Molecular Therapy
Original Article

Rational and computation-assisted engineering of a compact and efficient CRISPR-Cas12f genome editor

Jingtong Liu,[1,2,10] Shengzhou Wang,[1,10] Bei Wang,[2,10] Tonglu Yu,[1] Xuan Liu,[1] Miaomiao Li,[1] Min-Chi Ku,[3,4] Rui Wang,[1,5] Renjie Chai,[2,6,7] Wei V. Zheng,[8] and Yongming Wang[1,9]

[1]Center for Medical Research and Innovation, Shanghai Pudong Hospital, Fudan University Pudong Medical Center, School of Life Sciences, Fudan University, Shanghai 200438, China; [2]Spine Surgery Department, Nantong First People's Hospital, State Key Laboratory of Digital Medical Engineering, School of Life Sciences and Technology, School of Medicine, Advanced Institute for Life and Health, Jiangsu Province High-Tech Key Laboratory for Bio-Medical Research, Southeast University, Nanjing 210096, China; [3]Max-Delbrück-Center for Molecular Medicine in the Helmholtz Association (MDC), Robert-Rössle Strasse 10, 13125 Berlin, Germany; [4]DZHK (German Centre for Cardiovascular Research), Partner Site Berlin, Berlin, Germany; [5]Human Phenome Institute, Academy for Engineering and Technology, State Key Laboratory of Genetic Engineering, Greater Bay Area Institute of Precision Medicine (Guangzhou), Fudan University, Shanghai 200438, China; [6]Co-Innovation Center of Neuroregeneration, Nantong University, Nantong 226001, China; [7]Department of Neurology, Aerospace Center Hospital, School of Life Science, Beijing Institute of Technology, Beijing 100081, China; [8]Intervention and Cell Therapy Center, Peking University Shenzhen Hospital, Shenzhen, Guangdong 518036, China; [9]Shanghai Engineering Research Center of Industrial Microorganisms, Shanghai 200438, China

The CRISPR-Cas12f system is an ultracompact genome-editing platform, yet only a few orthologs exhibit robust activity in mammalian cells. Here, we systematically screened 23 Cas12f orthologs and identified two active nucleases, PspCas12f1 and TcCas12f1, capable of genome editing in human cells. Single guide RNA (sgRNA) scaffold optimization enhanced the basal activity of PspCas12f1. To further improve its performance, we combined structure-guided rational design with protein language model-assisted filtering. Candidate mutations predicted by SaProt were further screened based on structural proximity to the DNA-binding interface and electrostatic compatibility. This integrative strategy identified Q100R and E293R, whose combination yielded the optimized variant enPspCas12f1. enPspCas12f1 achieved genome-editing efficiencies comparable to SpCas9 across multiple endogenous loci while maintaining high specificity. Collectively, our results demonstrate that integrating protein language model-assisted filtering with structure-guided rational design provides an effective strategy for engineering PspCas12f1 and may facilitate the optimization of additional compact CRISPR nucleases.

INTRODUCTION

The CRISPR-Cas system is a well-known adaptive immune system found in archaea and bacteria.[1] It is categorized into two major classes and six types, which are further subdivided into more than 30 subtypes.[2] Class 2 systems, including type II (Cas9) and type V (Cas12), utilize a single multidomain effector protein and have been widely adopted as genome-editing tools.[3] In these systems, the CRISPR nuclease functions together with two noncoding RNAs: the CRISPR RNA (crRNA) and the *trans*-activating crRNA (tracrRNA).[3] These two RNAs can be artificially fused into a single guide RNA (sgRNA), thereby simplifying the editing machinery.[4] During genome editing, Cas nucleases first recognize a short protospacer adjacent motif (PAM) on the target DNA and locally unwind the DNA duplex. Complementary base pairing initiated at the PAM-proximal seed region promotes R-loop formation and, upon sufficient guide-target complementarity, triggers DNA cleavage to generate a double-strand break (DSB).[5] Subsequent repair of the DSB by the host cellular machinery leads to targeted sequence modifications.

To date, dozens of CRISPR-Cas genome editors have been developed.[6–15] Among them, the CRISPR-Cas12f system has drawn considerable attention owing to its ultracompact size (approximately

Received 1 April 2026; accepted 27 August 2026;
https://doi.org/10.1016/j.ymthe.2026.08.043.

[10]These authors contributed equally

Correspondence: Rui Wang, Center for Medical Research and Innovation, Shanghai Pudong Hospital, Fudan University Pudong Medical Center, School of Life Sciences, Fudan University, Shanghai 200438, China.
E-mail: wangr@fudan.edu.cn
Correspondence: Renjie Chai, Spine Surgery Department, Nantong First People's Hospital, State Key Laboratory of Digital Medical Engineering, School of Life Sciences and Technology, School of Medicine, Advanced Institute for Life and Health, Jiangsu Province High-Tech Key Laboratory for Bio-Medical Research, Southeast University, Nanjing 210096, China.
E-mail: renjiec@seu.edu.cn
Correspondence: Wei V. Zheng, Intervention and Cell Therapy Center, Peking University Shenzhen Hospital, Shenzhen, Guangdong 518036, China.
E-mail: zhengw2013@yeah.net
Correspondence: Yongming Wang, Center for Medical Research and Innovation, Shanghai Pudong Hospital, Fudan University Pudong Medical Center, School of Life Sciences, Fudan University, Shanghai 200438, China.
E-mail: ymw@fudan.edu.cn

Please cite this article in press as: Liu et al., Rational and computation-assisted engineering of a compact and efficient CRISPR-Cas12f genome editor, Molecular Therapy (2026), https://doi.org/10.1016/j.ymthe.2026.08.043

400–700 amino acids).[16] Cas12f nucleases typically recognize pyrimidine-rich PAM sequences and often exhibit low intrinsic editing efficiency.[17] Through extensive protein engineering and optimization of sgRNA scaffolds, several high-activity Cas12f variants, including Un1Cas12f1,[18–20] enAsCas12f1,[21] SpaCas12f1,[22] CnCas12f1,[23] and enOsCas12f1,[24] have been generated. Because each Cas12f nuclease possesses distinct PAM-recognition properties and target-sequence preferences, the continued discovery and engineering of additional Cas12f orthologs are crucial for broadening the CRISPR toolkit and enabling more versatile manipulation of complex genomes.

Electrostatic optimization has emerged as an effective strategy for enhancing nuclease activity. Introducing positively charged residues near DNA-binding interfaces strengthens electrostatic attraction to the negatively charged DNA backbone, thereby facilitating DNA unwinding,[25] stabilizing R-loop formation, and improving editing efficiency across Cas9 and Cas12 effectors. In parallel, advances in AI-powered protein language models (PLMs), such as ESM2[26] and SaProt,[27] provide new opportunities to predict potentially beneficial mutations. However, PLMs alone often suffer from limited precision and residue-type bias,[28,29] highlighting the need to develop practical strategies that effectively integrate computational prediction with rational, structure-guided design.[30]

In this study, we systematically screened 23 Cas12f nucleases and identified PspCas12f1 and TcCas12f1 as two active genome editors in mammalian cells. Through rational protein engineering and optimization of the PspCas12f1 sgRNA scaffold, we further developed an enhanced variant, enPspCas12f1, that achieves genome-editing efficiencies comparable to SpCas9 while maintaining high specificity. Our work expands the Cas12f toolbox with a highly active compact genome editor and illustrates the utility of combining PLM-assisted filtering with structure-guided rational engineering for optimizing Cas12f nucleases.

RESULTS

Identification of active Cas12f orthologs for mammalian genome editing

To identify Cas12f orthologs suitable for genome editing, we first selected 19 representative nucleases spanning distinct evolutionary branches based on the phylogenetic analysis reported in the initial characterization of the Cas14 (Cas12f) family.[16] We then used AsCas12f1 and SpaCas12f1, two nucleases that had been experimentally demonstrated to mediate DNA cleavage in bacteria,[17] as reference sequences to search the NCBI database for homologous proteins, leading to the identification of four additional Cas12f1 orthologs, namely AdCas12f1, PspCas12f1, Pt3Cas12f1, and TcCas12f1. These 23 nucleases range from 400 to 743 amino acids in length (Table S1). Phylogenetic analysis grouped these orthologs into seven distinct clades, among which the previously reported Cas12f nucleases active in mammalian cells are distributed across three separate clades (Figure S1A). Each identified CRISPR-Cas12f locus encodes a Cas12f nuclease, a CRISPR array, and a putative tracrRNA, predicted based on complementarity to the repeat sequence (Figure S1B).

For each ortholog, we constructed an sgRNA by fusing the 5′ end of the direct repeat to the 3′ end of the corresponding tracrRNA. The resulting sgRNA scaffolds contained 2–8 stem-loop structures (Figures S2 and S3).

We next evaluated the genome-editing activities of these orthologs using a GFP-activation reporter assay.[8] In this system, a target sequence, including a 5-bp random sequence followed by a protospacer, was inserted into the GFP coding sequence, introducing a frameshift mutation. When an active Cas12f nuclease induces insertions or deletions (indels) at the target site, the GFP open reading frame is restored in a subset of cells, producing GFP-positive signals (Figure 1A). The reporter construct was integrated into HEK293T cells via lentiviral transduction. After co-transfection of Cas12f and sgRNA expression plasmids, GFP-positive cells were detected for Un1Cas12f1 (used as a positive control), as well as for PspCas12f1 and TcCas12f1 (Figures 1B and S4), indicating that these two nucleases are active in mammalian cells.

To confirm targeted editing, GFP-positive cells were isolated and subjected to deep sequencing of the target site. The results revealed clear indel formation at the expected loci for all three active nucleases (Figures S5A–S5D). WebLogo and PAM wheel analyses of the sequencing data showed that Un1Cas12f1 recognizes a TTTR PAM (R = A or G), consistent with previous findings.[19] In contrast, PspCas12f1 recognizes a TTC PAM, while TcCas12f1 recognizes a CTTR PAM (Figures 1C and 1D).

To further evaluate the genome-editing activities of PspCas12f1 and TcCas12f1 in mammalian cells, we compared the editing activities of PspCas12f1, SpaCas12f1, and OsCas12f1 across 13 endogenous genomic loci. Among these targets, nine contained a CTTC PAM, and four contained a TTTC PAM. Overall, PspCas12f1 showed strong editing activity across the tested loci, showing significantly higher editing efficiencies than SpaCas12f1 at seven loci and outperforming OsCas12f1 at nine loci (Figure S6).

In addition, we compared the genome-editing activities of TcCas12f1 and OsCas12f1 at 14 endogenous loci harboring a CTTA PAM. TcCas12f1 likewise exhibited genome-editing activity, with significantly higher editing efficiencies than OsCas12f1 at 10 of the 14 loci (Figure S7). Collectively, these results demonstrate that both PspCas12f1 and TcCas12f1 are capable of mediating genome editing in mammalian cells (Figures S8A and S8B). Given that PspCas12f1 recognizes the simpler TTC PAM and can therefore target a broader range of genomic sites, we selected PspCas12f1 for subsequent engineering.

