing SysTem) and pSTAGE-McBEST (multiplexed cytosine Base Editing SysTem) are plasmids used in the DSB-independent tools, introducing C-to-T substitutions at one or multiple loci, respectively. Data are shown as means $\pm$ SD. For statistical details, see [Dataset S1](#) and [SI Appendix](#).

To exploit these pathways, accordingly, we designed five subsystems using programmable TnpB endonucleases (Fig. 1D).

As illustrated in Fig. 1D, pSTAGE-NHEJ is a subsystem that engages the native, compromised nonhomologous end joining (NHEJ) repair pathway, leading to the generation of random-sized insertions and deletions (indels) surrounding the target sequence. This strategy enables rapid gene disruption and efficient deletion of genes or gene clusters, useful for functional gene screening (3, 18). By coexpressing a DNA ligase D homolog from *S. carneus*, we can reconstitute the compromised NHEJ pathway (3), resulting in the pSTAGE-ligD subsystem, which enables more precise DSB repair. The pSTAGE-HDR and pSTAGE-BGC subsystems leverage the homology-directed repair (HDR) pathway, enabling precise genome editing through the use of exogenous homologous templates. These subsystems facilitate targeted gene or large fragment deletions, replacements, and site-specific mutations. Except these DSB-dependent editing systems, we also developed pSTAGE-cBEST and pSTAGE-McBEST, which are capable of performing base editing at single or multitarget sites, respectively.

Validation of STAGE with Different DSB Repair Machineries in *Streptomyces*. To assess the robustness of the system for genome editing, we systematically evaluated its editing efficiency under different DSB repair conditions. We selected SCO5087, which encodes the alpha subunit of a polyketide beta-ketoacyl synthase, the enzyme essential for the blue-pigmented actinorhodin biosynthesis in *S. coelicolor*. Disrupting SCO5087 results in a nonblue phenotype, indicating a loss of actinorhodin production. The STAGE constructs targeting 20-nt sites adjacent to the 5'-TTGAT TAM in SCO5087 were introduced into *E. coli* ET12567/pUZ8002 for conjugation. After conjugation, randomly selected colonies were analyzed for mutations (SI Appendix, Fig. S1).

Colony PCR and Sanger sequencing confirmed that no detectable indels occurred when using the pArch.1 system in *S. coelicolor* M145 (Fig. 1C), indicating a lack of in vivo genome editing activity. This deficiency is most likely due to insufficient production of functional reRNA scaffolds. Supporting this interpretation, previous in vitro studies have shown that the TnpB ortholog from *Alicyclobacillus macrorosporangioides* (AmaTnpB) possesses intrinsic RNA processing activity, enabling the generation of reRNA molecules required for target DNA cleavage (22). Notably, only weak but specific DNA cleavage was observed in gel-based assays when the full-length AmaTnpB operon, harboring a nonoverlapping reRNA and guide sequence structurally analogous to our pArch.1, was expressed in the absence of exogenous AmaTnpB protein. This guide-dependent cleavage was clearly distinct from the nonspecific degradation patterns seen with size-matched random negative controls. Supplementation with purified AmaTnpB protein increased DNA cleavage efficiency, but the activity still remained lower than in reactions where both reRNA and protein were provided separately. These findings suggest that, in *Streptomyces*, the pArch.1 configuration does not support sufficient reRNA biogenesis, which likely accounts for the absence of detectable genome editing activity in vivo.

Conversely, the pArch.2 system, wherein TnpB and reRNA were expressed separately, achieved an editing efficiency of approximately 30% under the native NHEJ pathway. This efficiency was notably higher than that of pArch.3, which lacks HDV ribozyme at the 3' spacer region (Fig. 1C). We subsequently used pArch.2 (renamed pSTAGE-NHEJ) to construct TnpB-mediated systems incorporating reconstituted NHEJ (pSTAGE-ligD) and HDR (pSTAGE-HDR) pathways (SI Appendix, Fig. S2). Notably, optimizing the NHEJ pathway by coexpressing a ligD gene increased editing efficiency from 30 to 44% in *S. coelicolor* M145 (Fig. 2D). Colony PCR and Sanger sequencing revealed a range of

random-sized indels under relative and reconstituted NHEJ conditions (Fig. 2C and SI Appendix, Fig. S3). These results align with the error-prone nature of the NHEJ (23) and our previous observation (3). Analysis of these indels showed that deletions were more frequent than insertions and substitutions, resembling the mutational profiles observed in human cells (10, 24). Importantly, when homologous recombination templates were provided, pSTAGE-HDR induced precise edits, resulting in editing efficiencies exceeding 71%, significantly surpassing the error-prone edits generated by NHEJ (Fig. 2B–D). As expected, no blue pigments were seen in the successfully edited colonies (Fig. 2A).

To further compare STAGE with pCRISPR-Cas9 (3), we targeted adjacent sites within SCO5087 under identical conditions. Under native NHEJ, both systems exhibited comparable efficiencies (Fig. 2D). However, under reconstituted NHEJ and HDR, STAGE showed slightly lower efficiency than Cas9, based on our previously reported data (3). To validate the versatility of STAGE, we applied pSTAGE for genome editing in a different *Streptomyces* strain, *S. lividans* TK24, and observed successful edits (Fig. 2E). Additionally, we observed successful gene editing with pSTAGE-NHEJ at multiple sites in *S. coelicolor* M145 (Fig. 2F and SI Appendix, Table S3). These results demonstrate that STAGE is a promising, versatile genome editing platform for *Streptomyces*, offering a compact and effective alternative to existing editing systems.

STAGE Is Capable of Deleting Complete BGCs. We next investigated whether pSTAGE could facilitate the deletion of large DNA fragments, like a complete BGC, which is a critical requirement for *Streptomyces* genome engineering. To evaluate its capability, we targeted BGCs of actinorhodin (*act* BGC, 17.3 kb) and undecylprodigiosin (*red* BGC, 31.7 kb) (Fig. 2G and SI Appendix, Table S3). First, we optimized the length of homologous arms (HAs, also referred to as homologous templates) for efficient large-fragment deletions. We constructed pSTAGE-based editing plasmids with HAs of approximately 1.0 kb, 1.5 kb, 2.0 kb, and 2.5 kb flanking *act* BGC (designed as pSTAGE-*act*BGC-1.0 kb, pSTAGE-*act*BGC-1.5 kb, pSTAGE-*act*BGC-2.0 kb, and pSTAGE-*act*BGC-2.5 kb, respectively), while a plasmid lacking HAs served as a control (SI Appendix, Fig. S4). Following introduction of these *act*BGC-editing plasmids into *S. coelicolor* M145, we observed that exconjugant numbers were lower compared to those transformed with the control plasmid (lacking HAs). Randomly selected exconjugants were restreaked onto antibiotic plates, and after 14 d of incubation, we found that increasing HA length correlated with a higher proportion of strains that failed to secrete blue pigment (SI Appendix, Fig. S4A). To evaluate editing efficiency and mutant purity, colony PCR analysis was performed using primers that anneal outside the HAs and within the target BGC for pSTAGE-*act*BGC-1.0 kb and pSTAGE-*act*BGC-1.5 kb (SI Appendix, Fig. S4B). Deletion efficiency increased from 0 to 50% when HAs were extended from ~1.0 kb to 1.5 kb, although mixed-genotype mutants remained prevalent (SI Appendix, Fig. S4C).

Considering the impact of plasmid size on transformation efficiency, we adopted the 1.5 kb HAs strategy for *red* BGC deletion, achieving an efficiency of ~11 to 16% (Fig. 2H). These findings confirm that our compact STAGE system, integrating TnpB and reRNA, is functional in *Streptomyces* and can be effectively applied for BGC-scale genome editing.

STAGE-Mediated C-G-to-T-A Base Editing in *Streptomyces*. DSB-dependent genome editing poses risks of genome instability and chromosomal rearrangements in *Streptomyces* (25–27). Base editing offers a precise and DSB-independent alternative, allowing targeted single-nucleotide modifications using DNA-binding

