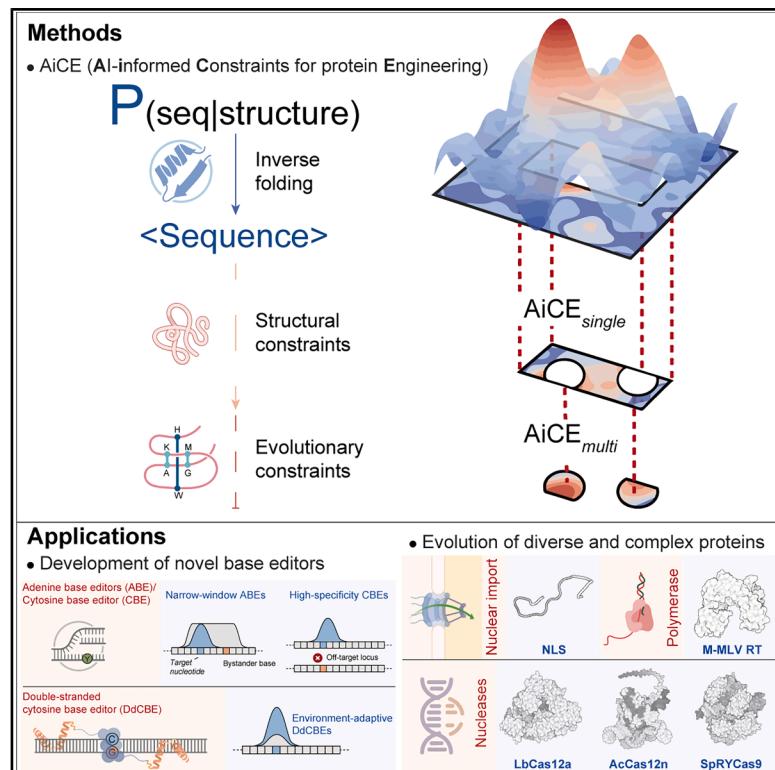


Advancing protein evolution with inverse folding models integrating structural and evolutionary constraints

Graphical abstract



Authors

Hongyuan Fei, Yunjia Li, Yijing Liu,
Jingjing Wei, Aojie Chen, Caixia Gao

Correspondence

cxgao@genetics.ac.cn

In brief

An AI-informed approach integrates generic protein inverse folding models with structural and evolutionary constraints to efficiently identify high-fitness mutations, enabling the development of advanced base editors and demonstrating broad scalability for artificial protein evolution.

Highlights

- Protein inverse folding models can effectively predict high-fitness mutations
- Structural and evolutionary constraints improve AiCE-driven protein evolution
- AiCE enables the development of precise and efficient base editors
- AiCE supports engineering of proteins with varying sizes, structures, and functions



Article

Advancing protein evolution with inverse folding models integrating structural and evolutionary constraints

Hongyuan Fei,^{1,3} Yunjia Li,^{1,2,3} Yijing Liu,^{1,2} Jingjing Wei,¹ Aojie Chen,^{1,2} and Caixia Gao^{1,2,3,4,*}

¹New Cornerstone Science Laboratory, Center for Genome Editing, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China

²College of Advanced Agricultural Sciences, University of Chinese Academy of Sciences, Beijing, China

³These authors contributed equally

⁴Lead contact

*Correspondence: cxgao@genetics.ac.cn

<https://doi.org/10.1016/j.cell.2025.06.014>

SUMMARY

Protein engineering enables artificial protein evolution through iterative sequence changes, but current methods often suffer from low success rates and limited cost effectiveness. Here, we present AI-informed constraints for protein engineering (AiCE), an approach that facilitates efficient protein evolution using generic protein inverse folding models, reducing dependence on human heuristics and task-specific models. By sampling sequences from inverse folding models and integrating structural and evolutionary constraints, AiCE identifies high-fitness single and multi-mutations. We applied AiCE to eight protein engineering tasks, including deaminases, a nuclear localization sequence, nucleases, and a reverse transcriptase, spanning proteins from tens to thousands of residues, with success rates of 11%–88%. We also developed base editors for precision medicine and agriculture, including enABE8e (5-bp window), enSdd6-CBE (1.3-fold improved fidelity), and enDdd1-DdCBE (up to 14.3-fold enhanced mitochondrial activity). These results demonstrate that AiCE is a versatile, user-friendly mutation-design method that outperforms conventional approaches in efficiency, scalability, and generalizability.