Optimization of spacer length enhances PspCas12f1 activity

We next sought to determine the optimal spacer length for PspCas12f1. A series of sgRNAs containing 17–25 nt spacers was designed to target exon 10 of the *PPP1R12C* gene. Five days after co-transfection of HEK293T cells with PspCas12f1 and each sgRNA expression plasmid, genomic DNA was extracted and subjected to

Please cite this article in press as: Liu et al., Rational and computation-assisted engineering of a compact and efficient CRISPR-Cas12f genome editor, Molecular Therapy (2026), https://doi.org/10.1016/j.ymthe.2026.08.043

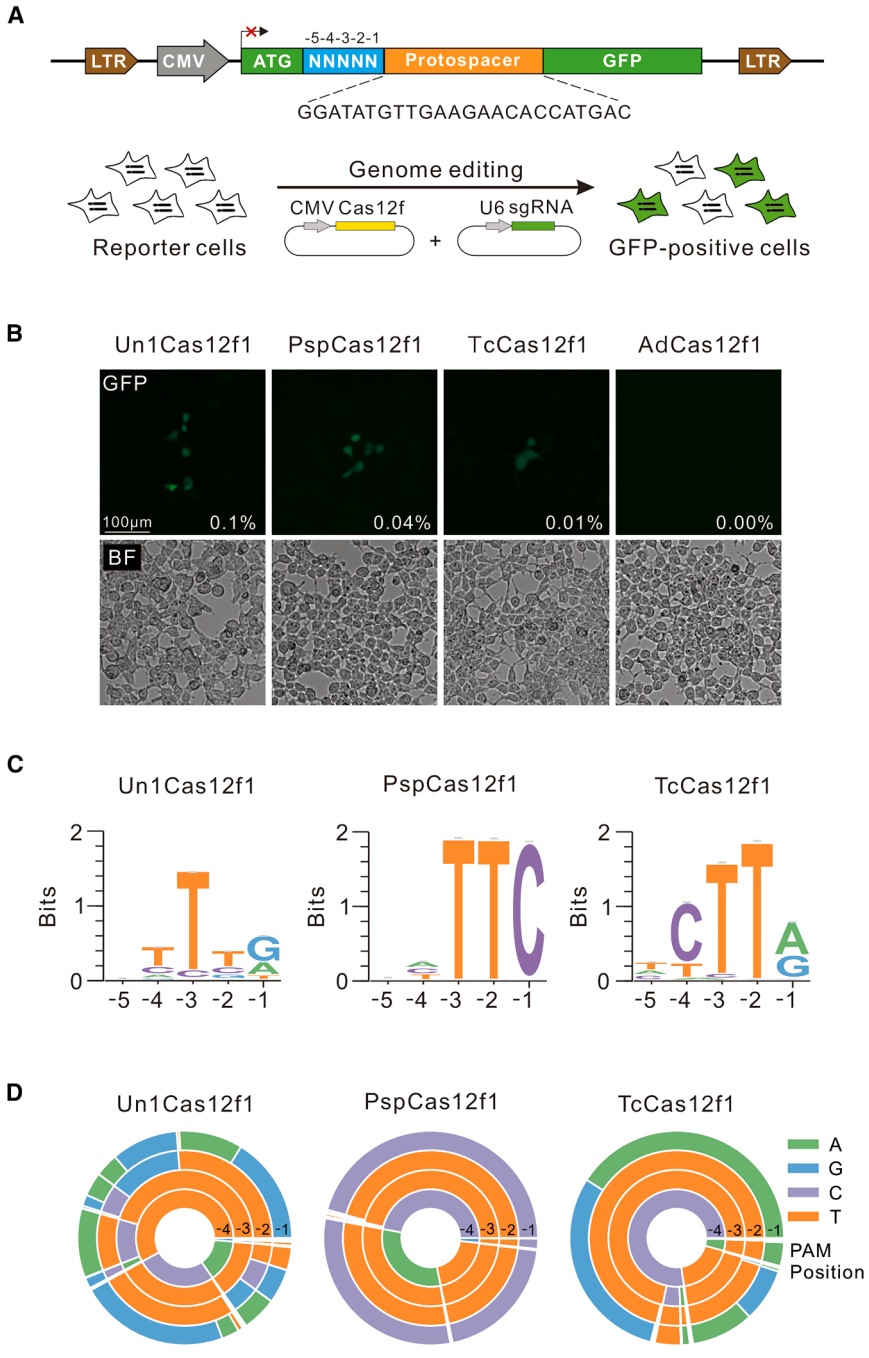


Figure 1. Identification of PspCas12f1 and TcCas12f1 as active Cas12f nucleases in mammalian cells using a GFP reporter

(A) Schematic of the GFP-activation reporter. The reporter contains a 5-bp randomized sequence followed by the target protospacer, both inserted between the ATG start codon and the GFP coding sequence. The reporter library was stably integrated into HEK293T cells via lentiviral transduction. Genome editing by active Cas12f enzymes restores the correct GFP reading frame, resulting in GFP-positive cells. (B) Quantification of GFP activation. Two Cas12f1 orthologs triggered robust GFP expression. The percentages of GFP-positive cells are shown. Un1Cas12f1 was used as a positive control. Scale bar, 100 μm. (C) PAM sequence logos generated from deep-sequencing data using WebLogo. (D) PAM wheel plots derived from deep-sequencing data. Numbers indicate PAM nucleotide positions relative to the protospacer.

targeted deep sequencing. The results revealed that the 20-nt spacer yielded the highest editing efficiency (Figure S9A). To validate this observation, we performed a similar analysis using sgRNAs targeting the 5′ UTR of the *EMX1* gene, with spacer length also varying from 17 to 25 nt. Consistent with the previous results, the 20-nt spacer again produced the strongest editing activity (Figure S9B). Based on these results, the 20-nt spacer was used in all subsequent experiments.

To further assess its performance, we compared PspCas12f1 with SpaCas12f1[22] and the engineered variant enOsCas12f1[24] across 12 endogenous sites. PspCas12f1 showed slightly higher activity than SpaCas12f1 but markedly lower efficiency than enOsCas12f1 (Figures S10A and S10B). These results suggest that the intrinsic activity of PspCas12f1 remains suboptimal and could benefit from further optimization.

Engineering of the PspCas12f1 sgRNA scaffold enhances editing efficiency

We first optimized the PspCas12f1 sgRNA scaffold. Previous studies have also shown that deleting specific stem loops within the sgRNA scaffold, correcting mismatches, or converting unstable A:U pairs into more stable G:C pairs can boost nuclease activity.[22–24] We therefore adopted similar scaffold modification strategies in this study.

We assessed the impact of truncating individual stem-loop structures on editing efficiency. The PspCas12f1 sgRNA scaffold forms five stem loops (SL1–SL5; Figure 2A). Eight truncation variants (V1.1–V1.8) were designed to target different stem loops, and their editing efficiencies were measured at three endogenous loci in HEK293T cells (Figure S11A). Deletion of any single complete stem loop derived from the tracrRNA (V1.1–V1.5), or removal of the complete stem loop between the tracrRNA and the repeat sequence (V1.6), reduced editing efficiency (Figures 2B and S11B). In contrast, the partial truncation represented by V1.7 increased activity at several loci,

Please cite this article in press as: Liu et al., Rational and computation-assisted engineering of a compact and efficient CRISPR-Cas12f genome editor, Molecular Therapy (2026), https://doi.org/10.1016/j.ymthe.2026.08.043

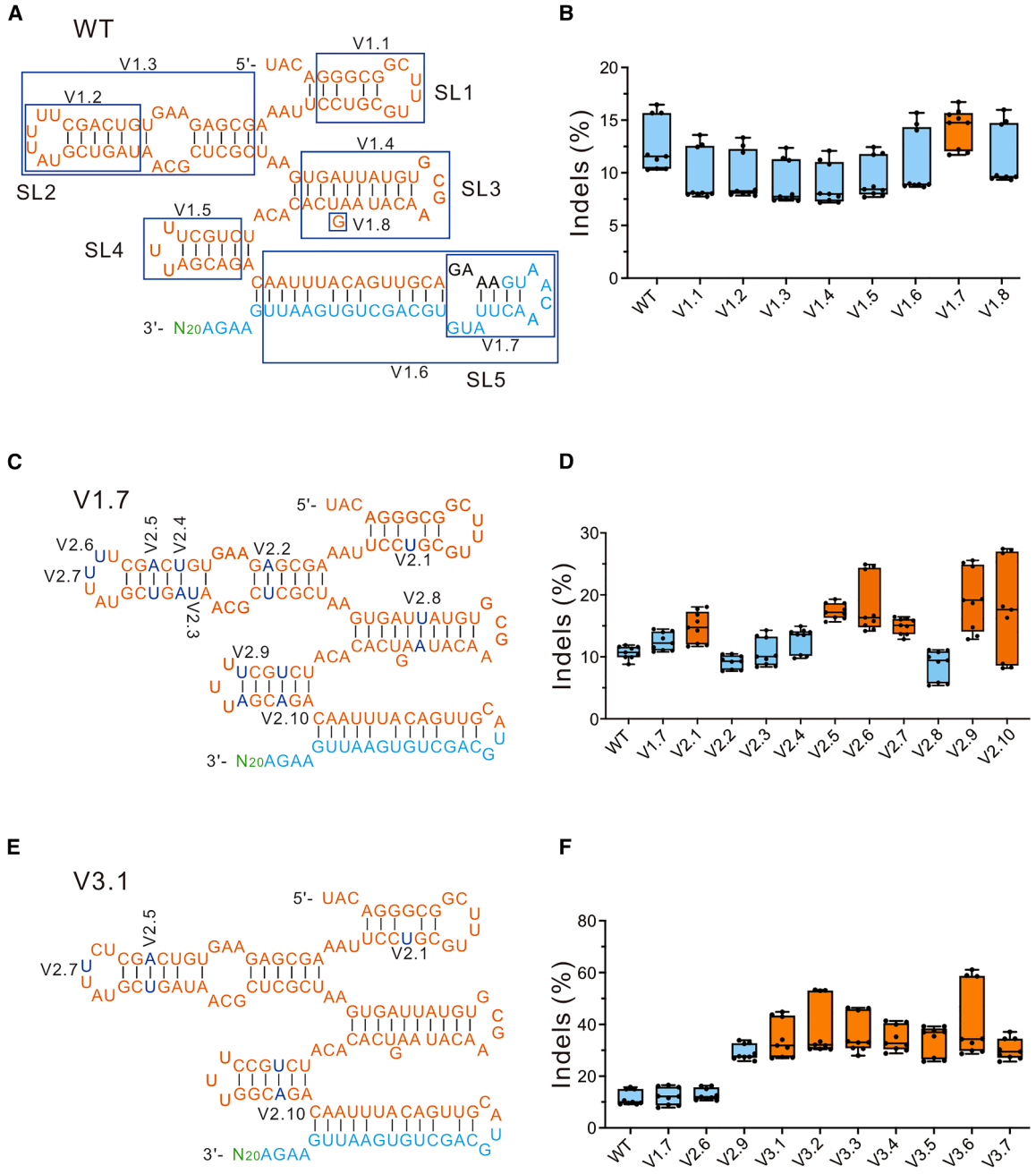


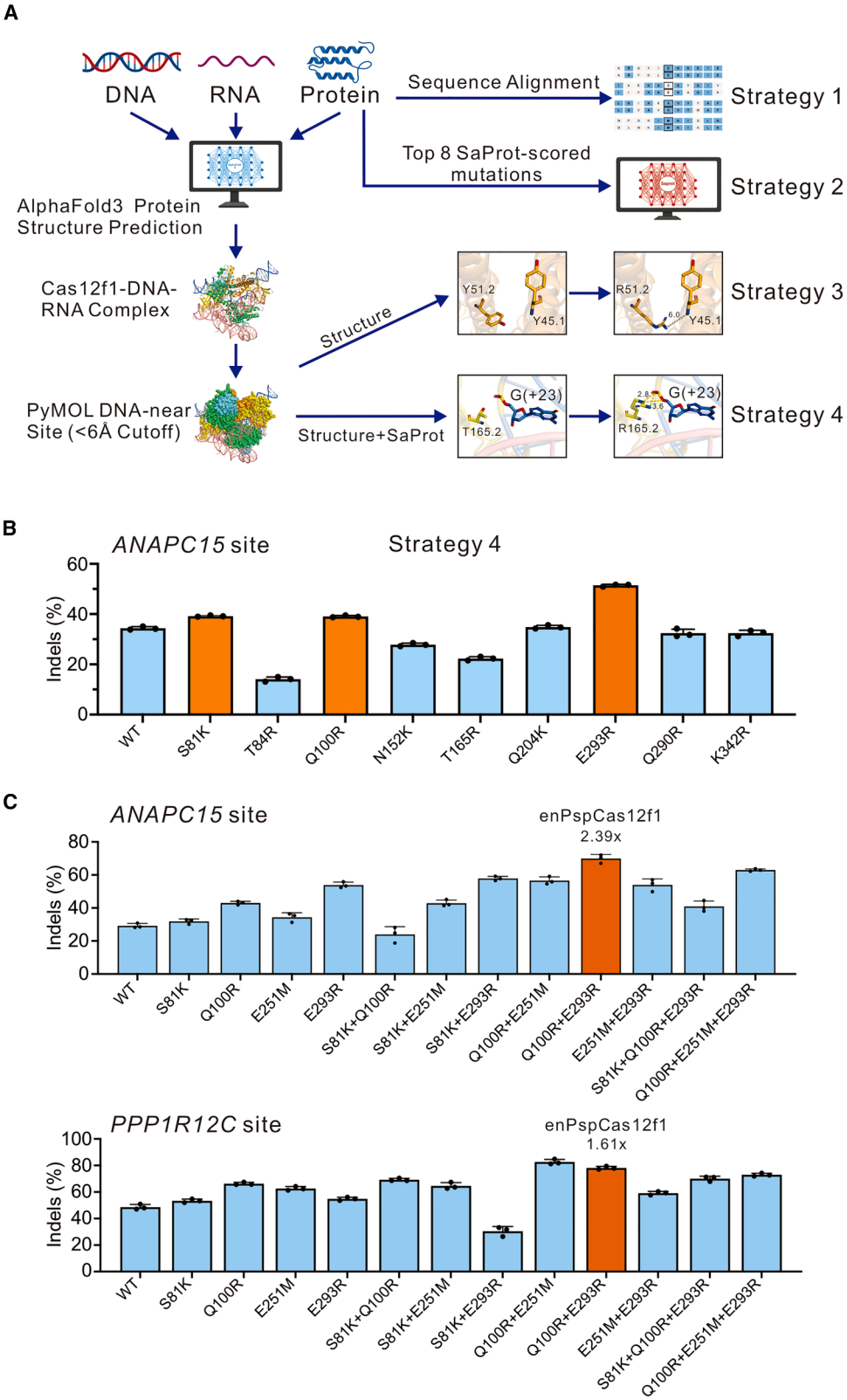
Figure 2. Rational sgRNA scaffold optimization enhances PspCas12f1 genome-editing efficiency