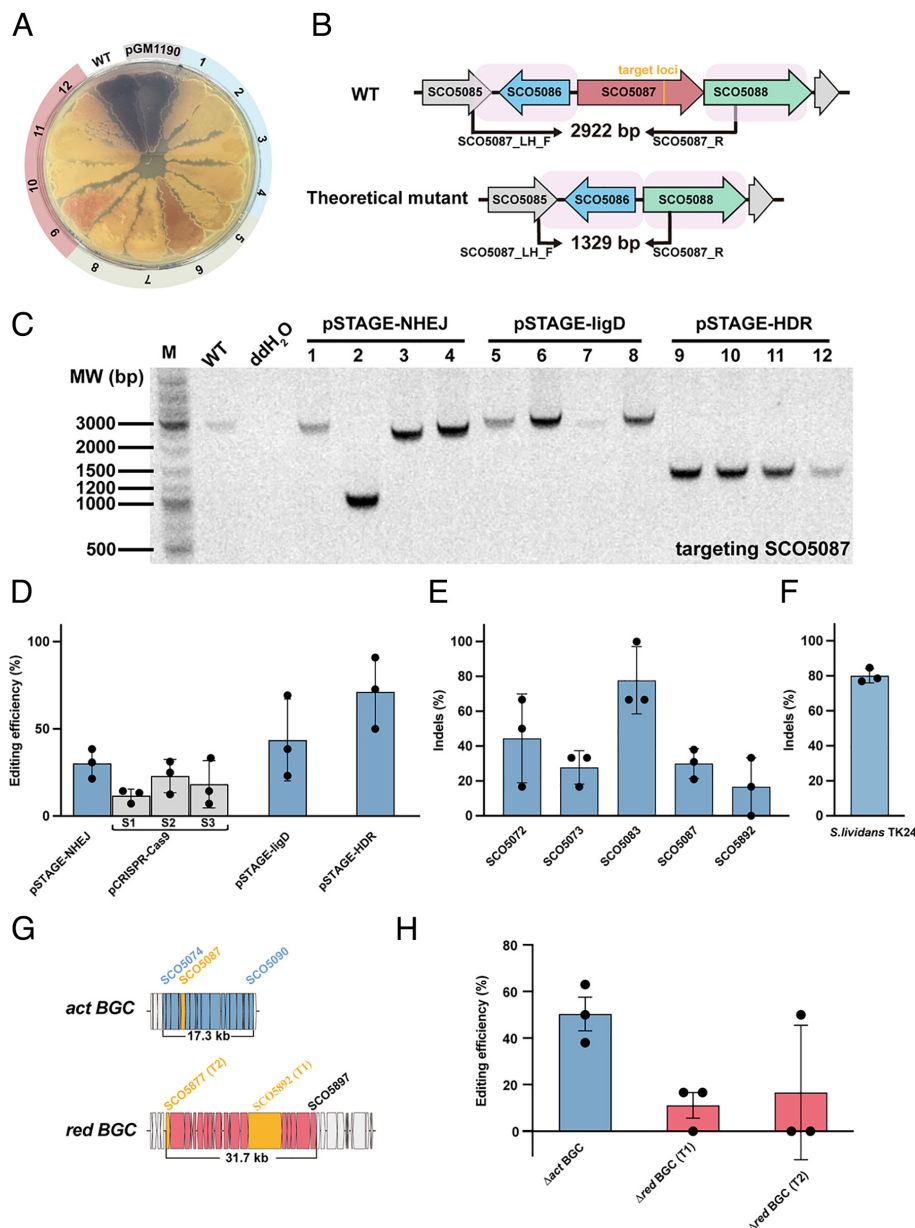


Fig. 2. Phenotypic and genotypic characterization of DSB-dependent STAGE-edited mutants. (A) Phenotypic comparison of *S. coelicolor* M145 (WT) and representative mutants generated using pSTAGE-NHEJ (1–4), pSTAGE-ligD (5–8), and pSTAGE-HDR (9–12), along with the vector control (pGM1190). (B) PCR-based strategy for evaluating pSTAGE-HDR editing efficiency. Primers were designed outside the upstream homologous template and within the downstream homologous template. Unedited loci yield a 2,922-bp PCR product (WT), while successful editing produces a 1,329-bp fragment. (C) Representative agarose gel electrophoresis of colony PCR products from *S. coelicolor* M145 (WT) and 12 randomly selected pSTAGE-edited colonies at the SCO5087 locus. PCR was performed using SCO5087_LH_F and SCO5087_R primers. (D) Assessment of editing efficiency of TnpB and Cas9 systems across different DNA repair pathways in *S. coelicolor* M145. Data are shown as means $\pm$ SD. (E) Target-specific editing efficiency of pSTAGE-NHEJ in *S. coelicolor* M145. Data are shown as means $\pm$ SD. (F) Editing efficiency of pSTAGE-NHEJ in *S. lividans* TK24. Data are shown as means $\pm$ SD. (G) Illustration of large-fragment deletions (17.3 kb and 31.7 kb) in *act* and *red* BGCs. Orange arrows indicate target genes recognized by the reRNA. (H) TnpB-mediated large-fragment deletions efficiency in *act* and *red* BGCs. Data are shown as means $\pm$ SEM. For statistical details, see [Dataset S1](#) and [SI Appendix](#).

proteins fused to deaminases. This approach employs scaffolds such as Cas9 nickase (D10A) or catalytically inactive Cas12 (dCas12), which tether cytosine or adenosine deaminases for single-base resolution genome editing (26, 28–31).

Here, we repurposed TnpB as a DNA binding scaffold for base editing and evaluated its activity in *S. coelicolor* M145. Unlike Cas9 and IscB nucleases, which utilize two distinct nuclease domains, TnpB relies on a single RuvC domain to cleave both DNA strands (10, 14, 32). The RuvC domain activity in TnpB is governed by three catalytic residues: D191, E278, and D361 (Fig. 3A). To create a catalytically inactive TnpB (dTnpB) that retains DNA binding ability while disabling dsDNA cleavage, we introduced alanine substitutions at these residues, generating

single, double, and triple mutants. Each variant was fused with an *S. coelicolor* codon-optimized rAPOBEC1 deaminase and uracil glycosylase inhibitor (UGI) at the N and C termini, respectively, forming a suite of dTnpB-based cytosine base editors (Fig. 3A).

To assess editing efficiency, we designed a reRNA cassette targeting SCO5087. All variants demonstrated robust C $\cdot$ G-to-T $\cdot$ A conversion, with varying efficiencies. Among them, dTnpB^{D191A} exhibited the highest conversion rate at both potentially editable cytosines ($97.01 \pm 1.72\%$ and $41.58 \pm 6.44\%$), whereas the triple mutant dTnpB^{D191A, E278A, D361A} (hereafter referred to as dTnpB^{DED}) showed the lowest conversion rate ($39.25 \pm 11.66\%$ and $1.60 \pm 1.77\%$) (Fig. 3B). Given its superior efficiency, the dTnpB^{D191A} was selected as the core enzyme for further optimization, leading to the development of

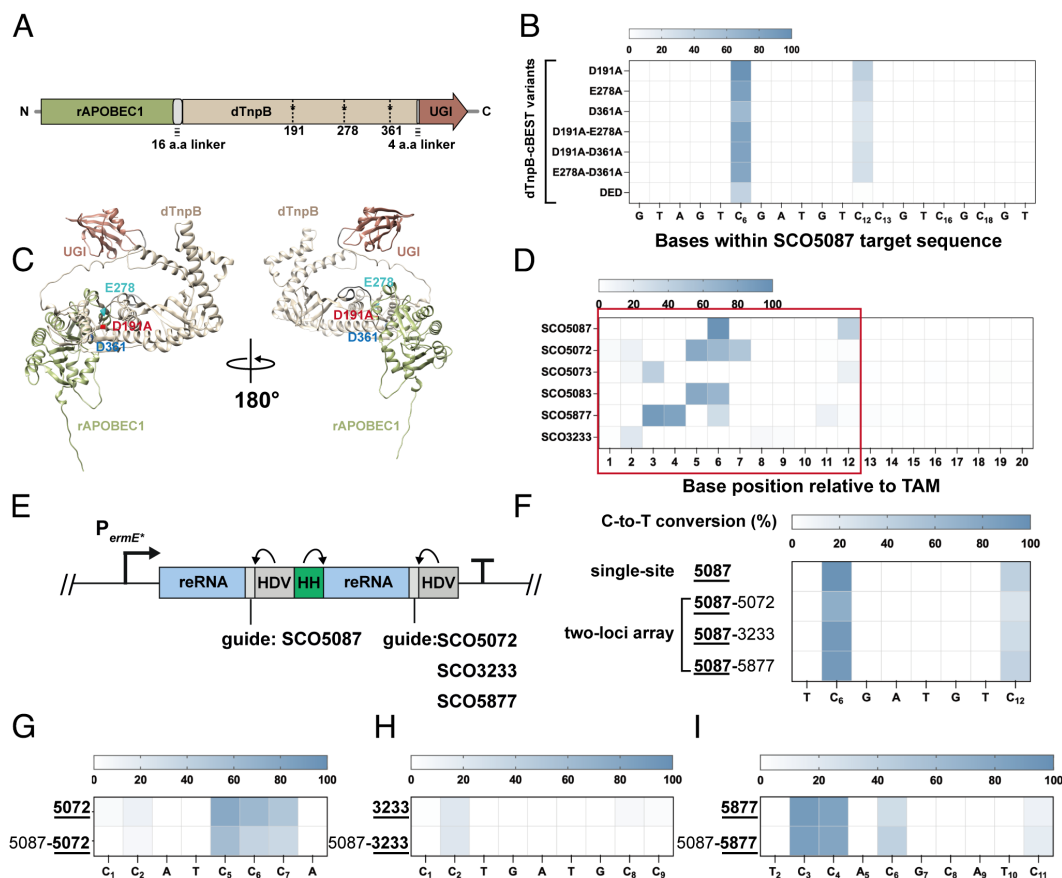


Fig. 3. Construction and characterization of pSTAGE-cBEST and pSTAGE-McBEST systems. (A) dTnpB-cytosine base editor configuration. dTnpB variants were generated by mutating three RuvC catalytic residues (D191, E278, and D361) to alanine. (B) Heatmap of C-G-to-T-A conversion rates for different dTnpB-cytosine base editor variants at the SCO5087 target site in *S. coelicolor* M145. (C) AlphaFold 3 (33)-predicted structural model of the rAPOBEC1-dTnpB^{D191A}-UGI fusion protein, highlighting domain architecture. The RuvC catalytic residues are shown as colored sticks. (D) Heatmap of C-G-to-T-A conversion rates across six endogenous sites edited using pSTAGE-cBEST. The base editing window spans positions 2 to 12, with peak activity at positions 3 to 7. (E) Schematic of the reRNA array used in pSTAGE-McBEST, where multiple reRNA units are concatenated in tandem and separated by a HH ribozyme within a single transcription cassette. (F–I) Heatmap of C-G-to-T-A conversion rates at SCO5087, SCO5072, SCO3233, and SCO5877 (redD) using pSTAGE-cBEST (single-site targeting) and pSTAGE-McBEST (two-locus multiplex editing). The top row of each heatmap represents the C-G-to-T-A conversion rates at individual target sites mediated by pSTAGE-cBEST. The subsequent rows depict the conversion rates at a given site when pSTAGE-McBEST is employed to simultaneously edit two distinct loci. Targeted sites displayed in the heatmaps are indicated by bold underlining. The proportions of colonies exhibiting single versus dual edits were originally presented in *SI Appendix, Figs. S6–S9*. Data are expressed as mean $\pm$ SD of three biological replicates. Base positions are numbered relative to the TAM (5'-TTGAT), with position 0 defined as the TAM. Conversion rates were analyzed using the BEAT program (34). For statistical details, see *Dataset S1* and *SI Appendix*.

