GENOME EDITING

Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity

Alexis C. Komor,[1,2,3]* Kevin T. Zhao,[1,2,3] Michael S. Packer,[1,2,3†] Nicole M. Gaudelli,[1,2,3] Amanda L. Waterbury,[1] Luke W. Koblan,[1,2,3] Y. Bill Kim,[1,2,3] Ahmed H. Badran,[1,2,3] David R. Liu[1,2,3‡]

We recently developed base editing, the programmable conversion of target C:G base pairs to T:A without inducing double-stranded DNA breaks (DSBs) or requiring homology-directed repair using engineered fusions of Cas9 variants and cytidine deaminases. Over the past year, the third-generation base editor (BE3) and related technologies have been successfully used by many researchers in a wide range of organisms. The product distribution of base editing—the frequency with which the target C:G is converted to mixtures of undesired by-products, along with the desired T:A product—varies in a target site–dependent manner. We characterize determinants of base editing outcomes in human cells and establish that the formation of undesired products is dependent on uracil N-glycosylase (UNG) and is more likely to occur at target sites containing only a single C within the base editing activity window. We engineered CDA1-BE3 and AID-BE3, which use cytidine deaminase homologs that increase base editing efficiency for some sequences. On the basis of these observations, we engineered fourth-generation base editors (BE4 and SaBE4) that increase the efficiency of C:G to T:A base editing by approximately 50%, while halving the frequency of undesired by-products compared to BE3. Fusing BE3, BE4, SaBE3, or SaBE4 to Gam, a bacteriophage Mu protein that binds DSBs greatly reduces indel formation during base editing, in most cases to below 1.5%, and further improves product purity. BE4, SaBE4, BE4-Gam, and SaBE4-Gam represent the state of the art in C:G-to-T:A base editing, and we recommend their use in future efforts.

INTRODUCTION

Traditional genome editing methods introduce a double-stranded DNA break (DSB) at a genomic target locus (*1*). The cellular response to a DSB lesion primarily proceeds through nonhomologous end joining (NHEJ) and related processes (*2*). Although NHEJ usually rejoins the two ends flanking the DSB, under typical genome editing conditions, DSBs are continuously reintroduced, eventually resulting in the accumulation of insertions and deletions (indels) or translocations at the site of the DSB and the disruption of the corresponding genomic locus (*3*). Actively dividing cells can also respond to DSBs by initiating homology-directed repair (HDR) in the presence of a donor DNA template containing homology to the regions surrounding the DSB, which allows researchers to more precisely and predictably manipulate genomes than is possible through NHEJ (*4*). HDR-dependent genome editing is limited by low efficiency arising from competition with NHEJ outcomes and from the dependence of HDR on mitosis (*5*).

We recently reported the development of base editing, which enables the direct, irreversible conversion of a C:G base pair to a T:A base pair in a programmable manner without requiring HDR or the introduction of a DSB (*6*). Base editors consist of a single-stranded DNA (ssDNA)–specific cytidine deaminase enzyme tethered to a catalytically impaired Cas9 protein (*6*–*9*). The Cas9 variant binds a genomic locus of interest, programmed by a corresponding guide RNA. Formation of the protein-RNA-DNA ternary "R-loop" complex (*10*) exposes a small (~5-nucleotide) window of ssDNA that serves as a substrate for the tethered cytidine deaminase enzyme. Any cytidines within this window are hydrolytically deaminated to uracils, resulting in G:U intermediates.

Base excision repair (BER) is the cell's primary response to G:U mismatches and is initiated by excision of the uracil by uracil N-glycosylase (UNG) (*11*). In an effort to protect the edited G:U intermediate from excision by UNG, we fused a 83–amino acid uracil glycosylase inhibitor (UGI) directly to the C terminus of catalytically dead Cas9 (dCas9) (*6*). To manipulate cellular DNA mismatch repair systems into preferentially replacing the G in the G:U mismatch with an A, we also reverted the Ala^{840} amino acid in dCas9 to His, enabling the Cas9 protein to nick the DNA strand opposite the newly formed uracil, resulting in much more efficient conversion of the G:U intermediate to desired A:U and A:T products (*6*). Combining these two engineering efforts resulted in BE3, a single protein consisting of a three-part fusion of the APOBEC1 cytidine deaminase enzyme tethered through a 16–amino acid linker to *Streptococcus pyogenes* Cas9 nickase [Cas9n(D10A)], which is covalently linked to UGI through a 4–amino acid linker (*6*). Since our initial report, the scientific community has used BE3 and related base editors for a wide variety of applications, including plant genome editing, in vivo mammalian genome editing, targeted mutagenesis, and knockout studies (*1*, *7*–*9*, *12*–*19*). We recently expanded the scope of base editing by reporting BE3 variants with altered PAM requirements (*7*), narrowed editing windows (*7*), reduced off-target editing (*9*), and small-molecule dependence (*20*).

At some loci, base editors such as BE3 give rise to undesired by-products in which the target C:G base pair is converted into a G:C or A:T base pair, rather than the desired T:A product (*1*, *12*, *13*, *15*, *16*). Here, we illuminate determinants of base editing product purity and establish that UNG activity is required for the formation of undesired by-products. By analyzing individual DNA sequencing reads, we discovered that blocking UNG access to the uracil intermediate is especially crucial

[1]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138, USA. [2]Howard Hughes Medical Institute, Harvard University, Cambridge, MA 02138, USA. [3]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA.
*Present address: Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, CA 92093, USA.
†Present address: Beam Therapeutics, Cambridge, MA 02142, USA.
‡Corresponding author. Email: drliu@fas.harvard.edu

for target loci in which a single C is within the editing window to minimize undesired products. Using these insights, we engineered fourth-generation base editors, BE4 (*S. pyogenes* Cas9-derived base editor) and SaBE4 (*Staphylococcus aureus* Cas9-derived BE4), that perform base editing with higher efficiency and greatly improved product purity compared to previously described base editors including BE3. Finally, we developed additional base editors—BE3-Gam, SaBE3-Gam, BE4-Gam, and SaBE4-Gam—that use the bacteriophage Mu dsDNA (double-stranded DNA) end-binding protein Gam to minimize the formation of undesired indels during base editing, and to further increase product purity with no apparent loss of activity.

RESULTS

UNG activity is required for by-product formation

We hypothesized that undesired base editing by-products arise during BER because of the formation and error-prone resolution of abasic sites within the uracil-containing DNA strand. This hypothesis predicts that the product purity of base editing in cells lacking UNG should be greatly improved. To test this prediction, we nucleofected HAP1 cells (a haploid human cell line) and HAP1 *UNG*$^{-}$ cells with plasmids encoding BE3 and single-guide RNAs (sgRNAs) targeting the *EMX1*, *FANCF*, *HEK2*, *HEK3*, *HEK4*, or *RNF2* locus (see Fig. 1B for target sequences). Three days after nucleofection, genomic DNA was extracted, and the target loci were amplified by polymerase chain reaction (PCR) and analyzed by high-throughput DNA sequencing (HTS). We define base editing product purity as the percentage of edited sequencing reads (reads in which the target C has been converted to a different base) in which the target C is edited to a T. The base editing product purity of BE3-treated HAP1 cells averaged 68 ± 6% (means ± SD for n = 3 biological replicates) across 12 target C's in the six loci. Remarkably, in HAP1 *UNG*$^{-}$ cells, all 12 target C's tested were base-edited with product purities of >98% (Fig. 1C). In addition, indel frequencies at all six tested loci decreased 7- to 100-fold upon UNG knockout (Fig. 1D). These data strongly implicate UNG activity as necessary for undesired product formation during base editing, consistent with a model in which abasic site formation and subsequent BER with error-prone polymerases lead to randomization of the target nucleotide and occasional strand breaks that result in indels.

