Nucleic Acids Research, 2024, 1–14
https://doi.org/10.1093/nar/gkae1066
Molecular and Structural Biology

Dimerization of the deaminase domain and locking interactions with Cas9 boost base editing efficiency in ABE8e

Pablo R. Arantes[1,†], Xiaoyu Chen[2,3,†], Souvik Sinha[1,†], Aakash Saha[1,†], Amun C. Patel[1], Matthew Sample[1], Łukasz Nierzwicki[1], Audrone Lapinaite[3,4,*] and Giulia Palermo[1,5,*]

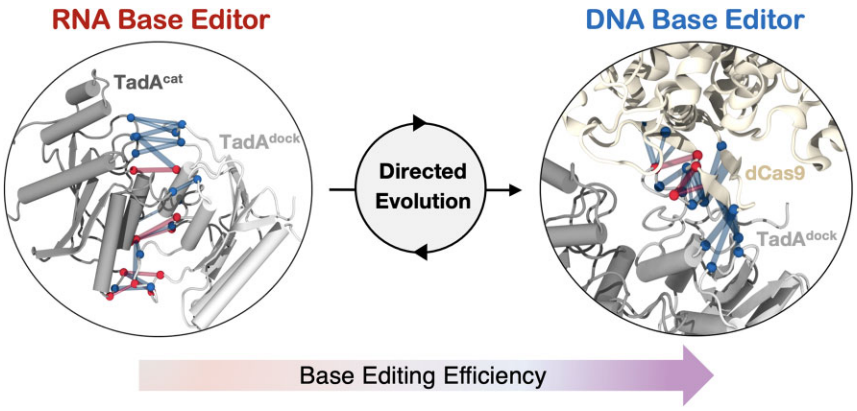
[1]Department of Bioengineering, University of California Riverside, 900 University Avenue, 92512 Riverside, CA, USA
[2]School of Molecular Sciences, Arizona State University, 551 E University Dr, Tempe, AZ 85281, USA
[3]Gavin Herbert Eye Institute - Centre for Translational Vision Research, University of California Irvine School of Medicine, 850 Health Sciences Rd, Irvine, CA 92617, USA
[4]Department of Ophthalmology, University of California Irvine School of Medicine, 850 Health Sciences Rd, Irvine, CA 92617, USA
[5]Department of Chemistry, University of California Riverside, 900 University Avenue, 92512 Riverside, CA, USA

[*]To whom correspondence should be addressed. Tel: +1 951 827 4303; Email: giulia.palermo@ucr.edu
Correspondence may also be addressed to Audrone Lapinaite. Tel: +1 510 384 2331; Email: alapinai@hs.uci.edu
[†]The first four authors should be regarded as Joint First Authors.

Abstract

CRISPR-based DNA adenine base editors (ABEs) hold remarkable promises to address human genetic diseases caused by point mutations. ABEs were developed by combining CRISPR-Cas9 with a transfer RNA (tRNA) adenosine deaminase enzyme and through directed evolution, conferring the ability to deaminate DNA. However, the molecular mechanisms driving the efficient DNA deamination in the evolved ABEs remain unresolved. Here, extensive molecular simulations and biochemical experiments reveal the biophysical basis behind the astonishing base editing efficiency of ABE8e, the most efficient ABE to date. We demonstrate that the ABE8e's DNA deaminase domain, TadA8e, forms remarkably stable dimers compared to its tRNA-deaminating progenitor and that the strength of TadA dimerization is crucial for DNA deamination. The TadA8e dimer forms robust interactions involving its R98 and R129 residues, the RuvC domain of Cas9 and the DNA. These locking interactions are exclusive to ABE8e, distinguishing it from its predecessor, ABE7.10, and are indispensable to boost DNA deamination. Additionally, we identify three critical residues that drive the evolution of ABE8e toward improved base editing by balancing the enzyme's activity and stability, reinforcing the TadA8e dimer and improving the ABE8e's functionality. These insights offer new directions to engineer superior ABEs, advancing the design of safer precision genome editing tools.

Graphical abstract



Introduction

The bacterial immune systems, CRISPR (clustered regularly interspaced short palindromic repeats)-Cas (CRISPR-associated proteins), have ushered in a new era of genome editing, offering powerful and easily customizable tools for precise genetic manipulation (1). These tools rely on the special

Received: August 8, 2024. **Revised:** October 16, 2024. **Editorial Decision:** October 18, 2024. **Accepted:** November 1, 2024

feature of noncoding RNA-guided CRISPR-Cas effector complexes: their ability to introduce double-stranded DNA breaks (DSBs) at specific genomic DNA loci in a programmable manner (2–4). The outcome of the CRISPR-Cas-mediated genome editing depends on which DNA repair pathway addresses the introduced DSB. Homology-directed repair (HDR) produces accurate DNA modifications, but it is inefficient since it is limited to actively dividing cells and requires a donor DNA template (5).

DNA base editing is one of the newest CRISPR-Cas technologies that allows direct and irreversible conversion of one base pair to another at a specific genomic site without relying on DSBs or HDR (6). This technology holds tremendous promise in treating a wide range of genetic diseases associated with single-nucleotide polymorphisms (SNPs) (7), due to its ability to precisely convert **A•T** base pairs to **G•C** (by adenine base editors, ABEs) and **C•G** to **T•A** (by cytosine base editors, CBEs) (8–13). ABEs are chimeric enzymes composed of a catalytically impaired Cas9 [Cas9 nickase (nCas9)] from *Streptococcuspyogenes* and a deoxyadenosine deaminase (Figure 1) (14). These machineries use the nCas9–guide RNA (gRNA) complex to locate DNA sequences of interest within the genome, and a deaminase enzyme to perform A-to-G modifications via deamination of deoxyadenosine to deoxyinosine at specific single-stranded DNA (ssDNA) sites.

The deaminase domain of ABEs' was evolved from an *Escherichia coli* transfer RNA (tRNA) adenosine deaminase (wt-TadA), which naturally deaminates adenosine in the anticodon loop of tRNA (9,15,16). During multiple rounds of directed evolution, mutations were introduced to the wt-TadA, leading to evolved TadA enzymes (TadA*). When conjugated with nCas9, these enzymes demonstrate increasing efficiency in deaminating deoxyadenosine in ssDNA across generations (measured as the number of **A•T** converted to **G•C** at different genomic sites) (9,17). Among the ABEs used for precision genome editing in various cells and organisms, ABE8e is one of the most efficient and widely used (with DNA base editing efficiency of ∼80–90%) (17,18). In ABE8e, the deaminase domain (TadA8e) contains a total of 20 mutations with respect to the wt-TadA and is fused to nCas9 likely forming homodimers in *trans* (Figure 1) (14,17). Compared to its closest predecessor, ABE7.10, ABE8e exhibits significant advancements. The additional eight mutations that were introduced to ABE7.10 to yield ABE8e not only enhance the efficiency of DNA editing in cells, soaring from ∼58% to ∼80–90% but also drastically improve the deamination rates (∼580-fold or more) measured *in vitro* under single turnover conditions (14,17).

The cryo-EM structure of ABE8e in a substrate-bound state captured dCas9 (catalytically dead Cas9 with D10A and H840A substitutions) fused with the TadA8e deaminase (Figure 1 and Supplementary Figure S1) (14). dCas9–gRNA complex binds the DNA by matching one strand with the gRNA (the target strand, TS) while the other non-target strand (NTS) is in a single-stranded state. TadA8e is captured as a dimer with one domain deaminating deoxyadenosine within the DNA NTS ($TadA^{cat}$, red) and another domain docking to the dCas9 protein ($TadA^{dock}$, pink). While this structure provides critical insights into the mechanism of DNA deamination, the biophysical rationale behind the ABE8e base editing efficiency remains unresolved. Additionally, it is perplexing how the eight amino acid substitutions lead to a sharp increase in the deamination efficiency of ABE8e compared to its predecessor ABE7.10. Nevertheless, this efficiency comes with elevated off-target editing rates (17). Thus, understanding the molecular basis of ABE8e's enhanced DNA base deamination is crucial for designing more precise genome editors and accelerating their translation for biomedical applications.

Here, we establish the biophysical basis behind the improved DNA base editing efficiency of the ABE8e enzyme through an integrative approach that combines extensive molecular dynamics (MD) simulations and enhanced sampling approaches with biochemical and biophysical experiments. An overall ensemble comprising multi-microsecond (μs) MD simulations was combined with ensemble Förster Resonance Energy Transfer (FRET) experiments, thermal stability assays, and single turnover kinetics.

We reveal that TadA8e dimerization and its unique juxtaposition to dCas9 are pivotal for the efficient DNA deamination in ABE8e. Indeed, the TadA8e dimer is more stable than that formed by its wt-TadA progenitor, which allows TadA8e to establish crucial interactions with DNA and dCas9, contributing to improved DNA deamination efficiency. Two residues within the TadA8e docking domain, R98 and R129, significantly increase their interactions with the dCas9–DNA complex in comparison to ABE7.10 and are required for efficient DNA deamination. Moving from ABE7.10 to ABE8e, the interplay between two critical substitutions, T111R and D119N, is shown to increase DNA deamination rates and balance the stability of the complex, while H122N contributes to stabilizing the TadA dimer, ultimately evolving the complex toward improved functionality. These findings hold significant implications for the design of ABE enzymes and open new directions for developing improved base editors.

Materials and methods

Structural models

Molecular simulations were based on the cryo-EM structure of the ABE8e complex (PDB 6VPC) (14) at 3.20 Å resolution, comprising a catalytically dCas9 from *S. pyogenes* and the TadA8e enzyme. Using this structure, which preserves the *bona fide* configuration of the TadA enzymes bound dCas9, four model systems were built: ABE0.1 (where TadA is WT), ABE7.10, ABE8e and an ABE8e including the TadA8e heterodimer, in which the catalytic domain is the evolved TadA8e and the docking domain is the wt-TadA (referred to as ABE8e hetero). The X-ray structure of wt-TadA (PDB 1Z3A (16), 2.03 Å resolution) was used to model the wt-TadA and heterodimer docking domain. The TadA7.10 catalytic domain was modelled by mutating the ABE8e structure (PDB 6VPC). Missing residues were added via AlphaFold2 (19). All systems were embedded in explicit waters, and Na^+ and Cl^- ions were used to reach physiological concentration of 150 mM. This resulted in periodic cells (∼133.68 × 148.43 × 224.94 $Å^3$) with ∼500 000 atoms each. Details are present in the Supplementary Text.

