








Improved cytosine base editors generated from TadA variants

Received: 16 August 2022

Accepted: 9 November 2022

Published online: 9 January 2023

 Check for updates

Dieter K. Lam ^{1,2}, Patricia R. Feliciano ^{1,2}, Amena Arif ¹, Tanggis Bohnuud¹, Thomas P. Fernandez¹, Jason M. Gehrke¹, Phil Grayson ¹, Kin D. Lee¹, Manuel A. Ortega¹, Courtney Sawyer¹, Noah D. Schwaegerle ¹, Leila Peraro¹, Lauren Young¹, Seung-Joo Lee¹, Giuseppe Ciaramella ¹ & Nicole M. Gaudelli ¹✉

Cytosine base editors (CBEs) enable programmable genomic C-G-to-T·A transition mutations and typically comprise a modified CRISPR–Cas enzyme, a naturally occurring cytidine deaminase, and an inhibitor of uracil repair. Previous studies have shown that CBEs utilizing naturally occurring cytidine deaminases may cause unguided, genome-wide cytosine deamination. While improved CBEs that decrease stochastic genome-wide off-targets have subsequently been reported, these editors can suffer from suboptimal on-target performance. Here, we report the generation and characterization of CBEs that use engineered variants of TadA (CBE-T) that enable high on-target C-G to T·A across a sequence-diverse set of genomic loci, demonstrate robust activity in primary cells and cause no detectable elevation in genome-wide mutation. Additionally, we report cytosine and adenine base editors (CABEs) catalyzing both A-to-I and C-to-U editing (CABE-Ts). Together with ABEs, CBE-Ts and CABE-Ts enable the programmable installation of all transition mutations using laboratory-evolved TadA variants with improved properties relative to previously reported CBEs.

Cytosine base editors (CBEs) are gene-editing enzymes capable of programmably introducing C-G-to-T·A base pair changes in the genomic DNA of living cells. This chemical conversion is achieved through enzyme-mediated hydrolytic deamination of cytosine to uracil, which is interpreted as thymine by DNA polymerases¹. To date, CBEs are typically composed of four distinct components: a naturally occurring cytidine deaminase (such as APOBEC, AID or CDA)², an impaired form of Cas9 capable of nicking the non-base-edited strand of DNA, one or more units of uracil glycosylase inhibitor (UGI) peptide and a nuclear localization sequence (NLS)^{1–3}. These components are typically covalently fused but may also be noncovalently assembled⁴. CBEs have been widely exploited for gene reversion and cellular engineering and have the potential to provide therapeutic benefits to patients living with debilitating genetic diseases or malignancies⁵.

Although high on-target DNA editing efficiency can be achieved with current CBE base editing tools², they can also cause genome-wide, stochastic, guide RNA (gRNA)-independent off-target editing^{6,7}. Next-generation CBE editors such as YE1 (ref. 8), BE4-PpAPOBEC⁹ and others¹⁰ have been reported to mitigate gRNA-independent off-target outcomes, but these editors use natural or lightly engineered variants of APOBEC deaminase and may suffer from decreased on-target editing performance^{2,9}. Additionally, in some sequence-specific contexts, APOBEC-based CBEs may lead to proximal editing adjacent to the targeted genomic sequence due to APOBEC's inefficient, but measurable, ability to accept double-stranded DNA (dsDNA) as a substrate¹¹.

Adenine base editors (ABEs) are gene-editing enzymes that programmably install A·T to G·C point mutations at targeted loci via a laboratory-evolved TadA deaminase that chemically converts adenine to inosine¹². Inosine base pairs with cytosine within the active site of

¹Beam Therapeutics, Cambridge, MA, USA. ²These authors contributed equally: Dieter K. Lam, Patricia R. Feliciano. ✉ e-mail: ngaudelli@beamtx.com

DNA polymerases resulting in an inosine to guanine mutation following DNA replication. Notably, ABEs cause low to no gRNA-independent off-targets and edit genomic DNA (gDNA) within a more precise window (positions -3–8, PAM 21–23), which may result in fewer guide-dependent off-targets as well as fewer bystander edits, relative to CBEs^{6,13}. Additionally, ABEs have not been reported to act on dsDNA.

To confer the favorable attributes of ABEs upon a CBE, we envisioned transforming TadA into an enzyme capable of robust cytidine deamination and subsequently generated an improved class of CBEs that uses TadA instead of a naturally occurring cytidine deaminase. Encouragingly, previous investigations have demonstrated ABEs' malleability toward low, but detectable C to T editing through the inclusion of UGI¹⁴. Indeed, through structure-guided design, an ABE variant has been reported, ABE-P48R-UGI, that enabled enhanced cytosine activity, relative to ABE7.10, but with high TC sequence specificity¹⁵. While these advancements represented progress toward our aims, we recognized that further engineering and evolution of TadA would be required to achieve therapeutically relevant C-G-to-T-A editing efficiencies with high product purity and without substrate sequence restrictions.

Starting with ABE8.20-m¹³ as a template for library generation, we conducted two rounds of directed evolution to generate base editor variants with improved C-G to T-A editing efficiencies and retention of adenine editing. We refer to these cytosine and adenine base editors (CABEs) utilizing TadA as 'CABE-Ts', and further developed and characterized these editors for C-G-to-T-A and A-T-to-G-C editing efficiencies in mammalian cells. With CABE-Ts in hand, we determined crystal structures of the TadA deaminase variants associated with these editors and performed structure-guided mutagenesis to create CBE-Ts, a distinct class of CBEs that use engineered TadA deaminases for high C-G to T-A conversion in gDNA with no appreciable levels of A-T to G-C editing. Relative to BE4, our CBE-Ts demonstrated comparable on-target editing efficiencies, had a more precise editing window, reduced guide-dependent off-target editing, and showed no detectable gRNA-independent genome-wide off-target editing. Furthermore, CBE-Ts demonstrated compatibility with orthogonal Cas enzymes, allowing for their potential application across a broader range of target sites. Finally, our CBE-Ts were highly active in primary cell types such as T cells and hepatocytes, thus validating their potential as an attractive gene-editing tool for therapeutic applications.

Results

Directed evolution of ABE for C-to-T editing

To alter the nucleobase substrate tolerance of ABE, we reasoned that we could selectively pressure ABE to increase its low, but detectable^{13,14}, C-G-to-T-A base editing capability through directed evolution and inclusion of UGI (to inhibit uracil repair by UNG³). First, we generated an ABE library chemically randomized in the TadA region of the editor (ABE8.20-m¹³ used as a template) or randomized via error-prone PCR (ABE8.19-m¹³ used as a template). The resulting -10-million-member library contained an average of three amino acid substitutions per

member. *Escherichia coli* were co-transformed with the ABE library, gRNA and a selection plasmid and were later challenged with lethal doses of antibiotic that were selected for ABE library members that performed C-G to T-A edits within a corresponding antibiotic resistance, restoring gene function (Fig. 1a–d and Supplementary Sequence 1). Sanger sequence analysis of surviving library members (Fig. 1e and Supplementary Fig. 1) revealed that the majority of antibiotic-selected clones contained amino acid substitutions at positions 27 and 49 of TadA. Because 19 of 20 variants contain at least one substitution in either position, we hypothesized that substitutions at these positions (E27H and I49K) located near the substrate binding pocket would induce conformational changes rendering TadA capable of binding and deaminating the cytosine nucleobase, which is notably smaller in size than adenine.

Of the surviving library members, 20 variants were characterized in mammalian cells for base editing outcomes and many variants identified from the first round of evolution demonstrated appreciable levels of C-G to T-A editing (for example, CABE-T1.2, avg 32.1%; CABE-T1.17, avg 34.9% across 22 genomic sites), with varying degrees of A-T to G-C editing retained (Fig. 1f and Supplementary Figs. 2–4). Architecturally, these base editors are comprised of a TadA variant covalently fused to the N-terminal end of a Cas9 nickase (nCas9, D10A) followed by two C-terminal UGI units and a nuclear localization tag (Fig. 1d). Accordingly, we refer to these dual A-T to G-C and C-G to T-A editors as CABE-T1s (Fig. 1c). CABE-T1s elicited an average of >25-fold C-G to T-A editing increase over ABE8.20-m at genomic sites tested (Fig. 1f and Supplementary Figs. 2–4).

To further increase the overall editing efficiency of CABE-T1, we created an -10-million-member CABE-T library on the background sequence of CABE-T1.2, a CABE-T1 that demonstrated robust C-G to T-A editing in mammalian cells (Fig. 1f). We required CABE-T1.2 library members to create two C-G to T-A reversions, in addition to two A-T to G-C edits for increased stringency in the selection, to survive antibiotic exposure at higher concentrations than in the previous round of directed evolution (Fig. 1b and Supplementary Sequence 2). The surviving library members, referred to as CABE-T2 variants, were sequence identified and evaluated for base editing efficiency and nucleobase substrate bias in mammalian cells (Fig. 1e,f and Supplementary Fig. 5). Overall, mammalian transfection experiments revealed an improvement in our CABE-T2s over CABE-T1s. For example, representative variants CABE-T2.6, CABE-T2.9 and CABE-T2.19 were able to achieve average maximum C-G to T-A editing rates of 53.0%, 53.6% and 49.4% across 22 genomic sites, respectively, while also maintaining various levels A-T to G-C editing (Fig. 1f and Supplementary Figs. 2, 3 and 6).

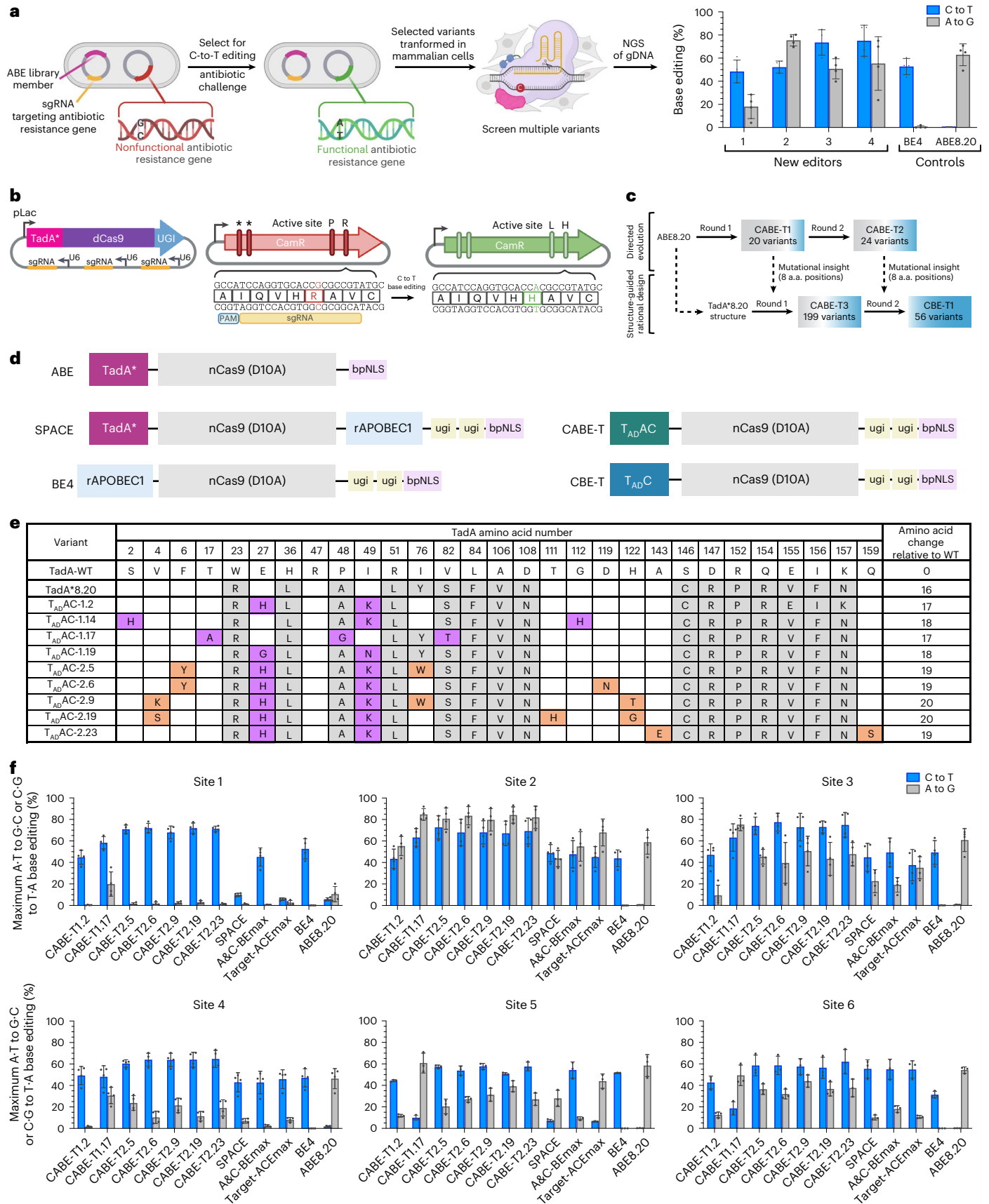
While CABEs have previously been reported in the literature, these tools have required the inclusion of both TadA*7.10 and rAPOBEC1 deaminases to enable adenine and cytosine base editing from a single full-length editor^{16–19}. The creation of CABE-Ts that use one TadA variant that acts on both DNA adenines and cytosines (T_{AD}AC) resulted in the

Fig. 1 | Directed evolution of CABE-T1 and CABE-T2. **a**, Schematic overview of directed evolution workflow to identify TadA variants capable of C-to-U deamination. **b**, Schematic representation of expression and selection plasmids used in directed evolution campaigns. Left: base editor expression plasmid encoding library member and sgRNA, right: selection plasmid encoding a nonfunctional antibiotic resistance gene. Reversion at targeted sites restores gene function. **c**, Overview of CABE-T editor development through directed evolution and protein engineering. **d**, Schematic representation of select base editor architectures. ABE¹² contains a laboratory-evolved TadA* deaminase, nCas9 (D10A) and a nuclear localization tag (bpNLS). Dual editor (for example, SPACE¹⁶) is comprised of an evolved TadA* deaminase, nCas9 (D10A), cytidine deaminase rAPOBEC1 along with two units of UGI and a bpNLS tag. CBEs (for example, BE4 (ref. 3)) are comprised of a naturally occurring cytidine deaminase (for example, rAPOBEC1), nCas9 D10A, two units of UGI and a bpNLS tag. CABE-Ts,

reported here, are comprised of a TadA variant capable of A-to-G and C-to-T editing (T_{AD}AC), nCas9 (D10A), two units of UGI and a bpNLS tag. CBE-Ts, reported here, are comprised of a TadA variant capable of C-to-T editing (T_{AD}C), nCas9 (D10A), two units of UGI and a bpNLS tag. **e**, Substitutions incorporated into TadA in selected CABE-T editors from directed evolution round 1 (CABE-T1) and round 2 (CABE-T2). Substitutions incorporated in TadA*8.20 related to wild-type (WT) TadA are highlighted in gray, and those identified in directed evolution round 1 and 2 are highlighted in violet and orange, respectively. Values in the last column represent the number of substitutions added to each variant compared to WT TadA. **f**, Maximum C-G to T-A and A-T to G-C conversion in HEK293T cells transfected with human expression plasmids encoding CABE-T1 and CABE-T2 editors or controls. Values and error bars reflect the mean and s.d. of $n = 3$ (sites 1–4) or 4 (sites 5 and 6) independent biological replicates performed on different days.

generation of a more compact base editor (~700 bp smaller) with superior dual base editing outcomes relative to previously described CABEs. For instance, CABE-T2.6 demonstrated -1.6-fold higher maximum C-G

to T·A and ~2.6-fold higher maximum A·T to G·C relative to SPACE¹⁶, A&C-BE_{max}¹⁷ and TargetACE_{max}¹⁸ (Fig. 1f and Supplementary Figs. 2 and 3).



editing efficiency and substrate specificity through structure-guided library design.

Following the first round of directed evolution, we structurally characterized three deaminases corresponding to CBE-T1 ($T_{AD}AC$ -1.17, $T_{AD}AC$ -1.14 and $T_{AD}AC$ -1.19) that generated appreciable levels of C·G to T·A base editing. Although the overall structures of these variants are similar to that of *TadA**8.20 (Fig. 2a), structural analyses revealed local structural changes that may explain the observed expanded substrate tolerance exhibited by CBE-T1 variants.

