

# A Potent CRISPR-Cas12I Double-Strand Break Gene Editor

Tomas Urbaitis,<sup>1,†</sup> Laima Trinkuniene,<sup>1,2,†</sup> Ieva Lenkaite,<sup>1</sup> Monika Petrauskite,<sup>1</sup> Renatas Krasauskas,<sup>1</sup> Migle Stitilyte,<sup>1,2</sup> Modestas Sabaliauskas,<sup>1</sup> Giedrius Sasnauskas,<sup>2</sup> Giedre Tamulaitiene,<sup>2</sup> Joshua K. Young,<sup>3</sup> Virginijus Siksnys,<sup>2</sup> and Giedrius Gasiunas<sup>1,\*</sup>

## Abstract

Recently, a new family of CRISPR-Cas12 endonucleases from an unexplored phylum of bacteria, *Armatimonadota*, was discovered. Named Cas12I, they are compact (800–900 aa), recognize a 5' C-rich protospacer adjacent motif, and present an N-terminal domain that stretches from the beginning to the end of the ribonucleoprotein-bound DNA target site, effectively locking it in place. Here, structure-guided rational design supplemented with AI-based large protein language model predictions was used to improve rates of DNA target cleavage of a family member, Asp2Cas12I. Compared to the wild-type, engineered variants exhibited an approximately 10-fold increase in double-strand break (DSB) editing efficiency in human cells with less target-to-target variation. Moreover, frequencies of editing were comparable to those of SpCas9 at overlapping target sites, and their DSBs efficiently corrected by homology-directed repair (39–56% of editing outcomes). Altogether, this study extends our understanding of CRISPR-Cas12 protein engineering and offers a potent new alternative for DSB-mediated genome editing in human cells.

## Introduction

Over the past decade, Types II and V clustered regularly interspaced short palindromic repeats (CRISPR)-associated (Cas) 9 and Cas12 endonucleases, respectively, have emerged as flexible and efficient genome editing tools.<sup>1,2</sup> Guided by a CRISPR RNA with ~20 bp complementary to one strand of a double-stranded (ds) DNA target adjacent to a short (1–4 bp) protospacer adjacent motif (PAM),<sup>3–5</sup> they offer unprecedented scalability and affordability for genome editing applications. As a result, they have been widely adopted to introduce site-specific genetic edits to further the understanding of gene function, generate new sequence variation, insert desirable traits, and correct disease-related alleles.<sup>6–14</sup>

Emerging innovations have sought to perfect their use and have attempted to remove any bottleneck that may limit their application. One constraint has been the length and composition of the PAM, as it restricts the genomic sequence space that can be targeted. This is further

compounded by the selection of unique targets to minimize the risk of off- or secondary target cleavage.<sup>2,15</sup> To address this, CRISPR-Cas endonucleases with different PAM recognition and improved targeting specificity have been identified or engineered.<sup>16–22</sup> Research to improve the intracellular delivery of CRISPR-Cas editing tools has exposed additional obstacles. For viral delivery, more compact editors with shorter coding sequences that fit within viral genomes have been coveted<sup>23,24</sup> and have helped propel the discovery of smaller CRISPR-Cas and transposon-related RNA-guided nucleases.<sup>25–27</sup> For methods that deliver editing reagents as short-lived transient doses of mRNA or ribonucleoprotein (RNP), the amount of RNA-guided reagent, particularly those from type V systems, needed to achieve efficient editing or the potency of the enzyme has been an underexplored area, necessitating further optimization.<sup>28–30</sup>

To help address these limitations, we engineered a Cas12I endonuclease from a bacteria species that is yet

<sup>1</sup>Caszyme, Vilnius, Lithuania; <sup>2</sup>Institute of Biotechnology, Vilnius University, Vilnius, Lithuania; and <sup>3</sup>Genome Center of Excellence, Corteva Agriscience™, Johnston, Iowa, USA.

<sup>†</sup>These authors contributed equally to this work.

\*Address correspondence to: Giedrius Gasiunas, Caszyme, Mokslininku st. 2, Vilnius LT-08412, Lithuania, E-mail: giedrius.gasiunas@caszyme.com



to be defined—*Armatimonadota* species (Asp)—as a potent DNA double-strand break (DSB) editing tool. AspCas12l proteins are compact (800–900 aa) and recognize a 5′ CCY PAM, providing a key counterbalance to the 5′ T-rich PAM recognition afforded by most other Cas12 nucleases.<sup>31,32</sup> To improve DSB editing efficiency, the ternary RNP-DNA target structure of a family member, Asp2Cas12l, was first solved to 2.51 Å using cryo-electron microscopy (cryo-EM). Guided by the structure, arginine amino acid substitutions were next introduced to fine-tune Asp2Cas12l electrostatic binding potential at DNA and sgRNA interfaces. DSB activity, measured as the frequency of insertion and/or deletion (indel) mutations resulting from cellular nonhomologous end-joining (NHEJ) repair, was improved by approximately 7-fold over the wild-type nuclease. Next, residues suggested by artificial intelligence (AI) algorithms to improve function were introduced and evaluated. In combination with structure-guided modifications, AI predictions bolstered Asp2Cas12l DSB activity and reduced target-to-target variation. When delivered as plasmid DNA, the best variant, M82, performed comparably to SpCas9 at overlapping sites and maintained DSB activity when delivered transiently as mRNA or RNP. Satisfied with its DSB potency, its utility for homology directed repair (HDR) applications was assessed in HEK293T cells where M82 boasted efficiencies of HDR-mediated editing ranging from 39% to 56% depending on target site and edit type. Finally, the target cleavage specificity of M82 was evaluated using TEG-seq and found to be similar to other Cas12 endonucleases.

## Materials and Methods

### Plasmid construction

Mammalian expression vector pCDNA3.1 containing a human codon-optimized Asp2Cas12l gene was synthesized to order by GenScript (Piscataway, NJ, USA). The expression construct harbored sequence encoding a SV40 nuclear localization signal (NLS) at the N-terminus and sequence encoding a nucleoplasmin NLS at the C-terminus of the Asp2Cas12l protein.

For the Asp2Cas12l mutant expression vector construction, DNA fragments encoding respective mutation were synthesized (Twist Biosciences) and cloned into the pCDNA3.1 Asp2Cas12l expression vector by restriction cloning.

To express single-guide RNAs (sgRNAs) that direct Asp2Cas12l to target sites, plasmids were synthesized that encode a U6 polymerase III promoter sequence, Asp2Cas12l sgRNA sequence, HDV ribozyme sequence, and a U6 terminator. At the 3′ end of each sgRNA

sequence, a variable spacer sequence was included. Plasmids encoding different spacer sequences of DNA were constructed by restriction cloning.

To produce mRNA *in vitro* transcription templates, Asp2Cas12l sequences were cloned to a plasmid vector optimized for mRNA synthesis. This optimized plasmid vector contains 5′ and 3′ human hemoglobin beta subunit (HBB) UTR sequences, T7 RNA polymerase promoter, and kanamycin resistance gene. To clone target genes into this optimized vector, Asp2Cas12l gene sequences were amplified by PCR, during which overlapping sequences from the destination vector were added; vector was also amplified by PCR, during which overlapping sequences of the target genes were added to the 5′ and 3′ ends of a linear PCR product.

### mRNA production

Plasmids encoding Cas12l mRNA templates were used as PCR templates to produce *in vitro* transcription templates for mRNA synthesis. During this PCR step, target sequences were amplified using universal primers. Reverse primer is elongated and used for 120 nt poly(A) tail addition. After PCR, products were confirmed by gel electrophoresis and purified from PCR reactions using New England Biolabs Monarch PCR and DNA Cleanup Kit (5 µg); concentrations were measured using spectrophotometry (NanoPhotometer<sup>®</sup> NP80, IMPLEN). Approximately 1 µg of each purified PCR product was then used as a template for *in vitro* transcription reaction performed using HiScribe<sup>®</sup> T7 mRNA Kit with CleanCap<sup>®</sup> Reagent AG (New England Biolabs); kit-provided UTP was not used; instead, N1-Methylspermidine (TriLink Biotechnologies or Thermo Fisher Scientific) was added to a final concentration of 5 mM. *In vitro* transcription reaction was run for 1 h at 37°C after which DNaseI and nuclease-free water were added to the reaction and incubated for 30 min at 37°C to remove template DNA. After this step, the IVT reaction product was purified using the Monarch RNA Cleanup Kit (500 µg). Parts of the purified samples were set aside to assess their quality. Sample concentrations were measured using spectrophotometry (NanoPhotometer<sup>®</sup> NP80, IMPLEN) and fluorimetry (Qubit RNA Broad Range Assay Kit and Qubit 4, Thermo Fisher Scientific). mRNA integrity was assessed by performing capillary electrophoresis using 4150 TapeStation System together with RNA Screen Tape reagents (Agilent). Synthesized mRNA was aliquoted and kept at −80°C until use.