Molecular dynamics simulations

MD simulations were performed using a simulation protocol tailored for RNA/DNA nucleases (20), which we also used in studies of CRISPR-Cas systems (21–23), embracing the use of the Amber ff19SB (24) force field, and the OL15 (25) and OL3 (26) corrections for DNA and RNA. The TIP3P (27) model was used for water molecules. We

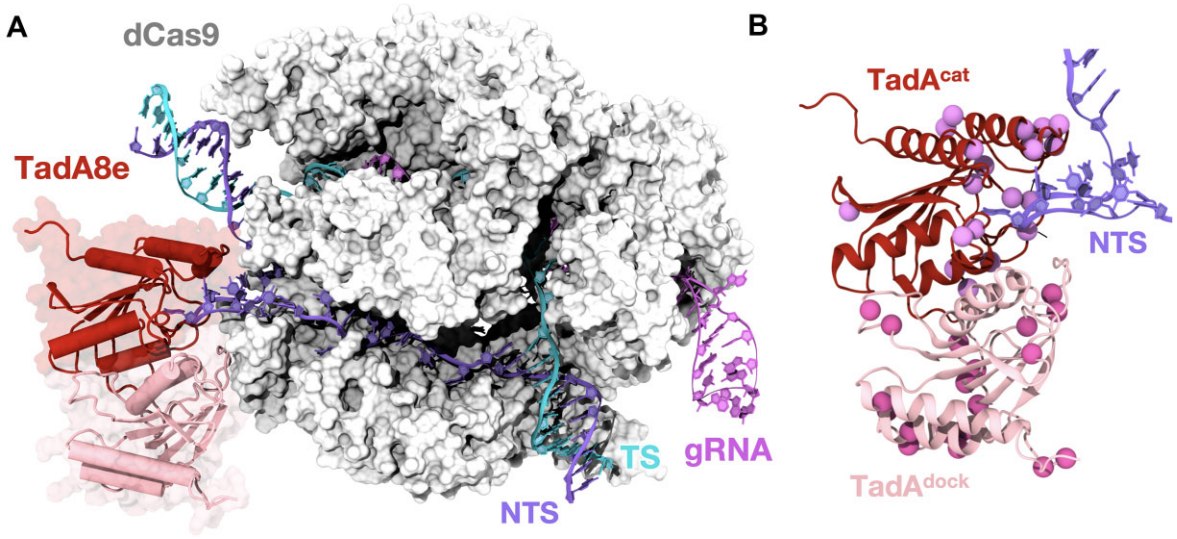


Figure 1. Overview of the adenine base editor ABE8e from *Streptococcus pyogenes*. (**A**) Cryo-EM structure of the catalytically dead Cas9 (dCas9) from *S. pyogenes* fused to the adenosine deaminase TadA8e (PDB 6VPC) (14). The fusion of dCas9 and TadA8e (i.e. ABE8e) is shown as a molecular surface. The dCas9 protein binds a gRNA (magenta), unwinds double-stranded DNA, base-pairs the gRNA with the DNA TS (cyan) and displaces the other NTS (violet). TadA8e forms a dimer with one domain deaminating deoxyadenosine within the DNA NTS [a catalytic domain, $TadA^{cat}$ (red)] and another domain docking between the $TadA8e^{cat}$ and dCas9 [a docking domain, $TadA^{dock}$ (pink)]. (**B**) Close-up view of TadA8e shows an array of amino acid substitutions (spheres, magenta) introduced to the wt-TadA through directed evolution.

re-parametrized the Zn^{2+} ion and the active site residues using Quantum Mechanical (B3LYP (28)/6–31G* (29)) calculations (Supplementary Table S1), following the protocol by Li and Merz (30). Upon energy minimization, the systems were heated from 0 to 100 K in two consecutive simulations in the canonical NVT ensemble of 5 ps each, imposing positional restraints of 100 kcal mol^{-1} $Å^2$ on the protein, the nucleic acids and the TadA active site. Further heating to 200 K occurred in a 100 ps run in the isothermal-isobaric ensemble (NPT), reducing restraints to 25 kcal/mol $Å^2$. Restraints were removed, heating to 300 K in a 500 ps NPT simulation. These simulations were performed using a 1 fs time step. The time step was subsequently increased to 2 fs for further equilibration and production runs. Hydrogen bonds were constrained with the SHAKE algorithm (31). After ∼600 ps of equilibration, ∼10 ns NPT simulations stabilized the systems' density around 1.01 g/cm^3. The temperature was kept constant at 300 K via Langevin dynamics, with a collision frequency $\gamma = 1$ ps^{-1}. The pressure was controlled in the NPT simulations by coupling the system to a Berendsen barostat (31) at a reference pressure of 1 atm and with a relaxation time of 2 ps. Subsequently, production runs were carried out in the NVT ensemble, reaching ∼3 μs for each system and in triplicate (i.e. ∼9 μs of aggregate sampling per system, totalling ∼36 μs across four systems). MD simulations replicates were performed starting from different configurations and velocities initialized according to the Maxwell-Boltzmann distribution at physiological temperature. These simulations were performed using the GPU-empowered version of AMBER 22 (32). Details are present in the Supplementary Text.

Umbrella sampling simulations

The Umbrella Sampling (US) (33) method was used to compute the free energy profiles associated with the dimerization of TadA enzymes. In this method, several simulations (US windows) are run in parallel with additional harmonic bias potential applied to selected Reaction Coordinates (RC):

$$V(RC) = \frac{k}{2}(RC(t) - RC^*)^2 \quad (1)$$

where $V(RC)$ is the value of the bias potential, k is a bias force constant, $RC(t)$ is the value of RC at given time t and RC^* is the reference value of RC. By using different RC^* values in each US window, one can sample the biased probability distribution $p_b(RC)$ along the whole RC range of interest. Thanks to a reweighting scheme providing unbiased probabilities, the free energy of a process along the *RCs* can be computed (*vide infra*).

We performed US simulations at 300 K using two RCs. RC1 was the inter-monomeric distance considering the center of mass of each TadA unit. RC2 was the difference in root-mean-square deviation (*RMSD*) of all the Cα atoms with respect to the experimental structures (i.e. PDB 1Z3A (16) for wt-TadA and PDB 6VPC (14) for TadA8e). The combination of these RCs enabled us to describe the dimerization process while keeping the experimental structures as a reference. Restraint forces were applied on the two RCs as follows. RC1 was discretized into bin sizes of 0.5 Å with a collective force constant k of 17.5 kcal/mol $Å^{-2}$. RC2 was divided into bin sizes of 0.5 Å *RMSD* difference with a collective force constant k of 25 kcal/mol $Å^{-2}$. Each window was run for ∼15 ns. Overall, we simulated 51 (RC1) * 31 (RC2) = 1 581 windows, resulting in a collective ensemble of ∼25 μs of sampling for each system. Two independent sets of two-dimensional (2D) US simulations were performed considering the wt-TadA and TadA8e. The free energy profiles were computed using the Variational Free Energy Profile (VFEP) method (34), which uses the maximum likelihood principle to construct free energy profiles. This method was applied upon removing the initial one-third of the US trajectories from each window, corresponding to the system's relaxation. Analysis of the minimum free energy pathway (MFEP) between the dimeric and monomeric states was performed using the MEPSA code (35), which employs

an approach similar to the Dijkstra algorithm to trace the MFEP between nodes of interest. The convergence of the free energy profiles was evaluated by plotting the MFEPs, considering the remaining ∼10 ns runs from each US window, for ∼5 ns through ∼10 ns (Supplementary Figure S2). The error estimation on the MFEPs was performed using the Monte Carlo bootstrap error analysis.

Alchemical free energy perturbation

We performed alchemical free energy perturbation (FEP) simulations using lambda (λ) dynamics (36), to compute the relative free energy change (i.e. $\Delta\Delta G$) in the ABE7.10 complex while introducing the ABE8e mutations (i.e. A109S, T111R, D119N, H122N, Y147D, F149Y, T166I, D167N). With our alchemical FEP simulations, we aimed to describe the effect of amino acid mutations on the thermal stability of the complex. Therefore, we utilized an alchemical FEP method to calculate the effect of mutations on the folding free energy (37,38), since changes in thermal stability result from modifications in the free energy of folding. Consistently with a thermodynamic cycle (details are present in the Supplementary Text), state *A* was evolved into state *B*, representing the initial and the final residues respectively, via a series of transition states following the alchemical free energy formalism. A lambda (λ) dynamic variable was used to define the transitioning thermodynamic states of the system, where $\lambda = 0$ and $\lambda = 1$ represent the initial and final states respectively. The distribution of the λ parameter was optimized using a gradient descent algorithm to maximize the probabilities of exchange between adjacent states (https://gitlab.com/KomBioMol/converge_lambdas) (39). In detail, the neighbouring λ-windows were allowed to exchange their configurations every 0.5 ps according to the Metropolis criterion, and the values of λ were optimized to achieve the acceptance rate of at least 15%. Since each intermediate λ state is technically a hybrid between the *A* and *B* end point states, we generated their dual coordinates and topologies using the PMX server (http://pmx.mpibpc.mpg.de/) (40). The relative free energy changes (i.e. $\Delta\Delta G$) for the transformation were computed using the Multistate Bennett Acceptance Ratio (MBAR) (41,42) method to integrate the free energies over the different λ values using flamel (https://github.com/alchemistry/flamel) (43,44). Using this method, we computed relative $\Delta\Delta G$ for each single-point mutation introduced through directed evolution in ABE7.10 leading to ABE8e (i.e. A109S, T111R, D119N, H122N, Y147D, F149Y, T166I, D167N). Alchemical FEP simulations were performed across the lambda values for each mutation, running ∼50 ns for each window and totalling ∼8 μs sampling. Details are present in the Supplementary Text.

Preparation of ABE-sgRNA samples

ABE proteins were expressed and purified as described previously by Lapinaite *et al*. (14). Single guide RNA (sgRNA) was transcribed overnight at 37°C using in-house prepared T7 polymerase and purified by using 7 M urea, 10% 19:1 polyacrylamide gel electrophoresis, followed by crush-and-soak extraction overnight at 4°C and ethanol precipitation as described previously. The ABE-sgRNA (RNP) complexes were formed by incubating the protein (Supplementary Table S2) with the refolded sgRNA at a molar ratio of 1:1.5 (protein to sgRNA) in RNP assembly buffer (20 mM Tris–HCl, 200 mM KCl, 5 mM $MgCl_2$, 5% [v/v] glycerol, 2 mM DTT, pH 7.5) at room temperature for 20 min. Details are present in the Supplementary Text.

Single-turnover *in vitro* DNA and RNA deamination assays

The single turnover DNA deamination reactions were conducted with 2 μM ABE RNPs in a deamination buffer consisting of 20 mM Tris–HCl (pH 7.5), 100 mM KCl, 5% (v/v) glycerol, 2.5 mM $MgSO_4$ and 2 mM dithiothreitol (DTT). These reactions were initiated by introducing fluorescently labelled double-stranded DNA (dsDNA) or hairpin RNA resembling anticodon hairpin of $tRNA^{Arg2}$ to achieve a final concentration of 7.5 nM. The reactions were incubated at 37°C. 20 μ, aliquots were transferred into 30 μl preheated water at 95°C at different time intervals (0, 1, 2, 5, 10, 20 and 60 min for ABE8e and ABE0.1 variants and 0, 1, 2, 5, 10, 20, 60, 180, 360 and 480 min for ABE 7.10 variants) and quenched by heating at 95°C for 5 min. For detection of DNA deamination the quenched and purified reactions were incubated with 20 units of EndoV (NEB) at 37°C for 1.5 h. For detection of RNA deamination, human EndoV (hEndoV; purified in-house) was added into the reaction at a final concentration of 0.1 μM. The reactions were incubated at 37°C for 1 h. Samples were resolved on 15% (v/v) denaturing (29:1) polyacrylamide gel. The gel was visualized by using the fluorescence imaging instrument Amersham Typhoon (GE Healthcare). The experiments were conducted in three separate replicates, and the levels of uncleaved and cleaved DNA or RNA were evaluated using ImageQuant TL Software from GE Healthcare. Details are present in the Supplementary Text.