INTRODUCTION

Protein engineering leverages the inherent flexibility of proteins—the capacity to modify their structure and function by altering their amino acid sequences.¹ By accelerating evolution far beyond natural rates, it enables the rapid generation of tailored protein variants.²

An ideal protein engineering strategy aims to achieve optimal engineering performance with minimal effort.¹ Current strategies, including structure-guided rational protein design and directed evolution, have achieved significant advances in engineering diverse proteins. However, functionally enhanced variants are rare in the protein fitness landscape, which defines the relationship between sequence variation and functional capacity under specific conditions.^{1,3,4} Existing strategies face difficulties in identifying rare advantageous mutations in proteins in a cost-effective manner (Figure 1A).^{1,5,6} Structure-directed rational protein design relies on empirical human expertise to tailor alterations in protein sequence to achieve desired functional changes but has a low success rate and risks becoming trapped in local fitness optima.^{7–9} Directed evolution, an iterative selection process, can be likened to uphill walks through rugged

fitness landscapes, facing potential evolutionary bottlenecks, high iteration costs, and difficulties in customizing mutations for different scenarios.^{10–13} In recent years, protein engineering methods using deep learning models have yielded excellent results.^{14–19} These methods explore vast protein sequence spaces and generate target mutations based on zero-shot or few-shot learning. While they hold great promise, they often require substantial computational resources for model training and transfer learning, have limited generalizability across diverse proteins, and depend on labor-intensive *in vitro* validation. These constraints present significant barriers for researchers lacking access to specialized infrastructure or expertise.

Generalized protein inverse folding models, such as ESM-IF1 and ProteinMPNN, are AI-driven sequence-generation models that implicitly learn the geometric and physical properties of protein backbones.^{20,21} They can predict high-confidence amino acid sequences that are structurally compatible with a given backbone, effectively performing the inverse of structure prediction. Because they are trained on natural protein structures and sequences, they can capture the complex distribution patterns of protein sequences shaped by evolutionary dynamics. Given this capability, inverse folding models may be directly applied