Targets with multiple editable C's exhibit higher product purity

We previously reported that base editing efficiency by BE3 can be lower for some (but not all) target C's that are immediately downstream of a G (*6*), consistent with the known sequence preference of APOBEC1 (Fig. 2B and fig. S2, A and B) (*21*). In an effort to efficiently edit these targets, we generated BE3 variants in which we replaced the APOBEC1 deaminase with CDA1 (to generate CDA1-BE3), AID (to generate AID-BE3), or APOBEC3G (to generate APOBEC3G-BE3), three ssDNA-specific cytidine deaminase enzymes with different sequence preferences (*22*). We transfected human embryonic kidney (HEK) 293T cells with plasmids encoding these BE3 variants and sgRNAs targeting the *EMX1*, *FANCF*, *HEK2*, *HEK3*, *HEK4*, or *RNF2* locus. Three days after transfection, genomic DNA was extracted, and the target loci were amplified by PCR and assessed for base editing using HTS. We observed more efficient editing of target C's that immediately follow a G with CDA1-BE3 and AID-BE3 than with BE3 (Fig. 2B, fig. S2, and tables S1 to S6). In general, CDA1-BE3 and AID-BE3 exhibited lower editing efficiencies than BE3 at target C's that do not follow a G (fig. S2). In contrast, APOBEC3G-BE3 exhibited unpredictable sequence preferences, with

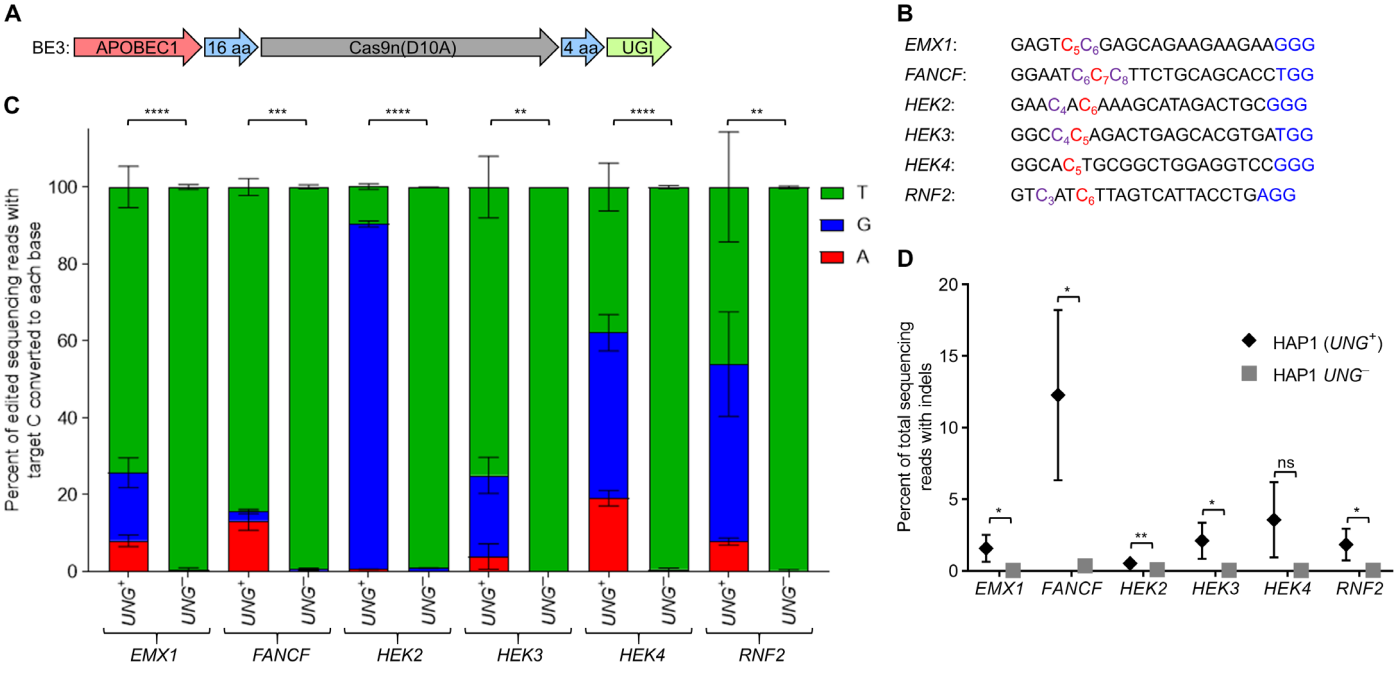
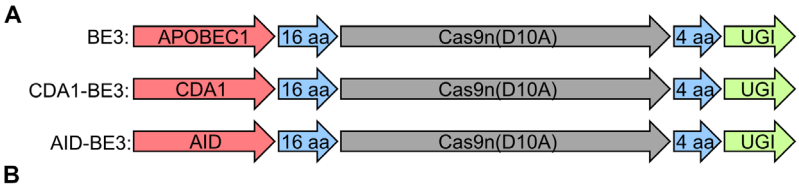


Fig. 1. Effects of knocking out *UNG* on base editing product purity. (**A**) Architecture of BE3. (**B**) Protospacers and PAM (blue) sequences of the genomic loci tested, with the target C's analyzed in (A) shown in red. (**C**) HAP1 (*UNG*$^{+}$) and HAP1 *UNG*$^{-}$ cells were treated with BE3, as described in Materials and Methods. The product distribution among edited DNA sequencing reads (reads in which the target C is mutated) is shown. See fig. S1 for C-to-T editing efficiencies, which generally varied between 15 and 45%. (**D**) Frequency of indel formation following treatment with BE3 in HAP1 or HAP1 *UNG*$^{-}$ cells. Values and error bars reflect the means and SD of three independent biological replicates performed on different days. ns (not significant), $P \geq 0.05$; *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$, by two-tailed Student's t test.

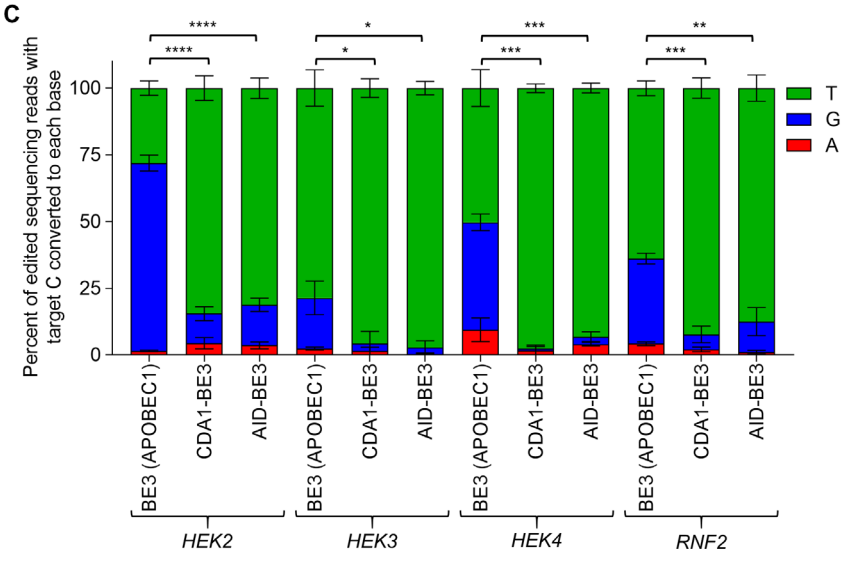
A



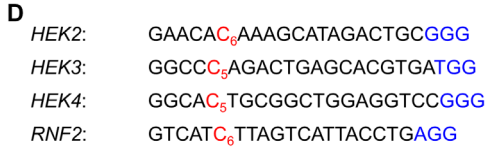
B

HEK4		G	G	C_3	A	C_5	T	G	C_8	G	G	C	T	G	G	A	G	G	T	G	G	G	G	G
BE3	A	0.1	0.0	0.1	99.9	8.4	0.1	0.0	0.1	0.0	0.1	0.0	0.0	0.0	0.0	99.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	C	0.0	0.0	99.0	0.0	41.9	0.0	0.0	99.3	0.0	0.0	99.8	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	G	99.9	99.9	0.0	0.0	24.5	0.0	100.0	0.0	100.0	99.9	0.0	0.0	100.0	100.0	0.1	100.0	100.0	0.0	100.0	100.0	100.0	100.0	100.0
	T	0.0	0.0	0.9	0.0	25.2	99.9	0.0	0.5	0.0	0.0	0.2	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0
CDA1-BE3	A	0.2	0.1	0.5	100.0	0.5	0.0	0.0	0.2	0.0	0.0	0.1	0.0	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	C	0.0	0.0	91.3	0.0	87.5	0.0	0.0	93.4	0.0	0.0	97.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	G	99.8	99.9	0.3	0.0	0.0	0.0	100.0	0.0	100.0	100.0	0.0	0.0	100.0	100.0	0.0	100.0	100.0	0.0	100.0	100.0	100.0	100.0	100.0
	T	0.0	0.0	7.8	0.0	12.0	100.0	0.0	6.3	0.0	0.0	2.9	100.0	0.0	0.0	0.0	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0
AID-BE3	A	0.3	0.4	0.3	100.0	0.9	0.0	0.0	0.9	0.0	0.0	0.1	0.0	0.0	0.0	99.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	C	0.0	0.0	81.5	0.0	71.3	0.0	0.0	80.7	0.0	0.0	97.9	0.1	0.0	0.0	0.2	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0
	G	99.7	99.6	0.1	0.0	0.7	0.0	100.0	0.1	100.0	100.0	0.0	0.0	99.9	100.0	0.0	100.0	100.0	0.2	99.9	100.0	100.0	100.0	99.9
	T	0.0	0.0	18.0	0.0	27.1	100.0	0.0	18.3	0.0	0.0	2.0	99.9	0.0	0.0	0.0	0.0	0.0	99.8	0.0	0.0	0.0	0.0	0.1