Thermal stability assay using CD spectroscopy

The sample was prepared and loaded into the 1 mm path length microcuvette (Hellma 110–1-40, style 110-QS). The measurements were performed using the Jasco J-815 spectropolarimeter. The ellipticity at the wavelength of 208 nm was recorded from 20°C to 95°C at the ramp rate of 1°C/min. The measurement was performed with an integration time of 2 s, using a bandwidth of 1 nm. The ellipticity was plotted as a function of temperature and fitted with a Boltzmann sigmoidal equation. The melting temperature, T_m, was determined by the inflection point.

Fluorescence resonance energy transfer (FRET) measurements

Briefly, the labelling reactions of ABEs were carried out in SEC buffer and contained 10 μM protein along with 200 μM Cy3- and 1.2 mM Cy5-maleimide (APExBIO) (1:6 ratio of donor to acceptor). The labelling reaction was initiated by gentle rocking in the dark for 2 h at room temperature, followed by overnight incubation at 4°C. Reactions were quenched by adding 10 mM DTT and unreacted dyes were removed via SEC. The purified labelled ABE samples were concentrated, flash-frozen in liquid nitrogen and stored at −80°C. All fluorescence measurements were conducted at room temperature in SEC buffer. Fluorescence measurements were performed on ABE-sgRNA samples containing various concentrations (400, 40 and 4 nM) of Cy3/Cy5-labelled ABE0.1 [wt-TadA(T133C/C141S)-dCas9(C80S/C574S)] or ABE8e [TadA8e(T133C/C141S)-dCas9(C80S/C574S)]. Fluorescence measurements were performed using a 10 mm path-

length quartz cuvette (Hellma Analytics), a QM-4/2005SE spectrofluorometer and a PTI 814 Photomultiplier, using 5 nm slit widths and 0.5 s integration time. For each sample, the fluorescence emission spectra were collected, in which the sample was excited at 500 nm and emission was recorded from 510 to 750 nm in 1 nm increments. Data were collected and processed using PTI's FeliX32 software. Experiments were performed at least three times. The FRET spectra obtained on samples at decreasing concentrations were normalized to the maximum Cy3 donor peak at 570 nm to eliminate the effect of sample dilution. Spectra were smoothed using the Savistsky–Golay method and plotted using Prism (GraphPad Software). Details are present in Supplementary Text.

Results

Multiple rounds of directed evolution have stabilized the dimeric state of TadA

ABEs have been developed to deaminate DNA by mutating the wt-TadA, a homo-dimeric enzyme with each monomer able to catalyse adenosine deamination in tRNA. The tRNA-binding pocket is formed by both monomers (16,45), suggesting that dimerization might be important for efficient tRNA deamination. To establish whether multiple rounds of directed evolution have impacted the dimerization of TadA and to characterize its role in DNA adenine base editing, we performed free energy simulations of TadA8e and compared it to wt-TadA. We used a two-dimensional Umbrella Sampling (2D US) (33) method to describe the dimerization process of the TadA enzymes along two reaction coordinates: the inter-monomeric distance considering the center of mass of each TadA unit and the difference in root-mean-square deviation (RMSD) of the Cα atoms with respect to the experimental structures (i.e. PDB 1Z3A (16) for wt-TadA and PDB 6VPC (14) for TadA8e). Combining these reaction coordinates enabled us to describe the dimerization process while keeping the experimental structures as a reference (details in the Materials and methods section). These runs were performed on the TadA dimers in solution, without the dCas9, describing the potential of TadA to dimerize per se. For both TadA variants, we rigorously collected the population states from dimer to monomer (and vice versa) through ∼25 μs of converged sampling (Supplementary Figure S2). The 2D free energy surfaces reveal that both wt-TadA and TadA8e reach a stable minimum that corresponds to the experimental dimeric state (Figure 2A), with TadA8e exhibiting a deeper minimum than the wt-TadA. Analysis of the minimum free energy paths (details in the Materials and methods section) also reveals that to escape the minimum of the dimeric state toward the monomer, the wt-TadA is favoured by > 10 kcal mol^{-1}, compared to TadA8e (Figure 2B). This indicates that TadA8e is more stable in the dimeric form than its wt-TadA predecessor.

To experimentally assess the dimerization of TadA in the presence of dCas9, we performed FRET experiments. We generated FRET constructs of the ABE0.1 (containing a single wt-TadA domain) and ABE8e (containing a single TadA8e domain), where all cysteines (except C87 and C90 which are present in the active site of TadA) were substituted to serine and T133 of TadA* was mutated to cysteine for site-specific labelling with maleimide derivatives of Cy3 and Cy5 fluorophores (Supplementary Figure S3A and Supplementary Table S1). The control labelling reaction of cysteine-free ABE8e confirmed site-specific labelling of the ABE8e FRET construct (Supplementary Figure S3B). The Cy3/Cy5-labelled ABE0.1 and ABE8e are properly folded (Supplementary Figure S3C), and the dye-labelling does not affect the deamination rates of ABE0.1 and ABE8e (Supplementary Figure S3D–G). Our ensemble FRET experiment shows that both wt-TadA and TadA8e enzymes exist in dimeric form when fused to dCas9 (similar to what was observed for the enzymes in solution through free energy simulations), as evidenced by the presence of a peak at ∼675 nm in the fluorescence emission spectra of Cy3/Cy5-labelled proteins when excited at 530 nm. To compare the dimerization affinity of wt-TadA and TadA8e, we conducted FRET measurements on dye-labelled samples at varying protein concentrations (Figure 2C). These measurements show that TadA8e forms highly stable dimers, with K_d values below 4 nM (likely in the picomolar range). In contrast, the wt-TadA has a propensity to dissociate, suggesting a higher K_d compared to TadA8e (likely in the nanomolar range), as evidenced by the decrease of Cy5 fluorescence emission intensity as the enzyme concentration decreases. This finding is in qualitative agreement with our computations, showing that the dimerization in TadA8e is stronger compared to the wt-TadA (Figure 2A and B). The discrepancy in the quantitative measure may stem from several factors, including the choice of reaction coordinates, water model and the time scales accessible in the simulations. Taken together, these findings suggest that multiple rounds of directed evolution have stabilized the dimeric state of TadA8e, which could critically contribute to their increasing base editing efficiency.

DNA–TadA–Cas9 interactions boost DNA deoxyadenosine deamination by ABE8e

To better understand the importance of TadA dimerization when fused to dCas9 and to characterize the interactions that the TadA enzymes establish with Cas9 and DNA, we performed extensive multi-μs MD simulations of the ABE complexes. We considered ABE0.1 (where TadA is WT), ABE8e and ABE7.10, the closest predecessor of ABE8e with ∼580-fold lower DNA deamination rate (14,17). The cryo-EM structure of the ABE8e complex (PDB 6VPC) (14), containing dCas9 and the TadA8e homodimer, was used as a template, as it preserves the *bona fide* configuration of the TadA enzymes fused to dCas9. The ABE0.1, ABE8e and ABE7.10 model systems, comprising ∼500k atoms, were simulated for ∼9 μs each (in three replicates of ∼3 μs). An in-depth analysis of our all-atom MD simulations was carried out to describe the interactions established between the TadA domains themselves, as well as those that the TadA enzymes form with dCas9 and the DNA. We computed the pairwise interactions among residues in the ABE complexes using the MM/GBSA method (46–48), assessing the relative interaction energies between different systems and comparing the gain/loss in interaction strength (details are present in Supplementary Text).

We first compared the ABE0.1 and ABE8e systems, using Sankey plots to report the difference in pairwise interaction energy (ΔE) between them, computed at the triad of interfaces between the TadA units, dCas9 and the DNA NTS (Figure 3A). In these plots, residues are connected through edges, whose width is proportional to ΔE. Interactions that gain strength in ABE8e ($\Delta E < 0$, blue) are compared to those that gain strength in ABE0.1 ($\Delta E > 0$, red). This analysis reveals an increased number of interactions that gain strength at the

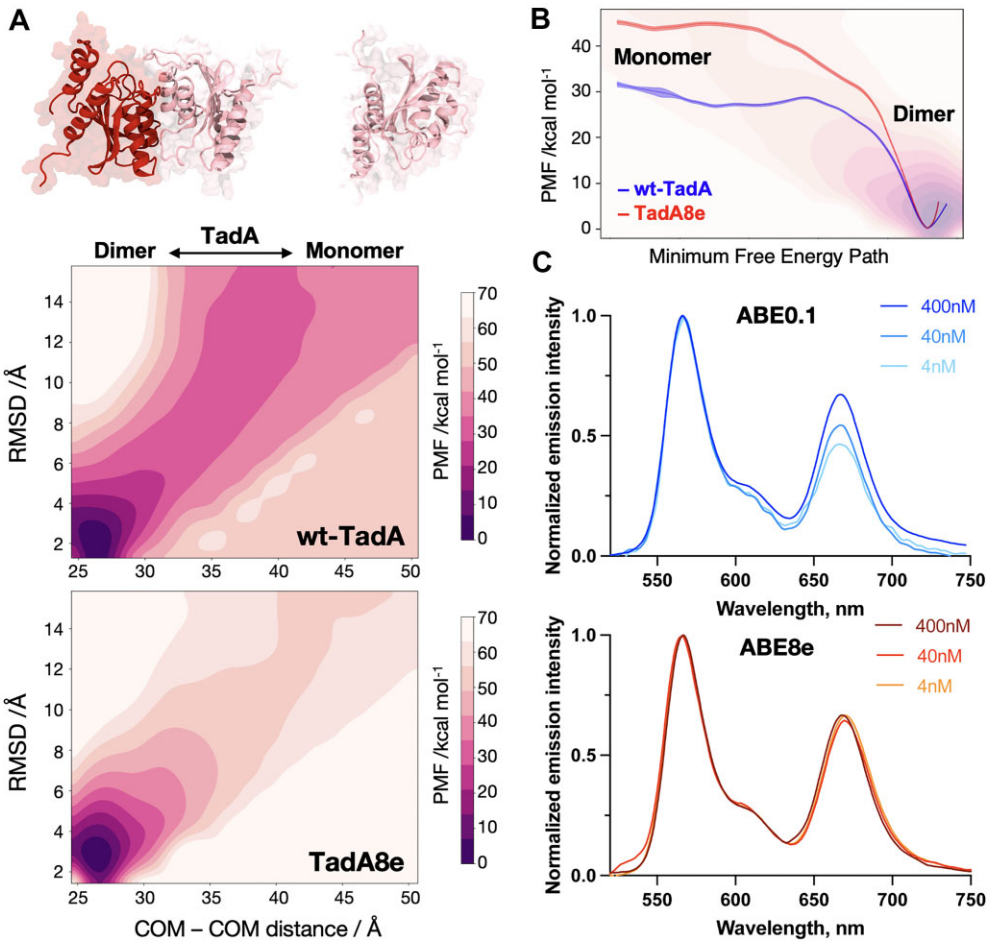


Figure 2. Dimerization of TadA enzymes. (**A**) Potential of TadA enzymes to dimerize in solution. The potential of mean force (PMF, in kcal mol^{-1}) for the dimerization of the wt-TadA (top) and TadA8e (bottom) enzymes is computed through 2-D Umbrella Sampling simulations along the inter-monomeric distance considering the center of mass of each TadA unit, and the root-mean-square deviation of the Cα atoms between the monomeric and dimeric states. The dimer and monomer states of TadA enzymes are indicated and shown on top of the plots. (**B**) Minimum free energy paths for the transition of the wt-TadA and TadA8e enzymes from monomer to dimer, computed from 2-D Umbrella Sampling simulations (see the Materials and methods section). The 2-D free energy surface of TadA8e dimerization is superposed in transparency to highlight the minimum corresponding to the dimeric state. The 2-D free energy surface of the TadA8e dimerization is shown in transparency to indicate the minimum corresponding to the dimeric state. (**C**) Fluorescence emission spectra of Cy3/Cy5-labelled ABE0.1 (with single wt-TadA, top) and ABE8e (with single TadA8e, bottom) across varying concentrations, normalized to the maximum Cy3 donor peak at 570 nm to eliminate the effect of sample dilution. The distinct reduction of FRET signal observed exclusively in the spectra of ABE0.1, as protein concentration decreases, provides a clear indication that TadA8e has a higher affinity for dimerization compared to wt-TadA.