Crystal structures of *TadA**8.20 and $T_{AD}AC$ -1.17 were determined in a complex with ssDNA substrate containing the adenine transition-state analog 2-deoxy-8-azanebularine (d8Az) (Fig. 2, Extended Data Figs. 1 and 2 and Supplementary Figs. 7–10). These two structures are highly similar, and the four $T_{AD}AC$ -1.17 substitutions (T17A, A48G, S82T and A142E) derived from evolution do not drastically alter substrate tolerance by changing the protein structure or the ssDNA binding mode (Fig. 2a,b). These findings correlate with $T_{AD}AC$ -1.17's relatively low C·G to T·A reversions at genomic site 5 compared to other variants in CBE-T1 (Supplementary Fig. 4). We hypothesize that the T82 side chain near the catalytic E59 residue (~ 4 Å) in the active site may have a role in increasing cytosine deamination by modulating proton transfer to or from E59 (Fig. 2c). Additionally, the hydrogen bonding between E142 and R153 may modulate ssDNA binding by stabilizing the $\alpha 6$ -helix, as exemplified by the interactions between F156 and dT(8) in the $T_{AD}AC$ -1.17 structure (Extended Data Fig. 2d).

Notably, the crystal structure of $T_{AD}AC$ -1.14 containing four substitutions (S2H, I49K, Y76I and G112H) reveals a structural difference in the loop between strands $\beta 4$ and $\beta 5$ (R107 to V130) on the right side of the active site cavity (Fig. 2a,d, Extended Data Fig. 3 and Supplementary Fig. 11). This loop contains a G112H substitution that dramatically alters its flexibility and conformation relative to *TadA**8.20 by introducing a bulky positively charged residue (Extended Data Fig. 3d). These structural changes may reshape the $T_{AD}AC$ -1.14 active site cavity to accommodate both adenines and cytosines (Fig. 2b and Extended Data Fig. 3f). Indeed, a comparison with the structure of *TadA**8.20 bound to ssDNA substrate shows that $T_{AD}AC$ -1.14 may engage ssDNA differently than *TadA* variants with strict adenine specificity (Fig. 2b and Extended Data Fig. 3d). We hypothesize that residue K49 within $T_{AD}AC$ -1.14 may contribute to the stabilization of protein–DNA interactions required for binding cytosine-containing ssDNA substrates due to its repositioning near nucleobase dC(10) (~ 4.5 Å) (Fig. 2d and Extended Data Fig. 3e).

In addition to the perturbations on the right side of the $T_{AD}AC$ -1.14 active site cavity, evaluation of the $T_{AD}AC$ -1.19 structure reveals that other substitutions (E27G and I49N) from our evolution caused major structural changes on the left side of the active site cavity (Fig. 2a,b, Extended Data Fig. 4 and Supplementary Fig. 12). These structural changes are likely caused by the E27G substitution, which results in the loss of essential hydrogen bonds between E27 and A48, I49 and G50. Because of these hydrogen bond losses, a conformational change occurred that reoriented residue E25 of $T_{AD}AC$ -1.19 to a similar position that was formerly occupied by residue E27 in *TadA**8.20 (Fig. 2e and Extended Data Fig. 4e). The displacement of E25 shortens the $\alpha 1$ -helix, changes the length and conformation of the loop between $\alpha 1$ and $\beta 1$ (D24 to P29) containing the E27G substitution and unfolds $\alpha 5$ - and $\alpha 6$ -helices (Fig. 2a,e and Extended Data Fig. 4), reshaping the $T_{AD}AC$ -1.19 active site cavity and potentially impacting target nucleobase binding within the active site (Fig. 2b).

Structure-guided design of CBE-T3s and CBE-Ts

Informed by the crystal structures described here (Fig. 2), we speculated that structural changes induced by substitutions in one of three distinct regions of *TadA**8.20 (E27G, S82T and G112H) were sufficient to alter substrate tolerance toward a cytosine (Fig. 3b). Thus, we hypothesized that combining amino acid substitutions from all three regions would yield a synergistic improvement in enhancing C·G to T·A editing.

To test this hypothesis, eight sites within the deaminase of CBE-T1 were selected for library construction, including substitutions at positions 27, 49, 82, 112 and 142 discussed above, plus two rationally selected sites at or near the active site of *TadA**8.20 (Fig. 3a,b) to generate the first combinatorial library containing 199 variants (CBE-T3). Each library member had 2–10 amino acid substitutions (~ 5.3 on average) in the deaminase of CBE-T3, and most library members encoded at least one amino acid substitution in all three of these regions (Fig. 3b,c and Supplementary Figs. 13 and 14). We identified several CBE-T3 variants, notably CBE-T3.1 and CBE-T3.155, that demonstrate dual C·G-to-T·A and A·T-to-G·C base editing activity at levels comparable to or higher than those from CBE-T1 or CBE-T2 (Fig. 3d and Supplementary Figs. 2, 3 and 15). Notably, by screening library members directly in mammalian cells for relative base editing activity, we were able to identify editors with a broad range of C·G to T·A and A·T to G·C editing ratios, including several variants (for example, CBE-T3.55, T3.153 and T3.154) capable of robust in C·G to T·A editing with minimal A·T to G·C editing (Supplementary Fig. 15).

Concordantly, to further increase overall C·G-to-T·A editing efficiency and optimize substrate specificity toward cytosine, we took CBE-T3.154, a base editor showing a strong preference for C·G to T·A editing (Supplementary Figs. 14 and 15) and combinatorially layered eight additional substitutions selected from the deaminases of CBE-T2s (Fig. 3 and Supplementary Figs. 5 and 6). These substitutions are located proximal to the DNA-binding pocket of the deaminase and their inclusion in CBE-T2s caused an overall increase in editing efficiency compared to CBE-T1s. We generated a 56-member library (Supplementary Fig. 16), screened them in mammalian cells via plasmid transfection, and found that all 56 variants achieved substantial C·G to T·A editing (69.2% averaged across all variants) but caused only low to undetectable levels of A·T to G·C editing (1.8% averaged across all variants at all sites tested; Fig. 3d and Supplementary Figs. 17 and 18). Therefore, we designate these CBEs containing *TadAs* acting on DNA cytosines ($T_{AD}C$) as CBE-Ts (Fig. 1d).

After observing the robust activity of our CBE-Ts in Hek293Ts, we were curious to evaluate how cytosine base editing outcomes of a representative subset of our 56 CBE-Ts compared to the previously published ABE-P48R-UGI¹⁵ editor at six genomic sites given the high degree of amino acid substitution per variant that was required to access our editors (Fig. 1e and Supplementary Fig. 16). Indeed, we found that our CBE-Ts greatly outperformed ABE-P48R-UGI in C·G to T·A editing efficiency, relative cytosine to adenine base editing product purity and substrate tolerance (Supplementary Fig. 19). Relative to ABE-P48R-UGI, CBE-Ts were not restricted to TC motifs, a limitation of the ABE-P48R-UGI editor, and therefore, we envision the CBE-Ts reported here to be more universally applicable (Supplementary Fig. 19).

To determine whether the $T_{AD}Cs$ present in our CBE-Ts were compatible with orthogonal Cas enzymes, we screened the base editing activity of a representative subset of our CBE-Ts, replacing the *Streptococcus pyogenes* D10A Cas9 nickase with *Staphylococcus aureus* Cas9 nickase (SaCas9, PAM: NGGRRRT), shown previously to have compatibility with ABE editors in mammalian cells^{13,20}. Indeed, we observed that $T_{AD}Cs$ are modular enzymes and are compatible with SaCas9 but elicit only modest C·G-to-T·A editing efficiencies, similar to BE4-SaCas9 variants, across the six genomic sites tested (Supplementary Fig. 20).

On-target characterization of CBE-Ts and CBE-Ts

To more deeply characterize CBE-Ts and CBE-Ts, we chemically synthesized gRNAs and in vitro transcribed (IVT) mRNAs encoding a representative subset of CBE-T and CBE-T editors and transfected them into HEK293T cells at both saturating and subsaturating doses of mRNA encoding the editor (Fig. 4a,b). For the CBE-T2s and T3s tested, we observed an average of 1.53-fold and 1.03-fold increase in maximum C·G to T·A editing and an average of 2.18-fold and 1.67-fold improvement in A·T to G·C editing relative to SPACE and A&C-BE_{max}, respectively

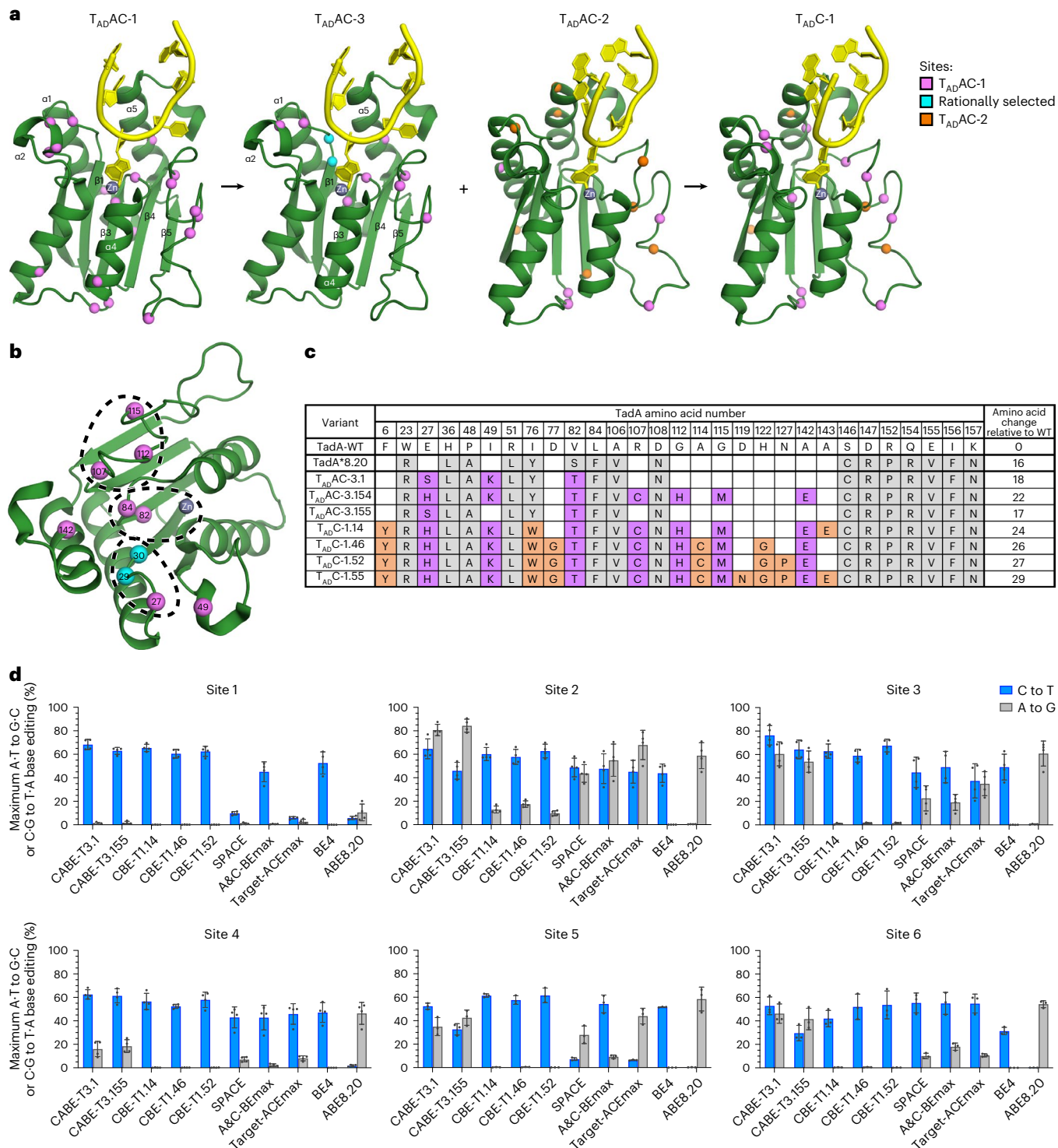


Fig. 3 | Structure-guided combinatorial screens. a, Structure-guided combinatorial library design workflow. For clarity, only one Tada*8.20 monomer (green) with ssDNA (yellow) is shown. The 23 substitution sites identified in the first round of evolution are shown as violet spheres in the first panel. Based on the structures of T_{AD}AC-1 variants (Fig. 2), 10 sites were selected, 8 from T_{AD}AC-1 (violet) plus 2 rationally selected (cyan), in the loop between α 1 and β 1, α 2-helix, active site, the loop between β 4 and β 5 and α 5-helix, to generate the T_{AD}AC-3 variants (second panel). The substitution sites in T_{AD}AC-2 from the second round of evolution are shown as violet (same sites identified in T_{AD}AC-1) and orange (new sites) spheres in the Tada*8.20 structure (third panel). To generate the T_{AD}C-1 variants, eight substitutions from the second round of evolution were added to T_{AD}C-3.154 (fourth panel). **b**, Another view of the Tada*8.20 structure (green) showing the substitution sites selected to generate T_{AD}AC-3 (violet

and cyan spheres). The dashed circles highlight the three regions that we hypothesized to be critical for altering substrate tolerance. **c**, Substitutions incorporated into selected TadAs from structure-guided screen round 1 (CABE-T3 editors; T_{AD}AC-3 deaminases) and round 2 (CBE-T1 editors; T_{AD}C-1 deaminases). Substitutions incorporated in TadA*8.20 relative to WT TadA are highlighted in gray, and those identified in directed evolution campaigns 1 and 2 are highlighted in violet and orange, respectively. Values in the last column represent the number of substitutions added to each variant compared to WT TadA. **d**, Maximum C-G to T-A and A-T to G-C conversion at targeted genomic loci in HEK293T cells transfected with human expression plasmids encoding CABE-T3 and CBE-T1 base editors or controls. Values and error bars reflect the mean and s.d. of $n = 3$ (sites 1–4) or 4 (sites 5 and 6) independent biological replicates performed on different days.

(Supplementary Fig. 21). Across all sites tested, we observed no significant difference in maximum editing outcomes for our characterized CBE-Ts relative to BE4 ($P = 0.30$, two-tailed Wilcoxon–Mann–Whitney U test) and remarkable differentiation from editing outcomes relative to the parent editor ABE8.20. Across all sites tested, our CBE-Ts resulted in an average 262-fold increase in C-G to T-A editing and a concordant 13-fold decrease in A-T to G-C editing relative to ABE8.20 across the editing window (Fig. 4a–c and Supplementary Figs. 2 and 3).

To confirm that our CBE-Ts and CBE-Ts proceeded through a C-to-U deamination mechanism, we employed an in vitro end-point deamination assay to evaluate a subset of editors as gRNA-programmed ribonucleoprotein (RNP) complexes acting on dsDNA substrate. In this assay, CBE-Ts and CBE-Ts resulted in an average of ~30% C-to-U substrate deamination after 24 h at the on-target site, compared to ~58% for BE4, with no detectable A-to-I deamination for the CBE-Ts evaluated (Extended Data Fig. 5a). In addition to C-to-U substrate deamination, CBE-Ts also produced up to 35% on-target A-to-I deamination. Altogether, these data provide orthogonal biochemical support for the C-to-U deamination activity of our CBEs and CBEs utilizing TadA variants. In a separate experiment, the C-to-U apparent deamination rate constant (k_{app} , also referred to as rate) of CBE-T1.14 RNP on dsDNA substrate was measured to be $0.014 \pm 0.006 \text{ min}^{-1}$, much slower than the rate of A-to-I deamination for ABE8.20 RNP on dsDNA ($0.17 \pm 0.06 \text{ min}^{-1}$; Extended Data Fig. 5b), while their nicking rate for the nontarget strand remained nearly identical (Extended Data Fig. 5b and Supplementary Data Figs. 22 and 23).

Despite having a slower deamination rate in vitro, CBE-Ts and BE4 produced comparable total deamination of target sites in cellular transfections conducted over 5 days (Fig. 4a,b). We hypothesized that, given enough time, total C-G to T-A editing or C-to-U deamination by CBE-Ts would reach levels comparable to the kinetically faster BE4. In agreement with this observation, extending deamination time to 24 h in vitro led to comparable total C-to-U deamination by CBE-Ts and BE4 (Extended Data Fig. 5a).

We next evaluated how the dose of delivered mRNA affects cytosine base editing outcomes by conducting mammalian cell transfections at subsaturating levels of mRNA (Supplementary Fig. 24). Under these conditions, CBE-Ts retained 55% to 70% maximum editing efficiency compared to saturating conditions and performed similarly or better relative to APOBEC-based CBEs on a per-site basis. CBE-T representative editors CBE-T1.14, CBE-T1.46 and CBE-T1.52 achieved average maximum C-G to T-A rates of 66% across eight genomic sites, compared to an average 59% C-G to T-A achieved by BE4, and an average of ~35% C-G to T-A achieved by YE1 (Fig. 4b and Supplementary Fig. 25). We also find that CBE-Ts cause similar levels of indel formation and C- to non-T edits (Supplementary Fig. 26 and Extended data Fig. 6a). Comparable product purity and indel outcomes relative to CBEs utilizing cytidine deaminases is likely due to the mechanisms of genomic uracil lesion repair, which is agnostic to how the lesion was created.