### Cell culture

HEK293T and ARPE-19 cells were cultured in Dulbecco's modified Eagle medium (DMEM) with GlutaMAX

(Thermo Fisher Scientific), supplemented with 10% fetal bovine serum (Thermo Fisher Scientific) and 1% Penicillin-Streptomycin (10,000 U/mL) (Thermo Fisher Scientific) at 37°C in 5% CO<sub>2</sub>. All cells were seeded into 96-well plates (Thermo Fisher Scientific) 1 day before transfection at a density of  $2 \times 10^5$ /mL, 90  $\mu$ L per well for HEK293T cells and of  $1 \times 10^5$ /mL, 100  $\mu$ L per well for ARPE-19 cells.

### Cell transfection

Cells were transfected using transfection reagent FuGENE<sup>®</sup> HD (Promega Corporation, Madison, WI, USA) according to the manufacturer's protocol. For each sample, 300 of DNA containing 43 fmol of plasmid encoding Asp2Cas12l and 43 fmol of plasmid encoding guide RNA was used. A 3:1 ratio of FuGENE<sup>®</sup> HD Transfection Reagent to DNA was used (0.3  $\mu$ L reagent:100 ng DNA per well). Cells were incubated at 37°C with 5% CO<sub>2</sub> for 4 days posttransfection before cell lysis and further genome editing evaluation.

For mRNA transfection, HEK293T cells were seeded into 96-well plates (Thermo Fisher Scientific) 1 day prior to transfection at a density of 18,000 cells per well. Cells were transfected using Lipofectamine MessengerMAX Transfection Reagent (Invitrogen) following the manufacturer's recommended protocol. Unless stated otherwise, for each well of a 96-well plate, a total amount of 100 ng RNA was used. This total amount consists of mRNA and sgRNA in a molar ratio of 1:4.

### RNP electroporation (HEK293T)

Two days before the electroporation, HEK293T cells were split and maintained as described above. On the day of the experiment, cells were washed using Dulbecco's phosphate saline buffer (DPBS) (Thermo Fisher Scientific) and trypsinized using TrypLE Express Enzyme (1X) (Thermo Fisher Scientific). Collected cells were resuspended in DPBS and diluted in SF Cell Line Nucleofector Solution (Lonza) supplemented with Supplement 1 (Lonza). RNPs were assembled in Nucleofector solution supplemented with Supplement 1 using 126 pmol of each purified nuclease and 160 pmol of sgRNA. RNP assembly reactions were incubated at room temperature for 20 min, then kept at 4°C until nucleofection. In total,  $2 \times 10^5$  cells per sample were added into the RNP assembly reaction, and the reaction was transferred to 16-well Nucleocuvette Strips (Lonza). Cells were electroporated using preprogrammed protocol CM-130 using 4D-Nucleofector (Lonza). After electroporation, contents of each well were seeded into wells containing

prewarmed DMEM with GlutaMAX (Thermo Fisher Scientific), supplemented with 10% fetal bovine serum (Thermo Fisher Scientific) in a 96-well plate (Thermo Fisher Scientific). Cells were then grown for 48 h at 37°C in 5% CO<sub>2</sub> before cell lysis and further genome editing evaluation.

### Asp2Cas12l-mediated HDR

HEK293T cells were electroporated with Asp2Cas12l-sgRNA or SpCas9 RNP complexes targeting the AAVS1 locus along with a dsDNA HDR template encoding the EGFP gene. 126 pmol of each purified Asp2Cas12l nuclease variant was mixed with 160 pmol of synthetic sgRNA. In the case of SpCas9, 50 pmol of EnGen<sup>®</sup> Sp Cas9 NLS (New England BioLabs) with 100 pmol sgRNA along with 500 ng of donor DNA. The mixtures were incubated for 15–20 min at room temperature. RNP electroporation performed as described earlier. The efficiency of EGFP knock-in was determined by MACS-Quant<sup>®</sup> Analyzer 16 Flow Cytometer (Miltenyi Biotec); flow cytometry data were analyzed with FlowJo<sup>™</sup> Software v10.10.0 (BD Biosciences). Single-nucleotide substitution knock-in efficiency was quantified by next-generation sequencing (NGS).

For the single-stranded oligo donor DNA-mediated HDR experiments, HEK293T cells were electroporated with Asp2Cas12l M82 nuclease-sgRNA RNP complex, targeting the AAVS1 or CD151 genes, along with 100 pmol of ssODN template encoding a 1 or 2 nt substitution, a 51 nt left homology arm, and 50 nt right homology arm. RNP electroporation performed as described earlier.

### Evaluation of genome editing

Cells were incubated at 37°C with 5% CO<sub>2</sub> for 4 days posttransfection before genomic DNA extraction. The cells were washed twice with 200  $\mu$ L prewarmed to 37°C 1X DPBS (Thermo Fisher Scientific) and resuspended in 25  $\mu$ L cell lysis solution containing 50 mM Tris-HCl, 150 mM NaCl, 0.05% Tween 20, pH 7.6 (Sigma Aldrich), and 0.2 mg/mL Proteinase K (New England Biolabs). Resuspended cells were incubated at 55°C for 1 h and 95°C for 15 min.

DNA fragments in the genomic DNA extracted by cell lysis surrounding each Cas12l target site were amplified by PCR. PCR reactions were performed using Q5 Hot Start High-Fidelity 2X Master Mix (New England Biolabs) according to the manufacturer's instructions. The reactions were set up using 1  $\mu$ L of the cell lysate and 0.5  $\mu$ M of each primer in a final reaction volume of 25  $\mu$ L.

For genome editing efficiency assessment by T7 Endonuclease I assay, 20  $\mu$ L of each PCR reaction was

combined with 3  $\mu\text{L}$  NEBuffer 2 (New England Biolabs) and 7  $\mu\text{L}$  of water before denaturation at 95°C for 5 min and re-annealing by temperature ramping from 95°C to 85°C at  $-2^\circ\text{C}/\text{s}$  followed by ramping from 85°C to 25°C at  $-0.1^\circ\text{C}/\text{s}$ . In total, 1  $\mu\text{L}$  of T7 Endonuclease I (New England Biolabs) was added to each reannealed sample, and cleavage reactions were incubated at 37°C for 20 min. Fragments were analyzed by performing gel electrophoresis using E-Gel Precast Agarose Gel Electrophoresis system (Invitrogen). Overall, 8  $\mu\text{L}$  of each sample was mixed with 7  $\mu\text{L}$  of E-Gel Sample Loading Buffer (Invitrogen), and then the whole volume was loaded to the well. Cleaved DNA fraction was evaluated by densitometric analysis using ImageJ software.

### illumina NGS

Cell lysates of transfected HEK293T cells were used to amplify genomic target regions, and fragments were extended by Illumina sequencing that included a unique index for each sample through two rounds of PCR. A total of 4  $\mu\text{L}$  of each lysate sample was used as a template in the primary PCR reaction. For the primary PCR custom primers were used that were complementary to the sequences surrounding the genomic targets and had noncomplementary “tails” with Illumina adapter sequences. Q5 HotStart 2x MasterMix (New England Biolabs) was used for the primary PCR, and the reaction was set up using 4  $\mu\text{L}$  of cell lysate as template and 0.2  $\mu\text{M}$  of each primer in a final volume of 25  $\mu\text{L}$ . The cycling conditions used were 98°C for 2 min 30 s, 24 cycles of 98°C for 30 s, 56.5°C for 30 s, 72°C for 25 s, and final extension at 72°C for 2 min. The primary PCR product was purified using SPRIselect (Beckman Coulter) magnetic beads and used for the secondary PCR.