TadA dimeric interface in TadA8e, compared to the wt-TadA (Figure 3A and Supplementary Figure S4). This is in line with our ensemble FRET measurements indicating that TadA8e forms more stable dimers than wt-TadA (Figure 2C). Interestingly, ABE8e also shows an increase in interactions that gain strength at the interface of the TadA8e docking domain and the DNA NTS, as well as the TadA8e docking domain and the RuvC domain of Cas9, with respect to ABE0.1. This underscores that ABE8e could strengthen the interactions at these interfaces as a possible strategy for improved catalytic efficiency.

We then compared ABE8e with ABE7.10, the closest predecessor of ABE8e with ∼580-fold lower DNA deamination rate (14,17) (Figure 3B and Supplementary Figure S5). We observed that, while the number of interactions that gain strength at the TadA dimeric interface is comparable, ABE8e increases the number of interactions that gain strength between the TadA docking domain, the DNA NTS and the RuvC domain of dCas9, compared to ABE7.10. In Figure 3C, the relative pairwise ΔE between ABE8e and ABE7.10 are plotted on the 3D structure at the TadA–dCas9–DNA triad of interfaces showing a noteworthy gain in interaction strength between the TadA docking domain, the two antiparallel β-sheets of the nCas9 RuvC domain (residues 1045 to 1062), and the DNA NTS. This may not only be attributed to the additional eight mutations in TadA8e but also to the fact that TadA8e forms a homodimer, while TadA7.10 is a heterodimer of an evolved TadA7.10 and the wt-TadA. We thereby conducted a comparison of the interaction gain/loss between the homodimeric and heterodimeric TadA8e bound to CRISPR-Cas9 (i.e. ABE8e versus ABE8e hetero) based on additional ∼9 μs of MD simulations of ABE8e including the TadA8e heterodimer. We observed that the homodimeric TadA8e maintains an elevated number of interactions with dCas9 and the DNA that gain strength, compared with its heterodimeric form (Supplementary Figure S6). This indicates that the homodimeric TadA8e establishes a strong network of interactions between its docking domain, the two antiparallel

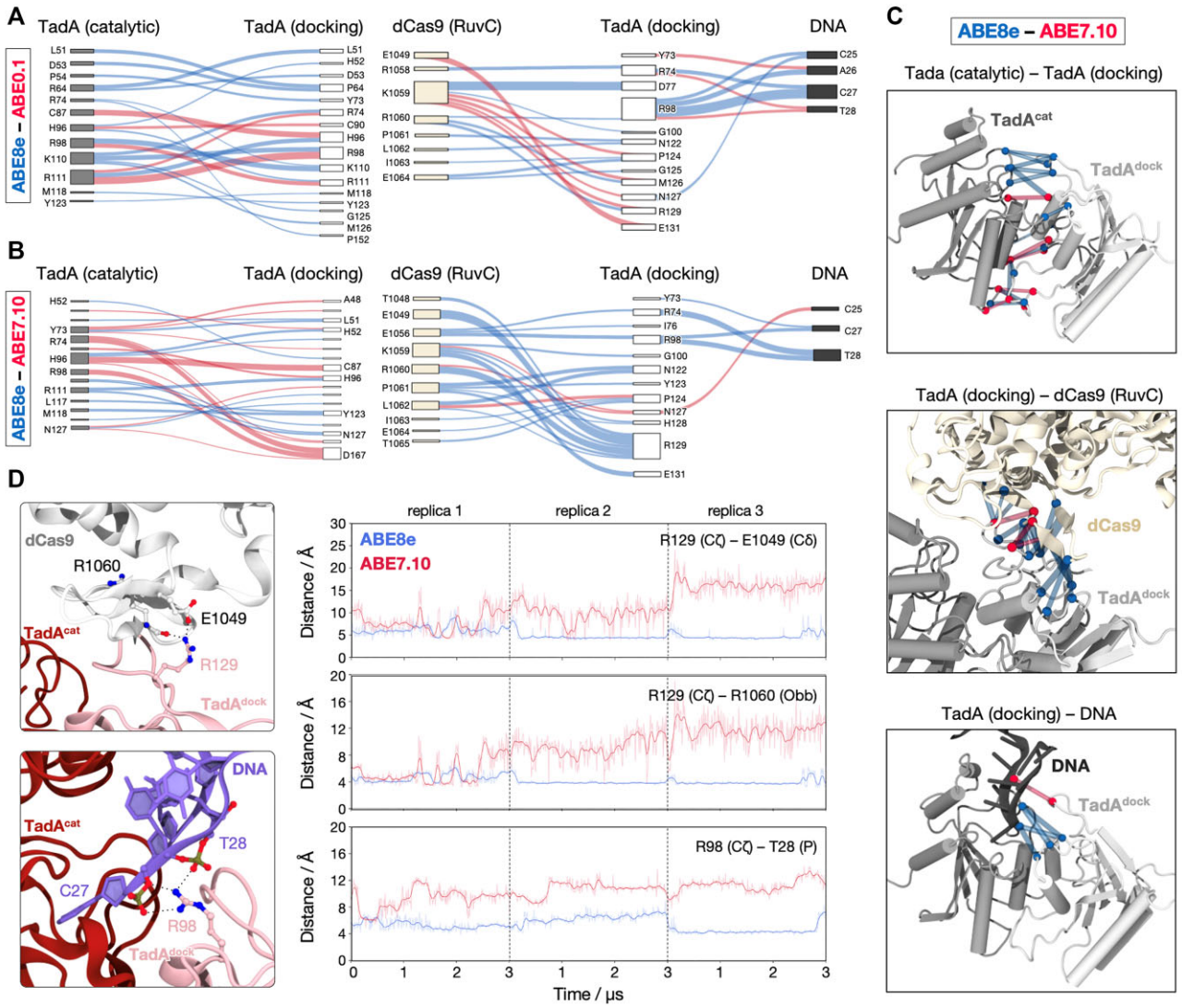


Figure 3. Gain/loss of interaction strength in ABE8e compared with ABE0.1 and ABE7.10. (**A** and **B**) Sankey diagrams of the relative change in pairwise interaction energy (ΔE) between ABE8e and ABE0.1 (A), and between ABE8e and ABE7.10 (B), computed at the interfaces between the TadA catalytic and docking domains (left), the TadA docking domain, the dCas9 RuvC domain and the DNA NTS (right). Sankey diagrams connect residues through lines the thickness of which is proportional to ΔE. Interactions that gain strength in ABE8e ($\Delta E < 0$, blue) are compared to those that gain strength in ABE0.1 and ABE7.10 ($\Delta E > 0$, red). A cut-off of 0.7 kcal/mol was used to filter interactions with substantial perturbation. Details are reported in the Supplementary Methods. (**C**) The relative pairwise ΔE between ABE8e and ABE7.10 are plotted on the three-dimensional structure of the complex and shown at the TadA–Cas9–DNA interface. Interactions that gain strength in ABE8e (blue) are compared to those that gain strength in ABE7.10 (red). Bonds are proportional to ΔE. (**D**) Time-evolution of the interactions established by R129 and R98 with the RuvC domain of dCas9 and the DNA NTS along MD simulations of the ABE8e (blue) and ABE7.10 (red) complexes. Data are reported for three simulation replicates of ~3 μs each.

β-sheets of RuvC and the DNA NTS that solidify the juxtaposition of TadA8e with the Cas9–DNA complex as a unique feature of ABE8e.

The Sankey diagrams comparing ABE8e and ABE7.10 (Figure 3B) pinpoint that R129 strengthens the interactions with several residues of the RuvC domain in ABE8e compared to ABE7.10. Concomitantly, R98 binds multiple bases of the NTS while also interacting with dCas9 E1053.

A time-evolution analysis of the interactions shows that in ABE8e, R129 unequivocally interacts with the E1049 side chain and the R1060 backbone carbonyl of dCas9, while these interactions are comparatively less stable in ABE7.10 (Figure 3D). R98 binds the NTS phosphate backbone by persistently interacting with thymine 8 (T8) in ABE8e, while this interaction is mostly absent in ABE7.10. Interestingly, R98 loosely binds E1056 of dCas9 in both ABE8e and ABE7.10 (Supplementary Figure S7), suggesting that R98 mainly binds to DNA in ABE8e. Taken together, the TadA8e docking domain notably strengthens its interactions with both Cas9 and the DNA NTS, suggesting that it could critically contribute to the 580-fold increased DNA deamination efficiency in ABE8e compared to ABE7.10.

We thereby performed a mutational analysis and evaluated the significance of the TadA8e docking domain (R98 and R129) interactions with the DNA and Cas9 for *in vitro* DNA deoxyadenosine deamination. This is accomplished by first designing a construct that contains two TadA8e units (the unit adjacent to the Cas9 construct is regarded as the C-terminal domain of TadA8e, and the other unit constitutes the N-terminal domain) fused to dCas9. Each TadA8e is en-

coded by different DNA sequences while retaining the same amino acid sequences, which allows us to introduce mutations specifically to only one TadA8e unit at a time. To distinguish which TadA8e acts as a catalytic unit, we inactivated each TadA8e one at a time by substituting the catalytic E59 with A (Supplementary Figure S8A). DNA deoxyadenosine deamination kinetics of these ABE8e variants in single-turnover conditions yielded identical observed deamination rates (k_{obs}), which are half of the rate observed for ABE8e* when both TadA8e units were active (Supplementary Figure S8B and C). This indicates that either of the TadA8e domains can deaminate deoxyadenosine within the NTS.