(A) Predicted secondary structure of the wild-type (WT) PspCas12f1 sgRNA. The tracrRNA sequence is shown in orange, the CRISPR repeat sequence in light blue, and truncated stem-loop (SL) regions in subsequent variants are indicated by blue boxes. (B) Quantitative comparison of genome-editing efficiencies of sgRNA truncation variants across three target sites. (C) Predicted secondary structure of sgRNA_V1.7. Positions containing engineered nucleotide substitutions are marked in dark blue. (D) Quantitative comparison of editing efficiencies for the WT sgRNA, sgRNA_V1.7, and its substitution variants across three target sites. (E) Predicted secondary structure of sgRNA_V3.1. Four substituted nucleotides are highlighted in dark blue. (F) Quantitative comparison of editing efficiencies for the WT sgRNA, sgRNA_V1.7, sgRNA_V2.6, sgRNA_V2.9, and seven combinatorial substitution variants across three target sites. Indel frequencies were quantified by targeted deep sequencing.

whereas deletion of the SL3 "bubble" (V1.8) again diminished activity. Based on these results, we selected V1.7 for subsequent optimization.

Using version V1.7 as the backbone, we introduced base substitutions at sites of scaffold mismatches and at selected A:U pairs to generate 10 point-mutant variants, V2.1–V2.10, and measured their

Please cite this article in press as: Liu et al., Rational and computation-assisted engineering of a compact and efficient CRISPR-Cas12f genome editor, Molecular Therapy (2026), https://doi.org/10.1016/j.ymthe.2026.08.043



(legend on next page)

Please cite this article in press as: Liu et al., Rational and computation-assisted engineering of a compact and efficient CRISPR-Cas12f genome editor, Molecular Therapy (2026), https://doi.org/10.1016/j.ymthe.2026.08.043

editing efficiencies at three endogenous loci in HEK293T cells (Figures 2C and S12A). The specific changes included U-to-C substitutions to promote pairing with G (V2.1 and V2.3), conversion of A:U pairs to G:C pairs to strengthen base pairing (V2.2, V2.4, V2.5, V2.8, and V2.10), and replacement of consecutive U residues with C to remove U runs that may trigger premature transcription termination or reduce transcription efficiency (V2.6, V2.7, and V2.9). Testing across three loci revealed that V2.1, V2.5–V2.7, and V2.9–V2.10 exhibited higher activity than V1.7 (Figures 2D and S12B).

We next combined beneficial point mutations to create seven sgRNA combination variants, V3.1–V3.7, and evaluated their editing efficiencies at the same three endogenous loci in HEK293T cells (Figures 2E and S13A). We first combined V2.6 and V2.9 to produce V3.1, thereby removing the contiguous U sequence. We then integrated mutations from V2.1, V2.5, V2.7, and V2.10 into the V3.1 scaffold to generate V3.2–V3.7. Specifically, V3.2 comprises V2.6, V2.9, and V2.10; V3.3 comprises V2.5, V2.6, and V2.9; V3.4 comprises V2.5, V2.6, V2.9, and V2.10; V3.5 comprises V2.1, V2.5, V2.6, V2.9, and V2.10; V3.6 comprises V2.5, V2.6, V2.7, V2.9, and V2.10; and V3.7 comprises V2.1, V2.5, V2.6, V2.7, V2.9, and V2.10. Across the three tested loci, V3.6 displayed the highest editing activity compared with the other combination variants (Figures 2F and S13B). These results demonstrate that rational redesign of the sgRNA scaffold substantially enhances PspCas12f1 activity, laying the foundation for further protein optimization.

Structure-guided protein engineering assisted by PLM filtering markedly improves PspCas12f1 activity

To further improve the performance of PspCas12f1, we designed and experimentally evaluated four complementary protein-engineering strategies based on sequence homology, PLM prediction, and structure-guided rational design (Figure 3A). Strategy 1 relied exclusively on sequence homology to introduce previously reported activity-enhancing AsCas12f1 substitutions at locally conserved homologous positions in PspCas12f1.[31] Strategy 2 relied exclusively on the PLM SaProt[27] to prioritize the top eight predicted functional substitutions. Strategy 3 was developed independently of SaProt predictions and relied exclusively on structural information to optimize residues at the protein dimer interface. In contrast, strategy 4 integrated SaProt filtering (SaProt score > 0) with structure-guided design based on AlphaFold3-predicted models. Specifically, strategy 4 targeted DNA-proximal residues within 6 Å that did not directly interact with the phosphate backbone. These residues were individually substituted with arginine or lysine to introduce new electrostatic interactions with the DNA.

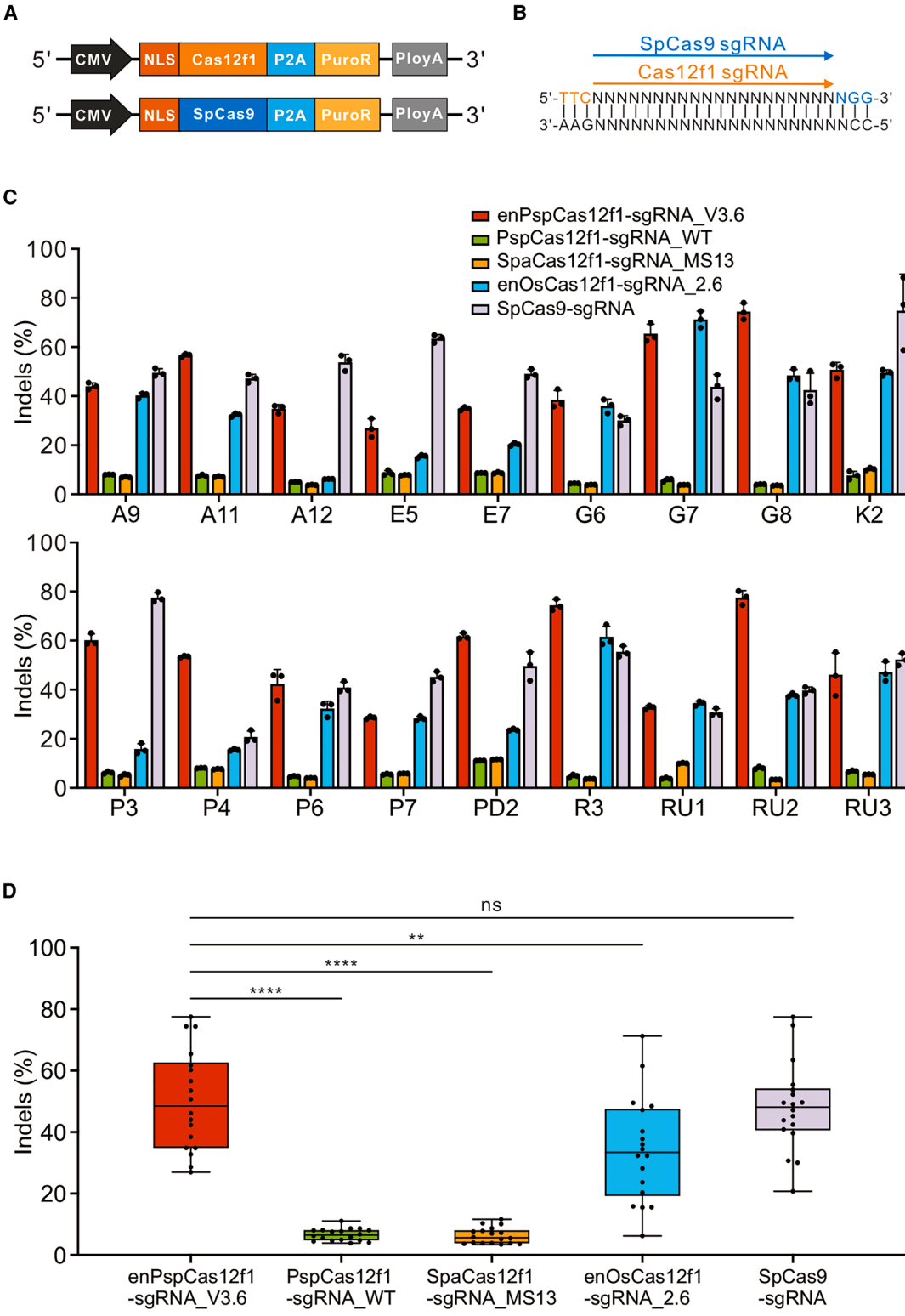
Mutations generated from these four independent engineering strategies were introduced into PspCas12f1 and experimentally evaluated for genome-editing activity at the *ANAPC15* locus in HEK293T cells. Using strategy 1, we designed 22 amino acid substitutions based on sequence homology and identified only one beneficial variant, E251M, that exhibited higher genome-editing activity than wild-type (WT) PspCas12f1 (Figure S14A). Likewise, the eight substitutions prioritized based solely on SaProt scores in strategy 2 failed to substantially improve editing activity (Figure S14B). Strategy 3, which focused on structure-guided optimization of the protein dimer interface, also failed to identify beneficial variants, as all 17 substitutions generated using this strategy exhibited lower editing activity than the WT (Figure S14C).

In contrast, strategy 4, which combined SaProt-based filtering (SaProt score > 0) with structure-guided rational design, identified three beneficial substitutions: S81K, Q100R, and E293R (Figure 3B). Specifically, strategy 4 targeted residues located adjacent to, but not directly contacting, the DNA backbone. Introducing positively charged substitutions at these positions may establish new electrostatic interactions without disrupting existing protein-DNA contacts, thereby enhancing editing activity. Among the four strategies, strategy 4 achieved the greatest overall improvement in PspCas12f1 activity (Figure S14D), demonstrating that integrating PLM prediction with structure-guided rational design enriches functionally beneficial mutations more effectively than either approach alone. Collectively, these findings establish an effective framework for engineering compact CRISPR nucleases by integrating PLM prediction with mechanistically informed structure-guided rational design.