Secondary PCR was performed using 2X Q5 High-Fidelity Hot Start PCR Master Mix (New England Biolabs) with custom primers (synthesis ordered from Integrated DNA Technologies) containing Illumina sequences and i7 and i5 indexes (on reverse and forward primers, respectively). Cycling conditions for secondary PCR were as follows: 98°C for 2 min, 10 cycles at 98°C for 10 s, 60°C for 30 s and 72°C for 2 min, followed by final extension at 72°C for 2 min, when using 2X Q5 High-Fidelity Hot Start PCR Master Mix (New England Biolabs) and custom primers, or 98°C for 30 s, 10 cycles at 98°C for 10 s, 65°C for 1 min 15 s, followed by final extension at 65°C for 5 min. The secondary PCR products were purified using SPRIselect (Beckman Coulter) magnetic beads, and their quantity and quality were checked using spectrophotometry (NanoPhotometer<sup>®</sup> NP80, IMPLLEN) and fluorimetry (Qubit 1X dsDNA HS Assay, Qubit 1X dsDNA

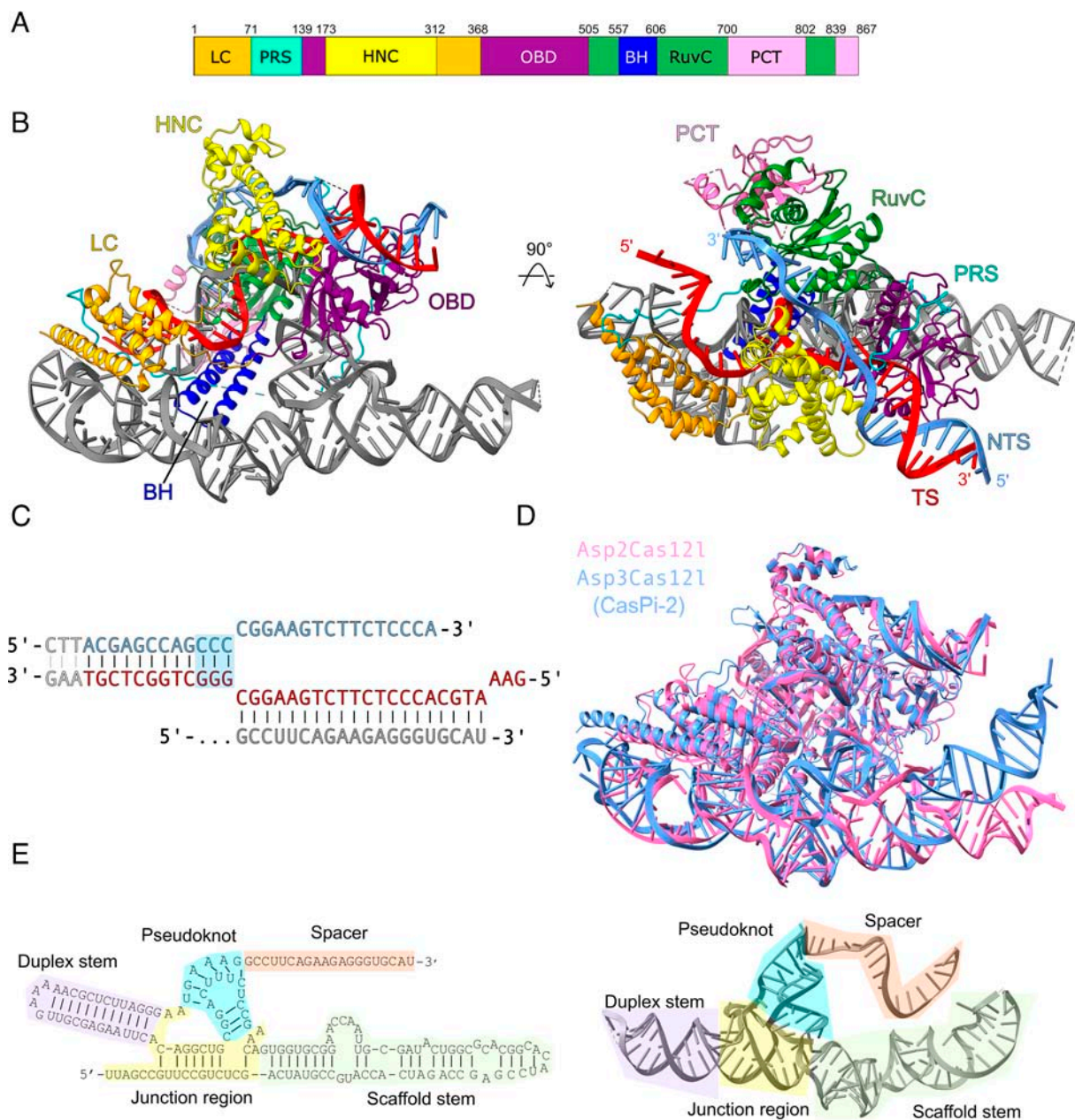
BR Assay, and Qubit 4, Thermo Fisher Scientific). The purified samples were pooled in an equimolar ratio and size selection performed using SPRIselect (Beckman Coulter) magnetic beads. The resulting library quantified using NEBNext Library Quant Kit for Illumina (New England Biolabs). Final library pool was prepared for deep sequencing according to Illumina’s specifications. Paired-end sequencing was performed using the MiSeq Reagent Kit v2 (300 cycles) (Illumina) on the MiSeq System (Illumina) with 7% PhiX v.3 (Illumina). All sequencing data analysis was done using Geneious Prime 2025.0. Reads were trimmed and filtered using BBDuk and mapped to the reference sequence with a 50 nt quantification window set around the target site. Reads that had differences from the reference sequence in this window were included in genome editing efficiency calculations.

## Results

### Cryo-EM structure of Asp2Cas12l-sgRNA-dsDNA complex

The cryo-EM structure of wild-type Asp2Cas12l in complex with sgRNA and its DNA substrate was determined at a resolution of 2.51 Å (see Supplementary Data S1). To maximize R-loop formation, a 16 nt long nontarget strand without complementarity to the target strand was utilized (Fig. 1A–C, Supplementary Fig. S1, Supplementary Table S1). The structure revealed extensive similarity to a related nuclease, Asp3Cas12l, which was published during the preparation of this study (labeled CasPi-2 in the referred publication) (Fig. 1D).<sup>31,32</sup> At the sequence level, these proteins shared 43.6% amino acid identity, however, the ternary complexes superimposed with an root-mean-square deviation (RMSD) distance of 1.022 Å over 600 pruned residue pairs in ChimeraX.<sup>33</sup> Protein structure comparison *via* the Dali server confirmed the high structural similarity with a Z-score of 38.3.<sup>34</sup> In general, functional domains of both proteins were arranged similarly. Minor differences included several flexible and thus unresolved stretches of amino acids in the Asp2Cas12l structure, a slightly different path of the proline-rich string (PRS) as it encircled the ternary complex, and a segment containing two antiparallel beta-strands in the helical and NTSB chimera (HNC) domain of Asp3Cas12l, which was not present in Asp2Cas12l (Fig. 1D). Therefore, we adopted the same naming convention for Asp2Cas12l as reported earlier (Fig. 1A–B).<sup>32</sup>

The sgRNA of Asp2Cas12l and Asp3Cas12l also formed similar architecture motifs with a difference observed in the positioning of the crRNA repeat and tracrRNA antirepeat duplex relative to the rest of the complex (Fig. 1D, Supplementary Fig. S2a). Defined



**FIG. 1.** The structure of the Asp2Cas12l-sgRNA-DNA ternary complex. **(A)** Domain organization of the Asp2Cas12l nuclease; LC—lock-catch domain, PRS—proline-rich string, HNC—helical and nontarget strand binding chimera domain, OBD—oligonucleotide binding domain, BH—bridge helix, RuvC nuclease domain, and PCT—C-terminal domain. Amino acid position boundaries of respective domain elements designated above the schematic. **(B)** Overall structure of the Asp2Cas12l-sgRNA-DNA ternary complex; protein domains colored as in **(A)**, DNA substrate nontarget strand in blue, target strand in red, sgRNA in gray; disordered regions indicated in dashed lines. **(C)** Schematic of the DNA substrate and sgRNA spacer; **(D)** Asp2Cas12l structure (blue) superimposed on CasPi-2 structure (pdb: 7YOJ) (pink). **(E)** Secondary structure schematic (left) and ternary structure (right) of the sgRNA with structural elements designated and highlighted.

by a tetrapartite structure, the sgRNA for Asp2Cas12l comprised a junction region that linked the other RNA folds, a large scaffold stem that interfaces with the HC domain, a duplex stem rising from the crRNA repeat and tracrRNA antirepeat interactions, and an RNA pseudoknot structure that orients the spacer into the DNA binding pocket for R-loop formation (Fig. 1E, Supplementary Fig. S2b).

### Rational engineering of CRISPR-Cas12l

DSB-stimulated indel editing with Asp2Cas12l and its relatives was previously shown to be less efficient and more target site dependent compared to SpCas9 in human cells.<sup>32</sup> To ameliorate this, the Asp2Cas12l-sgRNA-DNA structure was first analyzed for amino acids whose atoms were less than 0.4 Å away from the negatively charged phosphodiester backbone of the DNA target. Each position was then substituted with a positively charged arginine residue (if not already arginine) to increase its electrostatic attraction to the target site. To improve RNP complex formation, positions shown to be near the sgRNA were similarly modified. In all, 34 positions were altered (Fig. 2A).