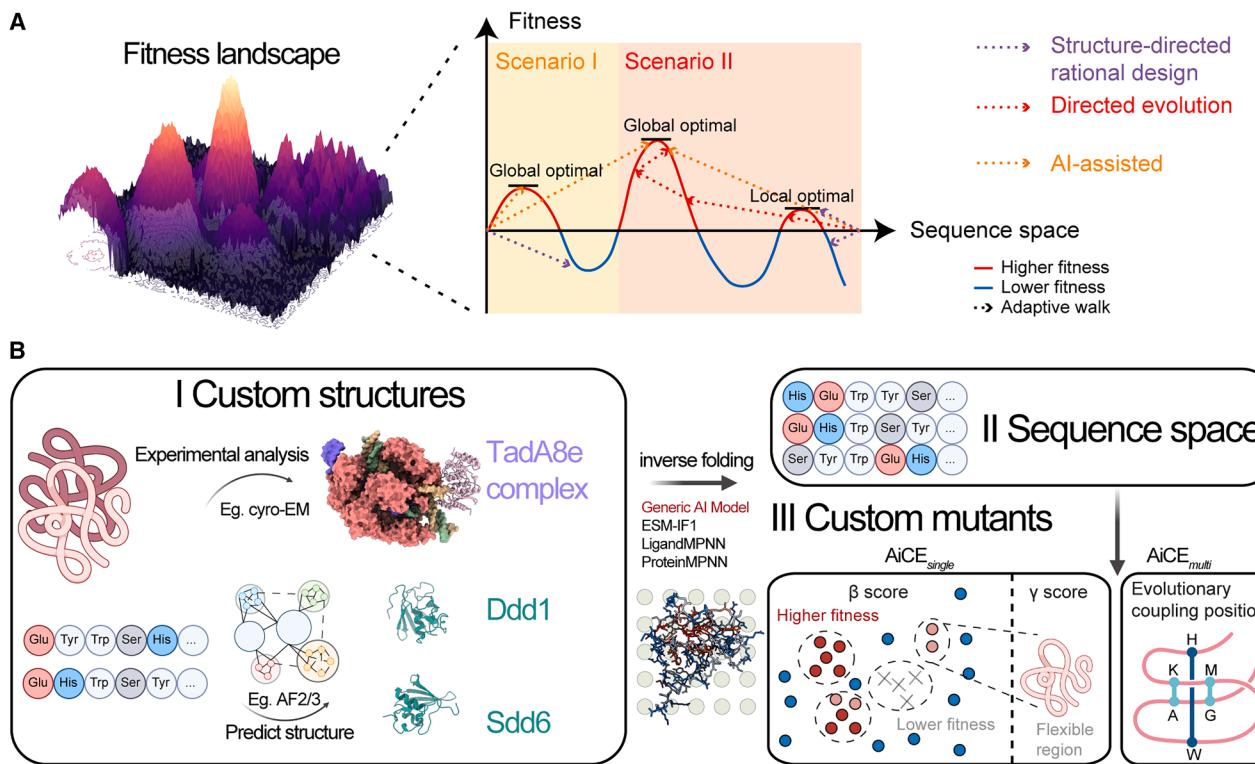


Figure 1. AiCE as an AI-informed approach for protein engineering

(A) Conceptual overview of structure-based, evolution-based, and AI-assisted protein engineering approaches.

(B) Schematic of the AiCE approach for the rational design of single and multiple mutations using protein inverse folding models.

to protein engineering without the need for additional AI model training, offering a more accessible approach to protein engineering. Indeed, it has recently been shown that simply sampling from inverse folding model outputs is sufficient to identify high-fitness (HF) mutations and achieve antibody evolution.²² However, despite these advances, previous studies have only explored conceptually the application of inverse folding models to the evolution of small proteins. Extending these models to large, functionally complex proteins, such as those used in genome editing, presents additional challenges due to functional synergies between enzymes, their interactions with nucleic acid, and endogenous repair systems.^{23,24} Moreover, doubts persist regarding the success rates of both single and combinatorial mutation designs.

We hypothesize that integrating structural and evolutionary constraints can further unlock the potential of inverse folding models for protein engineering. To this end, we present AI-informed constraints for protein engineering (AiCE), an approach that optimizes protein function by incorporating structural and evolutionary constraints into the process of AI-assisted mutation nomination (Figure 1B). AiCE comprises two modules: AiCE_{single}, which nominates single substitutions by sampling mutations from both global and flexible regions of inverse folding outputs using appearance-based screening, and AiCE_{multi}, which identifies combinatorial mutations by leveraging molecular co-evolution principles, where a change at one locus influences selection pressure at another.^{25–28}

In this study, we demonstrate that AiCE enables more efficient and accessible HF mutation nomination than existing deep learning models trained on structural information, thus highlighting the impact of applying structural and evolutionary constraints. Importantly, AiCE is model architecture-agnostic and can optimize both simple protein structures and complex functional enzymes. We have evaluated AiCE across eight diverse protein engineering tasks, including deaminases, nucleases, reverse transcriptases (RTs), and a nuclear localization sequence (NLS), achieving HF mutation prediction success rates ranging from 11% to 88%. Building on these results, we developed a series of precise and efficient base editors, including enABE8e with a smaller editing window, enSdd6-CBE with significantly improved fidelity, and enDdd1-DdCBE, which improves mitochondrial editing efficiency by a factor of 14.3. These editors yielded consistently improved functionality across various genomic target sites and cell types. Our results show that AiCE is an effective, user-friendly protein engineering strategy that reduces computational costs, manual intervention, and high-throughput screening requirements while improving protein engineering efficiency.

RESULTS

Prediction of HF substitutions using structure-constrained inverse folding models

Our goal was to assess the effectiveness of generalized inverse folding models in directing protein evolution and, if effective, to

develop a new protein engineering approach that eliminates the need for custom models or multiple iterations. For this purpose, we developed AiCE, a mutation design framework that leverages predicted or experimentally resolved structural backbones of target proteins. AiCE nominates potential HF substitutions by sampling amino acid sequences generated by generic inverse folding models, requiring no task-specific model training. The workflow consists of several key steps:

1. Using pre-trained protein inverse folding models to generate structurally compatible sequences based on a given protein backbone.
2. Sampling inverse-folding sequences and determining the most frequently occurring amino acid substitution at each position i . The appearance rate $f_x(i)$ of an amino acid x_i at position i is calculated as

$$f_x(i) = \frac{1}{M} \sum_{j=1}^M 1\{x_j = x\}$$

where x represents an amino acid in the standard alphabet, M is the total number of sequences in the multiple sequence alignment, and $1\{\cdot\}$ is the indicator function, which equals 1 when the condition is true and 0 otherwise.