Following the original assumption in the ABE design that the C-terminal TadA domain is catalytic (9), and to match the constructs commonly used in the genome engineering and our computational model (Figure 1) (14), we opted for the ABE8e*A construct, where the N-terminal domain is the docking unit and the C-terminal is the catalytic unit (Figure 4A and Supplementary Figure S8A). We then generated two mutants with the R98A substitution in either the TadA8e catalytic domain [ABE8e*A (R98A^{cat})] or the docking domain [ABE8e*A (R98A^{dock})] and measured their DNA deoxyadenosine deamination kinetics under single-turnover conditions (Figure 4B and Supplementary Figure S9A). In agreement with our computational analysis, R98A substitution in the TadA8e docking domain significantly reduced the observed DNA deamination rate, whereas the same substitution in the catalytic unit did not affect the DNA deamination rate (Figure 4B). The circular dichroism (CD) spectrum of ABE8e*A (R98A^{dock}) aligns well with that of ABE8e*A (Supplementary Figure S9B), indicating that the decrease in DNA deamination activity is attributable to the disruption of R98 interactions with the DNA NTS and/or the RuvC domain of dCas9, rather than protein misfolding. This supports our hypothesis that the docking unit of TadA8e contributes to the DNA deamination rate of the catalytic TadA8e. To assess whether R98 in the docking TadA8e interacts with the DNA or Cas9, we created two versions of ABE8e*A, introducing the R98E mutation in the docking TadA8e [ABE8e*A (R98E^{dock})] (to disrupt TadA8e-Cas9 RuvC interaction) and the same substitution in the docking TadA8e with the complementary E1056R mutation in dCas9 [ABE8e*A (R98E^{dock}/E1056R^{Cas9})] (to reinstitute TadA8e–Cas9 RuvC interaction). Single-turnover kinetics showed that the DNA deamination rates and efficiencies were similar for both mutants (Supplementary Figure S9B–D), supporting our computational observation that R98 in the docking TadA8e primarily binds to the NTS of DNA instead of forming strong interactions with Cas9 (Figure 3 and Supplementary Figure S7C).

Based on our computational analysis, R129 in the TadA8e docking unit anchors the E1049 residue in the Cas9 RuvC domain through charge-charge interactions (Figure 3D). To probe the role of R129 we adopted a similar approach, where we introduced R129E in the TadA8e catalytic unit [ABE8e*A (R129E^{cat})] and then in the docking domain [ABE8e*A (R129E^{dock})] (Figure 4A). In agreement with our computational analysis, R129E in the TadA8e docking domain hampers DNA deamination, while the activity stays unaltered when introduced in the catalytic unit (Figure 4C; Supplementary Figure S9E and F). To further elaborate the role of the interaction between R129 and the Cas9 residue E1049, we prepared a rescue variant with R129E in the TadA8e docking domain and E1049R in Cas9 [ABE8e*A (R129E^{dock}/E1049R^{Cas9})]. The measured deoxyadenosine deamination kinetics showed comparable activity as that of ABE8e*A (Figure 4C and Supplementary Figure S9E). Thus, our rescue substitution could successfully reinstate the ABE8e*A activity confirming that R129 and E1049 engage in charge–charge interactions and that the TadA8e-RuvC binding is important for efficient DNA deoxyadenosine deamination. Collectively, these findings provide compelling evidence that R98 and R129 residues of the TadA8e docking domain structurally bridge the juxtaposition of TadA8e with Cas9 and the DNA NTS and are necessary for efficient DNA deamination by ABE8e. As these interactions weaken in ABE7.10, they uniquely contribute to the enhanced activity of ABE8e.

Key substitutions introduced during ABE8e's evolution tune TadA stability for enhancing DNA deamination capability

TadA8e holds eight amino acid substitutions with respect to TadA7.10 (14,17). This contributes to a 580-fold increase in the rate of deoxyadenosine deamination in ABE8e with respect to ABE7.10, the highest gain across two consecutive generations during the directed evolution. To understand these mutations' role in the complex's stability and shed light on their role in the increased function in ABE8e, we performed free energy simulations. We employed an alchemical FEP method (36) designed to compute the impact of mutations on the folding free energy (37,38) (details in 'Materials and Methods' section and in the Supplementary Text), as changes in thermal stability are driven by variations in folding (38). We computed the relative free energy change in the stability of the ABE7.10 complex while introducing the ABE8e substitutions (i.e. A109S, T111R, D119N, H122N, Y147D, F149Y, T166I, D167N).

Our free energy simulations show that T111R and D119N alter the stability of the ABE7.10 complex, while A107S and F149Y produce moderate effects, and H122N, T166I and D167N result in negligible changes (Figure 5A and Supplementary Figure S10). The change in free energy of the complex ($\Delta\Delta G$) increases with the T111R mutation, destabilizing ABE7.10, while D119N provides stability to the complex as evidenced from negative $\Delta\Delta G$. To experimentally assess the effect of T111R and D119N on the stability of ABE7.10, we performed a thermal stability analysis using CD spectroscopy and extracted melting temperatures (T_m) for three TadA7.10 heterodimers (composed of wt-TadA and TadA7.10): TadA7.10, TadA7.10 (T111R) and TadA7.10 (T111R/D119N). Notably, the TadA7.10 (T111R) variant exhibited the lowest thermal stability, with a T_m of 55.4 ± 0.1°C, which is 15°C lower than the T_m of TadA7.10 (70.7 ± 0.3°C) (Figure 5B). However, the introduction of D119N to the TadA7.10 (T111R) variant restored its stability to the TadA7.10 level, elevating the T_m to 69.3 ± 0.9°C. This observation corroborates our alchemical FEP simulations, indicating that T111R destabilizes the complex, and contrarily D119N provides stability (Figure 5A). This agreement between our alchemical FEP and thermal stability assays stems from the fact that thermal stability is a key component of the folding free energy computed through our alchemical FEP approach.

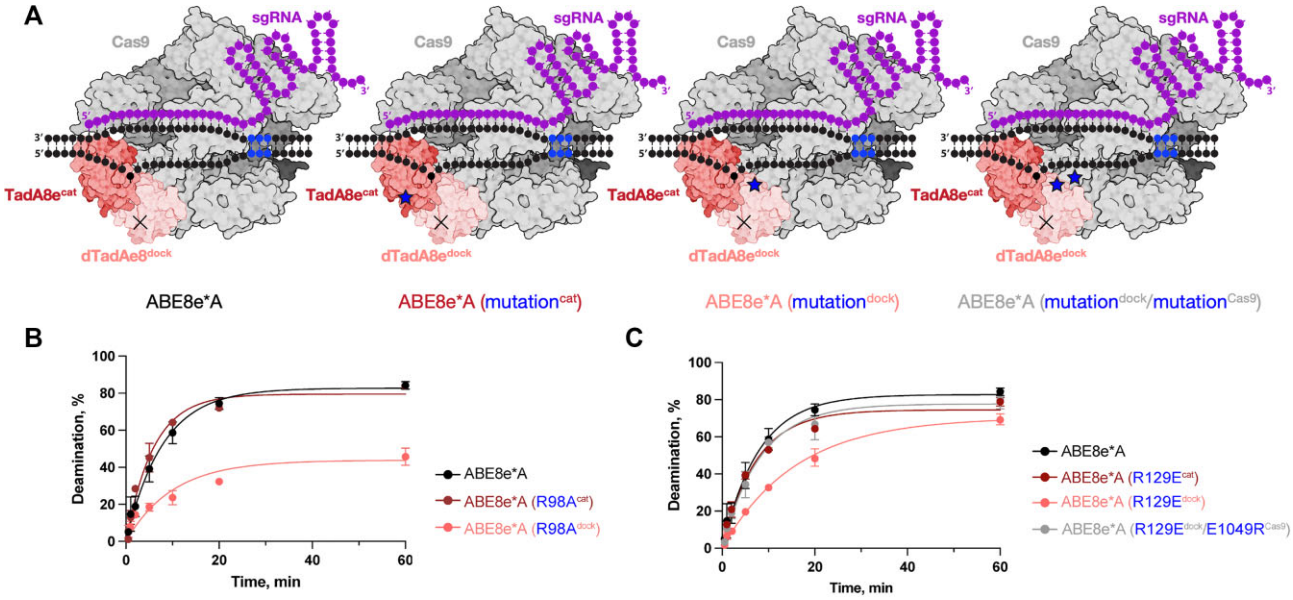


Figure 4. Key interactions between the docking domain of TadA8e with the DNA and Cas9 for efficient DNA deoxyadenosine deamination by ABE8e. (**A**) Model depicting the chosen construct, ABE8e*A, alongside the three different types of mutants designed to evaluate the role of the docking unit of TadA8e in efficient DNA deamination. The catalytic domain of TadA8e (TadA8e^{cat}) is coloured in dark red; the docking domain (TadA8e^{dock}) is pink. A blue star indicates the position of amino acid substitutions introduced to either the catalytic or docking TadA8e domains and the Cas9 RuvC domain. (**B**) Single-turnover kinetics of ABE8e*A and its variants with the R98A substitution introduced to either the catalytic (red) or the docking (pink) domains measured using fluorescently labelled dsDNA containing a single adenine within the editing window. The DNA deamination efficiency and rate of the ABE8e*A (R98A^{dock}) ($k_{obs} = 0.03 \pm 0.01$ min^{-1}) are significantly lower than those of the ABE8e*A ($k_{obs} = 0.12 \pm 0.01$ min^{-1}) and the ABE8e*A (R98A^{cat}) ($k_{obs} = 0.17 \pm 0.02$ min^{-1}) (data analysed with one-tailed *t*-tests, and $P = 0.0001$). (**C**) Single-turnover kinetics of ABE8e*A and its variants with R129E substitution introduced to either the catalytic (red) or the docking (pink) domains measured using fluorescently labelled dsDNA containing a single adenine within the editing window. The observed rates of DNA deamination by constructs ABE8e*A ($k_{obs} = 0.12 \pm 0.01$ min^{-1}) and ABE8e*A (R129E^{cat}) ($k_{obs} = 0.12 \pm 0.01$ min^{-1}) are identical. The DNA deamination rate ($k_{obs} = 0.06 \pm 0.01$ min^{-1}) and efficiency of the ABE8e*A (R129E^{dock}) are significantly reduced (data analysed with one-tailed *t*-tests, and $P = 0.0017$). The grey curve represents the single-turnover kinetics of the ABE8E*A mutant containing two amino acid substitutions: one in the docking TadA8e (R129E) and one in Cas9 (E1049R). The introduction of the Cas9 E1049R mutation increases the rate ($k_{obs} = 0.14 \pm 0.01$ min^{-1}) and the efficiency of DNA deamination to the levels of ABE8e*A. For all single-turnover kinetics assays the fraction (in %) of deaminated DNA was plotted as a function of time and fitted to a single exponential equation. Data are represented as the mean ± SD from three independent experiments.