We next assessed whether SaProt scores retained discriminative power within a structurally constrained candidate set, which is essential for the effectiveness of the integrated screening strategy. We analyzed 39 candidate mutations designed at 26 residues located within 6 Å of the DNA and classified mutations according to their SaProt scores (>0 or $\leq$0). The positive-score group exhibited a mean efficiency fold change of 0.743 and a beneficial mutation hit rate of 25%, compared to 0.476 and 4.3% for the non-positive group ($p = 0.0152$) (Figure S15). This analysis provides direct evidence that, even after structural proximity filtering, the SaProt score further enriches for functionally improved variants. Consequently, integrating PLM-based prediction as a secondary filter

Figure 3. Comparative evaluation of protein engineering strategies for improving PspCas12f1 activity

(A) Schematic overview of the rational protein engineering workflow, encompassing four complementary strategies for predicting beneficial protein variants. (B) Comparison of genome-editing efficiencies at the *ANAPC15* locus in HEK293T cells between WT PspCas12f1 and single-point mutants predicted by strategy 4. Blue bars represent the WT and individual single-amino-acid substitution mutants, whereas orange bars indicate variants exhibiting increased activity relative to the WT. Indel frequencies were quantified by targeted deep sequencing. (C) Editing efficiencies of WT PspCas12f1, activity-enhancing single-point mutants, and multi-mutation variants (double or triple mutants) generated from combinations of high-activity substitutions at the *ANAPC15* and *PPP1R12C* loci. Blue bars denote WT PspCas12f1 and different mutants, while orange bars highlight the best-performing double mutant (enPspCas12f1). Indel frequencies were measured by targeted deep sequencing. Data are presented as mean ± standard deviation (SD) from three independent experiments ($n = 3$).



(legend on next page)

Please cite this article in press as: Liu et al., Rational and computation-assisted engineering of a compact and efficient CRISPR-Cas12f genome editor, Molecular Therapy (2026), https://doi.org/10.1016/j.ymthe.2026.08.043

on top of structure-guided selection substantially enhances the discovery of beneficial mutations.

Identification and structural characterization of enPspCas12f1

We next tested whether beneficial mutations identified by the different strategies exhibited additive effects. Four mutations that increased editing activity, S81K, Q100R, E251M, and E293R, were combined to generate six double mutants and two triple mutants. The editing activities of all mutants were evaluated in parallel at the *ANAPC15* and *PPP1R12C* loci (Figure 3C). Combinations containing Q100R and E293R increased editing activity relative to WT PspCas12f1 by 2.39- and 1.61-fold at the two loci, respectively. Among the double- and triple-mutant combinations tested, the Q100R/E293R variant exhibited strong genome-editing activity across the evaluated target sites. Although it was not the highest-performing variant at every individual locus, it showed the best overall editing performance across the two tested loci. Therefore, the Q100R/E293R variant was selected for subsequent characterization and was designated as enPspCas12f1. Interestingly, adding S81K or E251M to the Q100R/E293R background to create triple mutants did not further enhance activity relative to the double mutant, suggesting a possible ceiling for additive effects or the introduction of conformational constraints by additional substitutions.

To delineate the structural basis underlying the enhanced activity conferred by the Q100R and E293R substitutions, we predicted the structure of the PspCas12f1-sgRNA-DNA complex using AlphaFold3 (Figure S16A). Structural modeling suggested that the Q100R substitution reorients the positively charged arginine side chain toward the DNA backbone in both protomers, resulting in shorter protein-DNA distances than those observed for the corresponding glutamine residues. The closer proximity of the arginine side chain is predicted to strengthen electrostatic interactions with the DNA phosphate backbone. In one protomer (R100.1), the substituted arginine was predicted to establish closer contacts with multiple nucleotides (Figure S16B), whereas in the other protomer (R100.2), the substituted arginine was predicted to form an additional hydrogen bond with the DNA (Figure S16C).

Similarly, replacement of E293 with arginine was predicted to generate an additional hydrogen bond between R293.1 and the DNA in one protomer (Figure S16D). In contrast, the corresponding residue in the second protomer (R293.2) remained spatially distant from the DNA and was not predicted to establish additional contacts (Figure S16E). Collectively, these structural predictions suggest that both the Q100R and E293R substitutions enhance protein-DNA interactions by increasing electrostatic complementarity between the protein surface and the DNA phosphate backbone and, in specific structural contexts, by introducing additional hydrogen-bonding interactions. These changes provide a plausible structural explanation for the enhanced genome-editing activity of enPspCas12f1.

We then compared enPspCas12f1 with PspCas12f1 across 30 endogenous loci. enPspCas12f1 showed significantly higher efficiency at every locus, and its mean editing activity across the 30 loci was 3.72 times that of PspCas12f1 (Figures S17A and S17B). Collectively, structure-guided mutagenesis yielded a double mutant (Q100R/E293R) with markedly improved genome-editing efficiency.

The engineered enPspCas12f1 achieves SpCas9-comparable genome-editing efficiency

To evaluate the combined effects of protein engineering and sgRNA optimization, we compared the genome-editing activities of enPspCas12f1 (sgRNA_V3.6), PspCas12f1 (sgRNA_WT), the previously reported SpaCas12f1 (sgRNA_MS13),[22] enOsCas12f1 (sgRNA_2.6),[24] and the widely used SpCas9. To ensure comparable expression among all nucleases, identical expression backbones were employed (Figure 4A). We selected 18 endogenous loci that met the PAM requirements of all Cas12f1 variants and SpCas9 and assessed editing efficiency in HEK293T cells (Figure 4B).

The results showed that enPspCas12f1 consistently outperformed PspCas12f1 and SpaCas12f1 at all tested loci (Figure 4C). On average, enPspCas12f1 (sgRNA_V3.6) exhibited a 7.60-fold increase in editing activity compared with PspCas12f1 (sgRNA_WT). Across different targets, enPspCas12f1 displayed efficiencies comparable to or higher than enOsCas12f1, depending on the locus. While its activity exceeded that of SpCas9 at several sites, it was slightly lower at others. Overall, enPspCas12f1 demonstrated SpCas9-comparable genome-editing efficiency (Figure 4D). These findings indicate that combining protein engineering and sgRNA scaffold optimization effectively transformed PspCas12f1 into a compact and highly active genome editor.

Notably, the data also revealed that each nuclease exhibits distinct sequence preferences. For example, enPspCas12f1 achieved 25% indel frequency at the E5 locus and 80% at the RU2 locus, whereas

Figure 4. enPspCas12f1 exhibits superior genome-editing efficiency compared with other Cas12f variants and SpCas9

(A) Schematic of the plasmid constructs used to express Cas12f1 variants (enPspCas12f1, PspCas12f1, SpaCas12f1, and enOsCas12f1) and SpCas9. All expression plasmids share the same backbone. NLS, nuclear localization signal; poly(A), polyadenylation signal. (B) Schematic representation of the shared target sites used for Cas12f1 variants (enPspCas12f1, PspCas12f1, SpaCas12f1, and enOsCas12f1) and SpCas9. PAM sequences recognized by SpCas9 are shown in blue, and PAM sequences recognized by Cas12f1 are shown in orange. (C) Comparison of editing efficiencies of enPspCas12f1, PspCas12f1, SpaCas12f1, enOsCas12f1, and SpCas9 across 18 endogenous loci in HEK293T cells. Editing outcomes were quantified as indel frequencies by targeted deep sequencing. Data are presented as mean ± standard deviation (SD) from three independent expriments ($n = 3$). (D) Boxplots of indel frequencies generated by enPspCas12f1, PspCas12f1, SpaCas12f1, enOsCas12f1, and SpCas9. Boxes represent the interquartile range, center lines indicate the median, and whiskers denote the minimum and maximum values. Each data point represents the mean indel frequency at one endogenous target site from three independent biological replicates. Statistical significance was determined with a threshold of $p < 0.05$ ($^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, and $^{****}p < 0.0001$).

Please cite this article in press as: Liu et al., Rational and computation-assisted engineering of a compact and efficient CRISPR-Cas12f genome editor, Molecular Therapy (2026), https://doi.org/10.1016/j.ymthe.2026.08.043

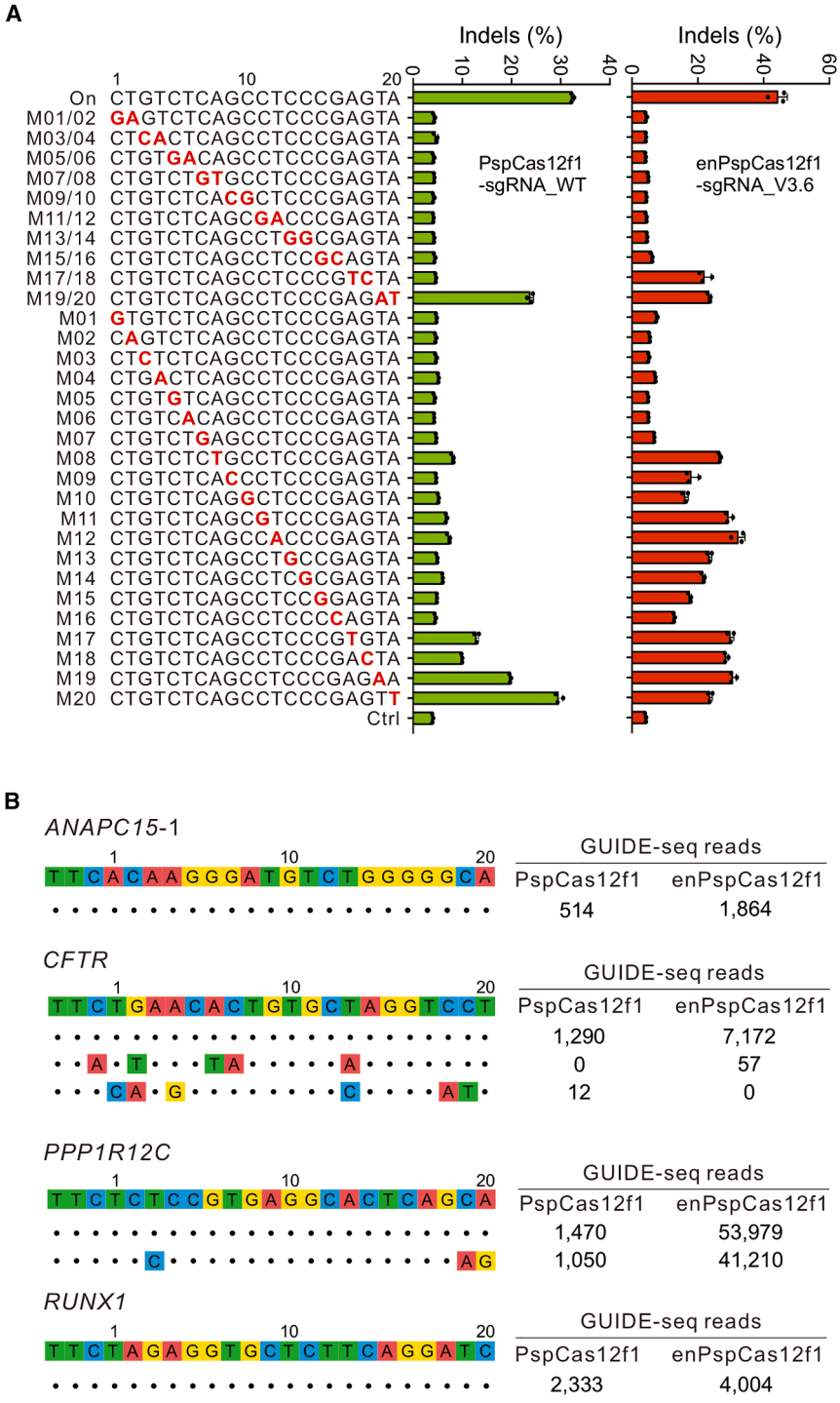


Figure 5. enPspCas12f1 maintains high genome-wide specificity despite enhanced activity

(A) Tolerance of PspCas12f1 and enPspCas12f1 to dinucleotide and single-nucleotide mismatches within the sgRNA. Mismatched nucleotides are shown in red. sgRNA activities were quantified by targeted deep sequencing. Data are presented as mean ± standard deviation (SD) from three independent experiments (n = 3). (B) Genome-wide off-target analysis of PspCas12f1 and enPspCas12f1 using GUIDE-seq. Target sequences for four endogenous sites are shown on the left, and corresponding read counts for on-target and off-target sites are shown on the right.

enOsCas12f1 showed 18% and 40%, and SpCas9 yielded 65% and 42%, respectively (Figure 4C). These results suggest that enPspCas12f1 provides a valuable alternative to existing editors, offering superior activity at specific genomic loci and expanding the flexibility of CRISPR-based genome engineering.