From this, we derive the corresponding appearance rates, $f_{wt}(i)$ and $f_{mut}(i)$, for the wild-type amino acid x_{i_wt} and its mutated counterpart x_{i_mut} . The highest mutation appearance rate at position i , denoted as $f_{max}(i)$, is then determined by

$$f_{max}(i) = \max \{f_{mut}(i) | f_{mut}(i) > f_{wt}(i)\}$$

The mutation type corresponding to $f_{max}(i)$ is denoted as x_{i_mut} .

3. Applying additional structural constraints based on flexibility and prioritizing single substitutions by selecting mutations according to global and flexible-region variations in $f_{max}(i)$.
4. Computing evolutionary coupling (EC) scores between different amino acids in the inverse-folding sequences and nominating multi-mutations based on a weighted scoring matrix derived from interactions between pairwise positions.

We first used 31 large-scale deep mutation scanning (DMS) datasets to evaluate whether HF mutations could be identified solely by direct sampling of inverse folding model outputs, without additional filtering. Here, HF mutations were defined as those ranking in the top 5th percentile of raw library data with experimental effect scores. At this threshold, the expected success rate of randomly designed HF mutations is $\leq 5\%$. The inverse folding models used in this analysis included ESM-IF1, LigandMPNN, and ProteinMPNN.²⁹ These libraries are derived from 27 distinct protein families or domains spanning viruses, bacteria, and eukaryotes (Table S1). They include three protein complex libraries and 28 monomeric protein libraries, covering a wide range of functions, including translational regulation

(infA),³⁰ post-translational modification (SUMO1),³¹ signal transduction (HSP90),³² tumor suppression (p53),³³ and immunity (Cas9).³⁴ This broad representation provides an unbiased evaluation of AiCE's performance. Overall, the predicted mutations had a high proportion of positive fitness outcomes (Figure 2A). The accuracy of HF mutation prediction without additional structural constraints was 12% across all models, with individual accuracies of 12%, 9%, and 12% (Figure 2G; Table S1). These results suggest that inverse folding models may effectively identify HF mutations, although further validation is required.

We further investigated potential factors that influence the accuracy of HF mutation prediction. Given the known link between protein function and structural dynamics³⁵ and that inverse folding models take structure as input, we hypothesized that structural context influences prediction accuracy. To investigate these effects, we categorized mutations nominated by the three inverse folding models into flexible and non-flexible regions based on their structural distribution and found that mutations located in flexible regions had significantly higher fitness ($p < 0.0001$; Figure 2A). To refine HF mutation predictions, we systematically tested different mutation appearance rate thresholds, using 0.1 as an initial cutoff across all DMS libraries. In 30 of the 31 DMS libraries, the optimal screening threshold varied across structural regions (Figure 2B), highlighting the benefit of incorporating structural constraints to improve HF mutation prediction accuracy.

Subsequently, we defined candidate mutations predicted by inverse folding that fall within the top 5th percentile of fitness as relatively HF (RHF) mutations and mapped their enrichment trends across a range of appearance rates (Figures 2C, 2D, and S1A). Structurally flexible regions tended to be enriched in HF and RHF mutations, with predictive accuracy surpassing that of non-flexible regions when mutation appearance rates exceeded 0.5 (Figures 2C and 2D). To reduce potential statistical biases due to the limited number of databases used, we extended our analysis to 29 additional DMS libraries, including six protein complexes (Table S1). The results confirmed that flexible regions had significantly higher prediction accuracy than non-flexible regions ($p < 0.0001$, Figures 2E and S1B). Logistic regression analysis further supported this trend by showing that predictions for flexible regions were 18% more likely to be classified as HF compared with those in non-flexible regions ($p < 0.001$, Figure S1C).