Interestingly, these observations align with several studies performed to understand the role of single amino acid substitutions that appear during directed evolution. Independent studies by Romero and Arnold (49), Tawfik *et al.* (50) and Shoichet *et al.* (51) demonstrated that destabilizing substitutions are accommodated during directed evolution to reshape functionality (i.e. providing improved or new function) while structural, stabilizing mutations to compensate for the loss of stability. We thereby monitored the single-turnover deamination rate of ABE7.10 and its mutants, ABE7.10 (T111R), ABE7.10 (D119N), and ABE7.10 (T111R/D119N) (Figure 5C and Supplementary Figure S11). Introducing T111R increases the deamination rate by ∼10-fold with respect to ABE7.10, indicating the functional role of this destabilizing mutation. D119N by itself has a slight negative effect on deamination, while the deamination rate of the ABE7.10 (T111R/D119N) is comparable to the rate of ABE7.10 (T111R), indicating that D119N does not contribute to the enhancement of DNA deamination rate, but rather stabilizes the ABE7.10. Hence, the destabilizing T111R substitution has a functional role, while D119N compensates for the loss of stability during evolution. The above-mentioned studies on directed evolution also indicated that mutations providing improved (or new) catalytic function often appear first and destabilize the complex. In the evolution of ABE8e from ABE7.10, the T111R mutation frequently appeared in several early variants (17), enhancing the ABE7.10 deamination rate according to our single-turnover kinetics. Since T111R reduces complex stability, the subsequent introduction of D119N compensates for this, facilitating the enzyme's evolution. Together, T111R and D119N introduce features that balance activity and stability, driving the evolution from ABE7.10 to the improved ABE8e.

To further understand the biophysical role of T111R and D119N in the activity of ABE8e, we analysed our multi-μs long simulations of ABE7.10 and ABE8e. Notably, the T111R and D119N substitutions reside in a solvent-exposed loop of the TadA8e catalytic domain (residues 108–125). This loop exhibits significantly reduced flexibility in ABE8e when compared to ABE7.10 (Supplementary Figure S12). This loss in flexibility is attributed to altered interactions involving D119N and a few other mutations. The backbone interaction between D119 and H122 in ABE7.10 is weakened in ABE8e and a new persistent interaction is established between N122 and R129 in ABE8e (Figure 5D and Supplementary Figure S13). We then evaluated whether the flexibility of this loop could alter the interactions between the TadA monomers (Supplementary Figure S14). We observe that H123Y in the TadA8e docking domain improves interaction with the L117 backbone (Figure 5E), with respect to H123 in TadA7.10. Additionally, the reduced flexibility of the loop reinforces the interactions between T111R of the catalytic

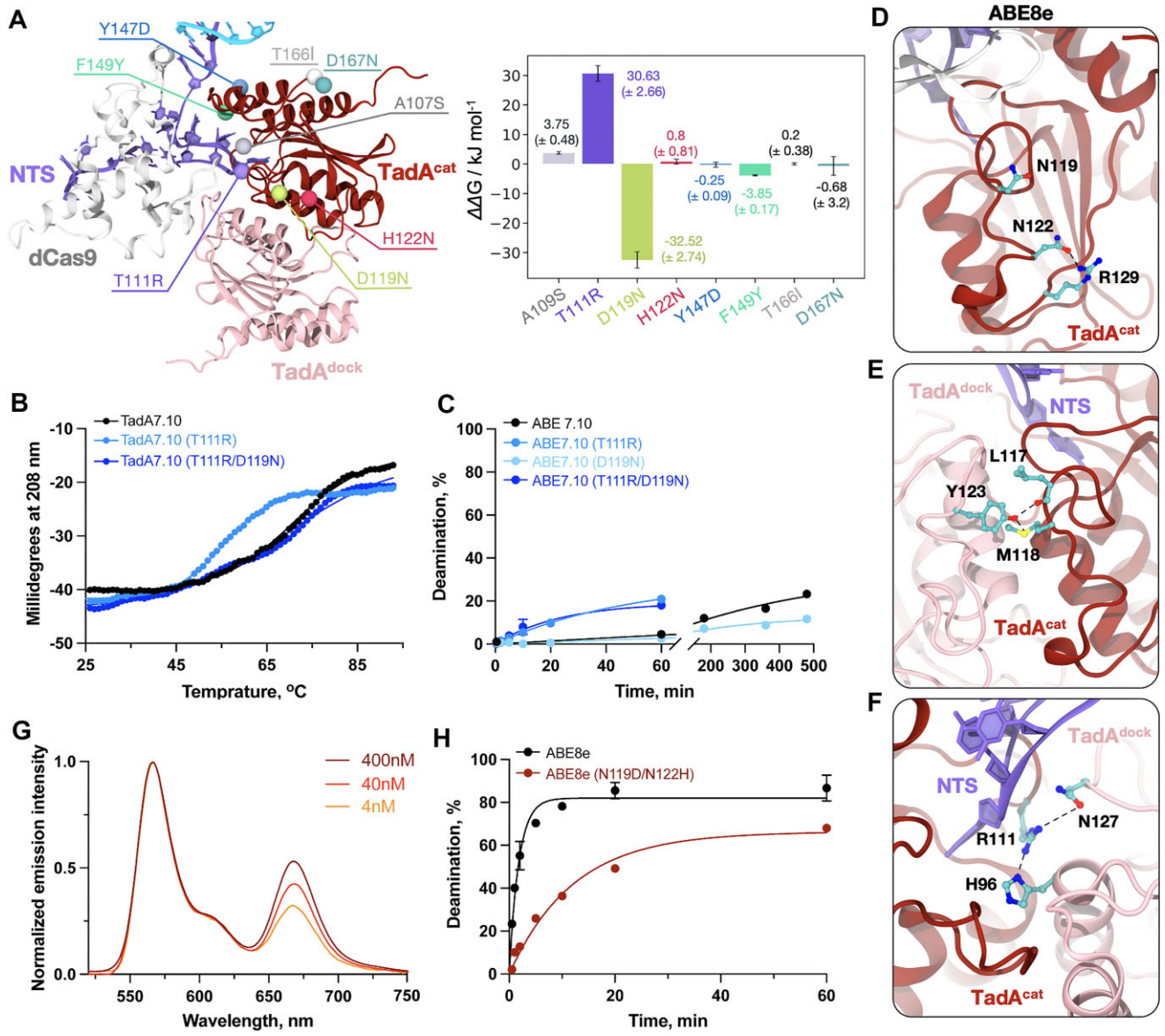


Figure 5. Functional role of amino acid substitutions introduced during directed evolution of ABE8e. (**A**) Relative free energy changes ($\Delta\Delta G$, in kJ mol^{-1}) in the stability of ABE7.10 while introducing the ABE8e mutations, computed through alchemical FEP simulations. The amino acid substitutions considered for alchemical FEP are shown on the three-dimensional structure of ABE8e on the right. (**B**) Thermal denaturation profiles of TadA7.10 variants [TadA7.10 (black), TadA7.10 (T111R) (blue) and TadA7.10 (T111R/D119N) (dark blue)] were measured by CD spectroscopy with a temperature change of 1°C/min. The thermal denaturation data were fitted to the Boltzmann sigmoidal equation and the T_m values for each TadA7.10 variant were determined by the inflection point and are 70.7 ± 0.3°C, 55.4 ± 0.1°C and 69.3 ± 0.9°C, respectively. (**C**) Single-turnover kinetics of ABE7.10 (black), ABE7.10 (T111R) (blue), ABE7.10 (D119N) (light blue) and ABE7.10 (T111R/D119N) (dark blue) measured using fluorescently labelled dsDNA containing a single adenine within the editing window. The observed rates of DNA deamination by these constructs are 0.0021 ± 0.0005 min^{-1}, 0.020 ± 0.003 min^{-1}, 0.0038 ± 0.0007 min^{-1}, 0.020 ± 0.003 min^{-1} and 0.042 ± 0.007 min^{-1}, respectively. The fraction (in %) of deaminated DNA was plotted as a function of time and fitted to a single exponential equation. Data are represented as the mean ± SD from three independent experiments. (**D–F**) Interactions established by the solvent-exposed loop of the TadA catalytic domain (residues 108–125) in ABE8e. N122 forms a stable interaction with R129 within the same loop (D). Y123 of the TadA8e docking domain interacts with L117 and M118 of the TadA catalytic domain (E). R111 of the TadA8e catalytic domain interacts with the H96 and N127 residues of the docking domain (F). Detailed analysis from MD simulations and comparison with ABE7.10 are reported in Supplementary Figures S13 and 14. (**G**) Fluorescence emission spectra of Cy3/Cy5-labeled ABE8e (N119D/N122H) across varying concentrations, with normalized Cy3 emission intensity, highlighting the changes in FRET signal. The observed distinct reduction in Cy5 fluorescence emission intensity (at ~675 nm), as protein concentration decreases, clearly indicates that N119D and N122H substitutions destabilize TadA8e dimer. (**H**) Single-turnover kinetics of ABE8e (black) and ABE8e (N119D/N122H) (red) measured using fluorescently labelled dsDNA containing a single adenine within the editing window. The observed rates of DNA deamination by these constructs are 0.56 ± 0.06 min^{-1} and 0.08 ± 0.01 min^{-1}, respectively. The fraction (in %) of deaminated DNA was plotted as a function of time and fitted to a single exponential equation. Data are represented as the mean ± SD from three independent experiments.

domain and the H96 and N127 residues of the docking domain (Figure 5F), which is also evidenced by a gain in the strength of these interactions in ABE8e versus ABE7.10 (Figure 3B). This promotes the engagement between the catalytic and the docking TadA8e units.

To experimentally assess the role of D119N and N122H in the dimerization of TadA, we performed ensemble FRET experiments. We generated FRET constructs of the ABE8e containing N119D and N122H substitutions as seen in its predecessors TadA7.10 as well as wt-TadA. The Cy3/Cy5-labelled ABE8e (N119D/N122H) is properly folded (Supplementary Figure S15A), and the dye-labelling does not affect the DNA deamination rate (Supplementary Figure S15B and C). Our ensemble FRET experiment shows that TadA8e (N119D/N122H) enzymes exist in dimeric form when fused to dCas9, as evidenced by the presence of a peak at ∼675 nm in the fluorescence emission spectra of Cy3/Cy5-labelled proteins when excited at 530 nm. However, the dimerization affinity of this TadA8e variant is lower compared to that of TadA8e. It appears to be similar to the affinity of the wt-TadA, as indicated by the reduction in FRET signal with decreasing protein concentrations (Figures 2C and 5G). Since N119D and N122H destabilize the TadA8e dimer, based on our observation that TadA dimerization is important for the enhanced activity of the ABE8e (Figure 2), we anticipate that the ABE8e variant with these substitutions—although they are distant from the catalytic site—will exhibit lower activity compared to ABE8e. Indeed, our single turnover kinetic assays show that the DNA deamination rate of ABE8e (N119D/N122H) ($k_{obs} = 0.08 \pm 0.01$ min^{-1}) is seven times lower than the rate of ABE8e ($k_{obs} = 0.56 \pm 0.06$ min^{-1}) (Figure 5H). Collectively, experimental and computational observations demonstrate that the D119N and H122N, introduced to the ABE8e during directed evolution reduce the flexibility of the loop and stabilize inter-molecular interactions. These findings agree with the previous observations demonstrating the importance of this loop in DNA deamination activity of first generations of TadA (52). Moreover, it aligns with the growing recognition of the importance of loop dynamics in protein evolution, with several studies indicating that loop motion can influence catalysis (53,54).