Mismatch tolerance and genome-wide off-target profiling of enPspCas12f1

We next assessed the specificity of PspCas12f1 and its engineered variant enPspCas12f1. To evaluate mismatch tolerance, we designed a panel of 10 sgRNAs containing dinucleotide mismatches relative to the *EMX1* target site. PspCas12f1 exhibited strong sensitivity to dinucleotide mismatches at positions 1–18, whereas enPspCas12f1 was sensitive to mismatches at positions 1–16 (Figure 5A). We further examined single-nucleotide mismatch tolerance using a set of 20 sgRNAs, each containing a single mismatch relative to the *EMX1* target site. PspCas12f1 was sensitive to single mismatches at positions 1–16, while enPspCas12f1 tolerated mismatches beyond position 7, indicating reduced sequence stringency (Figure 5A). Collectively, these results suggest that enPspCas12f1 achieves enhanced activity at the expense of target specificity.

To further investigate genome-wide specificity, we performed GUIDE-seq (genome-wide unbiased identification of double-stranded breaks enabled by sequencing) in HEK293T cells.[32] We designed nine sgRNAs targeting the *ANAPC15*, *RUNX1*, *CFTR*, *PPP1R12C*, *RNF2*, *EMX1*, and *DYRK1A* loci. Cas12f expression plasmids, sgRNA expression plasmids, and GUIDE-seq double-stranded oligodeoxynucleotides (dsODNs) were co-electroporated into HEK293T cells, and genomic DNA was harvested 5 days after electroporation for GUIDE-seq analysis. GUIDE-seq analysis revealed abundant read counts at all target sites for both nucleases. Genome-wide off-target analysis identified three off-target sites for PspCas12f1 and five off-target sites for enPspCas12f1 (Figures 5B and S18). Although enPspCas12f1 exhibited a modest increase in the number of detected off-target sites, the overall number of

Please cite this article in press as: Liu et al., Rational and computation-assisted engineering of a compact and efficient CRISPR-Cas12f genome editor, Molecular Therapy (2026), https://doi.org/10.1016/j.ymthe.2026.08.043

off-target events remained low, and no widespread nonspecific editing was observed. Collectively, these GUIDE-seq analyses indicate that enPspCas12f1 retains low off-target activity while substantially enhancing genome-editing activity, further supporting its potential as an efficient genome-editing tool with favorable editing specificity.

Therapeutic genome editing and AAV-mediated delivery of enPspCas12f1

We next evaluated the therapeutic potential of enPspCas12f1 in clinically relevant genome-editing applications. We selected two disease-relevant hearing-loss targets, *CIDEA* and *GSDMD*. Previous studies have identified *CIDEA* and *GSDMD* as promising therapeutic targets for hearing loss, as disruption of *CIDEA* reduces cochlear hair cell apoptosis,[33] whereas inhibition of *GSDMD*-mediated pyroptosis alleviates hair cell damage and protects auditory function.[34] We designed 12 endogenous target sites within each gene and compared the genome-editing efficiencies of PspCas12f1-sgRNA_WT and enPspCas12f1-sgRNA_V3.6 across these loci. Across all 24 endogenous target sites, enPspCas12f1-sgRNA_V3.6 consistently exhibited significantly higher editing efficiencies than PspCas12f1-sgRNA_WT. The highest editing efficiency reached 93% at the *CIDEA* locus and 83% at the *GSDMD* locus (Figures S19A–S19C). These results demonstrate that the engineered enPspCas12f1-sgRNA_V3.6 enables highly efficient genome editing at therapeutically relevant hearing-loss genes.

We next evaluated whether this enhanced editing capability could be effectively translated to an adeno-associated virus (AAV)-mediated delivery system. Given the compact size of enPspCas12f1, we constructed an all-in-one expression cassette in which enPspCas12f1 and its corresponding sgRNA were integrated into a single AAV vector. The expression cassettes were packaged into the AAV-DJ serotype for viral production. We subsequently generated 12 AAV-DJ vectors, each carrying an individual sgRNA targeting one of 12 endogenous genomic loci, and transduced three distinct human cell lines, including SH-SY5Y, A375, and C33A. Seven days after transduction, genome-editing efficiencies at each target site were quantified by targeted deep sequencing.

AAV-delivered enPspCas12f1 mediated efficient genome editing in all three human cell lines. In SH-SY5Y cells, editing efficiencies exceeded 60% at five target sites, with a maximum editing efficiency of approximately 76% (Figure S20A). In A375 cells, the highest editing efficiency reached 81% (Figure S20B). In C33A cells, editing efficiencies also exceeded 60% at five target sites, with a maximum editing efficiency of approximately 72% (Figure S20C). Although editing efficiencies varied among target sites and cell types, enPspCas12f1 consistently exhibited robust genome-editing activity across all three cellular models. Collectively, these findings demonstrate that enPspCas12f1 enables highly efficient genome editing of therapeutically relevant hearing-loss genes and can be effectively delivered using a single all-in-one AAV vector, highlighting its compatibility with AAV-mediated delivery and its strong potential for future *in vivo* therapeutic genome-editing applications.

Integration of PLM predictions with structural filtering enhances variant discovery

To provide a more comprehensive assessment of SaProt's standalone predictive performance, we expanded the experimental dataset by including additional missense mutants, yielding a total of 107 experimentally tested variants, and systematically evaluated the relationship between SaProt scores and experimental activity. Spearman's rank correlation between SaProt scores and functional readout was modest ($\rho = 0.245$, $p < 0.05$), and the area under the receiver operating characteristic (ROC) curve (AUC) was 0.637 (Figures S21A and S21B), indicating a basic level of predictive ability. Critically, SaProt's precision was low (0.1–0.3 across all recall levels; AUC-precision recall [AUC-PR] < 0.2), and it produced a high false-positive rate (47 out of 53 predicted functional mutations were incorrect; Figures S21C and S21D), collectively underscoring that SaProt alone is insufficient for reliably pinpointing improved variants.

Using this expanded dataset, we systematically evaluated candidate SaProt score cutoffs for variant prioritization. Statistical significance (minimum *t* test and Mann-Whitney *p* values; Figure S22A) together with effect size (maximum Cohen's d; Figure S22B) identified two local optima at score thresholds of 0 and 1.536. At a threshold of 0, the resulting groups were relatively balanced ($n = 64$ versus 43) and exhibited significantly different mean efficiency fold changes (Figure S22C). Although a threshold of 1.536 also yielded a statistically significant difference, it produced substantially imbalanced group sizes ($n = 86$ versus 21; Figure S22D). These analyses support the use of a SaProt score of 0 as a reliable threshold for variant prioritization, enabling efficient reduction of the candidate mutational space before subsequent structure-guided selection. Subsequent structure-guided selection, which prioritized residues proximal to the DNA-binding interface, further refined this pool and led to the identification of highly active mutations, including Q100R and E293R. Thus, integrating SaProt-based filtering with structural constraints provides an efficient strategy for discovering beneficial mutations.

DISCUSSION

CRISPR-Cas12f nucleases possess intrinsic advantages for *in vivo* delivery due to their exceptionally compact size; however, individual Cas12f orthologs often display distinct target sequence preferences and variable editing efficiencies. Several Cas12f nucleases, including Un1Cas12f1,[19] AsCas12f1,[35] SpaCas12f1,[22] enOsCas12f1,[24] and Ti1Cas12f1,[36] have been previously characterized and shown to mediate genome editing in eukaryotic cells. Here, we systematically analyzed 23 Cas12f nucleases and identified PspCas12f1 and TcCas12f1 as compact and active editors in mammalian cells, thereby expanding the functional repertoire of Cas12f-based genome-editing tools. Through the combined optimization of the sgRNA scaffold and rational protein engineering, we further developed enPspCas12f1, a substantially enhanced variant that achieved robust and broad editing activity across multiple endogenous loci. Notably, enPspCas12f1 outperformed both enOsCas12f1 and SpCas9 at several sites, underscoring its potential as a versatile and

delivery-friendly genome editor for both basic and therapeutic applications.

In protein engineering, we systematically benchmarked multiple strategies based on sequence homology, PLM predictions, and structure-guided rational design. Among these, strategy 4 achieved the greatest overall improvement by integrating SaProt-based prioritization with structure-guided selection of DNA-proximal residues and biophysically motivated electrostatic design. Introducing positively charged residues adjacent to the DNA-binding interface, without perturbing direct DNA contact, substantially enhanced editing efficiency. The combined substitutions Q100R and E293R exemplify this principle, markedly improving activity across multiple loci. These results parallel prior engineering efforts on Cas12f orthologs, such as optimized Un1Cas12f1, AsCas12f1, and enOsCas12f1 variants that enhance ribonucleoprotein stability or surface charge balance.[18,21,24] However, unlike earlier strategies that relied on consensus alignment[21] or random mutagenesis,[24] our rational, structure-based framework explicitly delineates the electrostatic principles underlying nuclease activation. The success of enPspCas12f1 underscores that fine-tuning charge distribution around the DNA-binding interface may represent a generalizable route to boosting the efficiency of compact CRISPR enzymes.

Our study also highlights the complementary roles of computational prediction and rational design, as prioritizing candidates supported by both approaches can substantially narrow the experimental search space. While general PLMs such as SaProt[27] serve as efficient initial filters, their systematic biases, particularly the underestimation of functionally important positively charged residues (R/K), limit their ability to independently identify highly active mutants. By integrating structural modeling, physicochemical constraints, and iterative experimental feedback, we improved predictive precision while reducing screening workload. This integrated strategy leverages the strengths of both approaches: PLMs provide broad coverage of sequence space, whereas structure-guided filtering focuses experimental efforts on mutations with a higher likelihood of functional improvement. Together, these complementary approaches enabled the efficient identification of beneficial mutations during the engineering of PspCas12f1 and may facilitate the future optimization of additional compact CRISPR nucleases. Despite these improvements, enPspCas12f1 recognizes a relatively strict TTC PAM, which remains more restrictive than the NGG PAM requirement of widely used genome editors such as SpCas9[37] or the broader PAM compatibility of recently engineered Cas12f variants.[24] This limited PAM compatibility reduces the density of accessible genomic target sites and may constrain its applicability at certain disease-associated loci.

In summary, through homolog discovery, sgRNA scaffold engineering, and AI-assisted structure-guided protein optimization, we have developed enPspCas12f1, a compact and efficient genome editor with broad potential for research and therapeutic applications. Beyond developing enPspCas12f1, this work provides a practical workflow for integrating PLM-assisted filtering with structure-guided rational engineering during the optimization of PspCas12f1. Future studies will determine the general applicability of this strategy to other compact CRISPR nucleases.

MATERIALS AND METHODS

Plasmid construction

pAAV-CMV-Cas12f/Cas9-P2A-Puro

To construct the pAAV-CMV-Cas12f/Cas9-P2A-Puro expression vectors, coding sequences for SpaCas12f1, enOsCas12f1 and SpCas9 were amplified by PCR from the plasmids pCMV-SpaCas12f1-gRNA_MS13 (Addgene #190877), pCBh_3xFlag-NLS-enOsCas12f1-NLS_pA_phU6_gRNAscaffold-spacer_pCMV_mCherry_pA (Addgene #197026), and pSpCas9(BB)-2A-Puro (PX459) V2.0 (Addgene #62988), respectively, using primer pairs SpaCas12f1-F/R, enOsCas12f1-F/R, and SpCas9-F/R. The pAAV backbone fragment was amplified from pAAV-CMV-SauriCas9-puro (Addgene #135965) using primers pAAV-F/R. The coding sequences of Un1Cas12f1, OsCas12f1, and 23 additional Cas12f homologs were human codon optimized, synthesized by HuaGene (Shanghai, China), and subsequently cloned into the pAAV backbone. Final pAAV-CMV-Cas12f and pAAV-CMV-Cas9 plasmids were assembled by recombining the PCR-amplified Cas12f/Cas9 fragments with the pAAV backbone using the NEBuilder HiFi DNA Assembly Kit (NEB, E5520). For AAV delivery, an all-in-one cassette encoding enPspCas12f1 together with the sgRNA_V3.6 scaffold was assembled into the same pAAV backbone to generate pAAV-enPspCas12f1-sgRNA_V3.6. The amino acid sequences of the Cas12f homologs are listed in Table S1, and the primer sequences are provided in Table S2.

pSK-hU6-Cas12f/Cas9 sgRNA

To generate pSK-hU6-Cas12f/Cas9 sgRNA expression plasmids, a DNA fragment containing the human U6 (hU6) promoter and the PspCas12f1 sgRNA scaffold was synthesized by HuaGene (Shanghai, China). This fragment was inserted into the pSK vector backbone that had been digested with XhoI and ClaI (pSK, Addgene #62540) using the NEBuilder HiFi DNA Assembly Kit to produce the basic pSK-hU6-PspCas12f1 sgRNA backbone for sgRNA expression in eukaryotic cells. Corresponding sgRNA scaffold fragments for other Cas12f homologs and for SpCas9 were synthesized and integrated into the pSK-hU6 backbone in an analogous manner. Specifically, the pSK-hU6-PspCas12f1 sgRNA plasmid was linearized by PCR amplification with primers PSK-F/R to generate the acceptor backbone, and each synthesized sgRNA scaffold fragment was assembled into this backbone using the same assembly kit to yield the final pSK-hU6-Cas12f/Cas9 sgRNA constructs. All sgRNA scaffold sequences are provided in Table S1. Target-specific sgRNA inserts were subsequently introduced at the two BbsI sites of the PSK-hU6-sgRNA plasmid.