pSTAGE-cBEST (cytosine Base Editing SysTem) (Fig. 3C). To define its editing window, we constructed five additional reRNAs targeting C-rich genomic sites. Results revealed that pSTAGE-cBEST exhibits an editing window spanning positions 2 to 12, with peak activity between positions 3 and 7 relative to the TAM site (position 0) (Fig. 3D). Furthermore, C-to-T conversion efficiency was influenced not only by position within the editing window but also by local sequence context, consistent with our previously characterized CRISPR-BEST sequence preferences (26) and the rAPOBEC1 deaminase (TC>CC>AC>GC) (35, 36) (*SI Appendix, Fig. S5*). Additionally, we note that the base editing activity of our construct is highly specific, consistently targeting only cytosines within the defined editing window and resulting exclusively in C-to-T conversions (*SI Appendix, Fig. S5*). Importantly, we did not observe any unintended substitutions, such as C-to-G or C-to-A transversions, which have been sporadically reported in previous studies of base editors (36).

Multiplexed Genome Editing with STAGE-McBEST. To enhance pSTAGE-cBEST for more efficient strain and metabolic engineering in *Streptomyces*, we developed pSTAGE-McBEST (Multiplexed cytosine Base Editing SysTem), enabling simultaneous genome editing at multiple loci. Multiple guide RNAs can be introduced in the vector in terms of separate expression cassettes (37–39) which

can be bulky, or encoded in tandem within a single expression cassette, relying on distinct processing mechanisms to release individual guide RNA units (26, 40, 41). Here, we opted for the latter, which is a more streamlined approach and leveraged self-cleaving ribozymes to process the reRNA array, incorporating Hammerhead (HH) ribozyme sequence as spacers between reRNA units (Fig. 3E). Upon transcription, HH and HDV ribozymes form distinct secondary structures, precisely self-cleaving at their 3' and 5' ends (40). This releases individual functional reRNAs, enabling pSTAGE-cBEST-mediated base editing at multiple loci simultaneously. To validate this multiplex editing approach, we constructed a dual-target reRNA array and compared pSTAGE-McBEST efficiency to that of the single-site system (Fig. 3E).

PCR genotyping and Sanger sequencing of randomly selected exconjugants confirmed that pSTAGE-McBEST efficiently converted C-G to T-A at both loci, achieving conversion rates comparable to single-site editing (Fig. 3F–I). Notably, some colonies exhibited edits at only one site (SCO5087), suggesting that incomplete or inefficient ribozyme cleavage may result in uneven reRNA release, compromising simultaneous base editing (40). Despite this, pSTAGE-McBEST achieved 75 to 100% editing efficiency at two loci (*SI Appendix, Figs. S6–S9*), demonstrating its feasibility for multiplex genome editing applications.

Impact of C-Terminal Truncation of TnpB on Genome Editing Activity in *Streptomyces*. Like Cas proteins, TnpB from *D. radiodurans* ISDr2 associated with its cognate reRNA to cleave dsDNA targets. Notably, the reRNA encoding region overlaps with the 3' end of the *tnpB* gene (residues 335 to 408

and -231G to -10U) (Fig. 4A). Previous studies suggested that the C-terminal domain (CTD) is dispensable for RNA-guided target DNA cleavage in TnpB proteins (42, 43) and in the close neighbor Cas12n nucleases (44). We hypothesized that truncating the 3' end of TnpB could generate a more compact

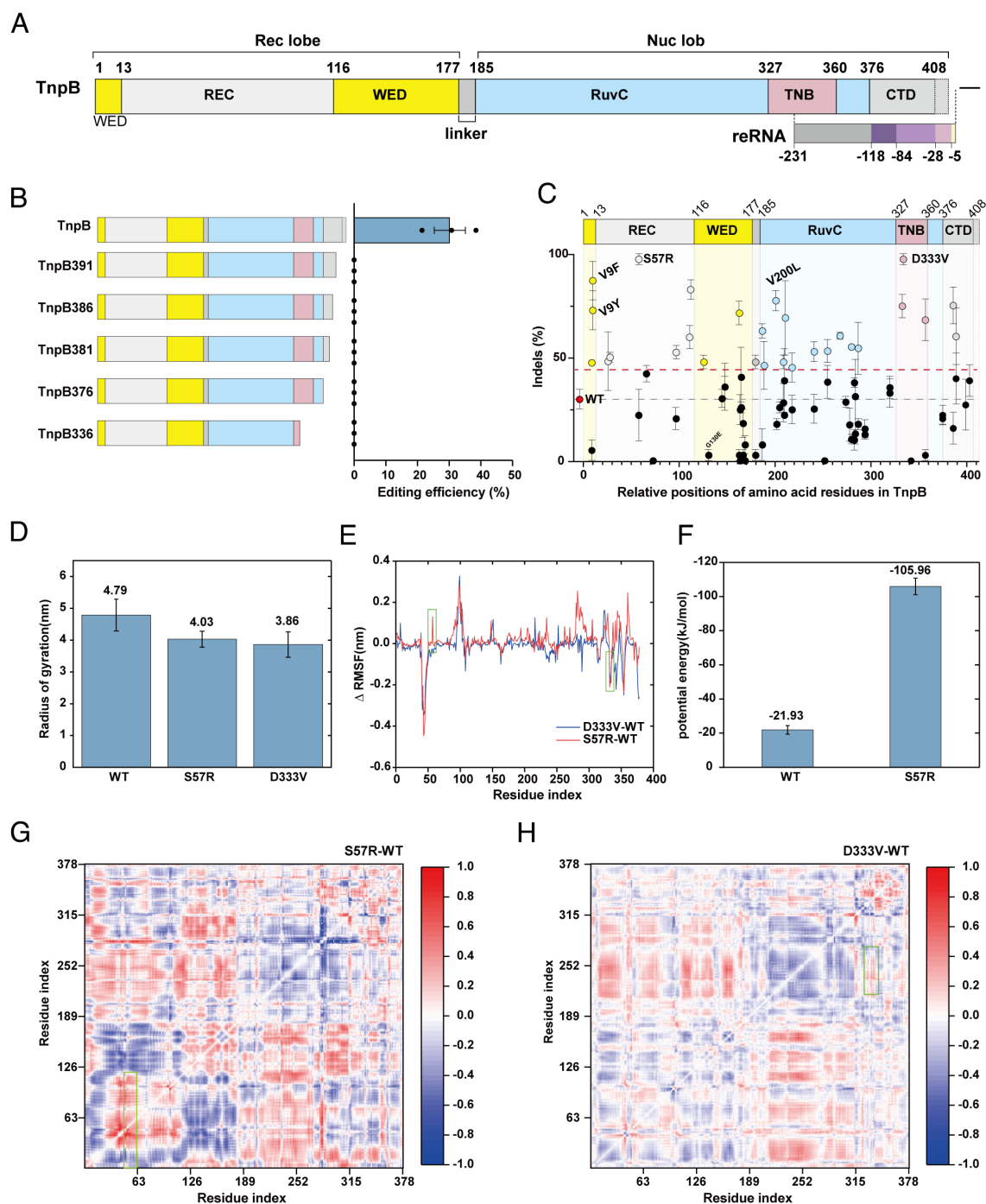


Fig. 4. Engineering of TnpB for enhanced genome editing. (A) Schematic of TnpB domain architecture, highlighting its CTD and associated reRNA locus. (B) Impact of C-terminal truncations on TnpB editing efficiency, underscoring the essential role of specific domains. Data are presented as mean $\pm$ SEM. (C) Comparative analysis of editing efficiency among TnpB variants with AI-guided mutations. The gray dashed line indicates wild-type editing efficiency, while the red dashed line represents 1.5 $\times$ wild-type efficiency. Data are presented as mean $\pm$ SEM. (D) Rg of the TnpB-reRNA-target DNA complex for WT (PDB: 8H1J), D333V, and S57R mutants, reflecting global conformational changes induced by the mutations. Data are presented as mean $\pm$ SD. (E) Differences in RMSF (Root Mean Square Fluctuation) between mutants and WT (mutant minus WT). The S57R mutation (red) increases local flexibility around residue 57 in the REC domain, while the D333V mutation (blue) decreases flexibility around residue 333. Regions with notable fluctuation changes are highlighted in green. (F) Interaction potential energy between protein and DNA before and after the S57R mutation, derived from MD simulations, indicating mutation-induced modulation of protein-DNA interaction strength. Data are presented as mean $\pm$ SD. (G) Differential DCC map for the S57R mutation relative to WT. Positive values indicate increased correlated motions; negative values, decreased correlations. (H) Differential DCC map for the D333V mutation relative to WT, highlighting altered long-range residue-residue communication and its potential impact on allosteric dynamics. For additional statistical and methodological details, please refer to Dataset S1 and SI Appendix.