Next, the indel editing efficiency of each of the 34 variants was evaluated. For this, plasmid DNA encoding each variant and a sgRNA targeting the *RunX1* (T1) and *WTAP* (T6) sites were co-delivered into HEK293T cells. Using T7 Endonuclease I to rapidly read out the target site mutation frequency, 13 of the 34 arginine substitutions enhanced indel rates for at least one of the tested target loci (Fig. 2B). These were next evaluated for additive effects by stacking substitutions located in different Asp2Cas12l domains in sets of two, three, or four (Fig. 2C). Exceptions included variants that combined S17R/V21R, S297R/T301R, or E762R/E766R (Fig. 2C). Based on the magnitude of improvement at the *RunX1* (T1) target, S17R/V21R/D142R/L253R, S297R/T301R/

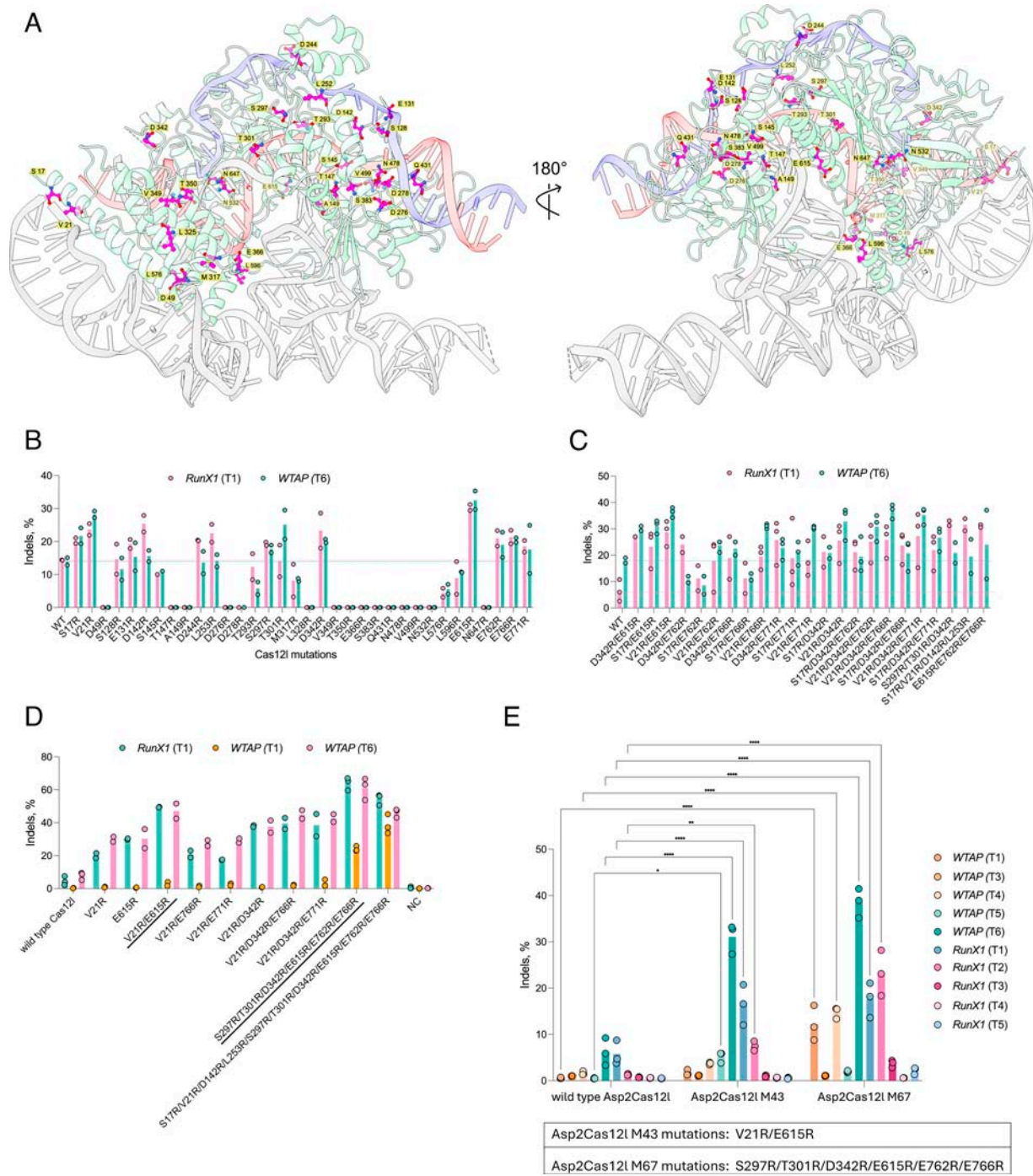
D342R, and E615R/E762R/E766R were selected for a third round of stacking (Fig. 2C). For this, S297R/T301R/D342R and E615R/E762R/E766R were combined as well as one variant that united all three (Fig. 2D). When compared with the wild-type and other best-performing variants from the first and second rounds of engineering using Ampli-Seq, both enzymes outperformed all others (Fig. 2D). Between the two, S297R/T301R/D342R/E615R/E762R/E766R exhibited the highest rates of editing at the *RunX1* (T1) and *WTAP* (T6) targets, while the one with 10 arginine substitutions had an advantage at the *WTAP* (T1) site (Fig. 2D). Given this, S297R/T301R/D342R/E615R/E762R/E766R (M67) was selected and evaluated on a larger collection of target sequences in *WTAP* and *RunX1* loci in HEK293T cells (Fig. 2E). The potent improvement afforded by just two alterations, V21R/E615R (M43), at the *RunX1* (T1) and *WTAP* (T6) targets also prompted its testing at additional sites. As averaged across ten targets, M43 and M67 produced a 3.5- and 7.1-fold improvement in indel editing, respectively, compared to wild-type Asp2Cas12l (Fig. 2E).

Close inspection of the amino acid changes in M67 revealed that three of the six arginine substitutions, S297R, T301R, and D342R, resided close to the DNA backbone of the target strand and were hypothesized to enhance target DNA engagement (Supplementary Fig. S3a,b). E615R was situated near the sgRNA pseudoknot, potentially stabilizing it (Supplementary Fig. S3c). Lastly, E762 and E766 were found in a disordered stretch of the C-terminal domain (Supplementary Fig. 3d) and posited to help orient the RuvC domain for optimal DNA target cleavage.

### Mutagenesis via large language model predictions

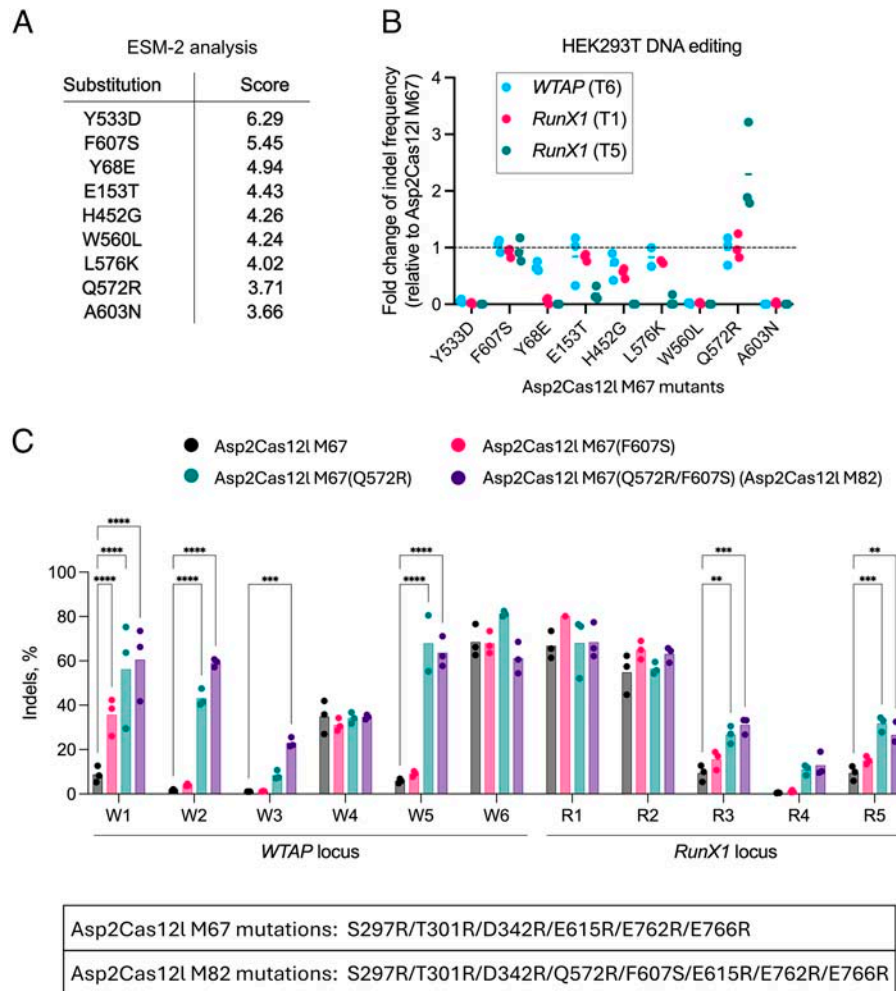
While Asp2Cas12l M67 enhanced rates of targeted indels at some sites, several remained recalcitrant to editing (Fig. 2E). To improve further, M67 was subjected to