Phylogenetic analysis of Cas12f orthologs

Amino acid sequences of all previously reported Cas12f orthologs with confirmed editing activity in mammalian cells, along with 48 additional uncharacterized Cas12f variants, were collected for phylogenetic analysis. Multiple sequence alignment of full-length Cas12f

Please cite this article in press as: Liu et al., Rational and computation-assisted engineering of a compact and efficient CRISPR-Cas12f genome editor, Molecular Therapy (2026), https://doi.org/10.1016/j.ymthe.2026.08.043

proteins was performed using MUSCLE implemented in MEGA11. Phylogenetic reconstruction was performed with the IQ-TREE Web Server using the best-fit evolutionary model, and branch support was estimated from 1,000 bootstrap replicates. The resulting tree was subsequently visualized and annotated in iTOL. Table S1 provides the protein sequences of all Cas12f homologs included in the analysis.

Identification of Cas12f orthologs

To identify additional Cas12f orthologs for functional evaluation, we first selected 19 representative Cas12f nucleases spanning distinct phylogenetic branches based on the evolutionary classification reported previously.[16] To expand this collection, the amino acid sequences of AsCas12f1 and SpaCas12f1 were independently used as query sequences to search the NCBI non-redundant protein database using the SmartBLAST algorithm with default parameters. Candidate proteins were ranked according to their sequence similarity to the query sequence, and proteins with the highest sequence similarity were prioritized for further evaluation. To avoid redundant sampling, proteins differing from the query sequence by only one amino acid were excluded, and when multiple highly similar homologs were identified, only a single representative ortholog with the highest sequence similarity was retained for subsequent analysis. Based on these inclusion and exclusion criteria, three orthologs (Pt3Cas12f1, TcCas12f1, and AdCas12f1) were selected from the AsCas12f1 search, and PspCas12f1, the highest-ranking non-redundant homolog identified from the SpaCas12f1 search, was selected for further experimental characterization. Together with the 19 representative Cas12f nucleases, these four selected orthologs constituted the panel of 23 Cas12f candidates evaluated in this study.

Identification of Cas12f tracrRNA

Putative tracrRNA sequences of Cas12f orthologs were first predicted using the previously reported script Cas12f_tracrRNA_Finder.pl and used as references.[24] The approximate 3′ ends of tracrRNAs were inferred based on complementarity between the tracrRNA 3′ region and the crRNA sequence. To determine the 5′ boundaries, candidate tracrRNA fragments of different lengths were linked to crRNA sequences via a short linker and subjected to secondary structure prediction using RNAfold (http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi). The integrity and stability of the predicted secondary structures were then evaluated to define the full-length tracrRNA sequences for each Cas12f ortholog. The predicted tracrRNA sequences were subsequently co-expressed with their corresponding Cas12f proteins for experimental validation.

Activity assay of sgRNA variants

To evaluate the editing activity of different sgRNA scaffold variants, the pSK-hU6 promoter backbone fragment was first amplified from the pre-constructed pSK-hU6-PspCas12f1 sgRNA plasmid using the primer pair sgRNA-F/R. All designed sgRNA scaffold variant sequences were synthesized by HuaGene (Shanghai, China). Each sgRNA scaffold variant was assembled into the pSK-hU6 backbone using the NEBuilder HiFi DNA Assembly Master Mix (New England Biolabs) to generate plasmids for expressing individual sgRNA variants in eukaryotic cells. The sequences of all sgRNA variants are listed in Table S3. Target-specific spacer sequences were then cloned into the two BbsI restriction sites within the pSK-hU6-sgRNA variant plasmids to obtain the final sgRNA expression constructs.

Cell culture and transfection

All cell culture reagents were obtained from Gibco. HEK293T and C33A cells were maintained in Dulbecco's modified Eagle's medium (DMEM), SH-SY5Y cells were maintained in DMEM/F-12 medium, and A375 cells were maintained in RPMI 1640 medium. All culture media were supplemented with 10% fetal bovine serum (FBS; heat inactivated at 56°C for 30 min) and 1% penicillin-streptomycin. Cells were incubated at 37°C in a humidified atmosphere containing 5% CO_2. For GFP-reporter experiments, reporter-containing cells were plated in 10-cm dishes and transfected at 50%–60% confluence using polyethyleneimine (PEI), with 10 μg of Cas12f expression plasmid and 5 μg of sgRNA expression plasmid per dish. For endogenous editing assays, cells cultured in 24-well plates were transfected at 50%–60% confluence with 500 ng of Cas12f plasmid and 300 ng of sgRNA plasmid per well. Puromycin selection was initiated 24 h after transfection at 1 ng/μL and maintained for 48 h. After selection, cells were returned to drug-free complete medium for continued culture. For genome-wide specificity experiments, cells were seeded in 6-well plates and grown to 80%–90% confluence prior to counting. 2×10^5 cells were mixed with 500 ng Cas12f expression plasmid, 500 ng sgRNA expression plasmid, and 1 μL annealed dsODNs, then electroporated. Following electroporation, cells were transferred to complete DMEM and cultured for the indicated period.

GFP-activation assay

As described previously,[38] the GFP-activation reporter cassette was stably introduced into HEK293T cells. The reporter contained a 5-bp randomized sequence followed by a 24-bp target sequence positioned between the ATG start codon and the GFP coding region, thereby disrupting the GFP reading frame and suppressing fluorescence. Cas12f-induced cleavage followed by DNA repair can generate indels that re-establish the reading frame, resulting in GFP expression. The GFP-activation system was used as a sensitive and high-throughput tool to screen Cas12f variants for editing activity and to identify PAM preferences and specificity profiles.

PAM identification

PAM preferences were profiled using the GFP-activation reporter library. The library contained a 5-nt randomized region positioned between the GFP start codon and coding sequence, with the sequence context 5′-NNNNNGGATATGTTGAAGAACACCATGAC-3′. Reporter library cells were plated in 10-cm dishes and, upon reaching 50%–60% confluence, transfected with 10 μg of Cas12f plasmid and 5 μg of sgRNA plasmid using PEI. Five days after transfection, GFP-positive cells were sorted on an MA900 flow cytometer. Genomic DNA from the enriched GFP-positive population was extracted using the DNeasy Blood & Tissue Kit (QIAGEN). Target regions were PCR amplified, sequencing libraries were prepared, and

Please cite this article in press as: Liu et al., Rational and computation-assisted engineering of a compact and efficient CRISPR-Cas12f genome editor, Molecular Therapy (2026), https://doi.org/10.1016/j.ymthe.2026.08.043

samples were subjected to next-generation sequencing. To extract PAM sequences, 15-bp flanking anchor sequences 5′-TTGTTTGC CACCATG-3′ and 5′-GTGAGCAAGGGCGAG-3′ were used to fix the target region. The anchor indices ATG and GGATATGTTG AAGAA were used to locate the region of interest. The 5-bp randomized sequence between these anchors was extracted from sequencing reads and visualized using WebLogo[39] and PAM wheel plots[40] to determine the PAM recognized by Cas12f.

Endogenous genome-editing assay

For endogenous editing, HEK293T cells cultured in 24-well plates were transfected at 50%–60% confluence with 500 ng of Cas12f expression plasmid and 300 ng of sgRNA expression plasmid per well using PEI. Puromycin selection (1 ng/μL) was initiated 24 h later and continued for 48 h. Cells were subsequently maintained in drug-free complete medium. Genomic DNA was prepared with QuickExtract from cells collected 5–7 days after transfection. Amplicon libraries were generated by two sequential PCRs, with Illumina-compatible adapters incorporated during the second amplification. After purification with the QIAquick Gel Extraction Kit (QIAGEN), the products were subjected to deep sequencing. Target sequences are provided in Table S4.

Cas12f-specificity assay

To assess the specificity of PspCas12f1 and enPspCas12f1, HEK293T cells cultured in 24-well plates were transfected at 50%–60% confluence with 500 ng of Cas12f expression plasmid and 300 ng of a mismatch-containing sgRNA plasmid. Mutations were introduced at positions 1–20 of the sgRNA as single- or double-nucleotide substitutions. Puromycin was applied 24 h after transfection at 1 ng/μL for 48 h. Drug-free medium was restored after selection. Genomic DNA was collected with QuickExtract 5–7 days after transfection. Sequencing libraries were produced by two rounds of PCR, with Illumina adapters incorporated during the second round, and purified using the QIAquick Gel Extraction Kit (QIAGEN) before high-throughput sequencing.

Test of sgRNA and Cas12f mutant activity

The activities of engineered sgRNA variants, including truncated, substituted, and combinatorial forms, as well as Cas12f protein mutants carrying single, double, or triple amino acid substitutions, were evaluated in 24-well plates. When the cells reached 50%–60% confluence, transfections were performed as follows. For sgRNA variants, 500 ng of the WT Cas12f expression plasmid was co-transfected with 300 ng of the corresponding sgRNA variant plasmid. For Cas12f mutants, 500 ng of the Cas12f mutant plasmid was co-transfected with 300 ng of the sgRNA_V2.9 plasmid. All transfections used PEI. At 24 h post-transfection, puromycin was added at a final concentration of 1 ng/μL for 48 h, and cells were then maintained in drug-free medium. Cells were collected on days 5–7, and genomic DNA was prepared using QuickExtract. Sequencing libraries were constructed by two sequential PCRs with Illumina adapters incorporated in the second PCR and purified with the QIAquick Gel Extraction Kit (QIAGEN) prior to deep sequencing.

GUIDE-seq

GUIDE-seq experiments were performed using nine sgRNAs targeting nine endogenous sites across seven genes, *ANAPC15*, *RUNX1*, *CFTR*, *PPP1R12C*, *RNF2*, *EMX1*, and *DYRK1A*, to evaluate genome-wide off-target activity of PspCas12f1 and enPspCas12f1. HEK293T cells were seeded in 6-well plates, grown to 80%–90% confluence, then collected and counted. For each reaction, 2×10^5 cells were mixed with electroporation solution and protector solution at a ratio of 4.5 to 1. To this mixture, 500 ng Cas12f expression plasmid, 500 ng sgRNA expression plasmid, and 1 μL annealed dsODN were added. The mixture was transferred to an electroporation cuvette and electroporated using a Lonza electroporation device with the HEK-293 program. After electroporation, cells were immediately transferred to complete medium and cultured for 5 days. Genomic DNA was purified with the DNeasy Blood & Tissue Kit (QIAGEN). Target loci were amplified by nested PCR, after which the products were gel purified and subjected to paired-end sequencing on an Illumina X Plus platform. After unique molecular identifier (UMI)-based duplicate removal, an average of 446,294 UMI-collapsed GUIDE-seq reads was retained per library, corresponding to a mean of 649 UMI-supported reads per detected integration site. Raw paired-end reads were processed with the previously described GUIDE-seq computational pipeline.[32] Reads were first demultiplexed according to sample-specific dual-index barcodes, and libraries containing fewer than 10,000 reads were excluded from downstream analysis. UMIs were extracted from the index reads and used to collapse PCR duplicates by retaining, for each genomic integration site, the dominant UMI species (minimum base quality $\geq$ 15; minimum UMI frequency $\geq$ 0.9 within each duplicate cluster). UMI-collapsed reads were aligned to the human reference genome (hg38) using BWA with default paired-end alignment parameters. Candidate off-target cleavage sites were identified by clustering dsODN integration events and matching them to genomic loci with up to six mismatches relative to the intended protospacer-PAM sequence. To eliminate background dsODN integration events, candidate sites detected in the negative control (Cas12f expression plasmid and dsODN without sgRNA) were removed using bedtools. Only candidate sites passing these filtering criteria were retained for downstream off-target analysis.