“supermini” variant with enhanced genome editing potential in *Streptomyces*.

To test this hypothesis, we designed five truncated TnpB variants: TnpB391, TnpB386, TnpB381, TnpB376, and TnpB336. Among these, TnpB376 lacked the entire CTD, while TnpB336 was truncated at the reRNA-overlapping (Fig. 4B). These variants were introduced into *S. coelicolor* M145, and their editing efficiencies at the SCO5087 locus were assessed. However, none of the truncated variants induced detectable mutations (Fig. 4B), indicating that the C-terminal region of TnpB is essential for efficient genome editing in *Streptomyces*, in contrast to prior in vitro studies (42).

We hypothesized that C-terminal truncation may affect the protein's stability, increasing susceptibility to aggregation and degradation in vivo. To evaluate this, we performed a comparative analysis of the aggregation propensity of full-length and truncated TnpB proteins (TnpB336, TnpB376, TnpB381 and TnpB386, TnpB391) using Aggrescan (SI Appendix, Table S6) (45, 46).

Among all tested variants, TnpB336 exhibited the most pronounced increase in aggregation propensity, with its average aggregation score (a3vSA) rising from -0.112 (full-length) to -0.092 , and the normalized number of hot spots per 100 residues (NnHS) increasing from 1.961 to 2.090. The total aggregation area (TA) also became substantially less negative (-24.801 vs. -41.267), and Na4vSS increased to -9.4 , indicating a greater intrinsic tendency toward aggregation, misfolding, and reduced solubility.

TnpB371 displayed the highest NnHS value (2.400), reflecting the greatest density of aggregation-prone regions among all variants. While its a3vSA (-0.111) was similar to the full-length TnpB, the combination of high NnHS and a less negative TA (-35.709) suggests spatial clustering of aggregation-prone residues, which may further exacerbate misfolding and degradation.

In contrast, the aggregation profiles of TnpB381, TnpB386, and TnpB391 were largely similar to the full-length TnpB, with only minor variations in a3vSA values (-0.111 to -0.113) and NnHS (~ 2.05 to 2.10). Their TA values, however, remained consistently less negative (-35.617 to -36.589) than full-length TnpB, indicating a subtle yet persistent increase in aggregation propensity. Although these changes are unlikely to cause dramatic protein instability, they may still affect expression yields or functional integrity under certain conditions.

Our results demonstrate that C-terminal truncation of TnpB increases aggregation propensity in a length-dependent manner. Shorter variants, such as TnpB336 and TnpB371, are especially susceptible to aggregation and likely degradation, while longer truncations (TnpB381–391) more closely resemble the full-length protein in aggregation behavior. These findings provide important guidance for the rational design of TnpB constructs for both biochemical and structural studies.

AI-Assisted Protein Engineering for Enhanced Editing Efficiency of STAGE. To enhance TnpB's editing efficiency, we employed PRIME, an AI-empowered deep learning model designed to guide protein engineering without requiring prior mutagenesis data (47). Using PRIME, we selected the top 80 predicted mutants for in vivo validation, with mutations primarily enriched in the RuvC and WED domains. Notably, these predictions aligned with findings from TnpB mutational scanning (43). More than half of the tested variants exhibited improved editing efficiency compared to wild-type TnpB (Fig. 4B).

Interestingly, the beneficial mutations were distributed across multiple domains, suggesting that diverse regions contribute to TnpB's cleavage activity. Several variants, including V200L,

K111S, V9F, D333V, and S57R, displayed over a threefold increase in activity, with D333V and S57R achieving nearly 100% editing efficiency (Fig. 4C). Statistical analysis across three biological replicates confirmed the significance of these improvements ($P < 0.01$, unpaired two-tailed t test). Further structural analysis revealed that the enhanced mutations (V200, K111, V9, D333, and S57) are located near RNA binding, heteroduplex formation, or TAM recognition sites, suggesting that these regions are critical for TnpB's interaction with reRNA or target DNA. These results align with TnpB mutational scanning data and previous studies on CRISPR-Cas12 enzymes, where mutations in positively charged residues near the guide RNA–DNA heteroduplex or protospacer-adjacent motif (PAM) have been shown to enhance enzymatic activity and alter specificity (43, 48, 49).

To elucidate the mechanistic basis for the observed enhancements, we performed 500-nanosecond molecular dynamics (MD) simulations for both S57R and D333V mutants, as well as the wild-type TnpB, in complex with reRNA and target DNA (SI Appendix, Methods). Both mutants exhibited a reduced radius of gyration (R_g) relative to wild-type, indicating formation of a more compact protein-nucleic acid assembly and suggesting strengthened intermolecular interactions (Fig. 4D). RMS fluctuation (ΔRMSF) analysis revealed that S57R significantly increased local flexibility at residue 57 within the REC domain, potentially accelerating nucleic acid recognition (Fig. 4E), and also substantially enhanced the calculated interaction potential energy between residue 57 and dsDNA, supporting enhanced stabilization of the protein–DNA complex (Fig. 4F). By contrast, D333V reduced flexibility near residue 333, implying local structural stabilization.

Dynamic cross-correlation (DCC) analyses showed that S57R enhances coordinated motions within the REC domain (Fig. 4G), while D333V strengthens positive interdomain correlations between the TNB and RuvC domains (Fig. 4H), suggesting an allosteric mechanism that promotes efficient DNA cleavage. Together, these data provide plausible structural and dynamic explanations for the increased activity conferred by these single-residue substitutions.

In addition, residue P282 is located at the boundary of the lid subdomain, which obstructs the RuvC active site from accessing the target strand (TS) (11). The P282V and P282I substitutions ranked 2nd and 16th, respectively, among the 80 predicted mutants and were hypothesized to enhance editing efficiency. EGFP knock-out assay in HEK293T cells confirmed our prediction, demonstrating that replacing P282 with small, hydrophobic residues (V and I) increase TnpB activity (43). However, in *Streptomyces* genome editing at SCO5087, the indel formation frequencies did not show a significant increase. The P282V mutation yielded 42%, while P282I only resulted in 10%, compared to 30% for wild-type TnpB. Conversely, mutations at S72K, H163A, E168Y, K251E, and V341F completely abolished editing activity, indicating their essential roles in DNA cleavage (Fig. 4C). Despite this, deep mutation screening (43) indicates that these mutants exhibit activity comparable to wild-type TnpB. These findings underscore the potential of AI-assisted protein engineering in optimizing TnpB for genome editing while offering valuable insights for further enhancing its efficiency and specificity.

Truncation and Engineering of reRNA to Enhance Genome Editing of STAGE. The cognate reRNA associated with TnpB is a 231-nt noncoding RNA. Previous studies have suggested that some regions of reRNA are redundant for cleavage (42) and that sgRNA engineering can enhance gene editing efficiency in related Cas enzymes (50). Building on these insights, we hypothesized

that truncation and engineering reRNA could create a more concise and efficient genome editing system.

To examine this, we constructed a series of TnpB-driven plasmids with truncated reRNAs targeting the SCO5087 locus (Fig. 4A and SI Appendix, Fig. S10) and evaluated their ability to direct TnpB-induced indels in *S. coelicolor* M145. Specifically, we designed five truncated mutants: Δ Trim1, Δ Trim2, Δ Stem1, Δ Stem2, and Δ PK, each involving deletions of distinct regions: 5' disordered region (-231G to -117U), triple-helix and PK formation region (-116G to -86C), stem1 region (-85G to -30A), stem2 region (-29U to -7G), and PK region (-6G to -1A) (Fig. 5A).

Among these, Δ Trim1 and Δ Trim2 maintained editing efficiencies comparable to full-length reRNA, while the remaining three truncated variants exhibited significantly reduced activity (Fig. 5B). This suggests that the deleted regions in Δ Trim1 and Δ Trim2 regions are not essential for reRNA function, whereas truncations in stem1, stem2, or PK disrupt critical secondary structures required for activity.

To elucidate the structural basis of these observations, we predicted the secondary structures of wild-type and truncated reRNAs using RNAfold WebServer. Both Δ Trim1 and Δ Trim2 preserved the core architectural features of the full-length reRNA, including nested stem-loop arrangements and a conserved 3'-end motif, which likely support TnpB recognition and activation (SI Appendix, Fig. S11). In contrast, Δ Stem1, Δ Stem2, and Δ PK disrupted these essential motifs, either by deleting key helices or adopting aberrant, non-native folds incompatible with TnpB loading (SI Appendix, Fig. S11).

Of particular interest, the Δ PK variant, removing only the final six nucleotides at the 3' end, was predicted by IPknot (51) to undergo extensive global structural rearrangement, resulting in multiple overlapping pseudoknot-like interactions, especially in the central and 3' regions. These long-range, non-native interactions likely reflect compensatory refolding that produces overly compact, nonfunctional conformations. While pseudoknot formation may enhance local stability, it can also distort the native folding pathway, mask functional motifs, or sterically hinder TnpB-reRNA assembly, thus abolishing activity.