C



D



E

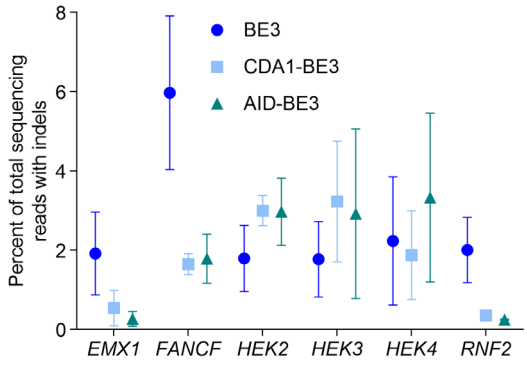


Fig. 2. Effects of multi-C base editing on product purity. (**A**) Architectures of BE3, CDA1-BE3, and AID-BE3. (**B**) Representative high-throughput sequencing data of BE3-, CDA1-BE3–, and AID-BE3–treated human HEK293T cells. The sequence of the protospacer is shown at the top, with the PAM in blue and the target C's in red, with subscripted numbers indicating their position within the protospacer. Underneath each sequence are the percentages of total sequencing reads with the corresponding base. The relative percentage of target C's that are cleanly edited to T rather than to non-T bases are much higher for cells treated with AID-BE3, which edits three C's at this locus, than for cells treated with BE3, which edits only one C. (**C**) HEK293T cells were treated with BE3, CDA1-BE3, and AID-BE3, as described in Materials and Methods. The product distribution among edited DNA sequencing reads (reads in which the target C is mutated) is shown. (**D**) Protospacers and PAM (blue) sequences of genomic loci studied, with the target C's analyzed in (B) shown in red. (**E**) Frequency of indel formation (see Materials and Methods) following the treatment in (A). Values and error bars reflect the means and SD of three independent biological replicates performed on different days. *P < 0.05; **P < 0.01; ***P < 0.001; ****P < 0.0001, by two-tailed Student's t test.

overall lower yields of C-to-T editing compared to BE3. These findings suggest that CDA1-BE3 and AID-BE3 may offer higher editing efficiencies over BE3 for some target 5′-GC-3′ sequences.

While analyzing these data, we noticed that the product purities of CDA1-BE3 and AID-BE3 were typically higher than those of BE3 at those sites for which CDA1-BE3 and AID-BE3 edited more C's than BE3 (Fig. 2). For example, at the *HEK4* locus, BE3 efficiently edits only a single C (the C not preceded by a G), but both CDA1-BE3 and AID-BE3 edit three C's (fig. S2). The product purity of BE3 at this locus is 50 ± 7% (means ± SD for n = 3 biological replicates), whereas the product purity of CDA1-BE3 and AID-BE3 are 97 ± 2% and 93 ± 2%, respectively. Moreover, *EMX1* and *FANCF*, edited by BE3 with product purities of 84 ± 3% and 91 ± 2%, respectively, contain multiple C's that are edited with comparable efficiency (fig. S3), whereas *HEK2* and *RNF2*, edited by BE3 with much lower product purities of 28 ± 3% and 64 ± 3%, respectively, contain multiple C's that are edited with unequal efficiencies (fig. S3). CDA1-BE3 and AID-BE3, which edit both C's within the *HEK2* locus with comparable efficiencies, exhibit much higher product purities at this locus (85 ± 5% and 81 ± 4%, respectively) (Fig. 2 and fig. S2C). We therefore ruled out the possibility that, at the *HEK2* and *RNF2* sites, the multiple C's are initially converted to U's by BE3 with comparable efficiency and then processed with different efficiencies by DNA repair systems; if this were the case, we would expect similar product distributions when these sites were treated with BE3 versus CDA1-BE3 or AID-BE3, rather than the different product distributions observed (Fig. 2C and tables S1 to S6). Instead, we hypothesized that an isolated G:U may

be more readily processed by UNG than clusters of G:U lesions. It is tempting to speculate that the processivity of the cytidine deaminase domain in BE3 (*6*, *23*) may increase the residence time of BE3 at loci containing multiple editable C's, thereby blocking access by UNG more effectively than at loci containing a single editable C.

We sought to further probe the relationship between product purity, the number of edited C's in individual sequencing reads, and UNG activity. To reveal the fate of base-edited DNA in the absence of explicit UNG inhibition, we removed the UGI component of BE3 to generate BE3B. We transfected HEK293T cells with plasmids encoding BE3 or BE3B and sgRNAs targeting the *EMX1*, *FANCF*, *HEK2*, *HEK3*, *HEK4*, or *RNF2* locus. As expected, given the role of UNG in diversifying base editing outcomes established above, the product purities at all target C's greatly decreased in BE3B-treated DNA compared with BE3-treated DNA, with the fraction of editing products containing non-Ts increasing by an average of 1.8 ± 0.4–fold (fig. S4D).

We analyzed individual DNA sequencing reads from HEK293T cells treated with sgRNAs targeting the multi-C sites *HEK2*, *HEK3*, and *RNF2* and either BE3 or BE3B. For each site, we designated the primary target C as the nucleotide modified most efficiently. Across all three sites, an average of 80 ± 10% of sequencing reads that contained an undesired C to non-T edit of the primary target C exhibited only that single base editing event (figs. S4 and S5). In contrast, across the same three multi-C sites, a much lower average of 32 ± 4% of sequencing reads containing a clean C-to-T edit of the primary target C exhibited only that single clean base editing event (figs. S4 and S5). In addition, the distribution of products for BE3B-treated *HEK4* DNA, a site that contains only one C within the editing window, roughly follows the ratio of 1:3:1 for A:G:T (fig. S4E). These observations collectively indicate that when a single cytidine in a given target is converted to U in the absence of UGI, it is processed efficiently by UNG-initiated BER to give a mixture of products.

These data are consistent with a model in which clustered G:U mismatches are processed differently than isolated G:U mismatches and are more likely to produce clean C-to-T edits. When only a single C-to-T editing event is desired, the abovementioned observations suggest that UNG inhibition is critical to minimize undesired by-products. However, when performing targeted random mutagenesis using dCas9-deaminase fusions, such as with TAM (*16*) and CRISPR-X (*12*), the abovementioned observations suggest that using BE3B on target sites with a minimum number of editable C's will maximize product mixtures.