Like ABEs, we show that CBE-Ts and CBE-Ts have a narrower editing window relative to BE4, with base edits restricted roughly to positions 3–8 in the protospacer (Fig. 4c and Supplementary Fig. 27). Additionally, we observe that CBE-Ts, relative to APOBEC-based CBEs, generate fewer bystander mutations because on average fewer Cs exist in the narrowed targetable window of the editor (Fig. 4d and

Supplementary Fig. 27). For therapeutic applications, we note this increase in base editing precision is an attractive feature when considering disease targets. Relatedly, while BE4 has been characterized to act on dsDNA proximal to the protospacer due to APOBEC's low, but detectable tolerances for dsDNA as a substrate, we do not observe this dsDNA editing activity with our CBE-Ts (Extended Data Fig. 6b).

Off-target evaluation of CBE-Ts and CBE-Ts on DNA

To characterize the gRNA-dependent DNA off-target editing of CBE-Ts and CBE-Ts, we performed mRNA transfections in cells with several gRNAs for which the gRNA off-target profile has been previously characterized with Cas9 and base editors^{12,13,21}. We find that CBE-Ts and CBE-Ts have lower gRNA-dependent off-target base editing frequencies at all sites examined relative to BE4 and BE4-PpAPOBEC, with 3.06-fold and 3.53-fold decreases in maximum C-G to T-A editing, respectively, and similar levels relative to that of mitigated off-target editors YE1 and BE4-PpAPOBEC-H122A (Extended Data Fig. 7 and Supplementary Figs. 28 and 29).

To evaluate the ratio of guide-independent base editing caused by CBE-Ts and CBE-Ts to BE4 (for C to T editing) or ABE8.20 (for A to G editing), we performed whole genome sequencing (WGS) of clonally expanded cells treated with mRNA encoding a base editor and quantified the relative C-G to T-A or A-T to G-C mutation rate as described before¹³ (Supplementary Fig. 30). We found that both CBE-Ts and CBE-Ts caused no significant elevation in genome-wide C-G to T-A SNVs relative to untreated samples, a pattern that was also reported for YE1 and BE4-PpAPOBEC-H122A (all $P > 0.05$; one-sided Mann–Whitney U test). In contrast, BE4 caused a mean fold-enrichment of 3.8 times for C-G to T-A edits over control ($P = 7.770e-05$; one-sided Mann–Whitney U test) and BE4-PpAPOBEC caused 1.5 times the mean fold-enrichment of C-G to T-A ($P = 0.00147$; one-sided Mann–Whitney U test) (Fig. 5a)^{6,13}. CBE-Ts and CBE-Ts also did not cause a significant elevation in genomic A-T to G-C SNVs (all $P > 0.05$; one-sided Mann–Whitney U test) (Fig. 5b) and stochastic deamination genome-wide was indistinguishable from untreated cells.

To illuminate the kinetic differences in ssDNA deamination, we measured single turnover, pseudo-first-order apparent deamination rate constants (k_{app}) of base editors lacking a guide RNA on ssDNA substrate in vitro. We measured the rate of C-to-U deamination by BE4 to be $0.78 \pm 0.02 \text{ min}^{-1}$, ~11-fold higher than the rate of A-to-I deamination elicited by ABE8.20 ($k_{app} = 0.071 \pm 0.005 \text{ min}^{-1}$) for the same ssDNA substrate (Fig. 5c and Supplementary Fig. 22). This difference in deamination rate of ssDNA further supports previous observations, and observations reported here, that APOBEC-based CBEs can stochastically deaminate single-stranded regions of the genome. We found that CBE-T1.14 catalyzed C-to-U deamination with k_{app} of $0.060 \pm 0.006 \text{ min}^{-1}$, a rate indistinguishable from that measured for A-to-I deamination by ABE8.20 (Fig. 5c and Supplementary Fig. 22) on the same ssDNA substrate. Notably, the catalytic residue remains unchanged between ABE8.20 and CBE-Ts (Figs. 2c, 3a and Extended Data Figs. 1d, 2, 3c).

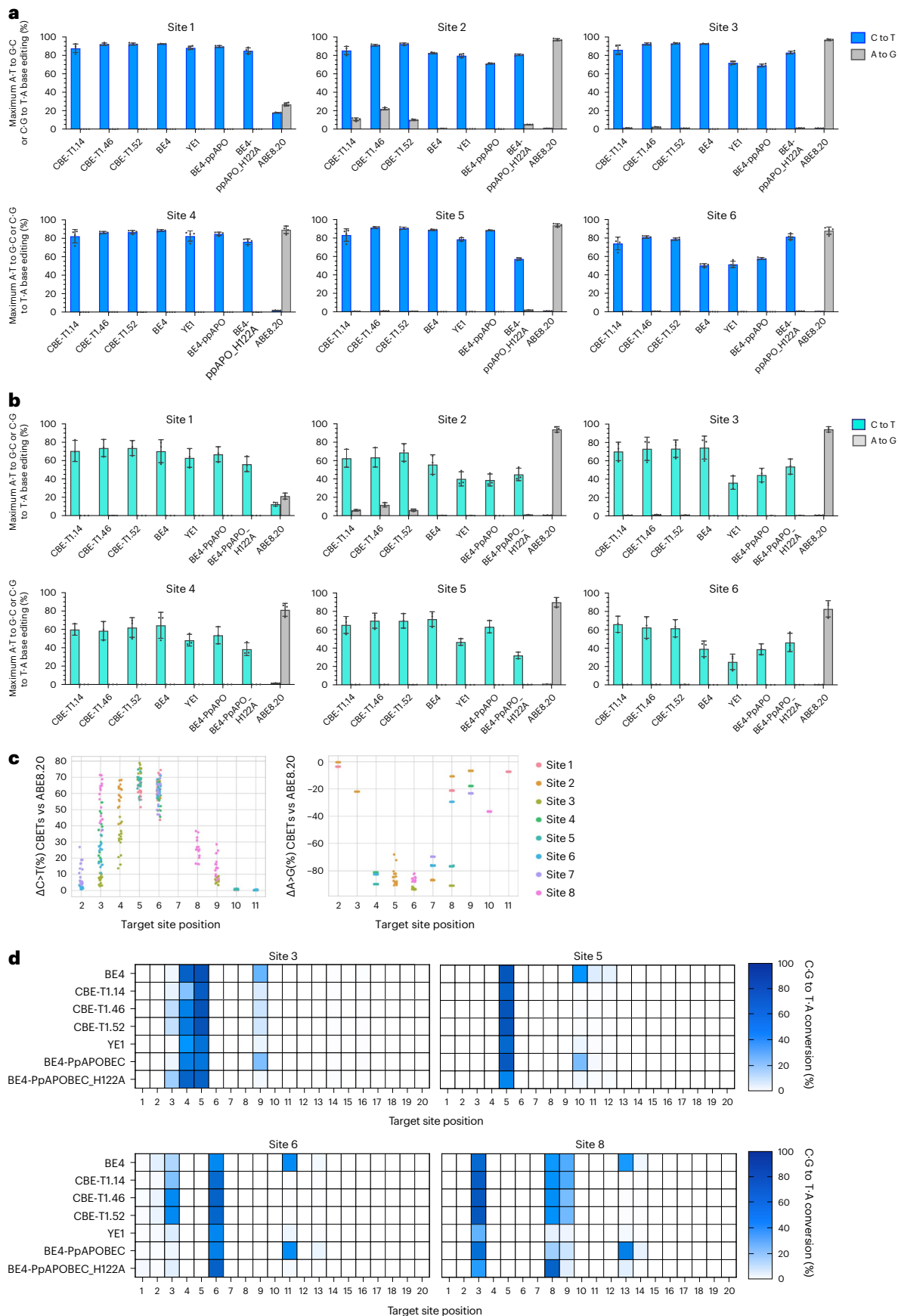
Application of CBE-Ts in primary cells

CBE-Ts have substantial potential for therapeutic use in gene reversion and silencing due to their improved properties relative to CBEs utilizing naturally occurring cytidine deaminases. To evaluate the

Fig. 4 | CBE-Ts elicit robust C-G to T-A conversions in human cells at levels comparable to or higher than BE4 with a narrower editing window.

a, b, Maximum C-G to T-A and A-T to G-C conversion in HEK293T cells transfected with mRNA encoding core CBE-T variants, plus controls, across eight targeted genomic loci via synthetic gRNAs at saturating (500 ng mRNA) (**a**) and subsaturating conditions (62.5 ng construct mRNA + 437.5-ng nontranslated carrier mRNA) (**b**). **c**, Percent change in C-G to T-A (left) and A-T to G-C (right) editing rates between CBE-T1 variants and ABE8.20 at each target site position

(PAM = positions 21–23) across eight genomic sites tested (sites 1 to 8; Supplementary Table 3). **d**, Median C-G to T-A conversion at each target window position as specified on the x-axis, with position numbering defined as the PAM designated as positions 21–23. Values for color maps were determined from mRNA transfections at saturating conditions. Values and error reflect the mean and s.d. of $n = 4$ (saturating conditions) or 3 (subsaturating conditions) independent biological replicates performed on different days.



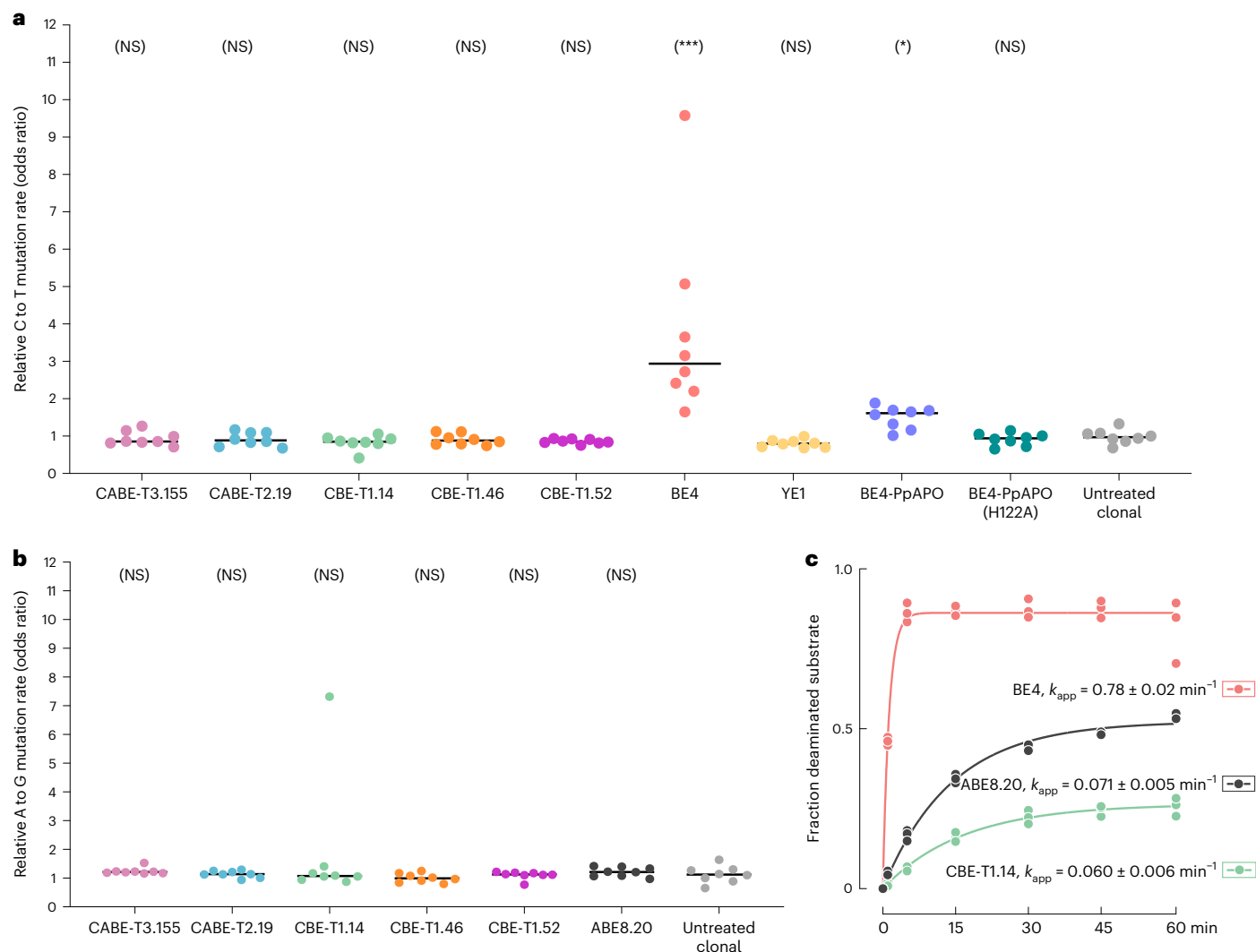


Fig. 5 | Guide-independent off-target evaluation of CBE-T and CBE-Ts.

a, Odds ratio plot for C to T mutations relative to all other mutation types in cells edited with CABE-T3.155, CABE-T2.19, CBE-T1.14, CBE-T1.46 and CBE-T1.52, BE4, YE1, BE4-PpAPO and BE4-PpAPO-H122A compared to untreated clonally expanded cells, with black bars representing the median odds-ratio for that treatment group ($P = 0.8359, 0.7473, 0.9476, 0.8089, 0.9751, 7.770 \times 10^{-5}, 0.9859, 0.00148$ and 0.7473 , respectively; one-sided Mann–Whitney U test). All $n = 8$ biologically independent single-cell expanded cell populations are shown for each condition. **b**, Odds ratio plot for A to G mutations relative to all other mutation types in cells edited with CABE-T3.155, CABE-T2.19, CBE-T1.14,

CBE-T1.46 and CBE-T1.52 and ABE8.20 compared to untreated clonally expanded cells, with black bars representing the median odds-ratio for that treatment group ($P = 0.1641, 0.4796, 0.5204, 0.8607, 0.5204$ and 0.2527 , respectively; one-sided Mann–Whitney U test). All $n = 8$ biologically independent single-cell expanded cell populations are shown for each condition. **c**, In vitro kinetics of A-to-I or C-to-U deamination of the same substrate presented as ssDNA to BE4, ABE8.20 and CBE-T1.14 in the absence of gRNA. Pseudo first-order apparent rate constants (k_{app}) obtained by fitting to a single exponential fit are reported (mean \pm s.d., $n = 3$ independent replicates). See gel source data.

editing potential of CBE-Ts in primary cells, we first assessed the ability of CBE-Ts to silence the expression of *PCSK9*, a target relevant to the therapeutic treatment of hypercholesterolemia²², in a long-lived primary human hepatocyte coculture system²³. Knock-down or knock-out of *PCSK9* gene results in lower levels of low-density lipoprotein (LDL) cholesterol in the blood and subsequently lowers the risk of heart disease^{24,25}. Indeed, promising results with splice-site targeting of *PCSK9* have been achieved with ABE8.8 in vivo²⁶. Similarly, we found that mRNA transfection of CBE-T1.46 with synthetic guide targeting the *PCSK9* gene in primary human hepatocytes achieved C-G to T-A base editing efficiencies that are comparable to or greater than BE4 at two *PCSK9* target sites that introduce the stop codon Q555X or disrupt pre-mRNA splicing at exon 4 (E4 splice; Fig. 6a). Evaluation of *PCSK9* protein levels via ELISA shows knockdown of *PCSK9* by CBE-T1.46 at levels comparable to or greater than BE4 at both target sites tested

(Fig. 6b). Concordantly, increases in total LDL receptor (LDLR) were observed for the CBE-T1.46 treated samples at both sites ($P < 0.05, < 0.01$), demonstrating the potential for CBE-T to generate a therapeutically relevant phenotypic effect (Fig. 6c).