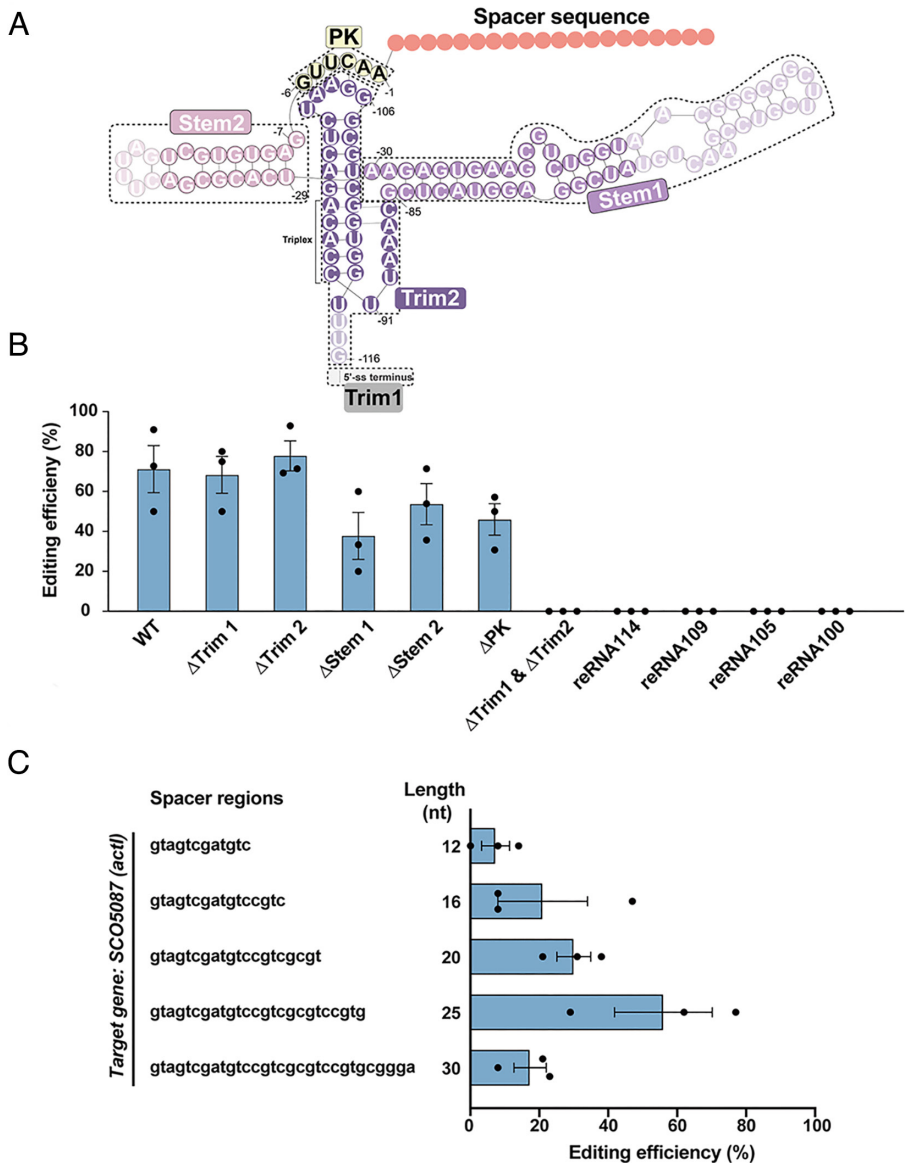


Fig. 5. Engineering of reRNA for enhanced genome editing. (A) Structural model of reRNA (PDB: 8EXA), illustrating functional regions and secondary structure. Unresolved RNA segments are shown in a lighter color. (B) Editing efficiency of truncated reRNAs variants, identifying regions critical for activity. (C) Impact of spacer length on TnpB editing efficiency, determining the optimal spacer length for genome editing. Data are shown as means $\pm$ SEM. For statistical details, see Dataset S1 and SI Appendix.

Taken together, these findings indicate that preservation of specific secondary structure elements, particularly hierarchically organized stem-loops and an accessible 3' end, is more critical for reRNA functionality than overall transcript length or thermodynamic stability. Moreover, the artificial introduction of pseudoknots, while potentially stabilizing the molecule, may interfere with the dynamic RNA–protein interactions required for efficient TnpB-mediated editing.

To further minimize reRNA, we progressively truncated the 5' end of Δ Trim1 down to position –86C, producing variants such as Δ Trim1+ Δ Trim2, reRNA100, reRNA105, reRNA109, and reRNA114. None of these minimized reRNAs exhibited detectable activity in *Streptomyces*. Structural predictions indicated that, while some core helices were retained, these minimized variants lacked the global architectural framework, including stem1 and distal 5' domains, necessary for long-range structural integrity (SI Appendix, Fig. S11). For example, Δ Trim1+ Δ Trim2 preserved central and 3' stem-loops but lacked extended 5' regions critical for tertiary structure stabilization. Similarly, reRNA100 adopted a compact structure, while reRNA105 and reRNA109 displayed reconfigured helices likely mispositioning essential recognition sites. Notably, reRNA114, although retaining an almost intact nested core, was inactive in *Streptomyces* but retained function in HEK293T cells (52), highlighting host-dependent requirements for RNA-mediated activation. These findings emphasize that localized secondary structure alone is insufficient for universal function; instead, effective reRNA designs must be evaluated within the appropriate host-specific structural and cellular context.

To assess sequence conservation, we analyzed full-length reRNA using the “Score a Sequence” model of Evo Designer (<https://arcinstitute.org/tools/evo/evo-designer>) (53). The results revealed that the truncated regions consist primarily of nonconserved nucleotides arranged contiguously. A similar pattern was observed in the sequence encoding a truncation within reRNA's Stem2 region, which maintains or enhances TnpB's genome editing activity (11). Removing the portion transcribed from the nonconserved DNA sequence within the 231-nt reRNA may provide an effective approach for obtaining truncated reRNA variants that could maintain or increase TnpB's genome editing activity.

Impact of Target DNA Length on Genome Editing Activity. To investigate the effect of spacer length on the genome editing efficiency of STAGE, we quantified the editing efficiency of pSTAGE-NHEJ plasmids with spacer length ranging from 12 to 30 nucleotides (nt). Our results revealed that a minimum spacer length of 16 nt is required for detectable gene editing activity in *Streptomyces*. Plasmids with a 12-nt spacer exhibited poor genome editing efficiency, aligning with previous in vitro findings (11). Further analysis showed that extending the spacer–protospacer pairing from 16 to 25 nt significantly enhanced editing efficiency (Fig. 5C), with the 25-nt spacers yielding the highest activity. However, when the spacer length was increased to 30 nt, editing efficiency declined sharply. These findings suggest that optimal spacer length is critical for maximizing STAGE-mediated genome editing, with 25 nt being the most effective length while excessively long spacers negatively impact efficiency.

Exploration and Engineering of TnpB Orthologs for Genome Editing in *Streptomyces*. Comprehensive genomic and metagenomic analysis have revealed that TnpB proteins are widespread in bacteria and archaea (9, 20). This suggests that, in addition to well-characterized IS200/IS605-associated TnpBs, other programmable TnpB endonucleases may exist within uncharted TnpB diversity.

To identify such orthologs, we analyzed closely related *tnpB* genes of ISDra2 using the Syntome section of MicroScope (54, 55) (Fig. 6A and SI Appendix, Fig. S12). We identified several orthologs with >50% protein sequence similarity to ISDra2 TnpB, including those from *D. radiodurans* R1, *Bifidobacterium catenulatum* DSM16992, *Bacillus cereus* 03BB102, *Fusobacterium periodonticum* ATCC33693, *Cyanospora* sp. PCC7120, *Nostoc* sp. PCC7120, *Nitrospirae bacterium* HCHbin1, *Clostridium tetani* E88, *Bacillus albus* PFYN01, and *Desulfotobacterium hafniense* PCP-1. Among them, TnpB* shared >80% sequence similarity with ISDra2 TnpB (SI Appendix, Fig. S13). Genetic structure analysis revealed that coding genes of both TnpB* and ISDra2 TnpB are transcribed in the same direction and located downstream of *tnpA*, though TnpB* overlaps with *tnpA*, whereas ISDra2 TnpB is adjacent to it (Fig. 6B). Structural modeling using AlphaFold 3.0 (33) suggested that TnpB* closely resembles ISDra2 TnpB, with TM-score > 0.5 and RMSD value around 1.3 to 1.5 Å, supporting their structural similarity (Fig. 6B and SI Appendix, Fig. S14). Given these observations, we hypothesized that TnpB* could function as an RNA-guided nuclease capable of dsDNA cleavage, similar to ISDra2 TnpB.

Defining the Cognate reRNA and TAM for TnpB*. To test this hypothesis, we first identified the cognate reRNA and TAM of TnpB*. Previous studies indicate that reRNAs overlap the 3' region of *tnpB* genes and are approximately 200 nt in size (20). Based on this, we selected a 231-nt sequence from the 3' region of TnpB* (designated reRNA*), which shares 85.71% sequence similarity with ISDra2 reRNA (SI Appendix, Fig. S15). Given the sequence and structural similarity between TnpB* and ISDra2 TnpB, we initially assumed that TnpB* would recognize the same TAM (5'-TTGAT). However, previous studies suggest that ISDra2 TnpB's TAM aligns with the annotated left cleavage site (C_L), which is essential for TnpA-mediated transposition (20). Using a computational model for predicting optimal TAM sequences of multicopy TnpB systems (20), we identified 5'-TTGAC as a potential alternative TAM for TnpB* (SI Appendix, Fig. S16).