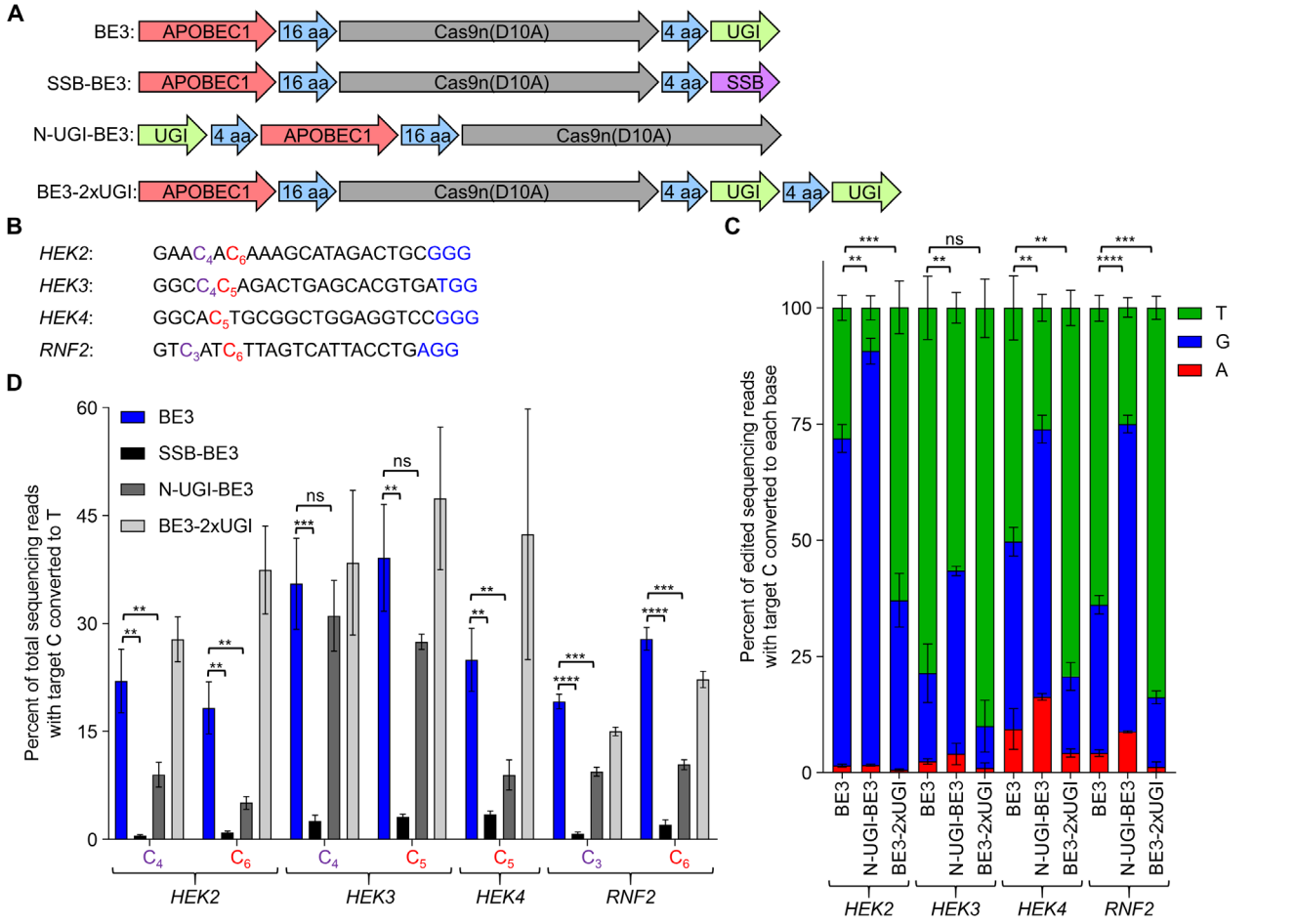


Fig. 3. Effects of changing the architecture of BE3 on C-to-T editing efficiencies and product purities. (**A**) Architectures of BE3, SSB-BE3, N-UGI-BE3, and BE3-2xUGI. (**B**) Protospacers and PAM (blue) sequences of genomic loci studied, with the target C's in (C) shown in purple and red, and the target C's in (B) shown in red. (**C**) HEK293T cells were treated with BE3, SSB-BE3, N-UGI-BE3, and BE3-2xUGI, as described in Materials and Methods. The product distribution among edited DNA sequencing reads (reads in which the target C is mutated) is shown for BE3, N-UGI-BE3, and BE3-2xUGI. (**D**) C-to-T base editing efficiencies. Values and error bars reflect the means and SD of three independent biological replicates performed on different days. ns, $P \geq 0.05$; *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$, by two-tailed Student's *t* test.

Optimization of BE3 architecture improves product purity

We sought to improve the product purity of base editing, especially for sites with a single C or with unevenly edited C's. The abovementioned findings suggest that optimizing BE3 to minimize access of UNG to the G:U intermediate should improve product purity. We first tried replacing the UGI component of BE3 with an ssDNA binding protein to yield SSB-BE3, reasoning that SSB may block the uracil-containing ssDNA portion of the R-loop from being accessed by UNG. We observed large decreases in base editing efficiency by SSB-BE3, with all seven C's across the four sites exhibiting an average of only 1.9 ± 0.5% C-to-T conversion (Fig. 3D). These data suggest that although the tethered deaminase enzyme is present at elevated effective molarity, SSB protects the ssDNA portion of the R-loop from the deaminase enzyme, impeding base editing.

Because the relative positioning of APOBEC, UGI, and UNG during steps that determine base editing outcomes is not known, we moved UGI to the N terminus of BE3 (N-UGI-BE3) in an effort to improve UNG inhibition. Moving UGI to the N terminus of BE3 resulted in an average decrease in C-to-T editing percentages of 2.3 ± 0.6–fold across all seven tested target C's compared to BE3 (Fig. 3D) and a decrease in overall product purity averaging 2.2 ± 0.5–fold at all four sites compared to BE3 (Fig. 3C). We concluded that the N-UGI-BE3 architecture is suboptimal for deaminase activity and may also impede UNG inhibition by UGI.

In contrast, appending an additional copy of UGI to the C terminus of BE3 (BE3-2xUGI) resulted in large increases in product purities relative to BE3 and C-to-T editing percentages comparable to those of BE3. Non-T editing products decreased by an average of 2.2 ± 0.8–fold across the four loci tested (Fig. 3C). These observations suggest that addition of a second copy of UGI substantially decreases the access of UNG to the G:U base editing intermediate, thereby greatly improving product purity.

Because these experiments revealed the sensitivity of base editing outcomes to the architecture of the components, we next optimized the linkers between BE3 components to further increase product purities and editing efficiencies. We varied the rAPOBEC1-Cas9n(D10A) linker from 16 amino acids (BE3) to 32 amino acids (BE3C) and the Cas9n(D10A)-UGI linker from 4 (BE3) to 9 (BE3D) to 16 amino acids (BE3E) (Fig. 4A). Non-T product formation on average decreased 1.3 ± 0.1–fold when the Cas9n(D10A)-UGI linker was nine amino acid residues in length (BE3D) instead of four amino acids (BE3) (Fig. 4D), with no apparent differences in C-to-T editing efficiencies (Fig. 4C). Increasing the rAPOBEC1-Cas9n(D10A) linker from 16 amino acids (BE3) to 32 amino acids (BE3C) elevated C-to-T editing efficiencies an average of 1.2 ± 0.1–fold at the *HEK2* locus (Fig. 4C). This locus was previously the most unevenly edited multi-C site tested (fig. S3), and extending this linker led to a reduction in preferential editing of C_6 over C_4 (the ratio of the percentage of sequencing reads that are edited at C_6 to that of C_4) from 2.6 ± 0.2–fold to 1.8 ± 0.1–fold. We reasoned that this longer linker may allow the deaminase better access to the ssDNA in the R-loop and result in more uniform deamination when multiple target C's are present in the base editing window. BE3C also exhibited comparable or improved base editing efficiencies and product purities at the other loci tested (Fig. 4, C and D).