We next evaluated the application of CBE-Ts to therapeutic T cell engineering. Autologous T cell therapies derived from TCR $\alpha\beta$ -expressing T cells are effective in treating some cancers, although the manufacture of these cell therapies on a per-patient basis can result in inconsistent products, high cost of goods and significant delays in patient treatment. Gene editing can be used to create universally compatible T cell therapies, generated from single donors for the treatment of many patients²⁷. Universally compatible T cell therapies require multigene silencing to eliminate expression of the T cell receptor to reduce the potential for graft-versus-host-disease (GvHD), and editing strategies to reduce or eliminate host rejection of the allogeneic

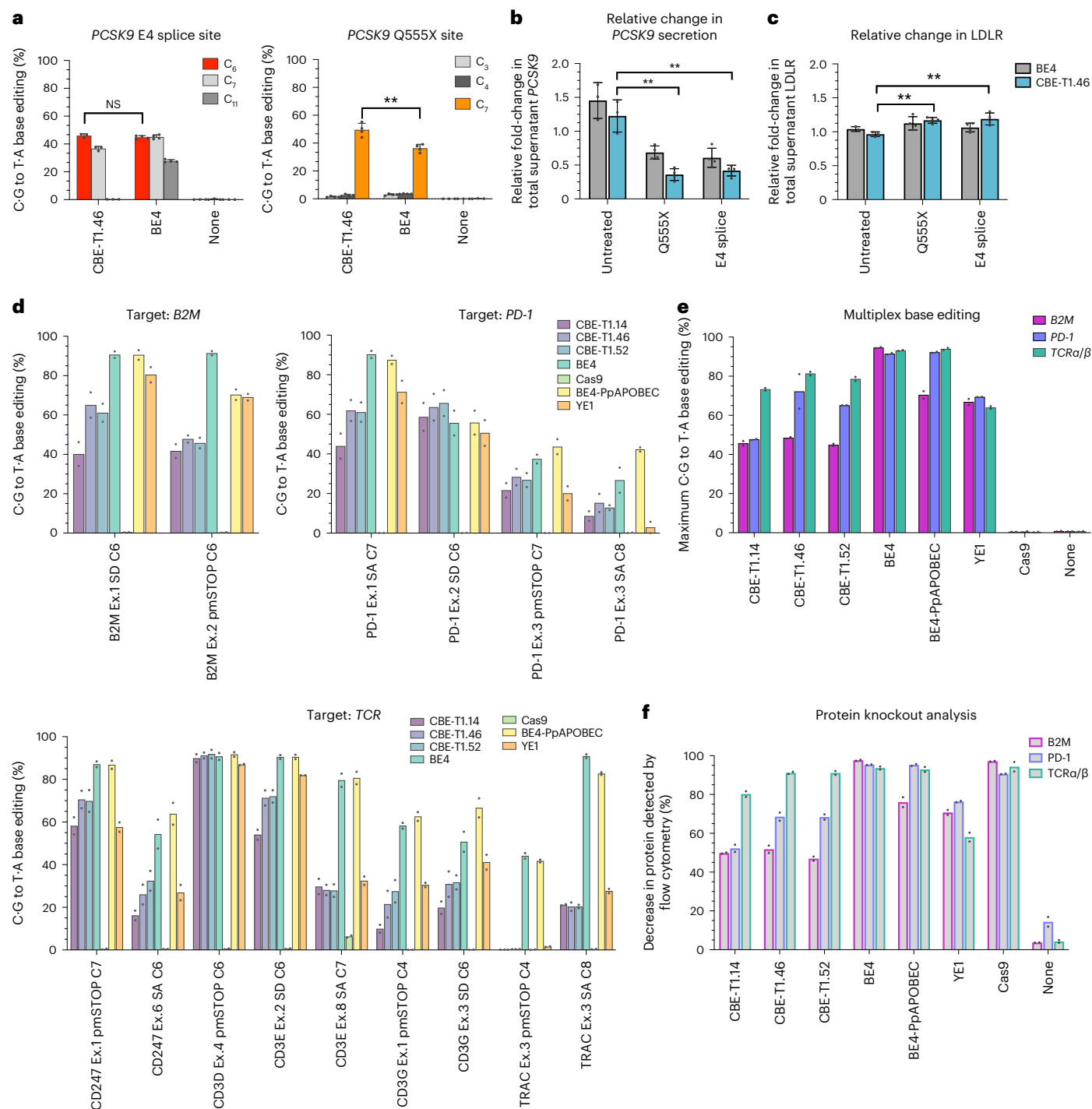


Fig. 6 | Evaluation of CBE-Ts and CBE-Ts in therapeutically relevant cell contexts. **a**, Gene-editing outcomes of primary human hepatocytes transfected with mRNAs encoding CBE-T1.46 and BE4 at three sites within the *PCSK9* gene (left, $P = 0.2882$ (NS); right, $P = 0.0017$ (double asterisk)). Positional edit within the propeptide indicated. **b**, Evaluation of relative change in *PCSK9* secretion between day 9 (collection time point) and day 0 (transfection time point) through ELISA. The targeted *PCSK9* site is indicated on the x axis. P values are as follows: Q555X versus untreated, $P = 0.001001$ (double asterisk); E4 splice versus untreated, $P = 0.001295$ (double asterisk). **c**, Relative change in LDL-R present in supernatant between day 9 and day 0 assessed by ELISA. P values are

as follows: Q555X versus untreated, $P = 0.001116$ (double asterisk), E4 splice versus untreated, $P = .009481$ (double asterisk). **d**, Gene-editing efficiencies from sgRNA screens in primary human T cells using CBE-Ts. X axis label indicates the targeted gene and target base within sgRNA. **e, f**, Percent C-G to T-A conversion (**e**) and surface protein loss (**f**) achieved by each base editor or control in multiplex-edited primary human T cells. Primary hepatocyte data were generated from $n = 3$ ('none'/untreated samples) or 4 (BE4 and CBE-T1.46 samples) independent biological replicates. T cell data were generated from $n = 2$ independent donors. Where applicable, statistical significance was computed via two-tailed unpaired t tests: NS, $P \geq 0.05$; * $P < 0.05$; ** $P < 0.01$.

T cells²⁸. To determine whether CBE-Ts could be used for T cell editing, we electroporated T cells with mRNA encoding CBE-Ts and gRNAs targeting genes coding for components of the T cell receptor, *B2M* or

CIITA and found that CBE-Ts yielded comparable or only slightly lower editing efficiencies compared to BE4 controls (Fig. 6d). Multiplexed CBE-T editing demonstrated comparable editing efficiencies compared

to single-plex editing, which resulted in corresponding levels of protein knock-down (Fig. 6e,f and Supplementary Fig. 31), demonstrating the potential of the CBE-T platform for therapeutic cellular engineering.

Discussion

Here we describe the development of two families of base editors, CBE-Ts and CBE-Ts, which use variants of TadA to catalyze the deamination of cytosines with either retention (CBE-Ts) or loss (CBE-Ts) of adenine deamination.

Over the course of ten total rounds of directed evolution and additional rounds of structure-guided design, TadA has matured to include over 29 substitutions in our most engineered CBE-Ts (Fig. 3c). Through X-ray crystallography, we show how the accumulation of substitutions impacts the shape of the active site cavity and may contribute to the accommodation of cytosine as substrate and the subsequent shift in specificity toward C-to-U deamination. The structures developed herein illuminate how amino acid substitutions in TadA influence gene-editing outcomes observed in cells.

The CBE-Ts and CBE-Ts reported here are precision base editors with highly mitigated guide-independent DNA off-target outcomes, fewer bystander edits and fewer guide-dependent DNA off-targets relative to previously reported CBEs due to the difference in kinetics of deamination of ssDNA by the TadA variant used in our CBE-T constructs. TadA-based CBE-Ts and CBE-Ts retain high on-target editing activity, enabling high gene editing efficiencies both in single- and multiplexed applications.

Finally, we show our CBE-Ts are active in therapeutically relevant cell types, including primary hepatocytes and primary T-cells, with editing outcomes similar or superior to what can be achieved with BE4. We demonstrate the ability of CBE-Ts to edit target sites in the *PCSK9* locus to reduce levels of secreted PCSK9 protein, as well as achieve high levels of multiplexed editing at T cell targets relevant for the generation of allogeneic CAR-T cells.

In summary, the development of TadA for use in highly efficient cytosine base editing represents an impactful advancement in the development of CBEs as therapeutic tools. Together with ABEs, CBE-Ts and CBE-Ts enable the programmable installation of all DNA transition mutations within living cells, separately or concurrently, through the use of laboratory-evolved and highly engineered TadA deaminases and consequently extend the potential therapeutic applications of cytosine base editing.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41587-022-01611-9>.

References

- Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).
- Anzalone, A. V., Koblan, L. W. & Liu, D. R. Genome editing with CRISPR-Cas nucleases, base editors, transposases and prime editors. *Nat. Biotechnol.* **38**, 824–844 (2020).
- Komor, A. C. et al. 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).
- Collantes, J. C. et al. Development and characterization of a modular CRISPR and RNA aptamer mediated base editing system. *CRISPR J.* **4**, 58–68 (2021).
- Rees, H. A., Minella, A. C., Burnett, C. A., Komor, A. C. & Gaudelli, N. M. CRISPR-derived genome editing therapies: progress from bench to bedside. *Mol. Ther.* **29**, 3125–3139 (2021).
- Jin, S. et al. Cytosine, but not adenine, base editors induce genome-wide off-target mutations in rice. *Science* **364**, 292–295 (2019).
- Zuo, E. et al. Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos. *Science* **364**, 289–292 (2019).
- Doman, J. L., Raguram, A., Newby, G. A. & Liu, D. R. Evaluation and minimization of Cas9-independent off-target DNA editing by cytosine base editors. *Nat. Biotechnol.* **38**, 620–628 (2020).
- Yu, Y. et al. Cytosine base editors with minimized unguided DNA and RNA off-target events and high on-target activity. *Nat. Commun.* **11**, 2052 (2020).
- Zuo, E. et al. A rationally engineered cytosine base editor retains high on-target activity while reducing both DNA and RNA off-target effects. *Nat. Methods* **17**, 600–604 (2020).
- Yang, L. et al. Engineering and optimising deaminase fusions for genome editing. *Nat. Commun.* **7**, 13330 (2016).
- Gaudelli, N. M. et al. Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage. *Nature* **551**, 464–471 (2017).
- Gaudelli, N. M. et al. Directed evolution of adenine base editors with increased activity and therapeutic application. *Nat. Biotechnol.* **38**, 892–900 (2020).
- Kim, H. S., Jeong, Y. K., Hur, J. K., Kim, J. S. & Bae, S. Adenine base editors catalyze cytosine conversions in human cells. *Nat. Biotechnol.* **37**, 1145–1148 (2019).
- Jeong, Y. K. et al. Adenine base editor engineering reduces editing of bystander cytosines. *Nat. Biotechnol.* **39**, 1426–1433 (2021).
- Grunewald, J. et al. A dual-deaminase CRISPR base editor enables concurrent adenine and cytosine editing. *Nat. Biotechnol.* **38**, 861–864 (2020).
- Zhang, X. et al. Dual base editor catalyzes both cytosine and adenine base conversions in human cells. *Nat. Biotechnol.* **38**, 856–860 (2020).
- Sakata, R. C. et al. Base editors for simultaneous introduction of C-to-T and A-to-G mutations. *Nat. Biotechnol.* **38**, 865–869 (2020).
- Li, C. et al. Targeted, random mutagenesis of plant genes with dual cytosine and adenine base editors. *Nat. Biotechnol.* **38**, 875–882 (2020).
- Richter, M. F. et al. Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. *Nat. Biotechnol.* **38**, 883–891 (2020).
- Tsai, S. Q. et al. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* **33**, 187–197 (2015).
- Chaudhary, R., Garg, J., Shah, N. & Sumner, A. PCSK9 inhibitors: a new era of lipid lowering therapy. *World J. Cardiol.* **9**, 76–91 (2017).
- Bhatia, S. N., Balis, U. J., Yarmush, M. L. & Toner, M. Effect of cell-cell interactions in preservation of cellular phenotype: cocultivation of hepatocytes and nonparenchymal cells. *FASEB J.* **13**, 1883–1900 (1999).
- Cohen, J. C., Boerwinkle, E., Mosley, T. H. Jr. & Hobbs, H. H. Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. *N. Engl. J. Med.* **354**, 1264–1272 (2006).
- Rao, A. S. et al. Large-scale phenome-wide association study of PCSK9 variants demonstrates protection against ischemic stroke. *Circ. Genom. Precis. Med.* **11**, e002162 (2018).
- Musunuru, K. et al. In vivo CRISPR base editing of PCSK9 durably lowers cholesterol in primates. *Nature* **593**, 429–434 (2021).
- Benjamin, R. et al. Genome-edited, donor-derived allogeneic anti-CD19 chimeric antigen receptor T cells in paediatric and adult B-cell acute lymphoblastic leukaemia: results of two phase 1 studies. *Lancet* **396**, 1885–1894 (2020).
- Liu, X. et al. CRISPR-Cas9-mediated multiplex gene editing in CAR-T cells. *Cell Res.* **27**, 154–157 (2017).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this

article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023

Methods

General methods

All molecular biology methods and cloning steps were performed as previously described¹³, including the utilization of USER enzyme (New England Biolabs, NEB, M5505L), Phusion U DNA Polymerase Green Multiplex PCR Master Mix (Thermo Fisher Scientific, F564L), Q5 Hot Start High-Fidelity 2X Master Mix (NEB, M0494L), Mach T1 competent cells (Thermo Fisher Scientific, C8681201) and ZymoPURE II Plasmid Midiprep kits (Zymo Research Corporation, D4201) in accordance with manufacturers' protocols. Amino acid sequences for base editors highlighted in this study can be found in Supplementary Sequences 3–31. Sequences of sgRNAs used to target genomic sites can be found in Supplementary Table 3. Representative CABE-Ts and CBE-Ts used in this study have been deposited on Addgene.

Generation of TadaA* and T_{AD}AC libraries for directed evolution

Synthetic libraries for directed evolution rounds one and two were obtained from Ranomics with the following specifications: evolution round one TadaA*8.20 library—each amino acid position of the TadaA*8.20 (from ABE8.20) sequence to be represented by all 20 amino acid substitutions at a frequency of 1–3 substitutions per library member (~10 million members). This library excluded all stop sequences and used only one codon per amino acid. This synthetic library was combined with a randomized library generated with error-prone PCR using TadaA*8.19 (ref. 13) as a template as previously reported in ref. 12. Evolution round two synthetic library—each amino acid position of the T_{AD}AC1.02 sequence to be represented by all 20 amino acid substitutions via a frequency of 2–3 substitutions per library member (~10 million members). These libraries were cloned into a bacterial expression plasmid containing dead Cas9 (dCas9 D10A and H840A) along with gRNAs targeting the chloramphenicol resistance gene through USER cloning.

Bacterial evolution of TadaA variants

Directed evolution of TadaA8.19 and TadaA8.20 library (directed evolution round one) and T_{AD}AC1.02 library (directed evolution round 2) was conducted as previously described in ref. 13 with the following changes: libraries of various TadaA* deaminase variants that are included in a bacterial plasmid containing TadaA*-dCas9-UGI editor architecture were challenged to revert edits in the chloramphenicol resistance gene to survive treatment with lethal doses of antibiotic drug. In the first round of directed evolution, the evolution library was a combination of an error-prone ABE8.19m TadaA* library and a synthetic ABE8.20m TadaA* library where each amino acid position is represented by all 20 substitutions at a frequency of 1–3 substitutions per library member. To overcome the antibiotic challenge, 2 C-to-T reversions (proline reversion and active site His reversion) were needed. In the second round of evolution, a synthetic library of CABE-T1.2 was used, which was generated with the specifications as the ABE8.20 TadaA* library but with 2–3 substitutions per library member. To overcome the antibiotic challenge, the same 2 C-to-T reversions plus 2 A-to-G STOP codon reversions were needed.

General HEK293T mammalian cell culture conditions

HEK293T cells (ATCC, CRL-3216) were cultured in DMEM + GlutaMAX (Gibco, 10569) supplemented with 10% (vol/vol) fetal bovine serum (Gibco, 10437) at 37 °C and 5% CO₂ in accordance with standard protocols from ATCC and as previously described.¹³

General HEK293T transfection conditions

For all transfections, HEK293T cells were seeded at a density of 3.0×10^5 cells per well in BioCoat poly-D-lysine coated 48-well plates (Corning, 356509) 16–22 h before transfection. Plasmid transfections were performed using Lipofectamine 2000 (Invitrogen, 11668-019) as previously described.¹³ Transfections with mRNA were performed using

Lipofectamine MessengerMAX in accordance with manufacturer protocols, with the following specifics: 500 ng (for saturating conditions) or 62.5 ng (subsaturating conditions) of mRNA encoding for editor or control and 100 ng of synthetic gRNA were combined in 12.5 µl total volume of OptiMEM serum reduced medium (Gibco, 31985). A 12.5 µl 1:12.5 (Lipo:OptiMEM) MessengerMAX mixture was then added to the mRNA/gRNA solution, and the entire contents were left to rest at ambient temperature for 15 min. For mRNA transfections at subsaturating conditions, 437.5 ng of carrier mRNA was also added to maintain equivalent amounts of transfected material. The entire 25 µl mixture was then used to treat the preseeded HEK293T cells. The sequences of sgRNAs used in this study are specified in Supplementary Table 3. Synthetic gRNAs for mRNA transfections have 5'/3' end-modifications as previously described.¹³

Targeted amplicon next-generation sequencing of DNA samples

After 4 d of incubation, gDNA from HEK293T cells was harvested from the cells using 100 µl of Quick Extract DNA Extraction Buffer (Lucigen, QE09050) in accordance with manufacturer protocols. For allogeneic T cells, 50 µl of Quick Extract DNA Extraction Buffer was used on 1×10^5 cells at 5–6 d post-transfections. Genomic DNA samples from mammalian cell samples were amplified with primers for site-specific genomic DNA amplification containing adapter sequences compatible with Illumina's TruSeq HT system (Adapter Read 1 sequence, AGATCG-GAAGAGCACACGTCTGAACTCCAGTCA; Adapter Read 2, sequence AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT). The sequences of these primers are listed in Supplementary Table 4. Specifically, 2 µl of gDNA was added to a PCR reaction mixture containing Phusion U Green Multiplex Master Mix (Thermo Fisher Scientific, F564L) and 0.5 µM of each forward and reverse primer. These amplicons were then barcoded using Q5 Hot Start High-Fidelity 2X Master Mix, where 2 µl of amplicon from the first round of PCR was added to the master mix containing 0.5 µM of each unique combination of forward and reverse barcode primer. Thermocycling conditions are as follows: 95 °C × 2 min of initial denaturation; 95 °C × 15 s of cycle denaturation; 62 °C × 20 s of annealing; 72 °C × 20 s of extension, with cycle repeats of 30 for the initial amplicon generation and 10 for barcoding. Barcoded amplicons were purified, size selected via gel electrophoresis and gel extracted using the Qiaquick Gel Extraction Kit (Qiagen, 28706×4), and the resultant DNA concentrations were evaluated with a NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific).