To further investigate these findings, we modeled the relationship between mutation appearance rates in different structural regions and prediction accuracy using machine learning and multilayer perceptron-based deep learning approaches. However, the limited dataset size and uneven data distribution resulted in suboptimal training performance. We therefore developed the AiCE_{single} module—which defines β and γ scores as structure-specific appearance rate-based screening thresholds—to directly identify HF mutations. This prediction method takes into account the differences in the distribution of HF mutations across different structural regions, in contrast to previous screening strategies based solely on global mutation appearance.²² Based on empirical validation (Figure 2E), we define $\beta = 0.8$ and $\gamma = 0.5$ as reference screening thresholds for “AiCE filtering.”

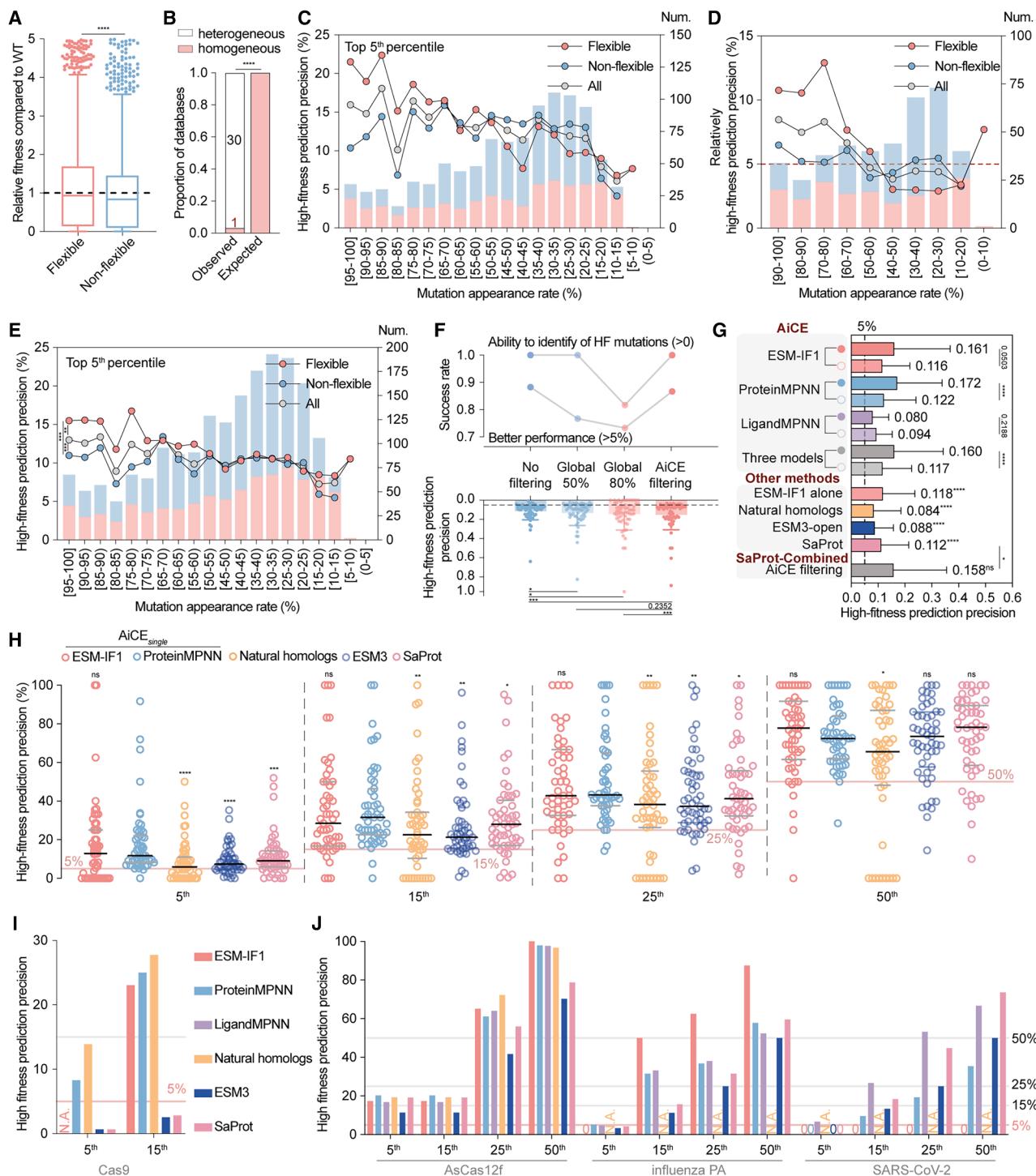


Figure 2. Performance and analysis of AiCE_{single} in predicting high-fitness mutations across various proteins