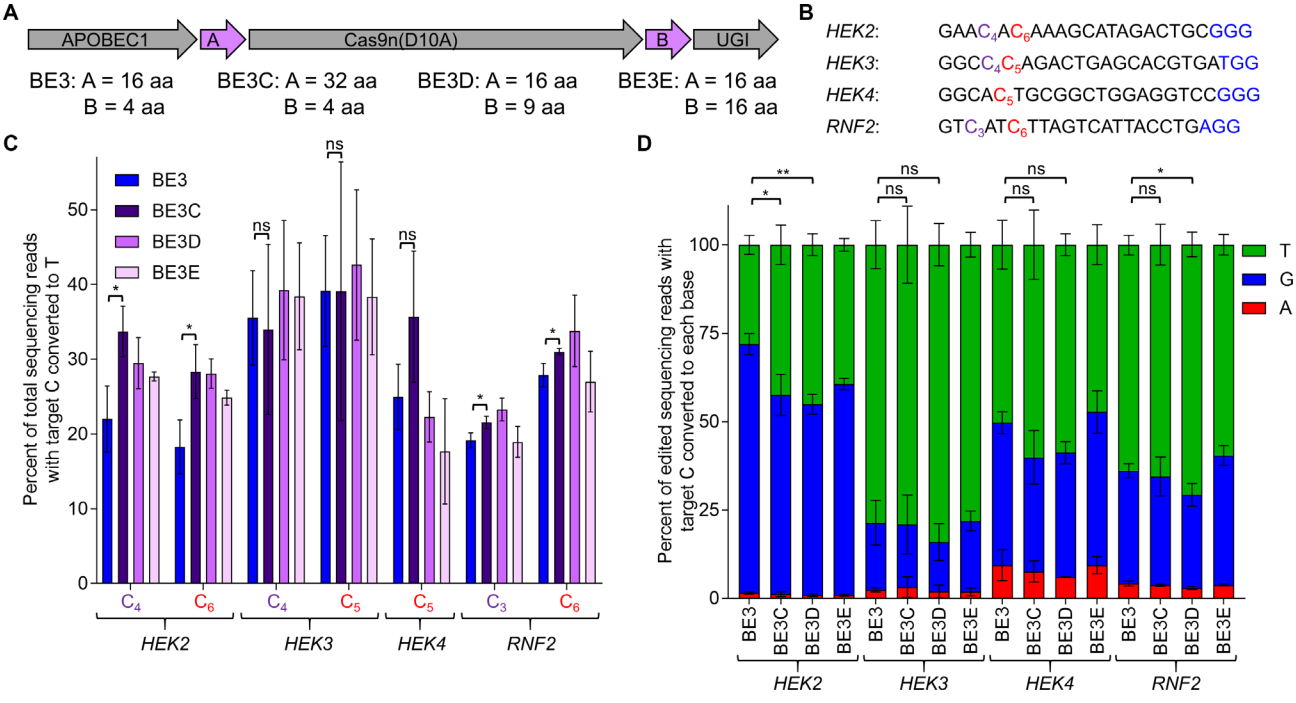


Fig. 4. Effects of linker length variation in BE3 on C-to-T editing efficiencies and product purities. (**A**) Architecture of BE3, BE3C, BE3D, and BE3E. (**B**) Protospacers and PAM (blue) sequences of genomic loci studied, with the target C's in (C) shown in purple and red, and target C's in (D) shown in red. (**C**) HEK293T cells were treated with BE3, BE3C, BE3D, or BE3E, as described in Materials and Methods. C-to-T base editing efficiencies are shown. (**D**) The product distribution among edited DNA sequencing reads (reads in which the target C is mutated) is shown for BE3, BE3C, BE3D, and BE3E. Values and error bars reflect the means and SD of three independent biological replicates performed on different days. ns, $P \geq 0.05$; *$P < 0.05$; **$P < 0.01$, by two-tailed Student's *t* test.

Generation of BE4, a C:G to T:A base editor with enhanced efficiency and product purity

We combined all three improvements—extending the rAPOBEC1-Cas9n linker to 32 amino acids, extending the Cas9n-UGI linker to 9 amino acids, and appending a second copy of UGI to the C terminus of the construct with another 9–amino acid linker—into a single base editor construct, BE4. We also cloned Target-AID, an alternative base editor construct reported by Nishida *et al.* (*8*), into the same plasmid backbone as BE4. We transfected HEK293T cells with plasmids encoding BE3, BE4, or Target-AID and sgRNAs targeting the *EMX1*, *FANCF*, *HEK2*, *HEK3*, *HEK4*, or *RNF2* locus. Three days after transfection, genomic DNA was extracted, and the target loci were amplified by PCR and analyzed by HTS. We observed an average increase in C-to-T editing efficiencies of 1.5 ± 0.3–fold across all 12 edited C's for BE4 relative to BE3 (Fig. 5C). Although the average efficiency of C-to-T editing for Target-AID at the same positions analyzed was 1.5 ± 0.5–fold lower than that of BE3 and 2.1 ± 0.5–fold lower than that of BE4, it is important to note that Target-AID, which uses the CDA1 deaminase, appears to have an editing window shifted relative to BE3 and BE4, with optimal editing around positions C_3 and C_4 (Fig. 5C). This shifted editing window makes comparisons of efficiency and product purity between Target-AID and BE3 or BE4 difficult because a given target C could lie in more optimal or less optimal position within the different editing windows, even when using the same guide RNA.

In addition to greater C-to-T editing efficiency, BE4 also exhibited substantially improved product purities relative to BE3 at all genomic loci tested, with an average decrease in non-T product formation of 2.3 ± 0.3–fold (Fig. 5D). As expected from further impeding BER, which can lead to indels (*24*), we also observed decreases in indel rates averaging 2.3 ± 1.1–fold across all six loci following BE4 treatment compared to BE3 (fig. S6A). Together, these results indicate that BE4 offers higher efficiencies of C-to-T editing, higher product purities, and lower indel rates than BE3 at all loci tested.

We tested whether the BE4 improvements could be integrated with *S. aureus* Cas9 (*25*) to generate SaBE4, which replaces the *S. pyogenes* Cas9n(D10A) with the smaller *S. aureus* Cas9n(D10A) and can access different targets because of its alternative PAM requirements. We transfected HEK293T cells with plasmids encoding SaBE3 (*7*) or SaBE4 and sgRNAs targeting the *FANCF*, *HEK3*, or *HEK4* locus. Consistent with the results comparing BE4 and BE3, we observed an average increase in C-to-T editing efficiencies of 1.4 ± 0.2–fold across all 10 edited C's for SaBE4 relative to SaBE3 (fig. S7A), with a 1.8 ± 0.5–fold average decrease in undesired non-T editing products (fig. S7C). These results indicate that the gains in base editing efficiency and product purity that arise from the BE4 enhancements also apply to base editors derived from other Cas9 homologs.

Fusion with Gam further reduces indel frequencies and improves product purity

For some genome editing applications, the formation of indels confounds research or poses safety risks. We therefore sought to further decrease indel frequencies that arise from base editing. We hypothesized

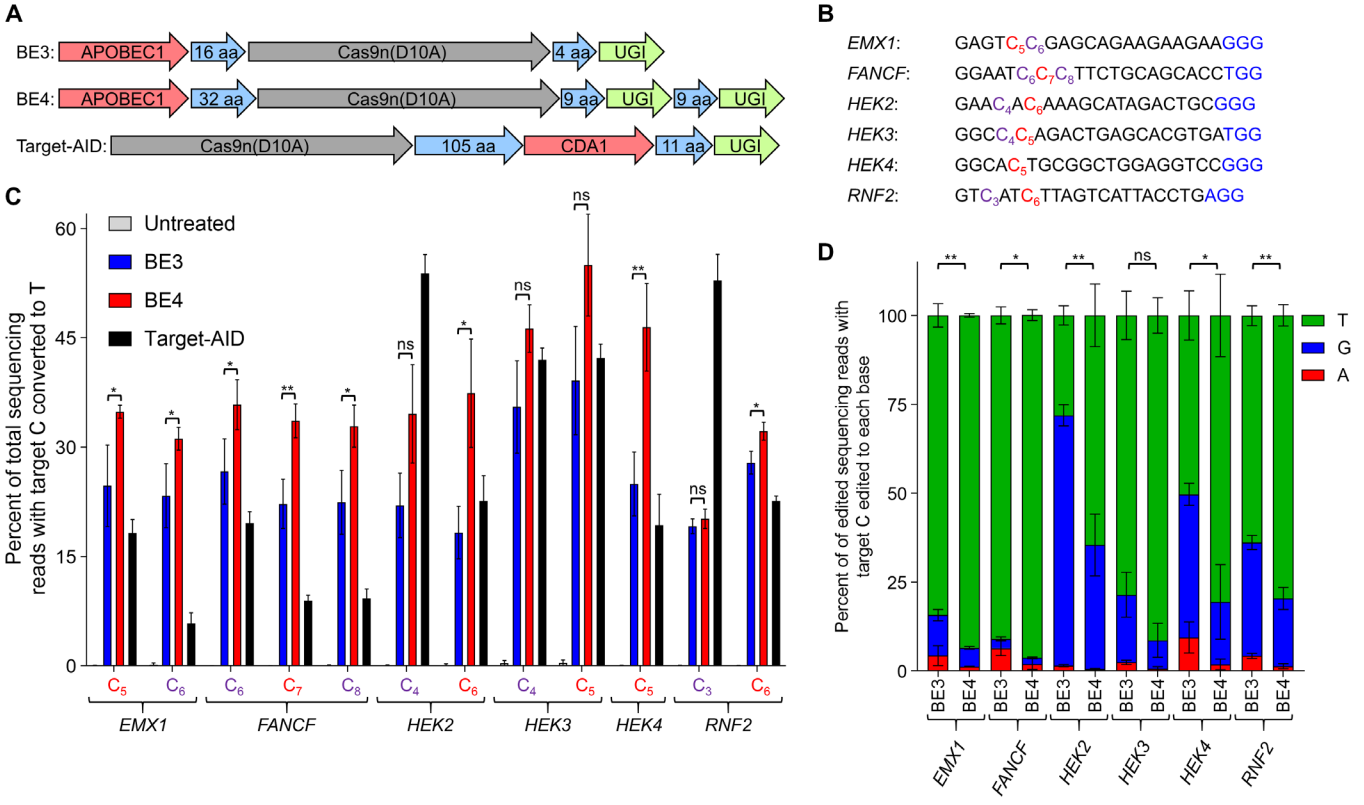


Fig. 5. BE4 increases base editing efficiency and product purities compared to BE3. (**A**) Architectures of BE3, BE4, and Target-AID. (**B**) Protospacers and PAM (blue) sequences of genomic loci studied, with the target C's in (C) shown in purple and red, and the target C's in (D) shown in red. (**C**) HEK293T cells were treated with BE3, BE4, or Target-AID, as described in Materials and Methods. C-to-T base editing efficiencies are shown. (**D**) The product distribution among edited DNA sequencing reads (reads in which the target C is mutated) is shown for BE3 and BE4. Values and error bars reflect the means and SD of three independent biological replicates performed on different days. ns, $P \geq 0.05$; *$P < 0.05$; **$P < 0.01$, by two-tailed Student's *t* test.