To evaluate TnpB* cleavage efficiency, we targeted SCO5087 (5'-TTGAT as TAM) and SCO5091 (5'-TTGAC as TAM) in *S. coelicolor* M145. However, no detectable TnpB*/reRNA*-mediated edits were observed (Fig. 6D), suggesting that either TnpB* lacks dsDNA cleavage activity or requires a different TAM. To further investigate, we assessed whether ISDra2 TnpB could edit SCO5087 using reRNA*. Surprisingly, TnpB with reRNA* showed significantly higher editing efficiency than with its native reRNA, achieving nearly 100% efficiency in *S. coelicolor* (Fig. 6D). A recent study assessed the mutational tolerance of the 116-nt reRNA scaffold via deep mutational scanning (43). Most nucleotide positions in reRNA* that differ from the reRNA scaffold (positions –1A to –116G) exhibited higher enrichment scores than wild-type reRNA, with nearly all variations transcribed from conserved nucleotides. This suggests that optimizing reRNA sites encoded by conserved nucleotides could enhance editing efficiency.

Additionally, we observed low-efficiency editing by TnpB* when guided by reRNA (pSTAGE-NHEJ*-2.0 and pSTAGE-HDR*-2.0). Given the flexible TAM requirement of TnpB (56), TnpB* likely recognizes a TAM sequence similar to that of TnpB, indicating that TnpB* retains some editing capability; further optimization is needed. Sequence analysis revealed significant differences between TAM recognition region of TnpB* and ISDra2 TnpB, potentially affecting target site recognition (Fig. 6C). To enhance TnpB* editing efficiency, we replaced its TAM recognition region (residues 40 to 100) with that of ISDra2 TnpB, generating an engineered variant, eTnpB*. Structural

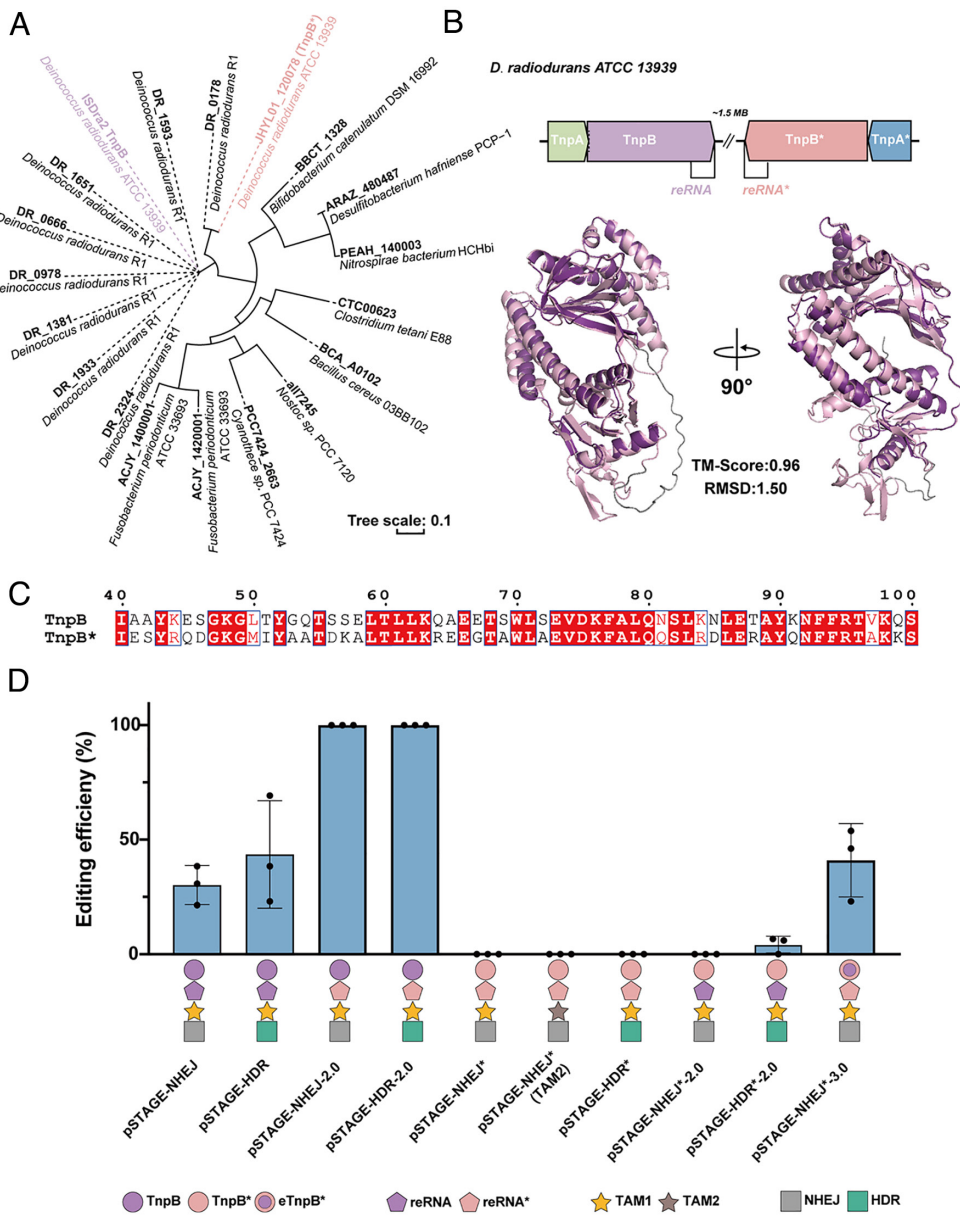


Fig. 6. Exploration of TnpB orthologs for *Streptomyces* genome editing. (A) Phylogenetic tree of TnpB orthologs with >50% sequence similarity, illustrating their evolutionary relationships. (B) Genomic organization of TnpB and TnpB* in *D. radiodurans* ATCC 13939, highlighting structural differences. (C) Partial sequence alignment of TnpB and TnpB*, generated using Clustal Omega and ESPrnt 3, with key conserved regions annotated. (D) Comparative analysis of editing efficiency for engineered TnpB and TnpB* systems targeting the SCO5087 locus in *Streptomyces*. pSTAGE-NHEJ*(TAM2) was used to edit SCO5091, with “TTGAC” defined as the TAM. Data are presented as means $\pm$ SD. For statistical details, see [Dataset S1](#) and [S1 Appendix](#).

modeling confirmed that eTnpB* closely resembles ISDra2 TnpB (*SI Appendix, Fig. S17*). To investigate the structural basis for the reduced activity of TnpB*, we modeled ternary complexes comprising TnpB-reRNA*-DNA, TnpB*-reRNA*-DNA, and engineered eTnpB*-reRNA*-DNA using AlphaFold 3, followed by electrostatic surface analysis with ChimeraX. The TnpB-reRNA*-DNA complex exhibited a largely neutral electrostatic profile around the TAM recognition interface. In contrast, TnpB* uniquely contains an aspartate residue at position 56 (Asp56), which introduces a localized negative charge at the DNA-binding surface. This negative patch likely generates electrostatic repulsion with the sugar-phosphate backbone of the target DNA, destabilizing TAM interaction and impairing cleavage efficiency. The domain-swapped variant eTnpB* effectively replaces this region, mitigating the unfavorable electrostatic effect and restoring efficient DNA binding and cleavage activity. These results demonstrate that even minor electrostatic mismatches within the TAM recognition region, despite high overall sequence similarity, can critically compromise TnpB* function, and that targeted domain swapping can efficiently rescue such defects (*SI Appendix, Fig. S18*).

Consistent with these findings, pSTAGE-NHEJ⁻-3.0, which harbors eTnpB^{*} and reRNA^{*}, exhibited an increased editing efficiency (~40%) at SCO5087 with a TAM of 5'-TTGAT, validating domain swapping as an effective strategy to rescue and enhance TnpB^{*} activity (Fig. 6D). Moreover, our observations are further supported by deep mutational scanning data (43), which indicated that the S56D mutation significantly impairs editing efficiency.

These findings confirm that TnpB* is capable of dsDNA cleavage, though its optimal TAM remains to be determined. The improved editing efficiency observed with TnpB combined with reRNA* and eTnpB* combined with reRNA* underscores the potential of reRNA engineering and TnpB domain swapping to enhance genome editing. Furthermore, the developed compact STAGE systems, including pSTAGE-NHEJ/HDR-2.0, pSTAGE-NHEJ*-3.0, and pSTAGE-NHEJ*-2.0, represent promising tools for genome editing in *Streptomyces*.

On-Target and Off-Target Analysis of TnpB-Mediated Editing. To further assess STAGE specificity, we analyzed on-target and off-target activity in *S. coelicolor* M145 via WGS. High-coverage WGS (100× depth) on 20 strains, including: 18 mutants edited using STAGE at the

SCO5087 locus; one mutant carrying the vector control (pGM1190); and one wild-type *S. coelicolor* M145 strain. We identified single-nucleotide polymorphisms (SNPs) and indels via breseq (57) (version 0.37.1 and 0.38.3) under default parameters. All expected on-target indels were accurately identified through sequencing and filtering (SI Appendix, Fig. S19 and Table S4). Additionally, in colonies lacking PCR amplicons, WGS revealed a large (>10,000 bp) deletion around the target site, suggesting that TnpB-mediated indels were precisely introduced at the intended genomic loci. These findings confirm STAGE's targeted editing accuracy, while also highlighting its potential for large-fragment deletions.