- (A) Relative value of single amino acid substitutions in flexible versus non-flexible regions, derived from 31 DMS libraries, based on direct sampling of inverse-folding model outputs without additional filtering.
- (B) Comparison of observed versus expected distributions of heterogeneous (white) and homogeneous (pink) selection thresholds across 31 DMS libraries. Statistical significance was assessed using the Fisher's exact test.
- (C) Predictive accuracy of high-fitness (HF) mutations at different mutation appearance rates across 31 DMS libraries. Line graphs show the predictive accuracy of HF mutations within these three regions, whereas columns depict the numbers of such mutations.

(legend continued on next page)

These values were selected to maximize precision while maintaining sufficient recall, thereby improving HF mutation nomination accuracy compared with global screening strategies. A comparative threshold assessment of the predictive accuracies of global-based screening and AiCE filtering revealed that while the predictive success of global screening correlated with mutation appearance rates, AiCE filtering was better able to identify HF mutations than global screening based on appearance rates (Figures 2F and S1D). To further validate $\text{AiCE}_{\text{single}}$, we performed a benchmarking analysis comparing its predictive performance against alternative models and methods, including ESM-IF1 alone, natural homolog-based predictions, SaProt, and ESM3, which incorporate structural information (Figure 2G).^{18,19} AiCE filtering outperformed all the other approaches, with AiCE -ProteinMPNN achieving the highest nomination accuracy—almost twice the accuracy of ESM3-open. The performance of AiCE -LigandMPNN was unsatisfactory, likely due to its smaller database, but it remained effective for nominating HF mutations. Interestingly, we also tested the effect of applying SaProt-derived scoring constraints and observed an improvement in prediction performance compared with unconstrained mutation nomination (Figure 2G). However, further incorporation did not provide a significant improvement over AiCE filtering. These results demonstrate that $\text{AiCE}_{\text{single}}$ is a simple yet highly effective mutation nomination approach that does not require additional model training, and incorporating structural constraints significantly improves prediction success, establishing $\text{AiCE}_{\text{single}}$ with AiCE filtering as a robust and efficient strategy for protein engineering.

Through a comprehensive analysis of protein libraries, we have shown that $\text{AiCE}_{\text{single}}$ with AiCE filtering effectively identifies HF mutations in a variety of evolutionary scenarios simply, quickly, and efficiently—attributes that are not achievable with traditional protein engineering methods (Figures 2H–2J; Table S1). Although AiCE performance varied slightly across models, most models were effective in nominating HF mutations using $\text{AiCE}_{\text{single}}$, indicating that it is a broadly applicable approach independent of model architecture. $\text{AiCE}_{\text{single}}$ enables HF mutation prediction for protein-protein complexes, protein-nucleic acid complexes, and large proteins (>600 residues), such as Cas9 and SARS-CoV-2 variants (Figures 2I and 2J).³⁶ Notably, our re-

sults revealed that inverse folding model outputs can even achieve strong HF mutation predictive performances for proteins that lack natural homologs, such as viral proteins (Table S1).

These conclusions align with prior experimental evidence, including the identification of key mutations such as Q18E in type II methyltransferase (haelliIM)³⁷ and M182T in TEM-1 β -lactamase (Bla) as highest fitness mutations.³⁸ We also identified an HF mutation, N199K, in the class II nuclease, Cas12,³⁹ as well as finding that the Q414A substitution in the receptor-binding domain (RBD) of the SARS-CoV-2 spike glycoprotein is associated with binding fitness.³⁶ $\text{AiCE}_{\text{single}}$ exhibited robust performance across many different scenarios. Specifically, we identified the Y36G substitution in mitogen-activated protein kinase 1 (MAPK1) as potentially HF and confirmed experimentally its ability to confer resistance to two distinct oncogenic MAPK1 inhibitors.⁴⁰ Another example is the successful evolution of breast cancer type 1 (BRCA1) susceptibility protein RING domain mutations across different functional scenarios.⁴¹ $\text{AiCE}_{\text{single}}$ demonstrated high predictive accuracy in identifying HF mutations across a broad range of structurally and functionally diverse proteins.