that most of the base editing–induced indels occur as a result of DNA-(apurinic or apyrimidinic site) lyase (AP lyase), a BER enzyme that converts abasic sites into ssDNA nicks (*24*). Because base editors nick the strand opposite the U, cleavage of the glycosidic bond by UNG followed by processing of the resulting AP site by AP lyase would result in a DSB, which promotes indel formation. This model is consistent with our observation of greatly reduced indel frequencies in UNG knockout cells (Fig. 1D). The Gam protein of bacteriophage Mu binds to the ends of DSBs and protects them from degradation (*26*), and has been repurposed to image DSBs in live mammalian cells (*27*). We reasoned that using Gam to bind the free ends of DSB may reduce indel formation during the process of base editing. We therefore fused the 174-residue Gam protein to the N terminus of BE3, SaBE3, BE4, and SaBE4 via the 16–amino acid XTEN linker to generate BE3-Gam, SaBE3-Gam, BE4-Gam, and SaBE4-Gam, respectively.

BE3-Gam and SaBE3-Gam decreased indel frequencies relative to BE3 and SaBE3 at all six and four genomic loci tested by an average of 1.7 ± 0.3–fold and 2.0 ± 1.0–fold, respectively (Fig. 6C and fig. S7D). C-to-T editing efficiencies for BE3-Gam and SaBE3-Gam were similar to those of BE3 and SaBE3, respectively (Fig. 6B). In addition, BE3-Gam and SaBE3-Gam also exhibited increased product purity relative to BE3 and SaBE3 at all genomic loci tested, with an average decrease in non-T product formation of 1.5 ± 0.1–fold and 2.3 ± 0.6–fold, respectively (Fig. 6D).

BE4-Gam exhibited greatly decreased indel frequencies relative to BE4, with an average decrease of 2.1 ± 0.9–fold (Fig. 6C). In general, indel frequency following BE4-Gam treatment is below 1.5%. Compared to SaBE4, SaBE4-Gam exhibited an average decrease in indel formation of 1.5 ± 0.8–fold (fig. S7D). We observed no apparent decreases in C-to-T editing efficiencies for BE4-Gam or SaBE4-Gam relative to BE4 and SaBE4, respectively (Fig. 6B), BE4-Gam offers overall editing/indel ratios that increase an average of 2.0 ± 1.1–fold across all six sites relative to BE4 (fig. S6B). Product purities of BE4-Gam are improved compared with BE4, with an average decrease in non-T product formation of 2.1 ± 0.7–fold (Fig. 6D).

Similarly, SaBE4-Gam exhibited an average decrease in non-T product formation of 1.7 ± 0.5–fold relative to SaBE4, with no apparent decrease in C-to-T editing efficiencies (Fig. 6). These data suggest that for sites that can be targeted by *S. aureus* Cas9, SaBE4-Gam provides the best combination of high C-to-T base editing efficiency, reduced indel formation, and increased product purity. Together, the abovementioned findings establish that the fusion of bacteriophage Mu Gam protein to decrease indel formation is compatible with multiple genome editing agents.

DISCUSSION

For base editing applications in which minimizing indel production is critical and Gam binding of DSBs is acceptable, BE4-Gam or SaBE4-Gam may be preferred BE4-Gam variants offer the lowest indel frequency and highest product purity among the base editors tested in this study. C-to-T editing efficiency/indel ratios increase as BE3 < BE3-Gam < BE4 < BE4-Gam across all six genomic loci (fig. S6B). We speculate that Gam may be inducing the death of DSB-containing

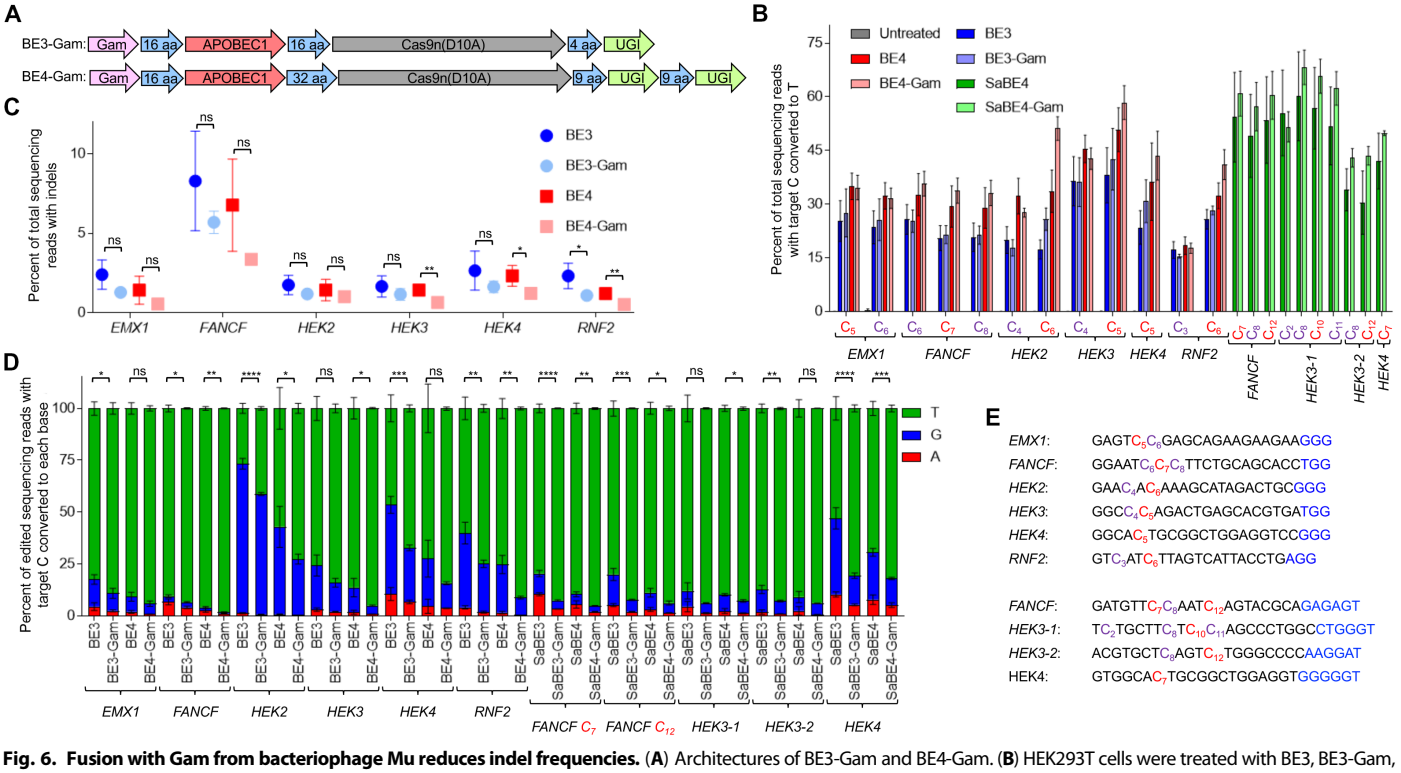


Fig. 6. Fusion with Gam from bacteriophage Mu reduces indel frequencies. (**A**) Architectures of BE3-Gam and BE4-Gam. (**B**) HEK293T cells were treated with BE3, BE3-Gam, BE4, BE4-Gam, SaBE3, SaBE3-Gam, SaBE4, or SaBE4-Gam, as described in Materials and Methods. C-to-T base editing efficiencies are shown. (**C**) Frequency of indel formation (see Materials and Methods) following the treatment in (B). (**D**) Product distribution among edited DNA sequencing reads (reads in which the target C is mutated). (**E**) Protospacers and PAM (blue) sequences of genomic loci studied, with the target Cs in (B) shown in purple and red, and the target Cs in (D) shown in red. Values and error bars of BE3-Gam, SaBE3-Gam, BE4-Gam, and SaBE4-Gam reflect the means and SD of three independent biological replicates performed on different days. Values and error bars of BE3, SaBE3, BE4, and SaBE4 reflect the means and SD of six independent biological replicates performed on different days by two different researchers. ns, $P \geq 0.05$; *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$, by two-tailed Student's *t* test.

cells, consistent with previous findings (*27*), thereby removing indels from the population of treated, surviving cells.