AAV packaging and transduction

HEK293T cells were plated in 15-cm dishes and transfected with PEI at approximately 70% confluence. Each dish received 40 μg of total plasmid DNA, comprising 10 μg pAAV-enPspCas12f1-sgRNA_V3.6 expression plasmid, 10 μg pAAV-DJ packaging plasmid, and 20 μg pAAV-Helper plasmid. Twelve hours later, the medium was exchanged for DMEM containing 1% FBS and 1% penicillin-streptomycin. Viral supernatant was collected at 48 h and maintained at 4°C. At 96 h, a second supernatant collection was performed, and cells were pelleted by centrifugation at 1,000 rpm for 10 min at 4°C. The pellet was resuspended in 500 μL PBS and subjected to 3–4 freeze-thaw cycles, each consisting of 5 min in liquid nitrogen followed by 5 min at 37°C. After another centrifugation at 1,000 rpm for 10 min at 4°C, the resulting supernatant was pooled with the previously collected fractions. The combined suspension

Please cite this article in press as: Liu et al., Rational and computation-assisted engineering of a compact and efficient CRISPR-Cas12f genome editor, Molecular Therapy (2026), https://doi.org/10.1016/j.ymthe.2026.08.043

was passed through a 0.45-μm membrane, mixed with 5× PEG-8000, and incubated at 4°C overnight. Viral particles were pelleted the next day by centrifugation at 4,000 rpm for 2 h at 4°C, resuspended in PBS, and stored at −80°C. AAV genome titers were quantified by qPCR. SH-SY5Y, A375, and C33A cells were plated in 24-well plates and exposed to 25 μL of AAV per well at approximately 50% confluence, corresponding to an MOI of approximately 1×10^5. Cells were collected 7 days later. Genomic DNA was then extracted, and editing efficiencies were determined by targeted deep sequencing.

Statistics and reproducibility

Unless noted otherwise, experiments included at least three independent biological replicates; the exact replicate number for each experiment is specified in the corresponding figure legend. Data are reported as mean ± SD, and statistical analyses were carried out in GraphPad Prism 10. Student's *t* test was used for two-group comparisons, whereas one-way analysis of variance (ANOVA) was applied when more than two groups were compared. $p < 0.05$ was considered statistically significant, with significance levels denoted as $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, and $^{****}p < 0.0001$.

Protein engineering strategy design

Homology-based design (strategy 1)

PspCas12f1 and its homolog AsCas12f1 were aligned using CLUSTALX (v.1.83), and the complete pairwise sequence alignment is provided in Figure S23. Residue positions in PspCas12f1 were considered suitable for homology-based mutagenesis when the corresponding 11 amino acid sliding window, centered on the candidate residue and extending five residues in each direction, exhibited >50% sequence similarity between PspCas12f1 and AsCas12f1. This window size was selected to capture the local sequence context surrounding each residue while preserving sufficient evolutionary conservation for reliable homology-based comparison. An 11-residue window spans approximately three turns of an α helix (∼3.6 residues per turn) or a typical β strand segment, thereby encompassing the local structural environment most likely to influence residue identity and function. Candidate positions meeting the sequence-similarity criterion were subsequently evaluated with reference to previously reported mutational studies of AsCas12f1.[31] The corresponding AsCas12f1 substitutions were ranked according to their experimentally determined fold changes in genome-editing activity relative to WT AsCas12f1. Substitutions that enhanced AsCas12f1 activity were prioritized, with variants showing larger activity improvements receiving higher priority for introduction at the corresponding homologous positions in PspCas12f1. Based on the local sequence-similarity criterion and the reported activity rankings, 22 candidate substitutions were selected for experimental evaluation (Table S5).

PLM-based prediction (strategy 2)

The SaProtHub_v2 Web Toolkit was used to predict the functional effects of all possible single-point mutations in PspCas12f1 (Table S6). When multiple substitutions were predicted for the same residue, only the substitution with the highest predicted activity score was retained. The eight highest-scoring candidate mutations were subsequently selected for experimental evaluation.

Structure-informed rational design (strategies 3 and 4)

A structural model of the PspCas12f1 dimer bound to target DNA was generated using AlphaFold3 (https://alphafoldserver.com/welcome). Strategy 3 focused on the protein dimerization interface because Cas12f1 functions as a dimer. According to their local interfacial environment, candidate residues were substituted with amino acids that have the potential to enhance dimer formation and stability, including positively charged or potentially protonatable residues (K, R, and H), hydrophobic residues (L and M), an aromatic residue capable of both hydrophobic interactions and hydrogen bonding (Y), and an uncharged polar residue (Q). These substitutions were designed to probe different types of intermolecular interactions, including electrostatic interactions, hydrogen bonding, hydrophobic packing, and aromatic interactions, that may improve interfacial complementarity and dimer stability. Using this strategy, we generated 17 mutants for experimental evaluation.

For strategy 4, the DNA-binding interface was defined as residues located within 6 Å of the DNA phosphate backbone, with distances measured in PyMOL. In Rosetta-based protein design, residues within approximately 6–8 Å of a ligand are typically included in the design sphere for mutagenesis. Consistent with this principle, a recent engineering study of Cas12j-8 also used an 8 Å distance cutoff to define candidate residues for structure-guided optimization.[41] To prioritize residues most likely to directly influence protein-DNA interactions while maintaining a manageable experimental search space, we adopted a more stringent cutoff of 6 Å. A total of 26 residues were located within 6 Å of the DNA in the predicted structure, giving rise to 39 candidate mutations (Table S7). Among these, 16 candidate mutations with SaProt scores greater than 0 were retained as the initial candidate pool for strategy 4. These candidates were then subjected to additional filtering based on both predicted structural stability and structural analysis. Specifically, mutation-induced changes in folding free energy (ΔΔG) were calculated using FoldX (v.5.0), and only mutations with predicted ΔΔG values < 1.5 kcal/mol, indicating minimal effects on protein stability,[42] were retained. In addition, PyMOL was used to exclude residues that already formed hydrogen bonds with the DNA phosphate backbone in the WT structural model, thereby prioritizing mutations with the potential to establish new electrostatic interactions with DNA. Applying these complementary criteria reduced the candidate set from 16 to 9 mutations, which were subsequently selected for experimental evaluation. The complete list of candidate residues, their distances to the DNA, and their corresponding SaProt scores is provided in Table S7.

Statistical analysis of protein engineering experiments

For the protein engineering experiments, all engineered PspCas12f1 variants included in this analysis, together with WT PspCas12f1, were evaluated using the sgRNA_V2.9 scaffold at the *ANAPC15* target site under identical experimental conditions, except for the

subset of variants additionally tested at the *PPP1R12C* target site in Figure 3C, as indicated in the figure. Each variant was assessed using three independent biological replicates ($n = 3$). The efficiency fold change for each variant was calculated as the indel frequency of the corresponding mutant divided by the indel frequency of WT PspCas12f1 measured in the same experiment. All reported values represent the mean of three independent biological replicates. WT PspCas12f1 and all mutant variants were processed in parallel using identical transfection, cell culture, genomic DNA extraction, library preparation, and sequencing procedures to ensure direct comparability across all measurements.

All statistical analyses were performed using Python 3.12. A comprehensive retrospective analysis of the correlation between SaProt scores and experimental activity was conducted on all 107 missense mutants evaluated across different engineering strategies. Spearman correlation coefficients and associated *p* values were calculated using the spearmanr function from the *scipy.stats* module (v.1.13.1). Differences in editing efficiency between groups were assessed using independent-samples *t* tests. ROC curves, AUC, PR curves, and confusion matrices were generated using the *sklearn.metrics* module (v.1.5.2). Effect sizes were quantified using Cohen's d according to Equation 1:

$$\text{Cohen's d} = (M_2 - M_1) \Big/ \sqrt{(SD_1{}^2 + SD_2{}^2)/2}. \qquad \text{(Equation 1)}$$

Scatterplots and box plots were created using the *matplotlib* library (v.3.8.4). All efficiency fold changes are listed in Table S8.

Protein structural analysis

Structural visualization and figure preparation for all modeled structures were performed using the graphical interface of PyMOL. For the structural representations shown in Figures 3 and S16, the distances between the mutated residues and the DNA were displayed for both the WT and mutant structures. Mutation-induced DNA contacts and newly formed hydrogen bonds were annotated to facilitate structural interpretation of the observed changes in genome-editing activity.

DATA AND CODE AVAILABILITY

The data supporting the conclusions of this study are included in the article and its supplemental information. Raw datasets are provided in Table S9.

ACKNOWLEDGMENTS

This work was supported by grants from the National Key Research and Development Program of China (2021YFA0910602); the National Natural Science Foundation of China (82370254, 82070258, 82371568, and 82570728); the Science and Technology Research Program of Shanghai (24HC2810100 and 23ZR1426000); the Shenzhen Science and Technology Program (JCYJ20220818102818039); the Shenzhen Biomedical Industry Major Public Service Platform and Core Technology Research Special Support Plan (XMHT20220104048); the BYD Charity Foundation (PUSH-BYD JCYJ20220401 and PUSH-BYD JCYJ20220220); the New Quality Clinical Specialty Program of High-end Medical Disciplinary Construction in Shanghai Pudong New Area (grant no. 2025-PWXZ-03); the Medical Discipline Construction Program of Shanghai Pudong New Area Health Commission (the Emerging and Interdisciplinary Disciplines Program) (PWXx2025-07); and the Shanghai Pudong Hospital Talent Recruitment Research Startup Fund Funding Scheme (YJYJRC202308).

AUTHOR CONTRIBUTIONS

J.L., B.W., T.Y., and M.L. performed the experiments. J.L., S.W., and X.L. analyzed the data. Y.W. provided experimental guidance and wrote the manuscript. R.W., R.C., W.V.Z., and M.-C.K. guided the project.

DECLARATION OF INTERESTS

Y.W. and J.L. have submitted a patent application (2025115801320) on the basis of the results reported in this study.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.ymthe.2026.08.043.

REFERENCES

1. Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. Science *315*, 1709–1712. https://doi.org/10.1126/science.1138140.

2. Makarova, K.S., Wolf, Y.I., Iranzo, J., Shmakov, S.A., Alkhnbashi, O.S., Brouns, S.J.J., Charpentier, E., Cheng, D., Haft, D.H., Horvath, P., et al. (2020). Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. Nat. Rev. Microbiol. *18*, 67–83. https://doi.org/10.1038/s41579-019-0299-x.

3. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science *337*, 816–821. https://doi.org/10.1126/science.1225829.

4. Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., and Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819–823. https://doi.org/10.1126/science.1231143.

5. Nishimasu, H., Ran, F.A., Hsu, P.D., Konermann, S., Shehata, S.I., Dohmae, N., Ishitani, R., Zhang, F., and Nureki, O. (2014). Crystal structure of Cas9 in complex with guide RNA and target DNA. Cell *156*, 935–949. https://doi.org/10.1016/j.cell.2014.02.001.

6. Chatterjee, P., Lee, J., Nip, L., Koseki, S.R.T., Tysinger, E., Sontheimer, E.J., Jacobson, J.M., and Jakimo, N. (2020). A Cas9 with PAM recognition for adenine dinucleotides. Nat. Commun. *11*, 2474. https://doi.org/10.1038/s41467-020-16117-8.

7. Edraki, A., Mir, A., Ibraheim, R., Gainetdinov, I., Yoon, Y., Song, C.Q., Cao, Y., Gallant, J., Xue, W., Rivera-Pérez, J.A., and Sontheimer, E.J. (2019). A Compact, High-Accuracy Cas9 with a Dinucleotide PAM for In Vivo Genome Editing. Mol. Cell *73*, 714–726.e4. https://doi.org/10.1016/j.molcel.2018.12.003.