Data analysis of targeted amplicon next-generation sequencing

All targeted amplicon NGS data were analyzed using methods previously described, including the use of the following tools/software: trimmomatic (v0.39), bowtie2 (v2.35), samtools (v1.9) and bam-readcounts (v0.8).¹³

Data analysis of WGS data for guide-independent deamination

FASTQ files were aligned to the human genome (Gencode GRCh38v31 primary assembly) using BWA mem2 (bwa-mem2-2.2.1). Alignments were sorted by coordinates, merged if necessary, and duplicates were marked using Picard (v2.21.7) on default settings. Base-quality score recalibration was then performed using GATK (v4.1.4.1) to create a BAM file for input into LoFreq (v2.1.5) for variant calling. Bulk sample 1 was used as the normal sample and each clonally expanded cell was run as a separate tumor sample to identify somatic mutations specific to each cell. LoFreq was run with the '-min-cov 10' flag to require a minimum of ten times coverage at the variant site and somatic variants were analyzed from the somatic_final_minus-dbsnp.snvs output file, to remove common variants that were likely false positives.

For the odds ratio plots, a single representative cell from the untreated clonally expanded cells is required as a reference point to

compare with both the treated and untreated cells for both C-to-T and A-to-G deaminations. This cell was selected by ordering the untreated cells by proportion of A-to-G mutations and proportion of C-to-T mutations and selecting the one cell closest to the median for both metrics. N1 was in position 5/8 for C-to-T mutations and position 3/8 for A-to-G mutations, making it the best candidate for the reference cell across both CABE-T and CBE-T treatments.

Protein expression and purification

TadA*8.20 protein was cloned into a pET51b⁺ vector with His and SUMO tags at the N-terminus and expressed in *E. coli* BL21 Star (DE3) cells (NEB, C25271) in LB media. Cell cultures were grown at 37 °C with shaking at 240 rpm, and protein expression was induced by 0.5 mM IPTG when OD₆₀₀ reached 0.6. Cell culture was incubated with shaking at 18 °C overnight. Harvested cells were lysed by a high-pressure homogenizer in lysis buffer (25 mM Bis-Tris, 500 mM NaCl, 1 mM TCEP, 10% (vol/vol) glycerol, pH 6.0 and 1 mM PMSF), and the cell lysate was clarified by ultracentrifugation. Clarified lysates were loaded onto Ni-NTA agarose resin by batch binding for 1 h at 4 °C. The resin was washed with lysis buffer with 20 mM imidazole on a gravity flow column followed by elution with the lysis buffer supplemented with 50/100/250 mM imidazole. The eluted sample was incubated with Ulp1 while dialyzed in 25 mM Bis-Tris, 300 mM NaCl, 1 mM TCEP, 10% (vol/vol) glycerol and pH 6.0 overnight. The dialyzed sample was loaded onto Ni-NTA resin to remove uncleaved protein and Ulp1. The flowthrough from reverse Ni-NTA was loaded on a 5 ml Heparin HP column (Cytiva) and eluted using a 0–2 M NaCl gradient. Fractions containing TadA*8.20 protein were further purified by size exclusion chromatography on Superdex75 10/300 in 25 mM Bis-Tris, 300 mM NaCl, 1 mM TCEP, 10% (vol/vol) glycerol, pH 7.0. T_{AD}AC-1.14 protein was expressed with N-terminal His-tag in pET51b⁺ vector and purified as described above, except that Ulp1 tag cleavage and reverse Ni-NTA steps were omitted. T_{AD}AC-1.17 and T_{AD}AC-1.19 were cloned in pD881 vector (ATUM) with N-terminal His-tag and SUMO tag and expressed in *E. coli* BL21 cells (NEB). Protein expression was induced by 0.2% (wt/vol) rhamnose at OD₆₀₀ of 0.6, followed by incubation at 37 °C for 4 h. Purification was performed as described above. These deaminase variants were used for X-ray crystallography studies. All CBE-T base editor proteins used for biochemical studies were expressed and purified as described above with slight modifications.

Crystallization of TadA*8.20 with ssDNA

The crystallization condition of TadA*8.20 with ssDNA containing the adenine analog 2-deoxy-8-azanebularine (d8Az), 5'-G(1)C(2)T(3)C(4)G(5)G(6)C(7)T(8)d8Az(9)C(10)G(11)G(12)A(13)-3', was identified and optimized using a Mosquito robot (SPT LabTech) at 20 °C. Drops were prepared by mixing 1 µl of protein plus ssDNA solution (0.15 mM TadA*8.20 in 25 mM Bis-Tris, 300 mM NaCl, 1 mM TCEP, 10% (vol/vol) glycerol, pH 7 and 0.22 mM ssDNA with d8Az) and 1 µl of reservoir solution (27–29% (vol/vol) PEG 3,350, 0.22–0.26 M ammonium acetate, 0.1 M Tris pH 8.5), and equilibrated against 70 µl of reservoir solution. The crystals were transferred to a cryoprotectant solution (15% (vol/vol) glycerol, 29% (vol/vol) PEG 3,350, 0.26 M ammonium acetate, 0.1 M Tris pH 8.5) and flash-cooled in liquid nitrogen.

Crystallization of T_{AD}AC-1.17 with ssDNA

The crystallization condition of T_{AD}AC-1.17 with ssDNA containing the adenine analog 2-deoxy-8-azanebularine (d8Az), 5'-G(1)C(2)T(3)C(4)G(5)G(6)C(7)T(8)d8Az(9)C(10)G(11)G(12)A(13)-3', was identified and optimized using a Mosquito robot (SPT LabTech) at 20 °C. Drops were prepared by mixing 1 µl of protein plus ssDNA solution (0.15 mM T_{AD}AC-1.17 in 25 mM Bis-Tris, 300 mM NaCl, 1 mM TCEP, 10% (vol/vol) glycerol, pH 7 and 0.22 mM ssDNA with d8Az) and 1 µl of reservoir solution (4–8% (vol/vol) PEG 3,350, 8–10% Tacsimate pH 6) and equilibrated against 200 µl of reservoir solution. The crystals were transferred to a

cryoprotectant solution (12% (vol/vol) PEG 3,350, 10% (vol/vol) Tacsimate pH 6, 25% (vol/vol) glycerol) and flash-cooled in liquid nitrogen.

Crystallization of T_{AD}AC-1.14 without ssDNA

The crystallization condition of T_{AD}AC-1.14 without ssDNA (T_{AD}AC-1.14-holo) was identified and optimized using a Mosquito robot (SPT LabTech) at 20 °C. Drops were prepared by mixing 1 µl of protein solution (0.18 mM T_{AD}AC-1.14 in 25 mM Bis-Tris, 450 mM NaCl, 1 mM TCEP, 10% (vol/vol) glycerol, pH 7) and 1 µl of reservoir solution (1.8–2.0 M ammonium sulfate, 0.1 M HEPES pH 7.5) and equilibrated against 200 µl of reservoir solution. The crystals were transferred to a cryoprotectant solution (1.8 M ammonium sulfate, 0.1 M HEPES pH 7.5, 20% (vol/vol) glycerol) and flash-cooled in liquid nitrogen.

Crystallization of T_{AD}AC-1.19 without ssDNA

The crystallization condition of T_{AD}AC-1.19 without ssDNA (T_{AD}AC-1.19-holo) was identified and optimized using a Mosquito robot (SPT LabTech) at 20 °C. Drops were prepared by mixing 1 µl of protein solution (0.3 mM T_{AD}AC-1.19 in 25 mM Bis-Tris, 300 mM NaCl, 1 mM TCEP, 10% (vol/vol) glycerol, pH 7) and 1 µl of reservoir solution (6–12% (vol/vol) PEG 3,350, 0.3–0.5 M ammonium citrate tribasic pH 7.0) and equilibrated against 200 µl of reservoir solution. The crystals were transferred to a cryoprotectant solution (16% (vol/vol) PEG 3,350, 0.6 M ammonium citrate tribasic pH 7.0, 20% (vol/vol) glycerol) and flash-cooled in liquid nitrogen.

Data collection and structure determination of TadA*8.20 and T_{AD}AC-1 variants

Data collections were performed at the Frontier Microfocusing Macromolecular Crystallography (FMX) beamline of the National Synchrotron Light Source II or the ID30B beamline of the European Synchrotron Radiation Facility, or the BL13-XALOC beamline of the ALBA Synchrotron or the P13 beamline of the EMBL Hamburg at the PETRA III storage ring (DESY). Diffraction data were processed using XDS²⁹ and scaled using AIMLESS³⁰. The crystal structures of TadA*8.20, T_{AD}AC-1.17, T_{AD}AC-1.14 and T_{AD}AC-1.19 without or with ssDNA were determined by molecular replacement techniques implemented in Phaser³¹. For the TadA*8.20 structure, the coordinates of the *E. coli* TadA structure (Protein Data Bank (PDB) code: 1Z3A)³² were used to obtain the initial phases. For T_{AD}AC-1.17, T_{AD}AC-1.14 and T_{AD}AC-1.19 structures, the coordinates of the TadA*8.20 (this study) were used to obtain the initial phases. Following molecular replacement, simulated annealing was performed in phenix.refine³³ to remove model bias. The models were refined by iterative rounds of model building and the addition of water molecules using Coot³⁴. Refinement of the structures in phenix.refine used noncrystallographic symmetry restraints, positional and B-factor refinement, and TLS (translation, libration and screw) (except for T_{AD}AC-1.17 and T_{AD}AC-1.14). The crystals of TadA*8.20 and T_{AD}AC-1.17 are merohedrally twinned with twin fractions of 0.375 and 0.246 by Britton analyses (phenix.xtriage), respectively, and the twin law -h, -k, l was used in refinement. The data collection and refinement statistics are summarized in Supplementary Table 2. The residues and nucleotides visualized in the structures, of 167 residues and 13 nucleotides, are listed in Supplementary Table 5. Figures were created with PyMol Software (Schrodinger, 2010. The PyMOL Molecular Graphics System, Version 2.4.1.).

Biochemical characterization of deamination by ABEs, CABEs and CBEs

An sgRNA (mG*ma*mA*CACAAAGCAUAGACUGCGUUUUAGAGCU AGAAAUAGC

AAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGC ACCGAGUCGGUGCU*mU*mU*mU; modifications: m, 2'-O-methyl, and *, phosphorothioate linkage) was synthesized at Agilent Technologies and Integrated DNA Technologies (IDT). Substrate DNA was synthesized

at IDT: DNA strand undergoing deamination (TTCGGTGGCTCCGTCCTGTAACACAAAGCATAGACTGCCGGCGTTTGGTTGCTCTTCG) was labeled with 5' ATTO-647 fluorophore and a complementary DNA strand undergoing nicking by D10A-Nickase (CGAAGACAACAAAACGCCGGCAGTCTATGCTTTGTGTTACGGACGGAGCCACCGAA) was labeled with 5' 6-FAM fluorophore. For guide RNA-independent deamination, the ATTO-647 labeled single-strand DNA was used as is. For guide RNA-dependent deamination, dsDNA substrate was prepared by annealing the two strands, with twofold excess of the strand undergoing nicking (1:2 nmol). The duplexed DNA was purified by 7.5% Native-PAGE (29:1, acrylamide:bisacrylamide; Sigma). The acrylamide band containing the dsDNA was excised, crushed and rotated overnight in crush-and-soak buffer (400 mM NaCl and 25 mM EDTA) to elute the dsDNA. The eluted dsDNA was precipitated at -20°C for 2 h after adding 1 volume of 100% 2-propanol, followed by centrifugation at 20,000g for 30 min at 4°C . The DNA pellet was washed with 1 volume of 70% vol/vol ethanol and centrifuged at 20,000g for 30 min at 4°C . The pellet was air-dried at room temperature for 30 min and resuspended in water.

RNP complexes were formed by mixing the sgRNA and the appropriate base editor protein in a 1.5:1 molar ratio in 'RNP assembly and reaction buffer' (20 mM HEPES-KOH pH 7.4, 100 mM KCl, 5 mM MgCl_2 , 5% vol/vol glycerol, 2 mM TCEP) and incubating at room temperature for 20 min.

For single turnover kinetics of guide RNA-dependent dsDNA deamination in vitro, to 1- μM final concentration of RNP, a final concentration of 10 nM dsDNA substrate (prepared at 100 nM in RNP assembly and reaction buffer) was added to initiate deamination. The reaction was incubated at 37°C and aliquots of 5 μl were withdrawn at the indicated time intervals. The reactions were quenched in 50- μl quenching buffer (50 mM Tris-Cl, pH 8.5, 400 mM NaCl, 25 mM EDTA, 0.1% SDS, 1 μl thermolabile proteinase K (New England Biolabs, NEB P8111S) and 1 μl 15 mg ml^{-1} coprecipitant Glycoblue (Thermo Fisher Scientific, A9515)) for 15 min at 37°C . The thermolabile proteinase K was inactivated at 75°C for 15 min.

The quenched reaction time points were then precipitated with 2-propanol as described above. For detecting deaminated adenine (inosine) catalyzed by ABEs, the precipitated time points were treated with Endonuclease V as described previously in refs. 20,35. For detecting deaminated cytidine (deoxy-uridine) catalyzed by CBEs, the precipitated time points were treated with USER II (NEB, M5508L) according to manufacturer guidelines. The samples were mixed with equal volume of formamide gel loading buffer (95% formamide, 25 mM EDTA, 0.025% SDS and 0.025% bromophenol blue), heated to 98°C for 5 min and resolved on denaturing 7.5% Urea-PAGE (19:1, acrylamide:bisacrylamide; National Diagnostics). The reaction was monitored by scanning the gel sequentially with FAM followed by Alexa-647 settings using ChemiDoc Imaging System (Bio-Rad). The intensities of the un-cleaved and cleaved DNA were quantified using ImageJ 1.53 K. Data were fit to a single exponential decay in Prism 9 (GraphPad Prism, v9.4.0) to calculate apparent deamination rates (k_{app}). Nicking of the substrate DNA by D10A-Nickase of base editor, constant across all base editors assayed, was detected with the 6-FAM fluorophore and used as control to ensure uniformly active recombinant proteins.

For single turnover kinetics of guide RNA-independent ssDNA deamination in vitro, the reaction was set up as described above but with the following modifications: the base editor was not programmed with sgRNA and was incubated with the ATTO-647 labeled ssDNA strand.

For in vitro end-point deamination assay to compare deamination by ABE 8.20, BE4, CABE-T2.17, CABE-T3.155, CBE-T1.14 and CBE-T1.52, the deamination reaction was set up with 1- μM BE RNP and 10-nM dsDNA substrate as described above. Instead of time points, the whole reaction was quenched after 24 h and precipitated as described above. The precipitated reaction was resuspended in

water and split into four equal parts: untreated, treated with Endonuclease V as described, treated with USER II as described and treated with human Alkyl Adenine Glycosylase (hAAG; NEB 0313S) followed by AP Endonuclease I (APE1; NEB M0282L) according to manufacturer's instructions. The combination of hAAG and APE1 was used because of our experimental observation that G:U (product of cytosine deamination) is a substrate for EndoV, which was confirmed by NEB (<https://www.neb.com/tools-and-resources/selection-charts/dna-repair-enzymes-on-damaged-and-non-standard-bases>). EndoV, therefore, could not be used when comparing ABEs, CABEs and CBEs for relative A-to-I and C-to-U deamination activities. hAAG is more specific, and only produced detectable cleavage product for A-to-I but not for C-to-U deamination under the same experimental conditions and thus was used for such comparisons. Following these treatments, the samples were resolved on Urea-PAGE and data were quantified as described above.

mRNA production of CABE-T, CBE-T and controls used in HEK293T, T cells and primary human hepatocytes

The mRNAs used in this study were produced through in vitro transcription of expression plasmids encoding our editors and controls, in accordance with protocols previously described in ref. 13.