Collectively, these developments advance the state of the art in programmable C:G to T:A base pair conversion and thereby increase the utility and applicability of base editing. Our findings also suggest that Gam has the potential to be repurposed to minimize indel formation in other genome editing applications. Finally, relationships among uracil incorporation, UNG activity, and cellular DNA repair outcomes illuminated in this study may guide future efforts to understand or manipulate eukaryotic DNA repair.

MATERIALS AND METHODS

Cloning of plasmids

All plasmids in this study were generated by USER cloning using Phusion U Hot Start Polymerase (Thermo Fisher). Deaminase and SSB genes were synthesized as gBlocks Gene Fragments (Integrated DNA Technologies), and Target-AID was obtained from Addgene (plasmid #79620). Protein sequences are listed in Supplementary Sequences.

Cell culture

HEK293T (American Type Culture Collection CRL-3216) cells were maintained in Dulbecco's modified Eagle's medium plus GlutaMAX (Thermo Fisher) supplemented with 10% (v/v) fetal bovine serum (FBS) at 37°C with 5% CO_2. HAP1 (Horizon Discovery C631) and HAP1 *UNG*$^-$ (Horizon Discovery HZGHC001531c012) were maintained in Iscove's modified Dulbecco's medium plus GlutaMAX (Thermo Fisher Scientific) supplemented with 10% (v/v) FBS at 37°C with 5% CO_2.

Transfections

HEK293T cells were seeded on 48-well collagen-coated BioCoat plates (Corning) and transfected at approximately 75% confluency. Briefly, 750 ng of BE and 250 ng of sgRNA expression plasmids were transfected using 1.5 μl of Lipofectamine 2000 (Thermo Fisher Scientific) per well according to the manufacturer's protocol.

HAP1 and HAP1 *UNG*$^-$ cells were nucleofected using the SE Cell Line 4D-Nucleofector X Kit S (Lonza) according to the manufacturer's protocol. Briefly, 4×10^5 cells were nucleofected with 300 ng of BE and 100 ng of sgRNA expression plasmids using the 4D-Nucleofector program DZ-113.

HTS of genomic DNA samples

Transfected cells were harvested after 3 days, and the genomic DNA was isolated by incubating cells in lysis buffer [10 mM tris-HCl (pH 8.0), 0.05% SDS, proteinase K (25 μg/ml)] at 37°C for 1 hour followed by 80°C for 30 min. Genomic regions of interest were amplified by PCR with flanking HTS primer pairs, as previously described (*6*, *7*). PCR amplification was carried out with Phusion High-Fidelity DNA Polymerase (Thermo Fisher), according to the manufacturer's instructions and as previously described. Purified DNA was amplified by PCR with primers containing sequencing adaptors. The products were gel-purified and quantified using the Quant-iT PicoGreen dsDNA Assay Kit (Thermo Fisher) and KAPA Library Quantification Kit (KAPA Biosystems). Samples were sequenced on an Illumina MiSeq, as previously described.

Data analysis

Sequencing reads were automatically demultiplexed using MiSeq Reporter (Illumina), and individual FASTQ files were analyzed with a custom MATLAB script, as previously described (*6*). Each read was pairwise-aligned to the appropriate reference sequence using the Smith-Waterman algorithm. Base calls with a Q-score below 31 were replaced with N's and were thus excluded in calculating nucleotide frequencies. This treatment yields an expected MiSeq base-calling error rate of approximately 1 in 1000. Aligned sequences in which the read and reference sequence contained no gaps were stored in an alignment table from which base frequencies could be tabulated for each locus.

Indel frequencies were quantified with the previously described MATLAB script (*6*, *7*, *9*). Briefly, sequencing reads were scanned for exact matches to two 10–base pair (bp) sequences that flank both sides of a window in which indels might occur. If no exact matches were located, the read was excluded from the analysis. If the length of this indel window exactly matched the reference sequence, the read was classified as not containing an indel. If the indel window was two or more bases longer or shorter than the reference sequence, then the sequencing read was classified as an insertion or deletion, respectively.

To evaluate interdependency (linkage disequilibrium) between the base editing outcomes at the multiple target cytidines within an editing window, target site sequences from BE-treated cells were analyzed by a custom Python script (note S1). Briefly, sequencing reads were scanned for exact matches to two 7-bp sequences that flank each side of the protospacer. If the intervening region was not exactly 20 bp, then it was excluded from further analysis. The protospacer sequences were further filtered into four groups based on the identity of the nucleotide at the position with the most non-T editing outcomes (the primary target C). For each of these four groups as well as the entire pool, we tallied the nucleotide abundance at each of the 20 positions within the protospacer.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/3/8/eaao4774/DC1

fig. S1. Base editing efficiencies in *UNG* knockout cells.
fig. S2. CDA1-BE3 and AID-BE3 edit C's following target G's more efficiently than BE3.
fig. S3. Uneven editing in sites with multiple editable C's results in lower product purity.
fig. S4. Base editing of multiple C's results in higher base editing product purity.
fig. S5. Base editing of multiple C's results in higher base editing product purity at the *HEK3* and *RNF2* loci.
fig. S6. BE4 induces lower indel frequencies than BE3, and Target-AID exhibits similar product purities as CDA1-BE3.
fig. S7. SaBE4 exhibits increased base editing yields and product purities compared to SaBE3.
table S1. Base editing outcomes from treatment with BE3, CDA1-BE3, AID-BE3, or APOBEC3G-BE3 at the *EMX1* locus.
table S2. Base editing outcomes from treatment with BE3, CDA1-BE3, AID-BE3, or APOBEC3G-BE3 at the *FANCF* locus.
table S3. Base editing outcomes from treatment with BE3, CDA1-BE3, AID-BE3, or APOBEC3G-BE3 at the *HEK2* locus.
table S4. Base editing outcomes from treatment with BE3, CDA1-BE3, AID-BE3, or APOBEC3G-BE3 at the *HEK3* locus.
table S5. Base editing outcomes from treatment with BE3, CDA1-BE3, AID-BE3, or APOBEC3G-BE3 at the *HEK4* locus.
table S6. Base editing outcomes from treatment with BE3, CDA1-BE3, AID-BE3, or APOBEC3G-BE3 at the *RNF2* locus.
note S1. Python script to detect linkage disequilibrium in base editing outcomes at target sites with multiple target cytidines.
Supplementary Sequences. Amino acid sequences of CDA1-BE3, AID-BE3, BE3-Gam, SaBE3-Gam BE4, BE4-Gam, SaBE4, and SaBE4-Gam.

REFERENCES AND NOTES

1. A. C. Komor, A. H. Badran, D. R. Liu, CRISPR-based technologies for the manipulation of eukaryotic genomes. *Cell* **168**, 20–36 (2017).
2. A. J. Davis, D. J. Chen, DNA double strand break repair via non-homologous end-joining. *Transl. Cancer Res.* **2**, 130–143 (2013).