Discussion

ABEs offer significant promise for treating genetic diseases resulting from point mutations. By leveraging the precise targeting ability of CRISPR-Cas systems combined with the catalytic power of laboratory-evolved TadA enzymes, ABEs enable the conversion of A•T base pairs to G•C base pairs. This process is achieved through the targeted deamination of deoxyadenosine to deoxyinosine at specific ssDNA sites (7). Among the ABEs used for precision genome editing in various cells and organisms, ABE8e is one of the most efficient DNA base editors to date (17,18). It demonstrates a notable leap forward compared to ABE7.10, its closest precursor, dramatically enhancing both DNA base editing efficiency and the rate of DNA deamination. However, ABE8e is more prone to DNA off-target editing (17). A thorough understanding of this improvement is instrumental in advancing the development of efficient ABEs. Here, we employed an integrative approach where we combined extensive MD simulations with free energy methods and biochemical and biophysical experiments to establish the biophysical rationale behind the astonishing base editing efficiency of ABE8e, providing new directions for the development of more advanced ABEs.

ABEs have been developed to deaminate DNA by evolving the wt-TadA (14,17), a natural homo-dimeric enzyme that catalyses the deamination of tRNA with its active site formed by both monomers. The cryo-EM structure of ABE8e in a DNA-bound state reveals that TadA8e forms a similar dimer (Figure 1). This raises the question of how amino acid substitutions introduced during directed evolution impact TadAs' dimerization and how the latter affects DNA base editing. Using enhanced sampling simulations, viz., an Umbrella Sampling method, in combination with FRET measurements, we show that TadA8e forms more stable dimers than its progenitor wt-TadA (Figure 2). Interestingly, during the directed evolution of the wt-TadA deaminase toward TadA8e, the selective pressure was applied to select for efficient DNA base editing rather than for more stable dimers. This suggests that TadA dimerization strength co-evolved with the DNA base-editing property, highlighting the critical role of TadA dimerization in the DNA deamination activity.

In light of these findings, we sought to establish what makes the dimeric TadA8e an efficient DNA base editor when bound to the dCas9–DNA complex. Extensive MD simulations of ABE8e, in comparison with ABE7.10, its closest predecessor with ∼580-fold lower DNA deamination rate, and ABE0.1 (where TadA is wt) provide evidence that the TadA8e dimer establishes key interactions with both the DNA and dCas9, a phenomenon uniquely observed in ABE8e (Figure 3). Indeed, TadA8e forms a robust network of interactions involving its docking domain, the two antiparallel β-sheets of RuvC (residues 1049 to 1056), and the DNA NTS, consolidating the positioning of TadA8e within the dCas9–DNA complex. Specific residues within the TadA8e docking domain, R98 and R129, are found to form essential interactions with the DNA NTS and the two antiparallel β-sheets of RuvC (Figure 3). Single-turnover kinetics demonstrated that these interactions are necessary for efficient DNA deamination by ABE8e (Figure 4) and, considering that they are remarkably weaker in ABE7.10 (Figure 3), they critically contribute to the activity of ABE8e. This discovery holds significant implications for the developing new ABEs, particularly when employing orthogonal Cas9 or other Cas effectors, in which the antiparallel β-sheet of RuvC that binds TadA8e is absent or is chemically and structurally different. It highlights the necessity of engineering the Cas9-docking TadA interface for each distinct Cas effector to achieve optimal interactions and, in turn, efficient DNA base editing. This aligns with prior observations that fusing TadA8e to different Cas effectors results in less efficient DNA base editing when compared to the ABE8e (TadA8e fusion to the Cas9 from *S. pyogenes*) (17).

Reinforcing the interactions between TadA8e and the Cas effector could enable them to synergize more effectively, potentially also leading to a reduction in DNA off-target editing. Indeed, the fundamental process driving ABE8e's bystander editing is proposed to stem from TadA8e being a multiple-turnover enzyme and its flexible connection to Cas9 enabling it to modify nearby ssDNA regardless of whether Cas9 is bound to it or not (14,17,55). Hence, enhancing the interactions between TadA8e and the Cas effector could serve as a strategy to reduce ABE8e's bystander editing and Cas9-independent DNA off-target effects.

ABE8e holds eight amino acid substations in its TadA8e unit with respect to its closest predecessor ABE7.10,

contributing somehow to enhancing ABE8e's superior efficiency and rate of DNA deamination compared to ABE7.10. Our alchemical FEP simulations and thermal stability assays reveal that two substitutions, T111R and D119N, are key for the TadA8e stability and function. We show that T111R is a destabilizing mutation that provides TadA with an improved catalytic function (Figure 5). In contrast, D119N is a compensatory, stabilizing mutation and does not directly enhance TadA's activity. This dynamic interplay between destabilizing and stabilizing mutations aligns with the complex nature of enzyme evolution and adaptation. Biophysical studies of evolved enzymes have shown that mutations conferring improved or new catalytic function commonly appear early and destabilize the complex (49–51). As the adaptive process continues, proteins regain stability through other mutations with little to no impact on activity. Accordingly, moving from ABE7.10 to ABE8e, T111R appeared in most of the evolution batches, followed by D119N (17). This suggests that the interplay between T111R and D119N provides the balance between activity and stability necessary to evolve the complex toward improved function. This underscores the importance of employing diverse approaches when engineering enzymes, as mutations impacting the complex stability and activity are not readily identifiable through rational design alone.

In addition to TadA stabilization, the D119N in combination with H122N substitution plays a crucial role in stabilizing the TadA8e dimer by altering the dynamics of the 105–125 loop region, as evidenced by our MD simulations and ensemble FRET experiments. Since a stable TadA8e dimer is essential for forming interactions with Cas9 and DNA (56,57), reversing these substitutions destabilizes the TadA8e dimer, subsequently reducing the activity of ABE8e (Figure 5). This is in line with the significance of loop dynamics in protein evolution, as numerous studies suggest that loop motion directly or indirectly affect catalysis (53,54). Therefore, we propose that the 105–125 loop is an important region that could be further engineered to develop more efficient ABEs.

In summary, extensive molecular simulations, single turnover kinetics and biophysical experiments reveal that TadA dimerization and its juxtaposition to the dCas9–DNA complex are pivotal for the efficient DNA deamination in ABE8e. The TadA8e dimer exhibits greater stability than its wt-TadA predecessor, facilitating critical interactions with both DNA and Cas9, which enhances DNA deamination efficiency. By identifying the key residues within the TadA–Cas9–DNA interaction triad that enhance DNA base editing in ABE8e, our study provides valuable insights for engineering superior ABEs. These insights are crucial for developing ABEs with orthogonal Cas9 or other Cas effectors to reduce the editor size improving its delivery to cells, and to potentially reduce DNA off-target editing, thereby contributing to the advancement of safer and more efficient genome editing tools.

Data availability

Data are available from the corresponding authors upon reasonable request.

Supplementary data

Supplementary Data are available at NAR Online.

Acknowledgements

We thank Dr Marcia Levitus for her helpful discussions in setting up ensemble FRET measurements and providing access to the spectrophotometer.

Author contributions: P.R.A.: Data curation, Formal Analysis, Investigation, Methodology, Validation, Visualization, Writing—original draft. X.C.: Data curation, Formal Analysis, Investigation, Methodology, Validation, Visualization, Writing—original draft. Souvik Sinha: Data curation, Formal Analysis, Investigation, Methodology, Validation, Visualization, Writing—original draft. Aakash Saha: Data curation, Formal Analysis, Investigation, Methodology, Validation, Visualization, Writing—original draft. A.C.P.: Formal Analysis, Methodology, Visualization. M.S.: Formal Analysis, Methodology. L.N.: Formal Analysis, Methodology. A.L.: Conceptualization, Supervision, Project administration, Funding acquisition, Writing—original draft. G.P.: Conceptualization, Supervision, Project administration, Funding acquisition, Writing—original draft.

Funding

National Institutes of Health [R01GM141329 to G.P.; DP2GM149550 to A.L.]; National Science Foundation [CHE-2144823 to G.P.]; National Institutes of Health; Sloan Foundation [FG-2023-20431 to G.P.]; Camille and Henry Dreyfus Foundation [TC-24-063]; San Diego Supercomputing Center [MCB160059]; Pittsburgh Supercomputer Center [BIO230007]; Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) program; National Science Foundation [2138259, 2138286, 2138307, 2137603, 2138296]. Funding for open access charge: National Institutes of Health.

Conflict of interest statement

P.R.A., X.C., S.S., A.S., A.C.P., A.L. and G.P. are co-inventors on patent applications filed by the University of California, Riverside and Arizona State University relating to work in this manuscript.

References

1. Wang,J.Y. and Doudna,J.A. (2023) CRISPR technology: a decade of genome editing is only the beginning. *Science*, **379**, eadd8643.
2. Jinek,M., Chylinski,K., Fonfara,I., Hauer,M., Doudna,J.A. and Charpentier,E. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, **337**, 816–821.
3. Cong,L., Ran,F.A., Cox,D., Lin,S., Barretto,R., Habib,N., Hsu,P.D., Wu,X., Jiang,W., Marraffini,L.A., *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science*, **339**, 819–823.
4. Mali,P., Yang,L., Esvelt,K.M., Aach,J., Guell,M., DiCarlo,J.E., Norville,J.E. and Church,G.M. (2013) RNA-guided human genome engineering via Cas9. *Science*, **339**, 823–826.
5. Xue,C. and Greene,E.C. (2021) DNA repair pathway choices in CRISPR-Cas9-mediated genome editing. *Trends Genet.*, **37**, 639–656.
6. Anzalone,A.V., Koblan,L.W. and Liu,D.R. (2020) Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. *Nat. Biotechnol.*, **38**, 824–844.
7. Liao,J., Chen,S., Hsiao,S., Jiang,Y., Yang,Y., Zhang,Y., Wang,X., Lai,Y., Bauer,D.E. and Wu,Y. (2023) Therapeutic adenine base