8. Hu, Z., Wang, S., Zhang, C., Gao, N., Li, M., Wang, D., Wang, D., Liu, D., Liu, H., Ong, S.G., et al. (2020). A compact Cas9 ortholog from Staphylococcus Auricularis (SauriCas9) expands the DNA targeting scope. Plos Biol. *18*, e3000686. https://doi.org/10.1371/journal.pbio.3000686.

9. Kim, E., Koo, T., Park, S.W., Kim, D., Kim, K., Cho, H.Y., Song, D.W., Lee, K.J., Jung, M.H., Kim, S., et al. (2017). In vivo genome editing with a small Cas9 orthologue derived from Campylobacter jejuni. Nat. Commun. *8*, 14500. https://doi.org/10.1038/ncomms14500.

10. Ran, F.A., Cong, L., Yan, W.X., Scott, D.A., Gootenberg, J.S., Kriz, A.J., Zetsche, B., Shalem, O., Wu, X., Makarova, K.S., et al. (2015). In vivo genome editing using Staphylococcus aureus Cas9. Nature *520*, 186–191. https://doi.org/10.1038/nature14299.

11. Wang, S., Mao, H., Hou, L., Hu, Z., Wang, Y., Qi, T., Tao, C., Yang, Y., Zhang, C., Li, M., et al. (2022). Compact SchCas9 Recognizes the Simple NNGR PAM. Adv. Sci. (Weinh) *9*, e2104789. https://doi.org/10.1002/advs.202104789.

12. Wang, S., Tao, C., Mao, H., Hou, L., Wang, Y., Qi, T., Yang, Y., Ong, S.G., Hu, S., Chai, R., and Wang, Y. (2022). Identification of SaCas9 orthologs containing a conserved serine residue that determines simple NNGG PAM recognition. Plos Biol. *20*, e3001897. https://doi.org/10.1371/journal.pbio.3001897.

13. Wei, J., Hou, L., Liu, J., Wang, Z., Gao, S., Qi, T., Gao, S., Sun, S., and Wang, Y. (2022). Closely related type II-C Cas9 orthologs recognize diverse PAMs. Elife *11*, e77825. https://doi.org/10.7554/eLife.77825.

14. Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., Slaymaker, I.M., Makarova, K.S., Essletzbichler, P., Volz, S.E., Joung, J., van der Oost, J., Regev, A., et al. (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell *163*, 759–771. https://doi.org/10.1016/j.cell.2015.09.038.

15. Zhang, Y., Wei, J., Wang, H., and Wang, Y. (2024). Characterization of NiCas12b for In Vivo Genome Editing. Adv. Sci. (Weinh) *11*, e2400469. https://doi.org/10.1002/advs.202400469.

16. Harrington, L.B., Burstein, D., Chen, J.S., Paez-Espino, D., Ma, E., Witte, I.P., Cofsky, J.C., Kyrpides, N.C., Banfield, J.F., and Doudna, J.A. (2018). Programmed DNA destruction by miniature CRISPR-Cas14 enzymes. Science *362*, 839–842. https://doi.org/10.1126/science.aav4294.

17. Karvelis, T., Bigelyte, G., Young, J.K., Hou, Z., Zedaveinyte, R., Budre, K., Paulraj, S., Djukanovic, V., Gasior, S., Silanskas, A., et al. (2020). PAM recognition by miniature CRISPR-Cas12f nucleases triggers programmable double-stranded DNA target cleavage. Nucleic Acids Res. *48*, 5016–5023. https://doi.org/10.1093/nar/gkaa208.

18. Xu, X., Chemparathy, A., Zeng, L., Kempton, H.R., Shang, S., Nakamura, M., and Qi, L.S. (2021). Engineered miniature CRISPR-Cas system for mammalian genome regulation and editing. Mol. Cell *81*, 4333–4345.e4. https://doi.org/10.1016/j.molcel.2021.08.008.

19. Kim, D.Y., Lee, J.M., Moon, S.B., Chin, H.J., Park, S., Lim, Y., Kim, D., Koo, T., Ko, J.H., and Kim, Y.S. (2022). Efficient CRISPR editing with a hypercompact Cas12f1 and engineered guide RNAs delivered by adeno-associated virus. Nat. Biotechnol. *40*, 94–102. https://doi.org/10.1038/s41587-021-01009-z.

20. Park, S.J., Ju, S., Jung, W.J., Jeong, T.Y., Yoon, D.E., Lee, J.H., Yang, J., Lee, H., Choi, J., Kim, H.S., and Kim, K. (2025). Robust genome editing activity and the applications of enhanced miniature CRISPR-Cas12f1. Nat. Commun. *16*, 677. https://doi.org/10.1038/s41467-025-56048-w.

21. Wu, T., Liu, C., Zou, S., Lyu, R., Yang, B., Yan, H., Zhao, M., and Tang, W. (2023). An engineered hypercompact CRISPR-Cas12f system with boosted gene-editing activity. Nat. Chem. Biol. *19*, 1384–1393. https://doi.org/10.1038/s41589-023-01380-9.

22. Wang, Y., Wang, Y., Pan, D., Yu, H., Zhang, Y., Chen, W., Li, F., Wu, Z., and Ji, Q. (2022). Guide RNA engineering enables efficient CRISPR editing with a miniature Syntrophomonas palmitatica Cas12f1 nuclease. Cell Rep. *40*, 111418. https://doi.org/10.1016/j.celrep.2022.111418.

23. Su, M., Li, F., Wang, Y., Gao, Y., Lan, W., Shao, Z., Zhu, C., Tang, N., Gan, J., Wu, Z., and Ji, Q. (2024). Molecular basis and engineering of miniature Cas12f with C-rich PAM specificity. Nat. Chem. Biol. *20*, 180–189. https://doi.org/10.1038/s41589-023-01420-4.

24. Kong, X., Zhang, H., Li, G., Wang, Z., Kong, X., Wang, L., Xue, M., Zhang, W., Wang, Y., Lin, J., et al. (2023). Engineered CRISPR-OsCas12f1 and RhCas12f1 with robust activities and expanded target range for genome editing. Nat. Commun. *14*, 2046. https://doi.org/10.1038/s41467-023-37829-7.

25. Eggers, A.R., Chen, K., Soczek, K.M., Tuck, O.T., Doherty, E.E., Xu, B., Trinidad, M.I., Thornton, B.W., Yoon, P.H., and Doudna, J.A. (2024). Rapid DNA unwinding accelerates genome editing by engineered CRISPR-Cas9. Cell *187*, 3249–3261.e14. https://doi.org/10.1016/j.cell.2024.04.031.

26. Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., Smetanin, N., Verkuil, R., Kabeli, O., Shmueli, Y., et al. (2023). Evolutionary-scale prediction of atomic-level protein structure with a language model. Science *379*, 1123–1130. https://doi.org/10.1126/science.ade2574.

27. Su, J., Han, C., Zhou, Y., Shan, J., Zhou, X., and Yuan, F. (2023). SaProt: Protein language modeling with structure-aware vocabulary. Preprint at bioRxiv. https://doi.org/10.1101/2023.10.01.560349.

28. Jiang, K., Yan, Z., Di Bernardo, M., Sgrizzi, S.R., Villiger, L., Kayabolen, A., Kim, B.J., Carscadden, J.K., Hiraizumi, M., Nishimasu, H., et al. (2025). Rapid in silico directed evolution by a protein language model with EVOLVEpro. Science *387*, eadr6006. https://doi.org/10.1126/science.adr6006.

29. Wei, D., Cheng, P., Song, Z., Liu, Y., Xu, X., Huang, X., Wang, X., Zhang, Y., Shu, W., and Wei, Y. (2025). AI-guided Cas9 engineering provides an effective strategy to enhance base editing. Mol. Syst. Biol. *21*, 1563–1580. https://doi.org/10.1038/s44320-025-00142-0.

30. Fei, H., Li, Y., Liu, Y., Wei, J., Chen, A., and Gao, C. (2025). Advancing protein evolution with inverse folding models integrating structural and evolutionary constraints. Cell *188*, 4674–4692.e19. https://doi.org/10.1016/j.cell.2025.06.014.

31. Hino, T., Omura, S.N., Nakagawa, R., Togashi, T., Takeda, S.N., Hiramoto, T., Tasaka, S., Hirano, H., Tokuyama, T., Uosaki, H., et al. (2023). An AsCas12f-based compact genome-editing tool derived by deep mutational scanning and structural analysis. Cell *186*, 4920–4935.e23. https://doi.org/10.1016/j.cell.2023.08.031.

32. Tsai, S.Q., Zheng, Z., Nguyen, N.T., Liebers, M., Topkar, V.V., Thapar, V., Wyvekens, N., Khayter, C., Iafrate, A.J., Le, L.P., et al. (2015). GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. Nat. Biotechnol. *33*, 187–197. https://doi.org/10.1038/nbt.3117.

33. Zhang, S., Qiang, R., Zhang, Y., Wan, J., Tao, C., Dong, Y., Tang, X., Xu, L., Xiao, H., Lin, Y., et al. (2026). Cidea Targeting Protects Cochlear Hair Cells and Hearing Function From Drug- and Noise-Induced Damage. Adv. Sci. (Weinh) *13*, e17206. https://doi.org/10.1002/advs.202517206.

34. Xiao, L., Liu, J., Chen, Y., Xia, L., Jiang, Y., Chen, X., Zhou, T., Sun, Y., Lu, W., Wang, H., et al. (2025). Supporting cells orchestrate noise-induced hearing loss via a Gasdermin D-dependent signaling loop with hair cells. Nat. Commun. *16*, 11181. https://doi.org/10.1038/s41467-025-66152-6.

35. Wu, Z., Zhang, Y., Yu, H., Pan, D., Wang, Y., Wang, Y., Li, F., Liu, C., Nan, H., Chen, W., and Ji, Q. (2021). Programmed genome editing by a miniature CRISPR-Cas12f nuclease. Nat. Chem. Biol. *17*, 1132–1138. https://doi.org/10.1038/s41589-021-00868-6.

36. Sharrar, A., Arake de Tacca, L., Collingwood, T., Meacham, Z., Rabuka, D., Staples-Ager, J., and Schelle, M. (2023). Discovery and Characterization of Novel Type V Cas12f Nucleases with Diverse Protospacer Adjacent Motif Preferences. CRISPR J. *6*, 350–358. https://doi.org/10.1089/crispr.2023.0006.

37. Anders, C., Niewoehner, O., Duerst, A., and Jinek, M. (2014). Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. Nature *513*, 569–573. https://doi.org/10.1038/nature13579.

38. Hu, Z., Zhang, C., Wang, D., Gao, S., Ong, S.G., Wang, Y., and Zheng, W.V. (2021). A Highly Sensitive GFP Activation Assay for Detection of DNA Cleavage in Cells. Front. Cell Dev. Biol. *9*, 771248. https://doi.org/10.3389/fcell.2021.771248.

39. Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004). WebLogo: a sequence logo generator. Genome Res. *14*, 1188–1190. https://doi.org/10.1101/gr.849004.

40. Leenay, R.T., Maksimchuk, K.R., Slotkowski, R.A., Agrawal, R.N., Gomaa, A.A., Briner, A.E., Barrangou, R., and Beisel, C.L. (2016). Identifying and Visualizing Functional PAM Diversity across CRISPR-Cas Systems. Mol. Cell *62*, 137–147. https://doi.org/10.1016/j.molcel.2016.02.031.

41. Meng, R., Li, J., Wang, W., Liang, D., Li, Z., Mao, C., Li, Q., Zhang, Y., Chen, H., Tang, J., et al. (2025). Engineered Cas12j-8 is a Versatile Platform for Multiplexed Genome Modulation in Mammalian Cells. Adv. Sci. (Weinh) *12*, e02593. https://doi.org/10.1002/advs.202502593.

42. Gerasimavicius, L., Liu, X., and Marsh, J.A. (2020). Identification of pathogenic missense mutations using protein stability predictors. Sci. Rep. *10*, 15387. https://doi.org/10.1038/s41598-020-72404-w.