Isolation of single cells by FACS and whole-genome sequencing

HEK293T cells were transfected via Lipofectamine MessengerMAX (Thermo Fisher Scientific, LMRNA001) with control (Cas9, SPACE, etc.) or editor-encoding mRNA along with synthetic gRNA (special order from Axolabs) targeting a region in β -2-microglobulin (B2M). The sequence of this synthetic guide is as follows (Axolabs-specific syntax): ascusCACGCUGGAUAGCCUCCGUUUUAGAGCUAGAAUA GCAAGUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGU GGCACCGAGUCGGGUGCusususU

The disruption of B2M upon successful targeting by ABE, CBE or Cas9 at this site has been internally validated. Three days after transfection, cells were dissociated with TrypLE Express, washed with cell staining buffer (Biolegend, 420201) via centrifugation, and resuspended in cell staining buffer containing 1:100 of PE-conjugated antihuman B2M antibody (Biolegend, 316306). After 30 min of incubation on ice in the dark, cells were washed three times with cell staining buffer via centrifugation and strained into standard 5-ml FACS tubes.

Single cells gated as PE-negative were sorted into 96-well plates containing DMEM + 20% FBS + 100 units per ml penicillin/streptomycin (Thermo Fisher Scientific, 15140122). For untreated control, single cells were sorted by live only. Representative gating strategies are provided in Supplementary Fig. 30. After 12 d of culture, gDNA was harvested from cells using the Agencourt DNAdvanced kit (Beckman-Coulter, A48705) in accordance with manufacturer protocols. Confirmation of successful editing of each clone was achieved through targeted amplicon sequencing of the B2M amplicon encompassing the target site. Sequence confirmed gDNA was then submitted to Novogene for library preparation and WGS.

Isolation and culture of allogeneic human T cells

Human T cells were isolated from leukapheresis products (Leukopaks, HemaCare) by positive selection using CD4 and CD8 MicroBeads (Miltenyi, 130045101 and 130045201). T cells were frozen at $25-50 \times 10^6$ cells per ml of Cryostor CS10 (Stemcell Technologies, 1001061). For editing experiments, T cells were thawed in a water bath at 37°C and then allowed to rest overnight in ImmunoCult-XF T Cell Expansion Medium containing (Stemcell Technologies, 10981) 5% CTS Immune Cell SR, Glutamax, 10 mM HEPES, 1% Penicillin/Streptomycin (Thermo Fisher Scientific, 15140122). The next day, T cells were activated using 25 μl of ImmunoCult Human CD3/CD28/CD2 T Cell Activator (Stemcell Technologies, 10970) per ml of cells at 1×10^6 cells per ml plus 300 IU ml^{-1}

of IL-2 (CellGenix, 1420050). Fresh IL-2 was added to T cells every 2–3 d. T cells were cultured at 37 °C and 5% CO₂.

Electroporation of human T cells

T cells were transfected 72 h after activation. Cells were resuspended in P3 Primary Cell Nucleofector Solution containing Supplement 1 (Lonza, V4SP-3960). 1×10^6 T cells were edited with 1 µg of synthetic sgRNA (IDT) and 2 µg of editor mRNA in a total volume of 20 µl using P3 96-well Nucleocuvette kit (Lonza, V4SP-3960). The three sgRNAs used are as follows: B2M Exon 2 (B2M Ex.2), pmSTOP C6, CD247 pomSTOP C7 and PD-1 Ex.1SA C7 are specified in Supplementary Table 3. T cells were electroporated with the 4D-Nucleofector system (Lonza, AAF-1003B and AAF-1003S) using program DH-102. All experiments were performed with two independent T cell donors. For NGS analysis, 1×10^5 T cells per condition at each timepoint were pelleted, supernatant was removed and pellets were resuspended in 50 ml of QuickExtract DNA Extraction buffer (Lucigen, QE09050) and transferred to a PCR plate for targeted amplicon sequencing.

Flow cytometry of human T cells

Protein knockout was evaluated by flow cytometry 5–6 d post-editing. T cells were stained with fluorophore-conjugated antibodies for TCRα/β (Biolegend, 306718), β2M (Biolegend, 316304) and PD-1 (Biolegend, 367422) via 1:33 dilution in standard PBS. For PD-1 analysis by flow cytometry, T cells were treated with Cell Activation Cocktail (without brefeldin A) (Biolegend) overnight before staining. Events were collected using a MACSQuant Analyzer 16 (Miltenyi). Data were analyzed using the FlowJo software (v10.8.1)

Generation and maintenance of primary human hepatocytes

Cryogenically frozen primary human hepatocytes (BioIVT) were thawed and plated at a density of 3.5×10^5 cells per well on BioCoat Collagen I24-well plates (Corning, 354408) and maintained in CP Media supplemented with Torpedo Antibiotic Mix (BioIVT) in accordance with protocols provided by BioIVT. Once PHH monocultures were established overnight, generation of long-lived PHH cultures involved the additional coculturing of 3T3-J2 murine fibroblasts (Kerafast, EF3003) at 2.0×10^4 cells per well to the established PHH monocultures. PHH cocultures were maintained with media changes every 48 h throughout the duration of the study.

Transfection of primary human hepatocytes

PHH cocultures were transfected 48 h after coculture generation with 3T3-J2 murine fibroblasts. Transfections with mRNA were performed using Lipofectamine MessengerMax (Thermo Fisher Scientific, LMRNA003) in accordance with manufacturers' protocols, with the following optimized specifics: 1 µg (for saturating conditions) of mRNA encoding for editor and 333 ng of synthetic gRNA (Synthego) were combined in 30 µl of OptiMEM serum reduced medium (Gibco, 31985). A 30 µl 1:15 (Lipofectamine:OptiMEM) mixture was added to the mRNA/gRNA solution with the resulting final mixture left to rest at ambient temperature for 15 min. The entire 60-µl solution was used to treat a well of cocultured primary human hepatocytes. Each study condition was run in triplicate and transfection amounts used were scaled up accordingly. At 9 d post-transfection, the PHH cocultures were lysed with a solution of 10 mM Tris–HCl pH8.0 (Thermo Fisher Scientific, 15568025), 0.05% SDS (Thermo Fisher Scientific, 15553027) and 500 µg proteinase K (Thermo Fisher Scientific, EO0491) at a total of 200 µl per well. Once lysed, lysate was treated at 85 °C for 15 min to inactivate proteinase K. The sequences of sgRNAs used in this study are specified in Supplementary Table 3.

Protein assays of transfected primary human hepatocytes

PCSK9 protein knockdown quantification was assessed using a Human PCSK9 SimpleStep ELISA kit (Abcam, ab209884) by measuring secreted

PCSK9 concentration in supernatant collected every 48 h. Supernatant was ten times diluted using assay buffer, and the assay protocol was run in accordance with the manufacturer's protocol. LDL-R quantification was assessed using a Human LDL-R SimpleStep ELISA kit (Abcam, ab209884) by measuring secreted LDL-R protein in supernatant collected every 48 h. Both SimpleStep ELISA kits employ an affinity tag labeled capture antibody and a reporter conjugated detector antibody. The capture antibody and detector antibody bind to sample analytes, which are then immobilized to an anti-tag antibody coating the assay well. Both colorimetric ELISA assays are read at an absorbance of 450 nm.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Next-generation sequencing data underlying all experiments are deposited in the NCBI Sequence Read Archive (SRA) under submission project [PRJNA869750](https://www.ncbi.nlm.nih.gov/sra/PRJNA869750). The atomic coordinates and structure factors have been deposited in the PDB as entries: [8E2P](https://www.rcsb.org/entry/8E2P), [8E2Q](https://www.rcsb.org/entry/8E2Q), [8E2R](https://www.rcsb.org/entry/8E2R) and [8E2S](https://www.rcsb.org/entry/8E2S). Source Data are available for Figs. 1, 3–6, Extended Data Fig. 5–7 and Supplementary Figs 2–4, 6, 15, 17–21, 22–29 (including gel image source files). Source data are provided with this paper.

Code availability

All software tools used for data analysis are publicly available and were used in a manner as previously described in ref. [13](#).

References

- Kabsch, W. XDS. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 125–132 (2010).
- Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. D. Biol. Crystallogr.* **69**, 1204–1214 (2013).
- McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
- Kim, J. et al. Structural and kinetic characterization of *Escherichia coli* TadA, the wobble-specific tRNA deaminase. *Biochemistry* **45**, 6407–6416 (2006).
- Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 213–221 (2010).
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 486–501 (2010).
- Lapinaite, A. et al. DNA capture by a CRISPR–Cas9-guided adenine base editor. *Science* **369**, 566–571 (2020).

Acknowledgements

We acknowledge and thank M. Humes and B. Gantzer (Beam Tx) for automation support. We thank J. Decker and David Born (Beam Tx) for NGS and computational support. We acknowledge and thank R. Manoukian and L. Hardy (Beam Tx) for their FACS expertise and for sorting cells used in WGS experiments. We thank A. Arvind (Beam Tx) for her assistance with protein crystallization.

Author contributions

D.K.L., P.R.F., A.A. and M.A.O. conducted directed evolution, structural, biochemical and gene-editing experiments and wrote the manuscript. C.S., N.S. and K.D.L. conducted experiments. T.P.F., J.M.G. and L.P. conducted primary cell experiments, analyzed data and wrote the manuscript. T.B., P.G. and L.Y. analyzed sequencing data and conducted statistical analyses. S.-J.L. directed structural biology,

biochemistry and protein engineering work and wrote the manuscript. G.C. edited the manuscript. N.M.G. conceived and directed the research and wrote the manuscript.

Competing interests

All authors were employees of Beam Therapeutics when the work was conducted and are shareholders in the company. Beam Therapeutics has filed patent applications on this work.

Additional information

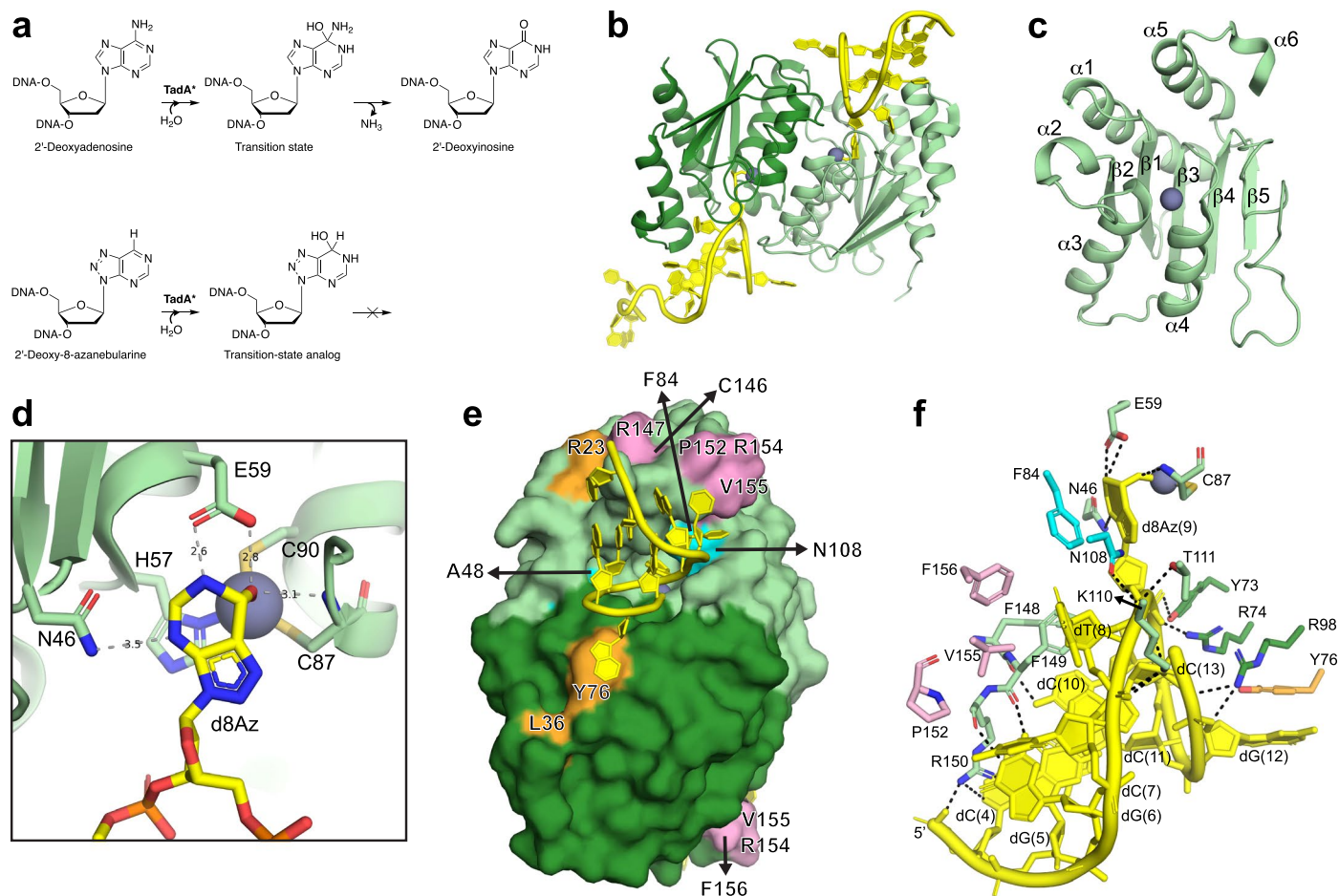
Extended data is available for this paper at <https://doi.org/10.1038/s41587-022-01611-9>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41587-022-01611-9>.

Correspondence and requests for materials should be addressed to Nicole M. Gaudelli.

Peer review information *Nature Biotechnology* thanks Sangsu Bae and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

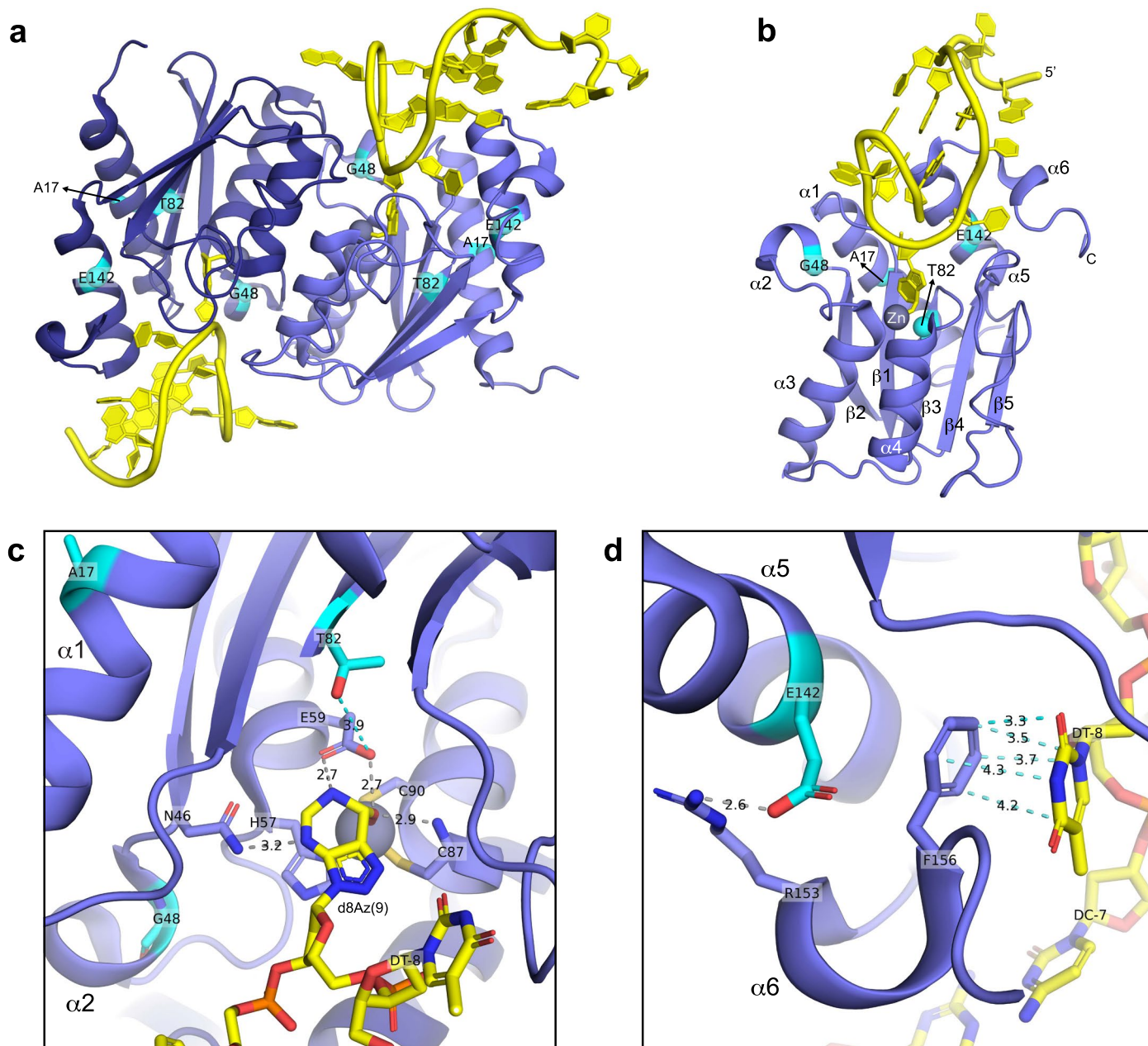
Reprints and permissions information is available at www.nature.com/reprints.