After filtering, we identified 57 SNPs and indels exclusively present in the mutant strains, using *S. coelicolor* A3(2) (GenBank: GCA_000203835.1) as the reference genome (SI Appendix, Table S5). To assess potential off-target effects, we analyzed the 100-nt flanking regions surrounding these SNPs and indels. Notably, none of these mutations occurred within 200 nt of a sequence perfectly matching the TAM and protospacer. Additionally, we examined potential TAM-adjacent sites (allowing up to one mismatch with the canonical TAM site). While some SNPs and indels were near potential PAM-like sequences, none of the adjacent sequences showed significant similarity (>50%) to the protospacer. These results indicate that no detectable off-target effects were induced by the STAGE system. In summary, WGS analysis of TnpB-edited mutants demonstrates that STAGE enables precise and efficient on-target genome editing, with no detectable evidence of unintended mutations under our experimental conditions. Nevertheless, we explicitly acknowledge the inherent limitations of WGS, such as its reduced sensitivity for detecting extremely rare variants or complex structural alterations.

Discussion

In this study, we introduced STAGE, a compact and versatile TnpB-based genome editing toolkit optimized for *Streptomyces*, which addresses key limitations of existing genome editing systems such as Cas9 and Cas12. Our findings highlight the potential of TnpB, an RNA-guided endonuclease derived from *D. radiodurans*, as a robust and efficient tool for genetic manipulation in *Streptomyces*. Through AI-driven optimization of the STAGE toolkit, including reRNA optimization and TnpB engineering, we achieved nearly 100% editing efficiency with negligible off-target effects.

One of the major contributions of this work is the development of STAGE-cBEST, a TnpB-mediated base editing system enabling precise C-G-to-T-A single-nucleotide conversions without introducing DSBs. This innovation addresses the risks of genomic instability and cytotoxicity often associated with DSB-dependent genome editing systems. Moreover, we successfully demonstrated the multiplexing capabilities of STAGE-cBEST, underscoring its potential for simultaneous editing of multiple loci, an essential feature for complex metabolic engineering and strain optimization. Systematic evaluation of STAGE across *Streptomyces* strains featuring three distinct DSB repair machineries confirmed its robustness and adaptability.

The toolkit's efficacy was validated across a range of applications, including gene knockouts, large DNA fragment deletions, and targeted base editing. The integration of AI-assisted protein engineering significantly enhanced TnpB's editing efficiency, with several engineered variants outperforming the wild-type protein. These results underscore the power of computational approaches in optimizing genome editing tools for specialized applications. In addition, exploration of TnpB orthologs and reRNA engineering broadened the system's targeting capabilities, paving the way for its wider adoption in genome editing. Furthermore, through deep genome comparison, we characterized a TnpB system from the same *D. radiodurans* genome. By implementing TAM recognition

domain swapping and reRNA hybridization, we unlocked its genome editing potential, further expanding the repertoire of programmable TnpB-based systems.

STAGE's ability to enable precise and efficient genetic modifications has profound implications for synthetic biology and metabolic engineering. By facilitating the systematic exploration and manipulation of BGCs, this toolkit unlocks opportunities for the discovery and production of novel natural products. The capacity to edit *Streptomyces* genomes with such precision and flexibility represents a significant advancement in natural product biosynthesis and industrial biotechnology.

Looking ahead, future efforts will focus on further refining the STAGE system through continuous AI-driven engineering and evolution of TnpB and reRNA, development of additional base editing modalities, expansion of TAM recognition profiles, and exploration of other compact programmable nucleases. Extending the toolkit's application to other actinomycetes and related microbial systems could further broaden its impact in natural product discovery and synthetic biology. In addition, the development of visual web-based tools for reRNA design would greatly facilitate broader adoption of the platform. The integration of AI-driven strategies, including rational redesign of TnpB, prediction of reRNA structure, and data-guided expansion of TAM sequences, will be central to the next generation of compact genome editors. These advances promise to further enhance the precision, flexibility, and programmability of systems like STAGE, unlocking possibilities in microbial genome engineering and biosynthetic innovation.

In conclusion, STAGE represents a transformative advancement in genome editing technology. Offering a compact, efficient, and precise genome engineering platform for *Streptomyces* research, this toolkit overcomes limitations of current systems and enables the unlocking of *Streptomyces*' vast genetic potential, fostering innovation in natural product discovery and metabolic engineering.

Methods

Detailed methods are provided in SI Appendix. The TnpB coding sequence was codon-optimized to *S. coelicolor*. Related plasmids were introduced into target *Streptomyces* via interspecies conjugation (25). Off-target effects were evaluated by WGS. PRIME (47) was used for TnpB engineering. MD simulations were carried out with the GROMACS 2025.1 software suite (58), and dynamics cross-correlation matrices were subsequently calculated for structural analysis.

Data, Materials, and Software Availability. Raw WGS data are available in NCBI's SRA database under accession number PRJNA1223038 (59). The datasets used in this study are provided in SI Appendix, Tables S1–S6 and Datasets S1 and S2.

ACKNOWLEDGMENTS. This work was supported by grants from the National Key Research and Development Program of China (2021YFA0909500, 2024YFA0917603), the National Natural Science Foundation of China (32170080 and 32370026), Shanghai Pilot Program for Basic Research-Shanghai Jiao Tong University (21TQ1400204), Science and Technology Commission of Shanghai Municipality (24HC2810200), and Open Funding Project of the State Key Laboratory of Microbial Metabolism (MMLKF22-03). Y.T. was also supported by Shanghai Municipality: Program for High-Level Overseas Talents Introduction and the Excellent Young Scientists Fund Program (Overseas). S.Y.L. was supported by the Development of platform technologies of microbial cell factories for the next-generation biorefineries project (2022M3J5A1056117) from the National Research Foundation supported by the Korean Ministry of Science and ICT. P.T. and L.H. were supported by the Computational Biology Key Program of Shanghai Science and Technology Commission (23JS1400600), Shanghai Municipal Education Commission (2024AIZD015), Shanghai Jiao Tong University Scientific and Technological Innovation Funds (21X010200843), Science and Technology Innovation Key R&D Program of Chongqing (CSTB2022TIAD-STX0017, CSTB2024TIAD-STX0032), the Student Innovation Center at Shanghai Jiao Tong University, Shanghai Artificial Intelligence Laboratory, and Shanghai Municipal Science and Technology Major Project.

Author affiliations: ^aState Key Laboratory of Microbial Metabolism, Joint International Research Laboratory of Metabolic and Developmental Sciences, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China; ^bSchool of Physics and Astronomy, and Shanghai National Center for Applied Mathematics (Shanghai Jiao Tong University Center), and Institute of Natural Sciences, Shanghai Jiao Tong University, Shanghai 200240, China; ^cShanghai Artificial Intelligence Laboratory, Shanghai 200030, China; ^dZhangjiang Institute for Advanced Study, Shanghai Jiao Tong University, Shanghai 200240, China; ^eMetabolic and Biomolecular Engineering National Research Laboratory, Department of Chemical and Biomolecular Engineering

(BK21 four), Korea Advanced Institute of Science and Technology, Daejeon 34141, Republic of Korea; ^fBioProcess Engineering Research Center, Korea Advanced Institute of Science and Technology, Daejeon 34141, Republic of Korea; and ^gGraduate School of Engineering Biology, Korea Advanced Institute of Science and Technology, Daejeon 34141, Republic of Korea

Author contributions: J.L., N.C., S.Y.L., and Y.T. designed research; J.L., N.C., Y.Q., P.T., L.Z., Z.Y., and L.H. performed research; J.L., N.C., Y.Q., S.Y., and Y.T. analyzed data; and J.L., N.C., S.Y.L., and Y.T. wrote the paper.