**FIG. 2.** Asp2Cas12l mutagenesis and HEK293T DNA editing. **(A)** positions of amino acids selected for mutagenesis (pink) in the Asp2Cas12l-sgRNA-DNA ternary structure. Asp2Cas12l protein colored in green, sgRNA—gray, DNA target strand—red, DNA nontarget strand—blue **(B)** comparison of indel frequencies of wild-type and Asp2Cas12l single amino acid mutants in HEK293T cells ( $n = 2$ ); dashed lines denote mean wild-type Asp2Cas12l indel frequency. **(C)** comparison of indel frequencies of wild-type and Asp2Cas12l variants containing combinations of beneficial mutations in HEK293T cells ( $n = 2$ ); dashed lines denote mean wild-type Asp2Cas12l indel frequency. **(D)** comparison of indel frequencies of wild-type and Asp2Cas12l variants in HEK293T cells determined by Illumina NGS ( $n = 2$ ). **(E)** comparison of indel frequencies of wild-type and select Asp2Cas12l variants on 10 genomic targets in HEK293T cells determined by Illumina NGS, bars represent the mean indel frequency from  $n = 3$  independent experiments. Two-way repeated measures analysis of variance (ANOVA) with Dunnett's test was used to assess statistical significance of Asp2Cas12l mutant variant results compared to wild-type Asp2Cas12l control. \*\*\*\* $p < 0.0001$ ; \*\* $p < 0.01$ ; \* $p < 0.05$ . In **(B)–(E)**, Asp2Cas12l nuclease variants and sgRNAs were delivered to HEK293T cells as DNA plasmids.



*in silico* mutagenesis, and the fitness of the resulting variants was predicted using the 15B parameter ESM-2 protein large language model paired with zero-shot learning.<sup>35,36</sup> For this, variant fitness was measured *via* a masked marginal scoring function, which outputs the log odds ratio of the mutated position, reflecting how likely that amino acid can be naturally found in a similar

sequence context.<sup>35</sup> Assuming that evolution tends to preserve function, a higher score increases the probability that the mutation will have a positive effect.<sup>37</sup> Based on this, the nine highest scoring zero-shot predictions were selected, individually introduced into M67, and tested for improvements in HEK293T cells (Fig. 3A). Q572R exhibited around a 2-fold higher mean indel frequency



**FIG. 3.** LLM-based Asp2Cas12l M67 mutagenesis. **(A)** Amino acid substitutions predicted by the ESM-2 protein language model, ranked by the attributed score. **(B)** Fold change of indel frequencies of Asp2Cas12l M67 mutants containing LLM-derived substitutions compared to Asp2Cas12l M67 in HEK293T cells, as determined by T7 Endonuclease I assay ( $n = 3$ ). **(C)** comparison of indel frequencies of Asp2Cas12l M67 and variants containing Q572R and/or F607S substitutions in HEK293T cells; bars represent the mean indel frequency from  $n = 3$  independent experiments. Two-way repeated measures ANOVA with Dunnett's test was used to assess significance of Asp2Cas12l mutant variant results compared to Asp2Cas12l M67 control. \*\*\*\* $p < 0.0001$ ; \*\*\* $p < 0.001$ ; \*\* $p < 0.01$ . In **(B)** and **(C)** Asp2Cas12l variants and sgRNAs were delivered to HEK293T cells as DNA plasmids.

compared to the parental M67 variant at the *RunX1* (T5) target and retained activity at *WTAP* (T6) and *RunX1* (T1) sites (Fig. 3B). Since F607S maintained similar editing efficiencies on all targets, it and Q572R were selected for testing across a wider collection of targets (Fig. 3C). The combination of Q572R and F607S retained similar indel efficiencies on target sites previously shown to be efficiently (>20%) edited (W4, W6, R1, R2) and increased indel rates at those that were not (W1, W2, W3, R3, R4, R5) (Fig. 3C). Q572R had a stronger impact than F607S, but in tandem and when averaged across targets, they outperformed Q572R

alone, resulting in the Asp2Cas12l M82 variant containing the combined mutations S297R/T301R/D342R/Q572R/F607S/E615R/E762R/E766R (Fig. 3C). Positioned in the bridge-helix domain and close to the phosphodiester backbone of the sgRNA spacer and PRS, Q572R was reasoned to stabilize sgRNA-mediated DNA target recognition similar to that posited for S297R, T301R, and D342R and/or promote PRS wrapping (Supplementary Fig. S4a). Located in the RuvC domain and pointing into the sgRNA pseudoknot, F607S was thought to modulate sgRNA binding (Supplementary Fig. S4b).

To further validate potency, the indel editing efficiency of Asp2Cas12l M67 and M82 was compared when delivered transiently as either RNP or mRNA into HEK293T cells. Here, indel rates were evaluated at *WTAP* (T6), as well as three previously untested target sites, either *CD34*, *CD151*, and *AAVS1* or *CD34*, *CD151*, and *VEGFA* for RNP and mRNA delivery experiments, respectively (Supplementary Fig. S5a, b). When nucleofected as RNP, M82 outperformed M67 and exhibited high rates of targeted indels (74–99%) at all four target sites (Supplementary Fig. S5a). Co-lipofection of M67 and M82 mRNA and sgRNA produced a similar outcome with M82, resulting in higher indel rates than M67 (Supplementary Fig. S5b).

### Benchmarking Asp2Cas12l M82 for DSB-mediated genome editing applications

To gauge the effectiveness of Asp2Cas12l M82 to introduce targeted indels, it was compared with SpCas9 at overlapping targets. For this, seven sites sharing the same PAM region (3' RGG for SpCas9 and 5' CCY for Asp2Cas12l) with each enzyme's sgRNA base pairing on opposite strands of the same target were selected. When delivered as plasmid DNA into HEK293T cells, three of the seven targets showed comparable indel rates between M82 and SpCas9 (Fig. 4A). Two of the remaining sites, *WTAP* T6 and *VEGFA* T10, were edited more efficiently with M82, while *WTAP* T4 and *VEGFA* T9 showed higher indel rates with SpCas9 (Fig. 4A). Altogether, the average rate of indel editing observed with Asp2Cas12l M82 (67.4%) was nearly identical to SpCas9 (64.8%) (Fig. 4A).

To support therapeutic applicability, the *VEGFA* gene associated with age-related macular degradation<sup>37,38</sup> in the human retinal cell line ARPE-19 was edited. In this case, mRNA encoding Asp2Cas12l M82 or SpCas9 was electroporated into cells, alongside the respective sgRNA. In ARPE-19 cells, Asp2Cas12l M82 exhibited similar

editing efficiencies to SpCas9, with up to 40% indel editing as measured by T7 Endonuclease I (Fig. 4B).