editing of human hematopoietic stem cells. *Nat. Commun.*, **14**, 207.
8. Komor,A.C., Kim,Y.B., Packer,M.S., Zuris,J.A. and Liu,D.R. (2016) Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature*, **533**, 420–424.
9. Gaudelli,N.M., Komor,A.C., Rees,H.A., Packer,M.S., Badran,A.H., Bryson,D.I. and Liu,D.R. (2017) Programmable base editing of T to G C in genomic DNA without DNA cleavage. *Nature*, **551**, 464–471.
10. Levy,J.M., Yeh,W.-H., Pendse,N., Davis,J.R., Hennessey,E., Butcher,R., Koblan,L.W., Comander,J., Liu,Q. and Liu,D.R. (2020) Cytosine and adenine base editing of the brain, liver, retina, heart and skeletal muscle of mice via adeno-associated viruses. *Nat. Biomed. Eng.*, **4**, 97–110.
11. Lam,D.K., Feliciano,P.R., Arif,A., Bohnuud,T., Fernandez,T.P., Gehrke,J.M., Grayson,P., Lee,K.D., Ortega,M.A., Sawyer,C., *et al.* (2023) Improved cytosine base editors generated from TadA variants. *Nat. Biotechnol.*, **41**, 686–697.
12. Neugebauer,M.E., Hsu,A., Arbab,M., Krasnow,N.A., McElroy,A.N., Pandey,S., Doman,J.L., Huang,T.P., Raguram,A., Banskota,S., *et al.* (2023) Evolution of an adenine base editor into a small, efficient cytosine base editor with low off-target activity. *Nat. Biotechnol.*, **41**, 673–685.
13. Chen,L., Zhu,B., Ru,G., Meng,H., Yan,Y., Hong,M., Zhang,D., Luan,C., Zhang,S., Wu,H., *et al.* (2023) Re-engineering the adenine deaminase TadA-8e for efficient and specific CRISPR-based cytosine base editing. *Nat. Biotechnol.*, **41**, 663–672.
14. Lapinaite,A., Knott,G.J., Palumbo,C.M., Lin-Shiao,E., Richter,M.F., Zhao,K.T., Beal,P.A., Liu,D.R. and Doudna,J.A. (2020) DNA capture by a CRISPR-Cas9-guided adenine base editor. *Science*, **369**, 566–571.
15. Wolf,J. (2002) TadA, an essential tRNA-specific adenosine deaminase from *Escherichia coli*. *EMBO J.*, **21**, 3841–3851.
16. Kim,J., Malashkevich,V., Roday,S., Lisbin,M., Schramm,V.L. and Almo,S.C. (2006) Structural and kinetic characterization of *Escherichia coli* TadA, the wobble-specific tRNA deaminase. *Biochemistry*, **45**, 6407–6416.
17. Richter,M.F., Zhao,K.T., Eton,E., Lapinaite,A., Newby,G.A., Thuronyi,B.W., Wilson,C., Koblan,L.W., Zeng,J., Bauer,D.E., *et al.* (2020) Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. *Nat. Biotechnol.*, **38**, 883–891.
18. Xiao,Y.-L., Wu,Y. and Tang,W. (2024) An adenine base editor variant expands context compatibility. *Nat. Biotechnol.*, **42**, 1442–1453.
19. Jumper,J., Evans,R., Pritzel,A., Green,T., Figurnov,M., Ronneberger,O., Tunyasuvunakool,K., Bates,R., Žídek,A., Potapenko,A., *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*, **596**, 583–589.
20. Sinha,S., Pindi,C., Ahsan,M., Arantes,P.R. and Palermo,G. (2023) Machines on genes through the computational microscope. *J. Chem. Theory Comput.*, **19**, 1945–1964.
21. Saha,A., Ahsan,M., Arantes,P.R., Schmitz,M., Chanez,C., Jinek,M. and Palermo,G. (2024) An alpha-helical lid guides the target DNA toward catalysis in CRISPR-Cas12a. *Nat. Commun.*, **15**, 1473.
22. Sinha,S., Molina Vargas,A.M., Arantes,P.R., Patel,A., O'Connell,M.R. and Palermo,G. (2024) Unveiling the RNA-mediated allosteric activation discloses functional hotspots in CRISPR-Cas13a. *Nucleic Acids Res.*, **52**, 906–920.
23. Pacesa,M., Lin,C.-H., Cléry,A., Saha,A., Arantes,P., Bargsten,K., Irby,M., Allain,F., Palermo,G., Cameron,P., *et al.* (2022) Structural basis for Cas9 off-target activity. *Cell*, **185**, 4067–4081.
24. Tian,C., Kasavajhala,K., Belfon,K.A.A., Raguette,L., Huang,H., Migues,A.N., Bickel,J., Wang,Y., Pincay,J., Wu,Q., *et al.* (2020) ff19SB: amino-acid-specific protein backbone parameters trained against quantum mechanics energy surfaces in solution. *J. Chem. Theory Comput.*, **16**, 528–552.
25. Galindo-Murillo,R., Robertson,J.C., Zgarbová,M., Šponer,J., Otyepka,M., Jurečka,P. and Cheatham,T.E. (2016) Assessing the current state of Amber force field modifications for DNA. *J. Chem. Theory Comput.*, **12**, 4114–4127.
26. Zgarbova,M., Otyepka,M., Sponer,J., Mladek,A., Banas,P., Cheatham,T.E. and Jurecka,P. (2011) Refinement of the Cornell et al. Nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. *J. Chem. Theory Comput.*, **7**, 2886–2902.
27. Jorgensen,W.L., Chandrasekhar,J., Madura,J.D., Impey,R.W. and Klein,M.L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, **79**, 926–935.
28. Becke,A.D. (1993) A new mixing of Hartree–Fock and local density-functional theories. *J. Chem. Phys.*, **98**, 1372–1377.
29. Frisch,M.J., Pople,J.A. and Binkley,J.S. (1984) Self-consistent molecular orbital methods 25. Supplementary functions for Gaussian basis sets. *J. Chem. Phys.*, **80**, 3265–3269.
30. Li,P. and Merz,K.M. (2016) MCPB.Py: a python based metal center parameter builder. *J. Chem. Inf. Model.*, **56**, 599–604.
31. Berendsen,H.J.C., Postma,J.P.M., van Gunsteren,W.F., DiNola,A. and Haak,J.R. (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, **81**, 3684.
32. Case,D.A., Aktulga,H.M., Belfon,K., Ben-Shalom,I.Y., Brozell,S.R., Cerutti,D.S., T.E. Cheatham,I., Cruzeiro,V.W.D., Darden,T.A., Duke,R.E., *et al.* (2022) AMBER 2022. University of California, San Francisco.
33. Kästner,J. (2011) Umbrella sampling. *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, **1**, 932–942.
34. Lee,T.-S., Radak,B.K., Pabis,A. and York,D.M. (2013) A new maximum likelihood approach for free energy profile construction from molecular simulations. *J. Chem. Theory Comput.*, **9**, 153–164.
35. Marcos-Alcalde,I., Setoain,J., Mendieta-Moreno,J.I., Mendieta,J. and Gómez-Puertas,P. (2015) MEPSA: minimum energy pathway analysis for energy landscapes. *Bioinformatics*, **31**, 3853–3855.
36. Knight,J.L. and Brooks,C.L. (2009) λ-dynamics free energy simulation methods. *J. Comput. Chem.*, **30**, 1692–1700.
37. Pitera,J.W. and Kollman,P.A. (2000) Exhaustive mutagenesis in silico: multicoordinate free energy calculations on proteins and peptides. *Proteins: Struc. Funct. Bioinf.*, **41**, 385–397.
38. Seeliger,D. and de Groot,B.L. (2010) Protein thermostability calculations using alchemical free energy simulations. *Biophys. J.*, **98**, 2309–2316.
39. Wieczor,M. and Czub,J. (2022) Gromologist: a Gromacs-oriented utility library for structure and topology manipulation. ChemRxiv doi: https://doi.org/10.26434/chemrxiv-2022-dhswc, 24 May 2022, preprint: not peer reviewed.
40. Gapsys,V. and de Groot,B.L. (2017) pmx webserver: a user friendly interface for alchemistry. *J. Chem. Inf. Model.*, **57**, 109–114.
41. Shirts,M.R. and Chodera,J.D. (2008) Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.*, **129**, 124105.
42. Matsunaga,Y., Kamiya,M., Oshima,H., Jung,J., Ito,S. and Sugita,Y. (2022) Use of multistate Bennett acceptance ratio method for free-energy calculations from enhanced sampling and free-energy perturbation. *Biophys. Rev.*, **14**, 1503–1512.
43. Klimovich,P.V., Shirts,M.R. and Mobley,D.L. (2015) Guidelines for the analysis of free energy calculations. *J. Comput. Aided Mol. Des.*, **29**, 397–411.
44. Wille,D., Jaeger,H., Schlaich,A., Wu,Z. and Beckstein,O. Flamel. San Francisco (CA): GitHub.
45. Losey,H.C., Ruthenburg,A.J. and Verdine,G.L. (2006) Crystal structure of Staphylococcus aureus tRNA adenosine deaminase TadA in complex with RNA. *Nat. Struct. Mol. Biol.*, **13**, 153–159.
46. Mongan,J., Simmerling,C., McCammon,J.A., Case,D.A. and Onufriev,A. (2007) Generalized Born model with a simple, robust molecular volume correction. *J. Chem. Theory Comput.*, **3**, 156–169.

47. Nguyen,H., Pérez,A., Bermeo,S. and Simmerling,C. (2015) Refinement of generalized born implicit solvation parameters for nucleic acids and their complexes with proteins. *J. Chem. Theory Comput.*, **11**, 3714–3728.
48. Nguyen,H., Roe,D.R. and Simmerling,C. (2013) Improved generalized born solvent model parameters for protein simulations. *J. Chem. Theory Comput.*, **9**, 2020–2034.
49. Romero,P.A. and Arnold,F.H. (2009) Exploring protein fitness landscapes by directed evolution. *Nat. Rev. Mol. Cell Biol.*, **10**, 866–876.
50. Tokuriki,N., Stricher,F., Serrano,L. and Tawfik,D.S. (2008) How protein stability and new functions trade off. *PLoS Comput. Biol.*, **4**, e1000002.
51. Wang,X., Minasov,G. and Shoichet,B.K. (2002) Evolution of an antibiotic resistance enzyme constrained by stability and activity trade-offs. *J. Mol. Biol.*, **320**, 85–95.
52. Rallapalli,K.L., Komor,A.C. and Paesani,F. (2020) Computer simulations explain mutation-induced effects on the DNA editing by adenine base editors. *Sci. Adv.*, **6**, eaaz2309.
53. Otten,R., Liu,L., Kenner,L.R., Clarkson,M.W., Mavor,D., Tawfik,D.S., Kern,D. and Fraser,J.S. (2018) Rescue of conformational dynamics in enzyme catalysis by directed evolution. *Nat. Commun.*, **9**, 1314.
54. Corbella,M., Pinto,G.P. and Kamerlin,S.C.L. (2023) Loop dynamics and the evolution of enzyme activity. *Nat. Rev. Chem.*, **7**, 536–547.
55. Chen,X., McAndrew,M.J. and Lapinaite,A. (2023) Unlocking the secrets of ABEs: the molecular mechanism behind their specificity. *Biochem. Soc. Trans.*, **51**, 1635–1646.
56. Zhu,H., Wang,L., Wang,Y., Jiang,X., Qin,Q., Song,M. and Huang,Q. (2024) Directed-evolution mutations enhance DNA-binding affinity and protein stability of the adenine base editor ABE8e. *Cell. Mol. Life Sci.*, **81**, 257.
57. Ranzau,B.L., Rallapalli,K.L., Evanoff,M., Paesani,F. and Komor,A.C. (2023) The wild-type tRNA adenosine deaminase enzyme TadA is capable of sequence-specific DNA base editing. *ChemBioChem*, **24**, e202200788,

Received: August 8, 2024. **Revised:** October 16, 2024. **Editorial Decision:** October 18, 2024. **Accepted:** November 1, 2024