1. K. Alam *et al.*, *Streptomyces*: The biofactory of secondary metabolites. *Front. Microbiol.* **13**, 968053 (2022).
2. L. Donald *et al.*, *Streptomyces*: Still the biggest producer of new natural secondary metabolites, a current perspective. *Microbiol. Res.* **13**, 418–465 (2022).
3. Y. Tong, P. Charusanti, L. Zhang, T. Weber, S. Y. Lee, CRISPR-Cas9 based engineering of actinomycetal genomes. *ACS Synth. Biol.* **4**, 1020–1029 (2015).
4. J. Zhang *et al.*, Efficient multiplex genome editing in *Streptomyces* via engineered CRISPR-Cas12a systems. *Front. Bioeng. Biotechnol.* **8**, 726 (2020).
5. H. M. Hua *et al.*, Low-toxicity and high-efficiency *Streptomyces* genome editing tool based on the miniature type V-F CRISPR/Cas nuclease AsCas12f1. *J. Agric. Food Chem.* **72**, 5358–5367 (2024).
6. Y. Zhao, G. Li, Y. Chen, Y. Lu, Challenges and advances in genome editing technologies in *Streptomyces*. *Biomolecules* **10**, 734 (2020).
7. Y. Lee *et al.*, CRISPR-aided genome engineering for secondary metabolite biosynthesis in *Streptomyces*. *J. Ind. Microbiol. Biotechnol.* **51**, kuae009 (2024).
8. D. G. Kim *et al.*, Engineered CRISPR-Cas9 for *Streptomyces* sp. genome editing to improve specialized metabolite production. *Nat. Commun.* **16**, 874 (2025).
9. H. Altae-Tran *et al.*, Diversity, evolution, and classification of the RNA-guided nucleases TnpB and Cas12. *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2308224120 (2023).
10. T. Karvelis *et al.*, Transposon-associated TnpB is a programmable RNA-guided DNA endonuclease. *Nature* **599**, 692–696 (2021).
11. G. Sasnauskas *et al.*, TnpB structure reveals minimal functional core of Cas12 nuclease family. *Nature* **616**, 384–389 (2023).
12. Z. Li *et al.*, Engineering a transposon-associated TnpB- ω RNA system for efficient gene editing and phenotypic correction of a tyrosinaemia mouse model. *Nat. Commun.* **15**, 831 (2024).
13. S. Karmakar *et al.*, A miniature alternative to Cas9 and Cas12: Transposon-associated TnpB mediates targeted genome editing in plants. *Plant Biotechnol. J.* **22**, 2950–2953 (2024).
14. K. S. Makarova *et al.*, Evolutionary classification of CRISPR-Cas systems: A burst of class 2 and derived variants. *Nat. Rev. Microbiol.* **18**, 67–83 (2020).
15. H. Altae-Tran *et al.*, The widespread IS200/IS605 transposon family encodes diverse programmable RNA-guided endonucleases. *Science* **374**, 57–65 (2021).
16. W. L. Yeo *et al.*, Characterization of Cas proteins for CRISPR-Cas editing in streptomycetes. *Biotechnol. Bioeng.* **116**, 2330–2338 (2019).
17. A. S. Meijers *et al.*, Efficient genome editing in pathogenic mycobacteria using *Streptococcus thermophilus* CRISPR1-Cas9. *Tuberculosis* **124**, 101983 (2020).
18. L. Li *et al.*, CRISPR-Cpf1-assisted multiplex genome editing and transcriptional repression in *Streptomyces*. *Appl. Environ. Microbiol.* **84**, e00827–18 (2018).
19. L. L. Tan *et al.*, Application of Cas12j for *Streptomyces* editing. *Biomolecules* **14**, 486 (2024).
20. G. Xiang *et al.*, Evolutionary mining and functional characterization of TnpB nucleases identify efficient miniature genome editors. *Nat. Biotechnol.* **42**, 745–757 (2023).
21. Y. Xu *et al.*, Reprogramming an RNA-guided archaeal TnpB endonuclease for genome editing. *Cell Discov.* **9**, 112 (2023).
22. S. P. Nety *et al.*, The transposon-encoded protein TnpB processes its own mRNA into omegaRNA for guided nuclease activity. *CRISPR J.* **6**, 232–242 (2023).
23. H. K. Jang, B. Song, G. H. Hwang, S. Bae, Current trends in gene recovery mediated by the CRISPR-Cas system. *Exp. Mol. Med.* **52**, 1016–1027 (2020).
24. M. Wang *et al.*, Hypercompact TnpB and truncated TnpB systems enable efficient genome editing *in vitro* and *in vivo*. *Cell Discov.* **10**, 31 (2024).
25. Y. Tong *et al.*, CRISPR-Cas9, CRISPRi and CRISPR-BEST-mediated genetic manipulation in streptomycetes. *Nat. Protoc.* **15**, 2470–2502 (2020).
26. Y. Tong *et al.*, Highly efficient DSB-free base editing for streptomycetes with CRISPR-BEST. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 20366–20375 (2019).
27. G. Hoff, C. Bertrand, E. Piotrowski, A. Thibessard, P. Leblond, Genome plasticity is governed by double strand break DNA repair in *Streptomyces*. *Sci. Rep.* **8**, 5272 (2018).
28. A. C. Komor, Y. B. Kim, M. S. Packer, J. A. Zuris, D. R. Liu, Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).
29. K. Nishida *et al.*, Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* **353**, aaf8729 (2016).
30. Y. Zhao *et al.*, Multiplex genome editing using a dCas9-cytidine deaminase fusion in *Streptomyces*. *Sci. China Life Sci.* **63**, 1053–1062 (2020).
31. T. P. Huang, G. A. Newby, D. R. Liu, Precision genome editing using cytosine and adenine base editors in mammalian cells. *Nat. Protoc.* **16**, 1089–1128 (2021).
32. G. Schuler, C. Hu, A. Ke, Structural basis for RNA-guided DNA cleavage by IscB- ω RNA and mechanistic comparison with Cas9. *Science* **376**, 1476–1481 (2022).
33. J. Abramson *et al.*, Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).
34. L. Xu, Y. Liu, R. Han, BEAT: A python program to quantify base editing from Sanger sequencing. *CRISPR J.* **2**, 223–229 (2019).
35. R. C. Beale *et al.*, Comparison of the differential context-dependence of DNA deamination by APOBEC enzymes: Correlation with mutation spectra *in vivo*. *J. Mol. Biol.* **337**, 585–596 (2004).
36. H. A. Rees, D. R. Liu, Base editing: Precision chemistry on the genome and transcriptome of living cells. *Nat. Rev. Genet.* **19**, 770–788 (2018).
37. S. Banno, K. Nishida, T. Arazoe, H. Mitsunobu, A. Kondo, Deaminase-mediated multiplex genome editing in *Escherichia coli*. *Nat. Microbiol.* **3**, 423–429 (2018).
38. R. E. Cobb, Y. Wang, H. Zhao, High-efficiency multiplex genome editing of *Streptomyces* species using an engineered CRISPR/Cas system. *ACS Synth. Biol.* **4**, 723–728 (2015).
39. J. Wang *et al.*, Engineered cytosine base editor enabling broad-scope and high-fidelity gene editing in *Streptomyces*. *Nat. Commun.* **15**, 5687 (2024).
40. N. S. McCarty, A. E. Graham, L. Studená, R. Ledesma-Amaro, Multiplexed CRISPR technologies for gene editing and transcriptional regulation. *Nat. Commun.* **11**, 1281 (2020).
41. C. M. Whitford *et al.*, Systems analysis of highly multiplexed CRISPR-base editing in *Streptomyces*. *ACS Synth. Biol.* **12**, 2353–2366 (2023).
42. R. Nakagawa *et al.*, Cryo-EM structure of the transposon-associated TnpB enzyme. *Nature* **616**, 390–397 (2023).
43. B. A. Thornton *et al.*, Latent activity in TnpB revealed by mutational scanning. bioRxiv [Preprint] (2025). <https://doi.org/10.1101/2025.02.11.637750> (Accessed 16 February 2025).
44. W. Chen *et al.*, Cas12n nucleases, early evolutionary intermediates of type V CRISPR, comprise a distinct family of miniature genome editors. *Mol. Cell* **83**, 2768–2780.e66 (2023).
45. M. Zalewski, V. Iglesias, O. Bárcenas, S. Ventura, S. Kmiecik, Aggrescan4D: A comprehensive tool for pH-dependent analysis and engineering of protein aggregation propensity. *Protein Sci* **33**, e5180 (2024).
46. O. Conchillo-Solé *et al.*, AGGRESAN: A server for the prediction and evaluation of “hot spots” of aggregation in polypeptides. *BMC Bioinformatics* **8**, 65 (2007).
47. F. Jiang *et al.*, A general temperature-guided language model to design proteins of enhanced stability and activity. *Sci. Adv.* **10**, eadr2641 (2024).
48. B. P. Kleinstiver *et al.*, Engineered CRISPR-Cas12a variants with increased activities and improved targeting ranges for gene, epigenetic and base editing. *Nat. Biotechnol.* **37**, 276–282 (2019).
49. H. Zhang *et al.*, An engineered xCas12i with high activity, high specificity, and broad PAM range. *Protein Cell* **14**, 538–543 (2023).
50. C. Dong, Y. Gou, J. Lian, SgRNA engineering for improved genome editing and expanded functional assays. *Curr. Opin. Biotechnol.* **75**, 102697 (2022).
51. K. Sato, Y. Kato, M. Hamada, T. Akutsu, K. Asai, IPknot: Fast and accurate prediction of RNA secondary structures with pseudoknots using integer programming. *Bioinformatics* **27**, i85–i93 (2011).
52. K. F. Marquart *et al.*, Effective genome editing with an enhanced ISDra2 TnpB system and deep learning-predicted ω RNAs. *Nat. Methods* **21**, 2084–2093 (2024).
53. G. Brixi *et al.*, Genome modeling and design across all domains of life with Evo 2. bioRxiv [Preprint] (2025). <https://doi.org/10.1101/2025.02.18.638918> (Accessed 21 February 2025).
54. D. Vallenet *et al.*, MicroScope: A platform for microbial genome annotation and comparative genomics. *Database* **2009**, bap021 (2009).
55. D. Vallenet *et al.*, MicroScope: An integrated platform for the annotation and exploration of microbial gene functions through genomic, pangenomic and metabolic comparative analysis. *Nucleic Acids Res.* **48**, D579–D589 (2019).
56. X. Feng *et al.*, Flexible TAM requirement of TnpB enables efficient single-nucleotide editing with expanded targeting scope. *Nat. Commun.* **15**, 3464 (2024).
57. D. E. Deatherage, J. E. Barrick, Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. *Methods Mol. Biol.* **1151**, 165–188 (2014).
58. M. J. Abraham *et al.*, GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).
59. J. Luo *et al.*, Whole-genome sequencing data of *Streptomyces coelicolor* M145 edited using the STAGE platform. NCBI Sequence Read Archive (SRA). <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1223038>. Accessed 9 August 2025.