Extended Data Fig. 1 | Crystal structure of Tada*8.20 in a complex with ssDNA containing the adenine transition-state analog 2'-deoxy-8-azanebularine.

a, The top panel depicts hydrolytic deamination of adenosine catalyzed by TadaA. The bottom panel shows the hydration of the adenine analog, 2'-deoxy-8-azanebularine (d8Az), forming the transition-state analog that gets trapped in the active site by coordinating with zinc. **b**, The overall structure of the Tada*8.20 functional homodimer (chain A in dark green; chain B in light green) bound to ssDNA (yellow). **c**, The overall structure of Tada*8.20 monomer. The monomer (light green) contains five β -strands ($\beta 1$ to $\beta 5$) and six α -helices ($\alpha 1$ to $\alpha 6$) that fold into a single domain with a central five-stranded β -sheet surrounded by α -helices. The zinc ion is shown as a gray sphere. **d**, Tada*8.20 active site with ssDNA-d8Az transition-state analog bound. The catalytic zinc (gray) ion is coordinated to a histidine residue (H57), two cysteine residues (C87 and C90), and the d8Az transition-state analog (yellow). The hydrogen bonds are shown as gray dashed

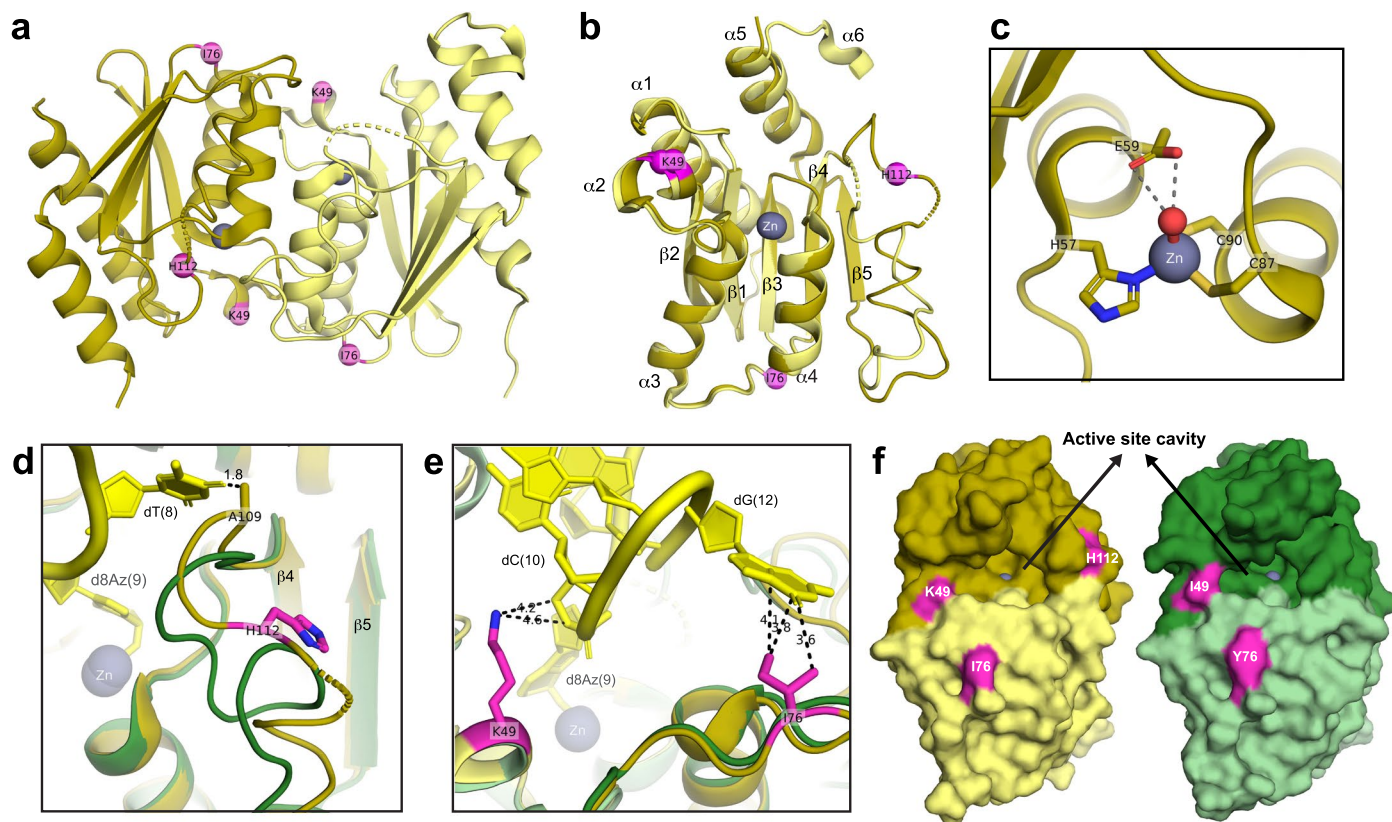
lines. **e**, The surface of Tada*8.20 dimer (dark and light green) bound to ssDNA (yellow), showing that ssDNA is bound in the active site deep cavity located at the protein dimer interface and interacts with residues from both monomers, including the substitutions I76Y, L84F, D108N, R152P, E155V, and I156F relative to the wild-type TadaA. The substitutions at the protein surface are shown in orange (W23R, H36L, R51L, I76Y, and A106V), at the C-terminal in pink (S146C, D147R, R152P, Q154R, E155V, I156F, and K157N), and at the active site in cyan (P48A, V82S, L84F, and D108N). **f**, Interactions between ssDNA (yellow) and Tada*8.20 active site residues. The residues from chains A and B are shown in dark and light green, respectively. The protein surface, C-terminal, and active site substitutions are shown in orange (chain A), pink (chain B), and cyan (chain B), respectively. The zinc ion is shown in a gray sphere. The hydrogen bonds are shown as black dashed lines. A stereo view is shown in Supplementary Fig. 7.



Extended Data Fig. 2 | Crystal structure of $T_{AD}AC-1.17$ in a complex with ssDNA containing the adenine transition-state analog 2-deoxy-8-azanebularine.

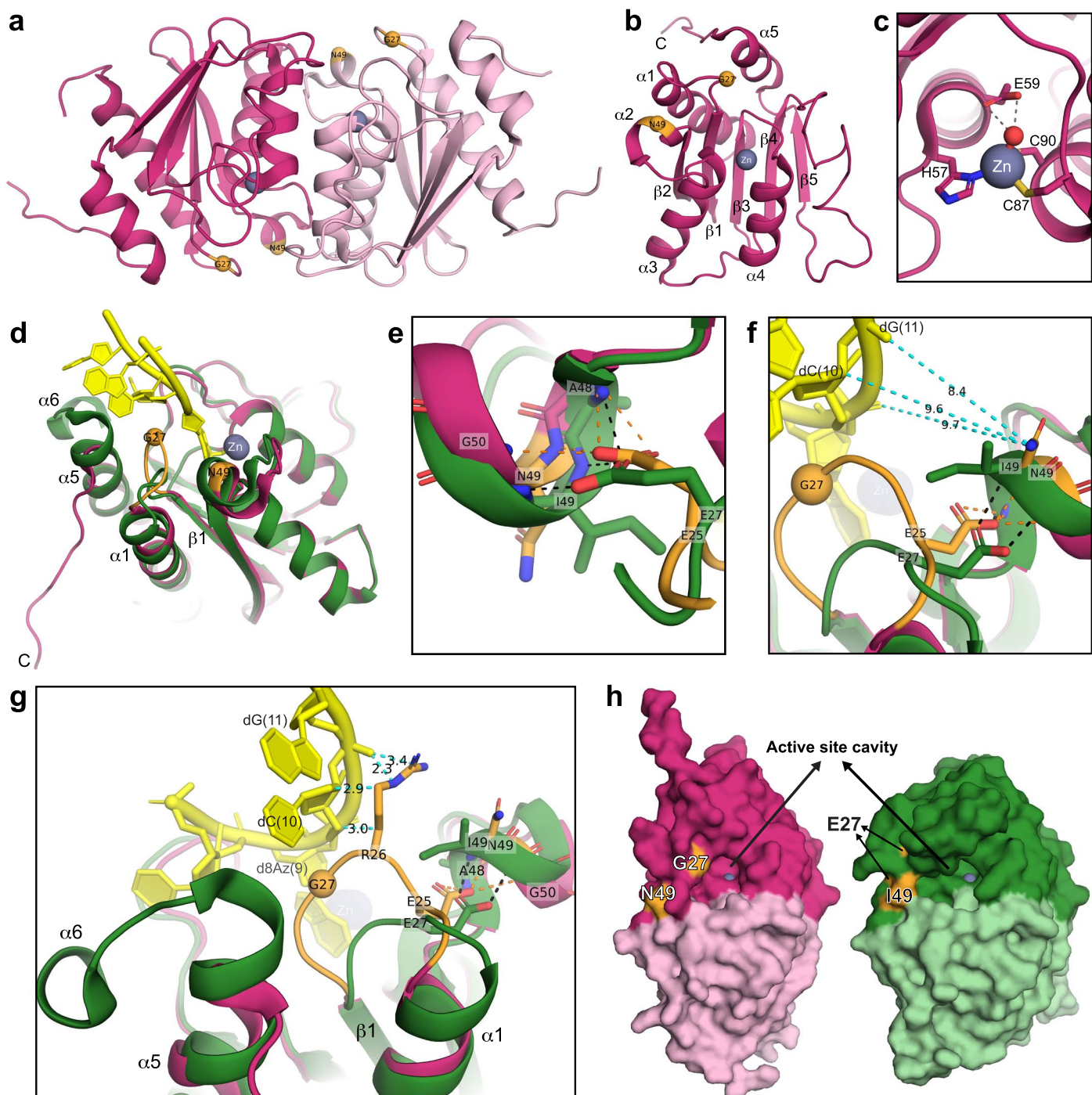
a, The overall structure of the $T_{AD}AC-1.17$ functional homodimer (chain A in dark blue; chain B in slate blue) with ssDNA (yellow) bound. The substitutions (T17A, A48G, S82T, and A142E) relative to $TadA^*8.20$ are shown in cyan spheres. **b**, The overall structure of $T_{AD}AC-1.17$ monomer (slate blue) in a complex with ssDNA (yellow). The monomer contains five β -strands ($\beta 1$ to $\beta 5$) and six α -helices ($\alpha 1$ to $\alpha 6$) that fold into a single domain with a central five-stranded β -sheet surrounded by α -helices. C and 5' represent the C-terminus and 5'-end of the ssDNA, respectively. **c**, $T_{AD}AC-1.17$ active site with ssDNA-d8Az transition-state

analog bound. The catalytic zinc ion (gray sphere) coordinates H57, C87, C90, and the d8Az transition-state analog (yellow). The T82 side chain (cyan) is near the catalytic E59 side chain (3.9-Å; cyan dashed line) and may play a role in deamination by donating/accepting a proton to/from E59. The residue A17 (cyan) is in $\alpha 1$ -helix at the protein surface. The residue G48 (cyan) is in $\alpha 2$ -helix at the substrate binding pocket. The H-bonds between the d8Az transition-state analog and protein residues are shown as gray dashed lines. **d**, The side chain of the E142, located in $\alpha 5$ -helix, H-bonds (gray dashed line) to the R153 side chain, located in $\alpha 6$ -helix, and helps stabilize the C-terminal $\alpha 6$ -helix to position the F156 side chain to interact (cyan dashed lines) with the pyrimidine base of dT(8).



Extended Data Fig. 3 | Crystal structure of $T_{AD}AC-1.14$ without ssDNA and structural comparisons with $Tada^*8.20$. **a**, The overall structure of the $T_{AD}AC-1.14$ functional homodimer (chain A in dark yellow; chain B in light yellow). The substitutions (I49K, Y76I, and G112H) relative to $Tada^*8.20$ are shown in magenta spheres. The residue H2 is disordered and not visualized in the structure. The zinc ion is shown as a gray sphere. The dashed line represents the partially disordered (A109 to A114) loop between $\beta 4$ and $\beta 5$ (R107 to V130). **b**, The overall structure of $T_{AD}AC-1.14$ monomer. The superposition between chains A (dark yellow) and B (light yellow) shows two different conformations for the loop between $\beta 4$ and $\beta 5$ (R107 to V130). **c**, $T_{AD}AC-1.14$ active site with water (red sphere) bound to the zinc ion (gray sphere). The residues H57, C87, and C90 coordinate with the zinc ion. The water molecule (red sphere) H-bonds (gray dashed lines) to the catalytic residue E59. In the first step of the $TadA$ reaction, this water is added to the substrate to form a transition state specie (Extended

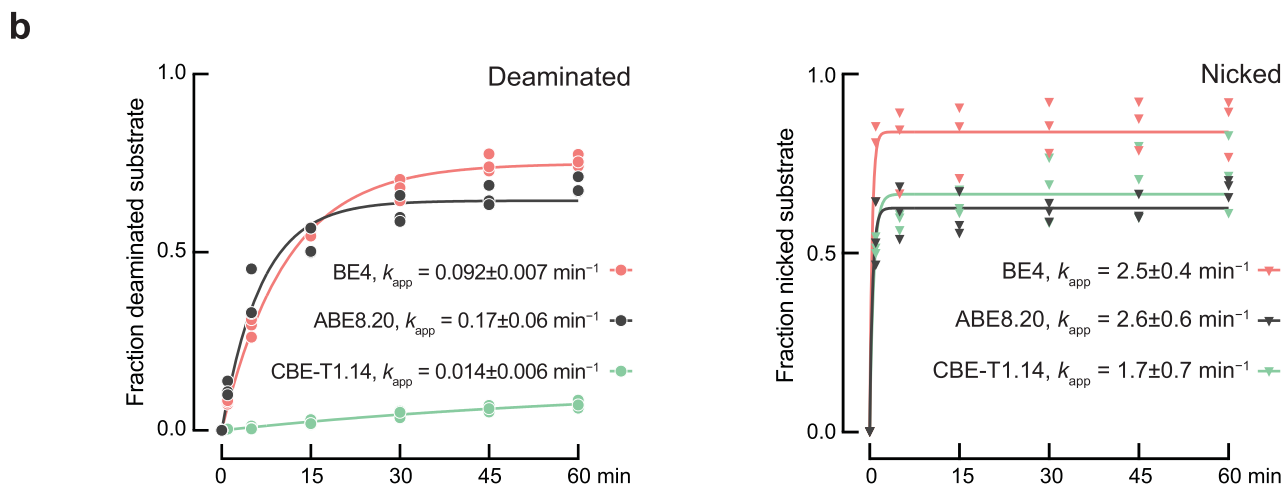
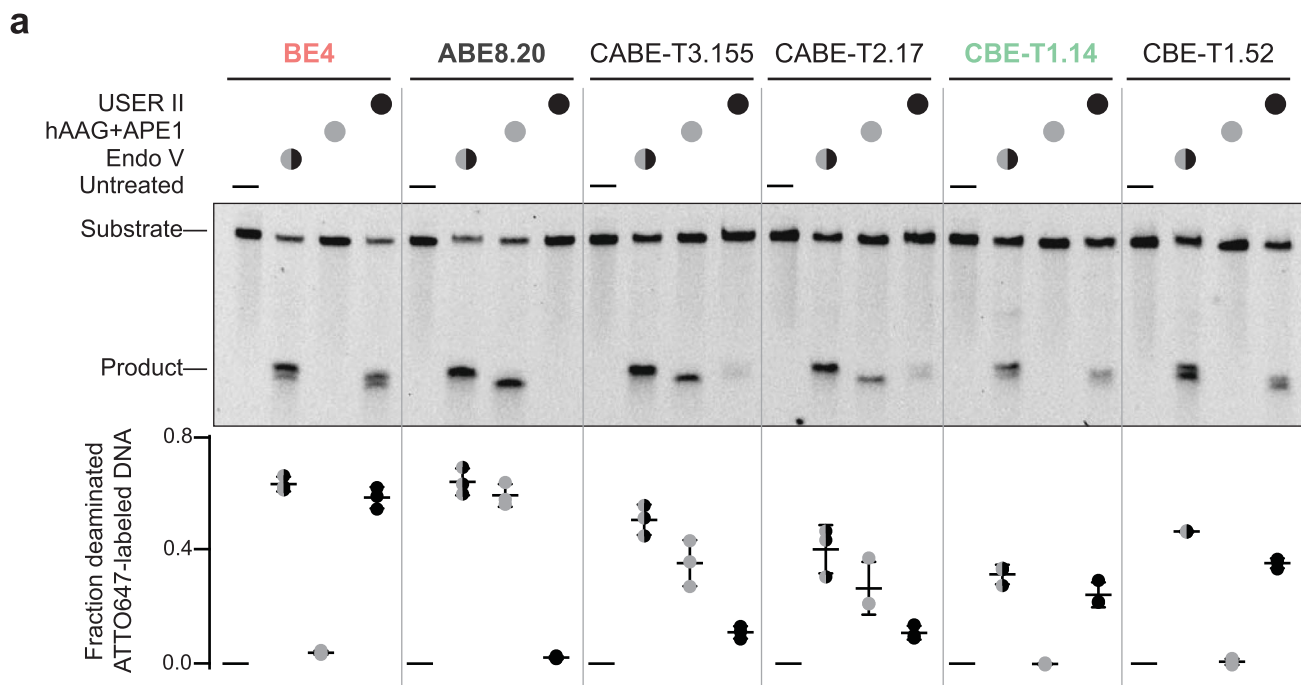
Data Fig. 1a). **d–f**, Structural comparisons between substrate-free $T_{AD}AC-1.14$ (dark yellow) and ssDNA-bound $Tada^*8.20$ (dark green and yellow) structures. The $T_{AD}AC-1.14$ loop between $\beta 4$ and $\beta 5$, which contains the substitution G112H (magenta), has a different conformation than $Tada^*8.20$ and may impact ssDNA (yellow) binding by making steric clashes between the residue A109 and the base dT(8) (1.8-Å), which is adjacent to the target base d8Az(9) (**d**). The substitution Y76I (magenta) may not affect ssDNA binding by conserving interactions (black dashed lines) with the base dG(12) (**e**). The substitution I49K positions the K49 side chain nearby the dC(10) backbone (~4.5-Å; black dashed lines) and may contribute to stabilizing the protein-DNA complex (**e**). The surface of $T_{AD}AC-1.14$ (dark and light yellow) shows that the novel conformation of the loop between $\beta 4$ and $\beta 5$ alters the shape of the active site cavity compared to $Tada^*8.20$ (dark and light green) (**f**).