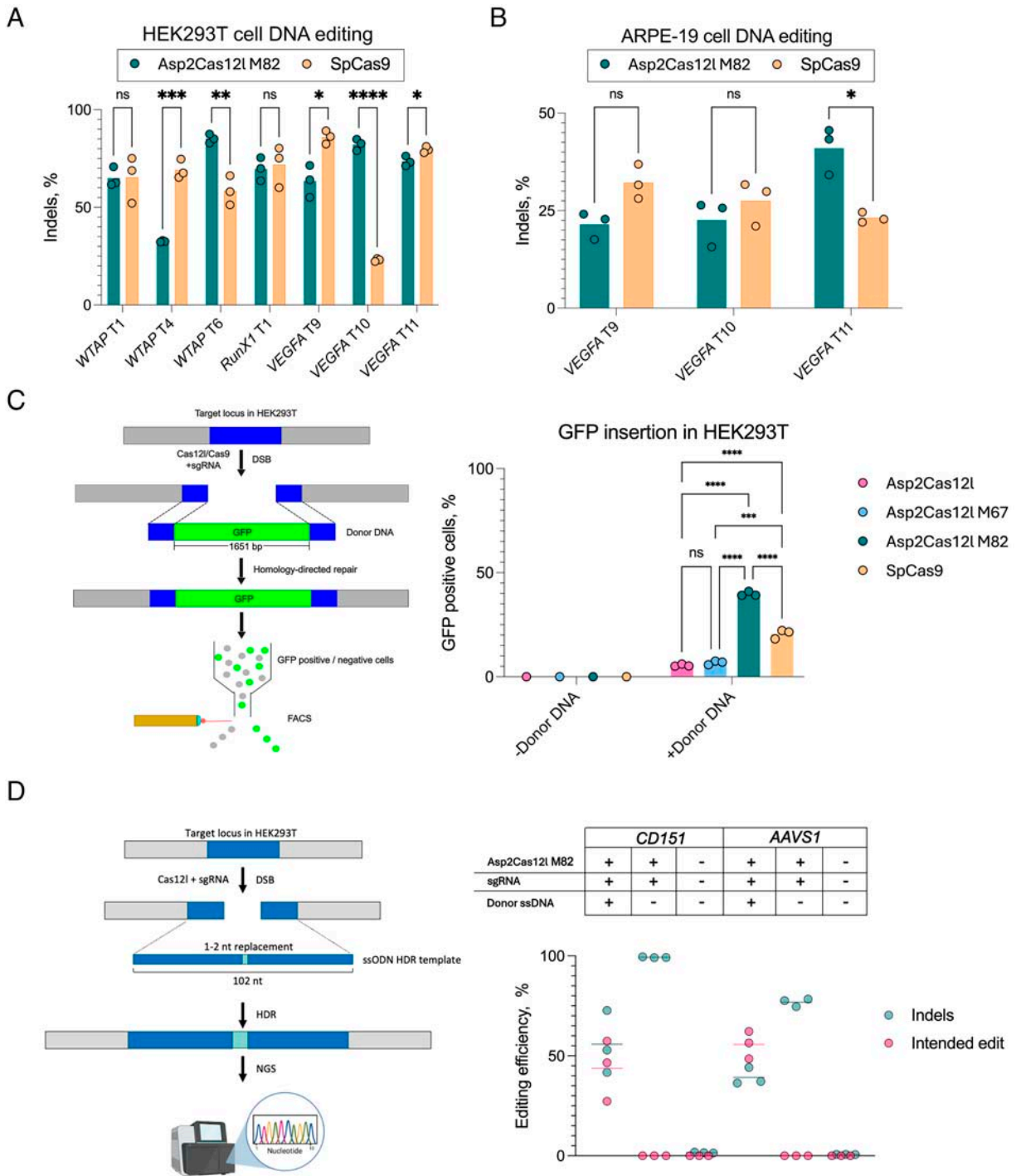
Next, the utility of Asp2Cas12l for HDR was evaluated. Initially, the safe harbor *AAVS1* locus in HEK293T cells was targeted for the site-specific insertion of a GFP expression cassette at overlapping SpCas9 and Asp2Cas12l sites (Fig. 4C). The exogenously supplied dsDNA repair template consisted of two 500 bp *AAVS1* homologous sequences flanking the DSB site for SpCas9 with the GFP gene nestled in between (Fig. 4C). To induce the DSB, cells were electroporated with either Asp2Cas12l or SpCas9 RNP and donor DNA. After 5 days, fluorescence-activated cell sorting (FACS) was performed to measure the frequency of HDR (Supplementary Fig. S6). Asp2Cas12l M82 facilitated the highest mean percentage of GFP-positive cells (39.2%) and thus site-specific gene insertion compared to wild-type Asp2Cas12l (6.08%), M67 (6.99%), and SpCas9 (22.2%) (Fig. 4C).

To confirm these observations, the efficiency of HDR was examined at two additional target sites for Asp2Cas12l M82. In these experiments, single-stranded (ss) DNA donor templates with 50 bp of homologous sequence flanking both sides of the target cut site containing 1–2 nt replacement(s) were used (Fig. 4D). HEK293T cells were electroporated with Asp2Cas12l M82 RNPs and donor DNA, and HDR efficiency was quantified with Ampli-Seq. Depending on the target, 44–56% of sequence reads showed the seamless incorporation of 1–2 nt polymorphism(s) while indel editing outcomes ranged from 39% to 56% (Fig. 4D).

### *In vitro* characterization of AspCas12l engineered variants

The biochemical attributes of Asp2Cas12l M43, M67, and M82 dsDNA target cleavage were next characterized *in vitro* (see Supplementary Data S1). Initially, this

**FIG. 4.** Asp2Cas12l M82 DNA editing validation and benchmarking. **(A)** Indel frequencies of Asp2Cas12l M82 and SpCas9 delivered as plasmid DNA in HEK293T cells. Multiple unpaired *t* tests were performed to assess significance between Asp2Cas12l M82 and SpCas9 results. \*\*\*\**p* < 0.0001; \*\*\**p* < 0.001; \*\**p* < 0.01; \**p* < 0.05; ns—nonsignificant, *p* > 0.05 (*n* = 3). **(B)** Indel frequencies of Asp2Cas12l M82 and SpCas9 delivered as mRNA in ARPE-19 cells. Multiple unpaired *t* tests were performed to assess significance between Asp2Cas12l M82 and SpCas9 results. Here \* denotes *p* < 0.05; ns—nonsignificant, *p* > 0.05 (*n* = 3). **(C)** DNA insertion into HEK293T *AAVS1* locus *via* HDR using Asp2Cas12l variants; left—experimental procedure schematic, right—fraction of GFP-positive cells determined *via* FACS, bars represent the mean percentage of GFP-positive cells from *n* = 3 independent experiments. Tukey's multiple comparison test was performed to assess significance between Asp2Cas12l variant and SpCas9 results. \*\*\*\**p* < 0.0001; \*\*\**p* < 0.001; ns—nonsignificant, *p* > 0.05 **(D)** DNA insertion *via* HDR using Asp2Cas12l M82 and ssDNA oligo donor DNA template; left—experimental procedure schematic, right—fraction of HEK293T DNA containing the intended edit encoded in the donor DNA template (pink) and stochastic indels (green), *n* = 3 independent experiments.



included the rate of linear dsDNA target cleavage at 37°C in the buffers established earlier for the wild-type enzyme (Supplementary Fig. S7). Surprisingly, the best DSB editor, Asp2Cas12l M82, did not exhibit the highest cleavage activity *in vitro* (Supplementary Fig. S7). To explore further, the effect of temperature, ionic concentration, and

pH on linear dsDNA target cleavage were evaluated (Supplementary Figs. S8, S9). In doing so, it was found that the wild-type enzyme exhibited the broadest thermal profile (Supplementary Fig. S8), the engineered variants showed the greatest tolerance to increasing ionic concentrations (NaCl) (Supplementary Fig. S9a), and all variants

functioned at pHs ranging from 5.5 to 8.0 (Supplementary Fig. S9b). Since the intracellular ionic composition of a HEK293 cell has been estimated to be  $\sim 150$  mM,<sup>39</sup> it was presumed to be a key factor behind the increased activity of the engineered variants.

#### Asp2Cas12l dsDNA target cleavage specificity

Next, dsDNA target cleavage specificity of Asp2Cas12l M82 was examined. Unbiased and with a sensitivity level equivalent to 0.0002–1% mutant reads by Ampli-Seq, TEG-seq was used to assess Asp2Cas12l M82 specificity (see Supplementary Data S1).<sup>40</sup> Briefly, HEK293T cells were electroporated with the corresponding Asp2Cas12l M82 RNP along with a DNA oligo duplex tag that was destined to be incorporated into intended and secondary target site DSBs. Next, genomic DNA was extracted, processed to enrich fragments containing the integrated tag *via* PCR, and sequenced using Ion Torrent chemistry. A custom analysis pipeline was then used to map sequence reads to the reference genome. Overall, four HEK293T sites were targeted by Asp2Cas12l M82 nuclease, and the potential for off or secondary target cleavage was evaluated (Fig. 5A, Table 1). Sequencing read depth, associated with putative secondary targets, varied between 0.06% and 3.7% across 48–149 loci, depending on the site (Table 1). Arising from low sequence read counts relative to the on-target, most of these were near or below the lower limit of TEG-Seq detection (Fig. 5A). As a control, TEG-seq was also performed with SpCas9 at a previously characterized target prone to secondary site cleavage in human cells, *AAVS1* (Fig. 5B, Table 1).<sup>41</sup> In line with previous results, a much larger proportion of sequencing reads could be attributed to SpCas9 secondary versus on-target cleavage, 62.31% and 37.69% of reads mapped to secondary sites and the on-target across 90 loci, respectively (Fig. 5B, Table 1). Secondary targets with highest read counts also matched sites that were more heavily edited in the earlier study (Fig. 5).<sup>41</sup>

To confirm these findings with an orthogonal approach, Cas-OFFinder was used to identify 29 putative secondary sites containing between three and nine mismatches for four on-targets, three in the *CD5* locus and one in the *VEGFA* gene. *CD5*-targeting sgRNAs were individually co-transfected along with a mRNA encoding M82 into HEK293T cells. The *VEGFA*-targeting sgRNA and M82 were delivered similarly into ARPE-19 cells. Ampli-Seq was then used to measure indel rates at the respective on-sites and secondary targets. While on-targets showed detectable frequencies of indels, none of the secondary sites accumulated indels at frequencies above the negative

control (DNA amplicons from untransfected cells) (Supplementary Fig. S10). Taken together, this suggests that the risk of Asp2Cas12l secondary target cleavage is low.