3. M. M. Vilenchik, A. G. Knudson, Endogenous DNA double-strand breaks: Production, fidelity of repair, and induction of cancer. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 12871–12876 (2003).
4. F. Liang, M. Han, P. J. Romanienko, M. Jasin, Homology-directed repair is a major double-strand break repair pathway in mammalian cells. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 5172–5177 (1998).
5. Y. Miyaoka, J. R. Berman, S. B. Cooper, S. J. Mayerl, A. H. Chan, B. Zhang, G. A. Karlin-Neumann, B. R. Conklin, Systematic quantification of HDR and NHEJ reveals effects of locus, nuclease, and cell type on genome-editing. *Sci. Rep.* **6**, 23549 (2016).
6. A. C. Komor, Y. B. Kim, M. S. Packer, J. A. Zuris, D. R. Liu, Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).
7. Y. B. Kim, A. C. Komor, J. M. Levy, M. S. Packer, K. T. Zhao, D. R. Liu, Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nat. Biotechnol.* **35**, 371–376 (2017).
8. K. Nishida, T. Arazoe, N. Yachie, S. Banno, M. Kakimoto, M. Tabata, M. Mochizuki, A. Miyabe, M. Araki, K. Y. Hara, Z. Shimatani, A. Kondo, T. Arazoe, N. Yachie, S. Banno, M. Kakimoto, M. Tabata, M. Mochizuki, A. Miyabe, M. Araki, K. Y. Hara, Z. Shimatani, A. Kondo, Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* **353**, aaf8729 (2016).
9. H. A. Rees, A. C. Komor, W.-H. Yeh, J. Caetano-Lopes, M. Warman, A. S. B. Edge, D. R. Liu, Improving the DNA specificity and applicability of base editing through protein engineering and protein delivery. *Nat. Commun.* **8**, 15790 (2017).
10. M. M. Jore, M. Lundgren, E. van Duijn, J. B. Bultema, E. R. Westra, S. P. Waghmare, B. Wiedenheft, U. Pul, R. Wurm, R. Wagner, M. R. Beijer, A. Barendregt, K. Zhou, A. P. L. Snijders, M. J. Dickman, J. A. Doudna, E. J. Boekema, A. J. R. Heck, J. van der Oost, S. J. J. Brouns, Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat. Struct. Mol. Biol.* **18**, 529–536 (2011).
11. L. H. Pearl, Structure and function in the uracil-DNA glycosylase superfamily. *Mutat. Res.* **460**, 165–181 (2000).
12. G. T. Hess, L. Frésard, K. Han, C. H. Lee, A. Li, K. A. Cimprich, S. B. Montgomery, M. C. Bassik, Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. *Nat. Methods* **13**, 1036–1042 (2016).
13. K. Kim, S.-M. Ryu, S.-T. Kim, G. Baek, D. Kim, K. Lim, E. Chung, S. Kim, J.-S. Kim, Highly efficient RNA-guided base editing in mouse embryos. *Nat. Biotechnol.* **35**, 435–437 (2017).
14. C. Kuscu, M. Parlak, T. Tufan, J. Yang, K. Szlachta, X. Wei, R. Mammadov, M. Adli, CRISPR-STOP: Gene silencing through base-editing-induced nonsense mutations. *Nat. Methods* **14**, 710–712 (2017).
15. Y. Lu, J. K. Zhu, Precise editing of a target base in the rice genome using a modified CRISPR/Cas9 system. *Mol. Plant* **10**, 523–525 (2017).
16. Y. Ma, J. Zhang, W. Yin, Z. Zhang, Y. Song, X. Chang, Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. *Nat. Methods* **13**, 1029–1035 (2016).
17. L. Yang, A. W. Briggs, W. L. Chew, P. Mali, M. Guell, J. Aach, D. B. Goodman, D. Cox, Y. Kan, E. Lesha, V. Soundararajan, F. Zhang, G. Church, Engineering and optimising deaminase fusions for genome editing. *Nat. Commun.* **7**, 13330 (2016).
18. Y. Zong, Y. Wang, C. Li, R. Zhang, K. Chen, Y. Ran, J.-L. Qiu, D. Wang, C. Gao, Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion. *Nat. Biotechnol.* **35**, 438–440 (2017).
19. Z. Shimatani, S. Kashojiya, M. Takayama, R. Terada, T. Arazoe, H. Ishii, H. Teramura, T. Yamamoto, H. Komatsu, K. Miura, H. Ezura, K. Nishida, T. Ariizumi, A. Kondo, Targeted base editing in rice and tomato using a CRISPR-Cas9 cytidine deaminase fusion. *Nat. Biotechnol.* **35**, 441–443 (2017).
20. W. Tang, J. H. Hu, D. R. Liu, Aptazyme-embedded guide RNAs enable ligand-responsive genome editing and transcriptional activation. *Nat. Commun.* **8**, 15939 (2017).
21. G. Saraconi, F. Severi, C. Sala, G. Mattiuz, S. G. Conticello, The RNA editing enzyme APOBEC1 induces somatic mutations and a compatible mutational signature is present in esophageal adenocarcinomas. *Genome Biol.* **15**, 417 (2014).
22. R. M. Kohli, R. W. Maul, A. F. Guminski, R. L. McClure, K. S. Gajula, H. Saribasak, M. A. McMahon, R. F. Siliciano, P. J. Gearhart, J. T. Stivers, Local sequence targeting in the AID/APOBEC family differentially impacts retroviral restriction and antibody diversification. *J. Biol. Chem.* **285**, 40956–40964 (2010).
23. L. Chelico, P. Pham, M. F. Goodman, Stochastic properties of processive cytidine DNA deaminases AID and APOBEC3G. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **364**, 583–593 (2009).
24. E. A. Kouzminova, A. Kuzminov, Patterns of chromosomal fragmentation due to uracil-DNA incorporation reveal a novel mechanism of replication-dependent double-stranded breaks. *Mol. Microbiol.* **68**, 202–215 (2008).
25. F. A. Ran, L. Cong, W. X. Yan, D. A. Scott, J. S. Gootenberg, A. J. Kriz, B. Zetsche, O. Shalem, X. Wu, K. S. Makarova, E. Koonin, P. A. Sharp, F. Zhang, In vivo genome editing using *Staphylococcus aureus* Cas9. *Nature* **520**, 186–191 (2015).
26. F. d'Adda di Fagagna, G. R. Weller, A. J. Doherty, S. P. Jackson, The Gam protein of bacteriophage Mu is an orthologue of eukaryotic Ku. *EMBO Rep.* **4**, 47–52 (2003).
27. C. Shee, B. D. Cox, F. Gu, E. M. Luengas, M. C. Joshi, L.-Y. Chiu, D. Magnan, J. A. Halliday, R. L. Frisch, J. L Gibson, R. B. Nehring, H. G. Do, M. Hernandez, L. Li, C. Herman, P. J. Hastings, D. Bates, R. S. Harris, K. M. Mille, S. M. Rosenberg, Engineered proteins detect spontaneous DNA breakage in human and bacterial cells. *eLife* **2**, e01222 (2013).

Acknowledgments

Funding: This work was supported by the Defense Advanced Research Projects Agency (HR0011-17-2-0049), the NIH (R01 EB022376 and R35 GM118062), the F-Prime Biomedical Research Initiative (A28161), and the Howard Hughes Medical Institute. A.C.K. is a Ruth L. Kirchstein National Research Service Awards Postdoctoral Fellow (F32 GM 112366-2). M.S.P. was an NSF Graduate Research Fellow and was supported by the Harvard Biophysics NIH Training Grant T32 GM008313. L.W.K. is an NSF Graduate Research Fellow and was supported by the Harvard Chemical Biology Program NIH Training Grant T32 GM095450. **Author contributions:** A.C.K. designed the research, performed experiments, analyzed data, and wrote the manuscript. K.T.Z. designed the research, performed experiments, and analyzed data. M.S.P. assisted with the data analysis. N.M.G. and A.L.W. assisted with the preparation of materials and execution of experiments, and L.W.K. and Y.B.K. assisted with the preparation of materials. A.H.B. provided intellectual input. D.R.L. designed and supervised the research and wrote the manuscript. All authors contributed to editing the manuscript. **Competing interests:** A.C.K., K.T.Z., M.S.P., A.L.W., L.W.K., Y.B.K., and D.R.L. have filed provisional patent applications on base editing through Harvard University. D.R.L. is a consultant and co-founder of Editas Medicine and Beam Therapeutics, companies that are developing genome editing therapeutics. M.S.P. is now a full-time employee of Beam Therapeutics. **Data and materials availability:** Plasmids encoding CDA1-BE3 (100804), AID-BE3 (100803), BE4 (100802), SaBE4 (100805), BE3-Gam (100807), BE4-Gam (100806), SaBE3-Gam (100810), and SaBE4-Gam (100809) are available from Addgene. High-throughput sequencing data have been deposited in the National Center for Biotechnology Information Sequence Read Archive database under accession code PRJNA397048. Correspondence and requests for materials should be addressed to D.R.L.

Submitted 25 July 2017
Accepted 4 August 2017
Published 30 August 2017
10.1126/sciadv.aao4774

Citation: A. C. Komor, K. T. Zhao, M. S. Packer, N. M. Gaudelli, A. L. Waterbury, L. W. Koblan, Y. B. Kim, A. H. Badran, D. R. Liu, Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity. *Sci. Adv.* **3**, eaao4774 (2017).

Science Advances

Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity

Alexis C. Komor, Kevin T. Zhao, Michael S. Packer, Nicole M. Gaudelli, Amanda L. Waterbury, Luke W. Koblan, Y. Bill Kim, Ahmed H. Badran and David R. Liu

Sci Adv **3** (8), eaao4774.
DOI: 10.1126/sciadv.aao4774