Extended Data Fig. 4 | Crystal structure of $T_{AD}AC-1.19$ without ssDNA and structural comparisons with $Tada^*8.20$.

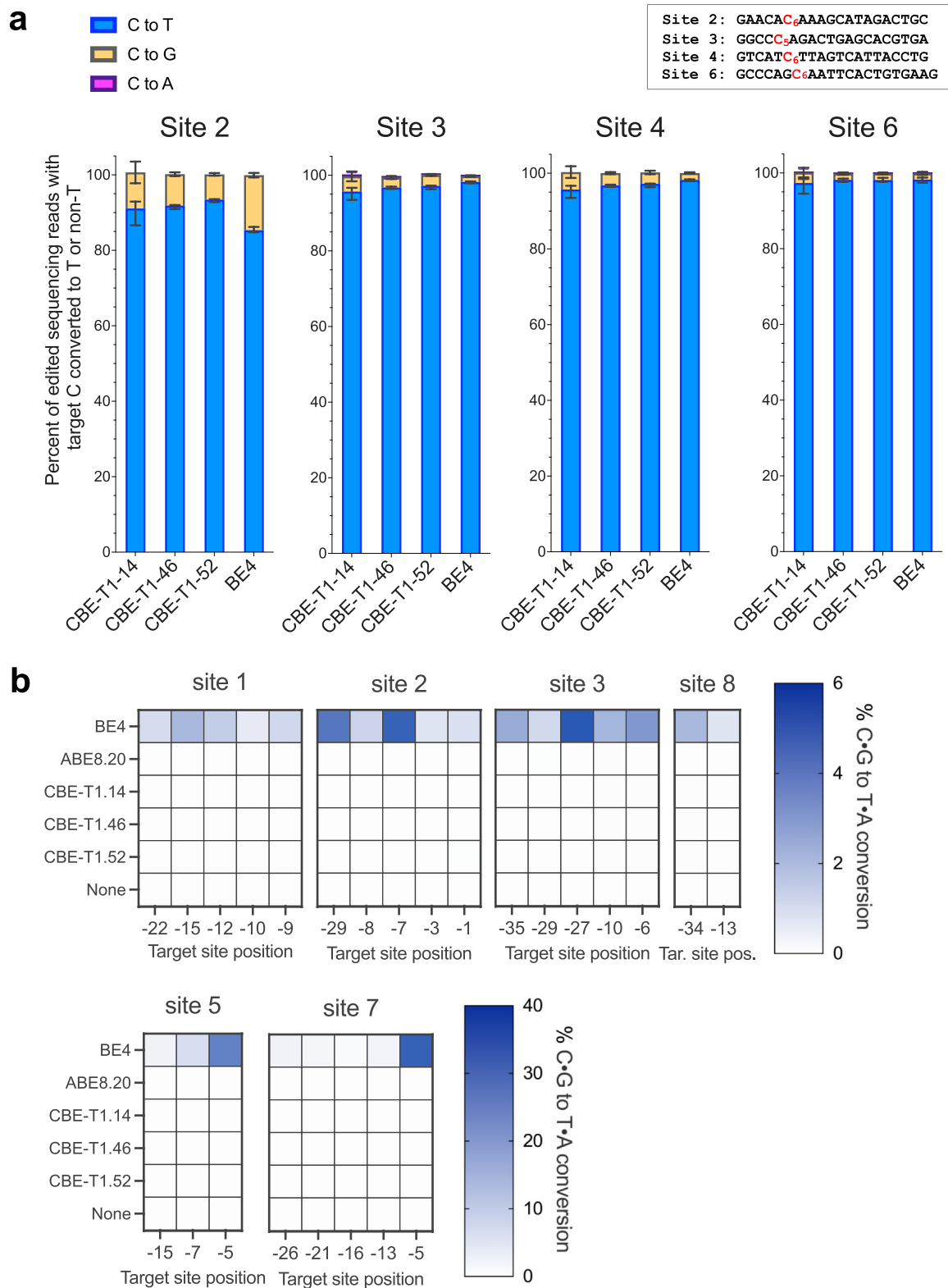
a, Overall structure of the $T_{AD}AC-1.19$ functional homodimer (dark and light pink). E27G and I49N substitutions relative to $Tada^*8.20$ are shown in orange spheres. Zinc ion is shown as a gray sphere. **b**, Overall structure of $T_{AD}AC-1.19$ monomer. C represents C-terminus. **c**, $T_{AD}AC-1.19$ active site with water (red sphere) bound to the zinc ion. H57, C87, and C90 coordinate with the zinc ion. The water molecule H-bonds (dashed lines) to the catalytic residue E59. **d–h**, Structural comparisons between $T_{AD}AC-1.19$ and $Tada^*8.20$ structures. **(d)** Superposition between substrate-free $T_{AD}AC-1.19$ (pink) and ssDNA-bound $Tada^*8.20$ (green, yellow) monomers, showing high structural similarity (RMSD of -0.9 \AA for all of the C α atoms). The main structural differences are in $\alpha 1$ -helix, the loop between $\alpha 1$ and $\beta 1$, and C-terminal $\alpha 5$ - and $\alpha 6$ -helices. **(e)** $Tada^*8.20$ has E27 side chain H-bonding (black dashed lines) to the main chains of A48, I49, and G50, and E27G substitution removes these interactions. To compensate for these critical contacts, $T_{AD}AC-1.19$ places E25 at a similar

position to the formerly occupied by E27 in $Tada^*8.20$ to make the same H-bonds (orange dashed lines) with A48, I49, and G50. E25 displacement shortens $\alpha 1$ -helix, and the loop between $\alpha 1$ and $\beta 1$ (orange), containing E27G substitution, is extended and adapts into a different conformation compared to $Tada^*8.20$ (**d**, **f**, and **g**). This results in partial unfolding of $\alpha 5$ -helix to prevent steric clash with this loop conformation and complete unfolding of $\alpha 6$ -helix (**d** and **g**). These structural changes alter the shape of the $T_{AD}AC-1.19$ active site cavity (**h**), affecting substrate binding in the active site (Fig. 2b). R26 present in $T_{AD}AC-1.19$ -loop (orange) would make close contacts (cyan dashed lines) with dC(10), adjacent to the target base d8Az(9), and dG(11) of $Tada^*8.20$ -ssDNA (**g**). I49N substitution positions the N49 side chain far from the ssDNA backbone (-9 \AA from dG(11); cyan dashed lines) (**f**), suggesting that a residue with a longer positively charged side chain like lysine would create additional contacts with the ssDNA, as observed in the $T_{AD}AC-1.14$ structure (Extended Data Fig. 3e).



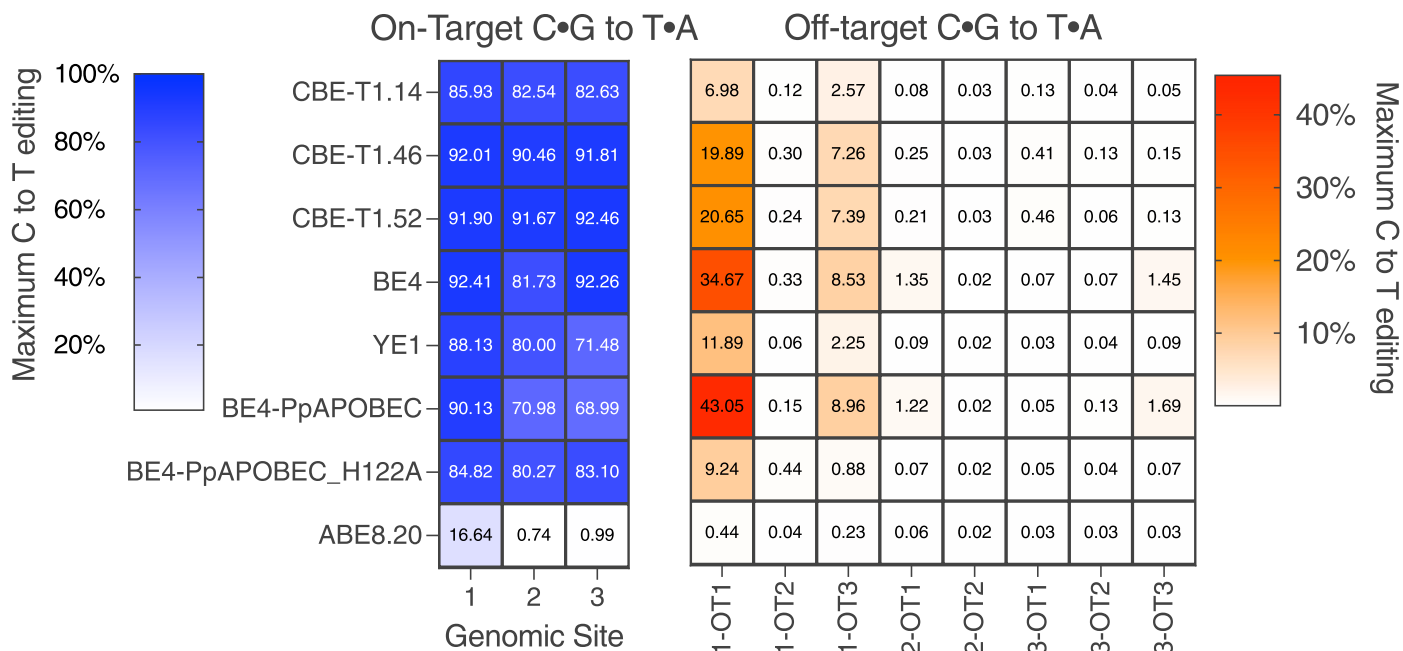
Extended Data Fig. 5 | Biochemical characterization of CBE-Ts and CBE-Ts. a, *In vitro* 24-hour end-point deamination assay to detect relative A-to-I (hAAG + APE1) and C-to-U (USERII) deamination by BE4, ABE8.20, CBE-Ts, and CBE-Ts programmed with the same guide and acting on the same dsDNA substrate. Endonuclease V, Endo V; human Alkyl Adenine DNA glycosylase, hAAG; Apurinic/aprimidinic Endonuclease I, APE1. Error bars represent standard deviation from the mean (plotted) of three independent replicates. Data

were normalized to untreated sample. Endo V detects both A-to-I and C-to-U deamination. **b**, Left, Single-turnover rates of A-to-I or C-to-U deamination of the same dsDNA substrate by BE RNP. Right, Single-turnover rates of nicking by BE RNP in the same experiment as shown on left. Pseudo-first order k_{app} rate constants obtained by fitting to single exponential are reported (mean \pm s.d., $n = 3$ independent replicates).



Extended Data Fig. 6 | Product purity of CBE-Ts relative to BE4. **a**, Product distribution of sequencing reads mapped as edited for core CBE-Ts and BE4, in which the specified target cytosine (highlighted in red) is mutated. Values were determined from transfection of HEK293T with mRNA at saturating conditions. Values and error bars reflect the mean and SD at $n = 3$ independent biological replicates performed on different days. **b**, Color map of maximum C•G to

T•A conversions outside and 5' of the protospacer target window. Target site positions where $>0.8\%$ C-to-T editing was detected for any editor are included. Values were determined from transfection of HEK293T with mRNA encoding editors or controls at saturating conditions, with $n = 4$ independent biological replicates performed on different days.



Extended Data Fig. 7 | Guide-dependent off-target evaluation of CBE-T. Color map of % maximum on-target C•G to T•A conversion at genomic sites and % maximum C•G to T•A conversion at their corresponding off-target sites

in HEK293T cells transfected with mRNA encoding editor (or control) plus synthetic sgRNA at saturating conditions. Median values were derived from n = 3 independent biological replicates performed on different days.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Illumina MiSeq Control Software (v3.1), Thermo Scientific Chromeleon (v7.2.9)

Data analysis

Illumina bcl2fastq (v2.20.0.422), trimmomatic (v0.39), bowtie2 (v2.35), samtools (v1.9), bam-readcounts (v0.8), R (v3.4.3), FlowJo (v10.6.1), bwa-mem2 (v2.2.1), Picard (v2.21.7), GATK (v4.1.4.1), LoFreq (v2.1.5), XDS (v2022Jan01), Aimless (v0.7.7), Phaser(v2.8), Phenix (v1.19.2-4158), Coot (v7766), PyMol (v2.5)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Next-generation sequencing data underlying all experiments are deposited in the NCBI Sequence Read Archive (PRJNA869750). The atomic coordinates and structure factors have been deposited in the Protein Data Bank as entries: 8E2P, 8E2Q, 8E2R, and 8E2S.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	<input type="text" value="No sample size calculations were performed in advance."/>
Data exclusions	<input type="text" value="No data were excluded from the study."/>
Replication	<input type="text" value="All measurements reported in the study were collected in at least three independent biological replicates, except for initial engineering screens"/>
Randomization	<input type="text" value="Randomization was not considered, as all major experimental variables were fully under the control of the experimenters."/>
Blinding	<input type="text" value="Blinding was not considered, as all data analysis was performed following uniform automated processes."/>

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

Methods

n/a	Involvement in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input type="checkbox"/>	<input checked="" type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used	<input type="text" value="B2M: Supplied by Biolegend. Clone 2M2. Catalog #316306. Lot #B335217 (WGS) and #B339354 (T cells)
TCRa/b: Supplied by Biolegend. Clone IP26. Lot #B316067
PD-1: Supplied by Biolegend. Clone NAT105. Lot #B349676"/>
Validation	<input type="text" value="All antibodies were validated by the suppliers by staining using isotype controls."/>

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)	<input type="text" value="Hek293T: American Type Cell Culture Collection (ATCC)"/>
Authentication	<input type="text" value="Done by supplier"/>
Mycoplasma contamination	<input type="text" value="Tested negative for mycoplasma contamination by the supplier, ATCC."/>
Commonly misidentified lines (See ICLAC register)	<input type="text" value="The cell line used is not listed as commonly misidentified."/>

Plots

Confirm that:

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	To assess editing efficiency, one million cells were taken from culture five days post electroporation and stained with the primary anti-human antibodies listed in the section above. The cells were incubated with antibodies according to manufacturer's instructions for 15 min at room temperature, washed, then analyzed on the flow cytometer.
Instrument	Attune NxT Flow Cytometer
Software	FlowJo (v10.8.1)
Cell population abundance	Not applicable, as no sorted populations were analyzed separately
Gating strategy	Linear FSC-A versus linear SSC-A was used to select the lymphocyte fraction from bulk cell suspension. Within the lymphocyte gate, FSC-A versus FSC-H was used to gate on single cells. Then within the live cell gate, a histogram gate was utilized on a negative control population to delineate the boundary between positive and negative cells. The values in the histogram represent surface protein knockout in the context of gene editing experiments.

- Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.