#### Discussion

Here, structure-guided rational engineering, zero-shot predictions from AI-based large language models trained on protein folding and stability, and the screening of multiple target sites in each round of engineering were applied to improve the DSB editing of Asp2Cas12l. In doing so, an improved variant, M82, was developed without laborious and expensive alternatives, like protein-directed evolution.<sup>42</sup> It contained eight amino acid substitutions compared to the wild-type parent, six from the structure-based rational engineering and two from AI-based suggestions. In combination, the alterations resulted in consistent and specific indel editing that rivaled those produced by SpCas9, independent of delivery method and human cell line used, with low target-to-target variability (Figs. 2E, 3C, 4A, B, 5; Supplementary Fig. S5a, b).

Close examination of the amino acid changes underlying Asp2Cas12l M82 were also suggestive of methods that may be used to streamline the improvement of other Cas12l orthologs or even Cas proteins. First, the observation that arginine substitutions enhanced DSB editing efficiency indicates that native Cas12l proteins may be deficient in sgRNA and DNA binding and that approaches aimed at improving these aspects may generally improve editing activity. Second, when M82 alterations were mapped (either through multiple sequence alignment of AlphaFold predicted structures) onto related Cas12l proteins, they were strongly conserved and therefore might represent evolutionary hotspots for the improvement of orthologs (Supplementary Fig. S11). Indeed, some Cas12l family members already exhibited positionally equivalent amino acid substitutions to those found in M82 (e.g., S297R, Q572R, E615R, E766R) (Supplementary Fig. S11). Finally, the masked marginal scoring from the 15B parameter ESM-2 protein large language model predictions lacked strong correlation with DSB editing efficiency in human cells (Fig. 3A, B). Since the model in essence predicts amino acid sequence plausibility based on orthologous sequence patterns rather than actual function, the small and relatively divergent evolutionary branch of known Cas12l nuclease family members may present a challenge for zero-shot predictions, and methods that remove this limitation may help future engineering campaigns.

In our study, it was observed that Asp2Cas12l M82 DSBs were also efficiently corrected by HDR (Fig. 4D, E; Supplementary Fig. S6). While this warrants further



**Table 1. Summary of Asp2Cas12l M82 secondary target activity evaluation via TEG-seq**

Nuclease	Target gene	Target sequence	Integration event	Sequencing reads	Read%	Note
Asp2Cas12l M82	<i>CD5</i>	AGTTGTCCGTAGCAGATGATTGAG	On-target	254764	99.94	
			Secondary target	154	0.06	Across 50 loci
	<i>AAVSI</i>	ACAGTGGGGCCACTAGGGACAGGA	On-target	144069	99.52	
			Secondary target	695	0.48	Across 54 loci
	<i>VEGFA</i>	CCTGGCCTCAGTTCCCTGGCAACA	On-target	175192	96.3	
			Secondary target	6475	3.70	Across 149 loci
<i>WTAP</i>	GTGGGAGAGTCTACACTTTCATAC	On-target	514837	98.78		
		Secondary target	6278	1.22	Across 48 loci	
SpCas9	<i>AAVSI</i>	GGGGCCACTAGGGACAGGAT	On-target	189286	37.69	
			Secondary target	312934	62.31	Across 90 loci

seems to occur naturally as part of their PAM-distal target cleavage mechanism, while for Cas9 one or more gRNA(s) need to be delivered to direct cleavage of the most prevalent NHEJ repair outcome(s).<sup>5,25,44–47</sup> In addition, the 5–9 nt 5' staggered cut produced by Asp2Cas12l target cleavage<sup>31</sup> may also enhance HDR, although other studies using paired SpCas9 nickases suggest that its contribution may be minimal.<sup>48,49</sup> Furthermore, the unique framework by which Cas12l enzymes encircle their dsDNA targets *via* a pair of interlocking helical bundles and a PRS of charged (e.g., arginine and lysine) residues<sup>32</sup> may also bias repair toward HDR, and additional studies are needed to understand its impact (if any) and underlying mechanism(s).

### Conclusions

Asp2Cas12l M82 presents itself as a potent new option for DSB-based genome editing applications. As highlighted herein, it can be delivered in a variety of different molecular formats, potentially enabling its introduction as RNP or mRNA either directly to cells or through engineered lipid nanoparticles (LNPs) or virus-like particles.<sup>50–52</sup> Depending on application, this has several benefits that include further reducing risk for secondary target cleavage, decreasing cellular toxicity, and alleviating potential regulatory concerns.<sup>53–56</sup> Moreover, it is compact in size (867 aa) making it compatible with many viral delivery systems, including adeno-associated viruses (AAVs).<sup>57,58</sup> Finally, its targets are efficiently corrected by HDR, offering a fresh perspective on traditional DSB-mediated genome editing.

### Acknowledgments

The authors would like to thank David Papp and Mudra Hegde (Thermo Fisher Scientific) for the assistance in setting up the TEG-seq experiments and performing the data analysis. Also, they thank Ieva Narkeviciute and Greta Busmaite (Vilnius University Hospital Santaros Klinikos) for performing the cell sorting procedures.

### Authors' Contributions

T.U., J.K.Y., and G.G. conceived the study. T.U., L.T., I.L., M.P., R.K., M.S., and M.S. performed the experiments. T.U., L.T., I.L., M.P., R.K., M.S., M.S., and G.G. analyzed the data. G.S. and G.T. determined and refined the cryo-EM structure. T.U. and G.G. wrote the article with contributions and guidance from J.K.Y.

### Author Disclosure Statement

T.U., J.K.Y., and G.G. have filed patent applications related to the article. T.U., L.T., I.L., M.P., R.K., and G.G. are employees of Caszyme. M. Stitilyte and M. Sabaliauskas were employees of Caszyme at the time of the study. Current affiliation of M. Stitilyte—Gensinta, Vilnius, Lithuania; M. Sabaliauskas—Thermo Fisher Scientific Baltics, Vilnius, Lithuania. J.K.Y. is an employee of Corteva Agriscience. V.S. is the chairman of Caszyme. V.S. and G.G. have a financial interest in Caszyme. The remaining authors declare that they have no conflicts of interest.

### Funding Information

Research was conducted as part of the execution of Project “Mission-driven Implementation of Science and Innovation Programmes” (No. 02-002-P-0001) “Safer and Smaller gene Editing tools for CURing genetic diseases of the Eye *in vivo* (SECURE),” funded by the Economic Revitalization and Resilience Enhancement Plan “New Generation Lithuania” and European Union “NextGenerationEU” initiative.

### Supplementary Material

Supplementary Figure S1  
 Supplementary Figure S2  
 Supplementary Figure S3  
 Supplementary Figure S4  
 Supplementary Figure S5  
 Supplementary Figure S6  
 Supplementary Figure S7  
 Supplementary Figure S8  
 Supplementary Figure S9

Supplementary Figure S10  
 Supplementary Figure S11  
 Supplementary Data S1  
 Supplementary Data S2  
 Supplementary Table S1

## References

- Barrangou R, Doudna JA. Applications of CRISPR technologies in research and beyond. *Nat Biotechnol* 2016;34(9):933–941.
- Anzalone AV, Koblan LW, Liu DR. Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. *Nat Biotechnol* 2020;38(7):824–844.
- Jinek M, Chylinski K, Fonfara I, et al. A Programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *Science* 2012; 337(6096):816–821.
- Gasiunas G, Barrangou R, Horvath P, et al. Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci USA* 2012;109(39); doi: 10.1073/pnas.1208507109
- Zetsche B, Gootenberg JS, Abudayyeh OO, et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* 2015;163(3): 759–771.
- Gurumurthy CB, Grati M, Ohtsuka M, et al. CRISPR: A versatile tool for both forward and reverse genetics research. *Hum Genet* 2016;135(9): 971–976.
- Shi J, Gao H, Wang H, et al. ARGOS 8 variants generated by CRISPR–Cas9 improve maize grain yield under field drought stress conditions. *Plant Biotechnol J* 2017;15(2):207–216.
- Chen J, Wang H, Bai J, et al. Generation of pigs resistant to highly pathogenic porcine reproductive and respiratory syndrome virus through gene editing of CD163. *Int J Biol Sci* 2019;15(2):481–492.
- Gao H, Mutti J, Young JK, et al. Complex trait loci in maize enabled by CRISPR-Cas9 mediated gene insertion. *Front Plant Sci* 2020;11: 535.
- Wang L, Yang Y, Breton C, et al. A mutation-independent CRISPR-Cas9–mediated gene targeting approach to treat a murine model of ornithine transcarbamylase deficiency. *Sci Adv* 2020;6(7):eaax5701.
- Frangoul H, Altshuler D, Cappellini MD, et al. CRISPR-Cas9 gene editing for sickle cell disease and  $\beta$ -thalassemia. *N Engl J Med* 2021;384(3): 252–260.
- Liu L, Gallagher J, Arevalo ED, et al. Enhancing grain-yield-related traits by CRISPR–Cas9 promoter editing of maize CLE genes. *Nat Plants* 2021; 7(3):287–294.
- Lorenzo CD, Debray K, Herwegh D, et al. BREEDIT: A multiplex genome editing strategy to improve complex quantitative traits in maize. *Plant Cell* 2023;35(1):218–238.
- Musunuru K, Grandinette SA, Wang X, et al. Patient-specific in vivo gene editing to treat a rare genetic disease. *N Engl J Med* 2025;392(22): 2235–2243.
- Kim D, Luk K, Wolfe SA, et al. Evaluating and enhancing target specificity of gene-editing nucleases and deaminases. *Annu Rev Biochem* 2019;88: 191–220.
- Gasiunas G, Young JK, Karvelis T, et al. A catalogue of biochemically diverse CRISPR-Cas9 orthologs. *Nat Commun* 2020;11(1):5512.
- Walton RT, Christie KA, Whittaker MN, et al. Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science* 2020;368(6488):290–296.
- Silverstein RA, Kim N, Kroell A-S, et al. Custom CRISPR–Cas9 PAM variants via scalable engineering and machine learning. *Nature* 2025; 643(8071):539–550.
- Nayfach S, Bhatnagar A, Novichkov A, et al. Customizing CRISPR–Cas PAM specificity with protein language models. *Nat Biotechnol* 2026; doi: 10.1038/s41587-025-02995-0
- Kleinstiver BP, Tsai SQ, Prew MS, et al. Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in human cells. *Nat Biotechnol* 2016;34(8): 869–874.
- Kleinstiver BP, Pattanayak V, Prew MS, et al. High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* 2016;529(7587):490–495.
- Slymaker IM, Gao L, Zetsche B, et al. Rationally engineered Cas9 nucleases with improved specificity. *Science* 2016;351(6268):84–88.
- Lino CA, Harper JC, Carney JP, et al. Delivering CRISPR: A review of the challenges and approaches. *Drug Deliv* 2018;25(1): 1234–1257.
- Wang D, Zhang F, Gao G. CRISPR-based therapeutic genome editing: Strategies and in vivo delivery by AAV vectors. *Cell* 2020;181(1): 136–150.
- Karvelis T, Bigelyte G, Young JK, et al. PAM recognition by miniature CRISPR–Cas12f nucleases triggers programmable double-stranded DNA target cleavage. *Nucleic Acids Res* 2020;48(9): 5016–5023.
- Altae-Tran H, Kannan S, Demircioglu FE, et al. The widespread IS200/IS605 transposon family encodes diverse programmable RNA-guided endonucleases. *Science* 2021;374(6563):57–65.
- Saito M, Xu P, Faure G, et al. Fanzor is a eukaryotic programmable RNA-guided endonuclease. *Nature* 2023;620(7974):660–668.
- Liu Z, Li Z, Li B. Nonviral delivery of CRISPR/Cas systems in mRNA format. *Adv NanoBiomed Res* 2022;2(11):2200082.
- Santos L, Mention K, Cavusoglu-Doran K, et al. Comparison of Cas9 and Cas12a CRISPR editing methods to correct the W1282X-CFTR mutation. *J Cyst Fibros* 2022;21(1):181–187.
- Oikemus S, Hu K, Shin M, et al. Identifying optimal conditions for precise knock-in of exogenous DNA into the zebrafish genome. *Development* 2025;152(12):dev204571.
- Urbaitis T, Gasiunas G, Young JK, et al. A new family of CRISPR-type V nucleases with C-rich PAM recognition. *EMBO Rep* 2022;23(12): e55481.
- Sun A, Li C-P, Chen Z, et al. The compact Cas $\pi$  (Cas12I) ‘bracket’ provides a unique structural platform for DNA manipulation. *Cell Res* 2023; 33(3):229–244.
- Meng EC, Goddard TD, Pettersen EF, et al. UCSF ChimeraX: Tools for structure building and analysis. *Protein Sci* 2023;32(11):e4792.
- Holm L. Dali server: Structural unification of protein families. *Nucleic Acids Res* 2022;50(W1):W210–W215.
- Lin Z, Akin H, Rao R, et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* 2023;379(6637): 1123–1130.
- Rives A, Meier J, Sercu T, et al. Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *Proc Natl Acad Sci U S A* 2021;118(15):e2016239118.
- Yiu G, Tieu E, Nguyen AT, et al. Genomic disruption of VEGF-A expression in human retinal pigment epithelial cells using CRISPR-Cas9 endonuclease. *Invest Ophthalmol Vis Sci* 2016;57(13): 5490–5497.
- Chung SH, Sin T-N, Dang B, et al. CRISPR-based VEGF suppression using paired guide RNAs for treatment of choroidal neovascularization. *Mol Ther Nucleic Acids* 2022;28:613–622.
- Liu B, Poolman B, Boersma AJ. Ionic strength sensing in living cells. *ACS Chem Biol* 2017;12(10):2510–2514.
- Tang P-Z, Ding B, Peng L, et al. TEG-Seq: An ion torrent-adapted NGS workflow for in Cellulo mapping of CRISPR specificity. *Biotechniques* 2018;65(5):259–267.
- Cameron P, Fuller CK, Donohoue PD, et al. Mapping the genomic landscape of CRISPR–Cas9 cleavage. *Nat Methods* 2017;14(6): 600–606.
- Wang Y, Xue P, Cao M, et al. Directed evolution: Methodologies and applications. *Chem Rev* 2021;121(20):12384–12444.
- Moreno-Mateos MA, Fernandez JP, Rouet R, et al. CRISPR-Cpf1 mediates efficient homology-directed repair and temperature-controlled genome editing. *Nat Commun* 2017;8(1):2024.
- Bodai Z, Bishop AL, Gantz VM, et al. Targeting double-strand break indel byproducts with secondary guide RNAs improves Cas9 HDR-mediated genome editing efficiencies. *Nat Commun* 2022;13(1):2351.
- Möller L, Aird EJ, Schröder MS, et al. Recursive Editing improves homology-directed repair through retargeting of undesired outcomes. *Nat Commun* 2022;13(1):4550.
- Karvelis T, Druteika G, Bigelyte G, et al. Transposon-associated TnpB is a programmable RNA-guided DNA endonuclease. *Nature* 2021; 599(7886):692–696.
- Yan WX, Hunnewell P, Alfonse LE, et al. Functionally diverse type V CRISPR-Cas systems. *Science* 2019;363(6422):88–91.

48. Ran FA, Cong L, Yan WX, et al. In vivo genome editing using *Staphylococcus aureus* Cas9. *Nature* 2015;520(7546):186–191.
49. Schiml S, Fauser F, Puchta H. The CRISPR/Cas system can be used as nuclease for in planta gene targeting and as paired nickases for directed mutagenesis in *Arabidopsis* resulting in heritable progeny. *Plant J* 2014;80(6):1139–1150.
50. Raguram A, Banskota S, Liu DR. Therapeutic in vivo delivery of gene editing agents. *Cell* 2022;185(15):2806–2827.
51. Finn JD, Smith AR, Patel MC, et al. A single administration of CRISPR/Cas9 lipid nanoparticles achieves robust and persistent in vivo genome editing. *Cell Rep* 2018;22(9):2227–2235.
52. Hamilton JR, Tsuchida CA, Nguyen DN, et al. Targeted delivery of CRISPR-Cas9 and transgenes enables complex immune cell engineering. *Cell Rep* 2021;35(9):109207.
53. Kim S, Kim D, Cho SW, et al. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res* 2014;24(6):1012–1019.
54. Woo JW, Kim J, Kwon SI, et al. DNA-free genome editing in plants with preassembled CRISPR-Cas9 ribonucleoproteins. *Nat Biotechnol* 2015;33(11):1162–1164.
55. Svitashv S, Schwartz C, Lenderts B, et al. Genome editing in maize directed by CRISPR-Cas9 ribonucleoprotein complexes. *Nat Commun* 2016;7:13274.
56. Liang X, Potter J, Kumar S, et al. Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection. *J Biotechnol* 2015;208:44–53.
57. Weiss T, Kamalu M, Shi H, et al. Viral delivery of an RNA-guided genome editor for transgene-free germline editing in *Arabidopsis*. *Nat Plants* 2025;11(5):967–976.
58. Wu Z, Yang H, Colosi P. Effect of genome size on AAV vector packaging. *Mol Ther* 2010;18(1):80–86.

*Received: February 17, 2026*

*Accepted: April 21, 2026*

*Online Publication Date: May 19, 2026*