$\text{AiCE}_{\text{multi}}$: A rational approach for designing HF multi-mutations based on EC constraints

A major challenge in protein engineering is negative epistasis, where combinations of individually beneficial mutations can reduce fitness or lead to loss of function.^{42,43} To address this issue, we developed a mutant screening module, designated $\text{AiCE}_{\text{multi}}$, which identifies potentially beneficial multi-mutations by recognizing sites that are strongly evolutionarily coupled. The core concept of $\text{AiCE}_{\text{multi}}$ exploits co-evolutionary analysis, using EC as the key selection criterion. This approach includes two main components: (1) a linkage disequilibrium-like analysis of simulated DNA sequences derived from inverse-folding results, detecting non-random co-occurrences of mutations at multiple sites, and (2) a direct statistical coupling analysis (SCA), which quantifies amino acid interdependencies based on evolutionary constraints.

To evaluate the performance of $\text{AiCE}_{\text{multi}}$, we analyzed the affinity landscapes of two antibodies, CR6261 and CR9114, that target different influenza hemagglutinin (HA) subtypes (H1 and H9 for CR6261 and H1 and H3 for CR9114).⁴⁴ These libraries

(D) Enrichment trends of RHF mutations at various mutation appearance rates. The dashed line represents the expected top 5th percentile enrichment. Other labeling is as in (C).

(E) Same as (C), but for 60 DMS libraries. The overall accuracy of HF mutation prediction remained stable as the mutation appearance rate increased, improving after 0.8. Under structural constraints, accuracy showed an upward bias in the flexible region after 0.5. Therefore, we recommend $\beta = 0.8$ and $r = 0.5$ as general parameters for $\text{AiCE}_{\text{single}}$ screening, defining it as “AiCE filtering.”

(F) Comparison of different filtering strategies for predicting HF mutations. “Ability to identify HF mutations” refers to prediction capacity, and “better performance” indicates that the prediction accuracy exceeds 5%.

(G) Comparison of different methods for predicting HF mutations. Solid circles represent the “AiCE filtering” strategy, and hollow circles indicate the “no filtering” strategy. “Other methods” show statistically significant differences relative to the $\text{AiCE}_{\text{single}}$ -three models.

(H) $\text{AiCE}_{\text{single}}$ prediction accuracies across percentile thresholds (5th, 15th, 25th, and 50th) based on 51 monomeric protein libraries. Each dot represents a DMS library, and black lines denote median and interquartile ranges. Theoretical prediction accuracies are indicated by pink reference lines.

(I) Performance of $\text{AiCE}_{\text{single}}$ in predicting HF mutations in Cas9. Gray reference lines indicate theoretical predictive accuracies.

(J) $\text{AiCE}_{\text{single}}$ prediction accuracies for HF mutations in protein-protein and protein-nucleic acid complexes. “N.A.” indicates cases lacking sufficient homologs for mutation prediction.

For (A) and (F)–(H), bars represent mean values \pm SD. p values from one-sided Mann-Whitney tests are shown: ^{ns} $p > 0.05$, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$, and $****p < 0.0001$.

See also Figure S1.

were selected for their comprehensive coverage of all combinations of mutations and were primarily used to assess the effects of combining mutations. For each dataset, we first calculated weighted mutation scores using different models and determined their correlation with the equilibrium dissociation constants (K_D) of the multi-mutations (Figures 3A, 3B, S2A, and S2B). The results showed varying degrees of correlation between weighted mutation scores and K_D , with differences between models influencing the correlations. LigandMPNN exhibited the strongest correlations across all four libraries, with Spearman's rank correlations ranging from 0.32 to 0.64 (Figures 3A and 3B). We then extracted positional combinations with high SCA scores, computed from inverse-folding sequences, or high LD (linkage disequilibrium) scores, computed from pseudo-DNA sequences derived from inverse-folding sequences. Candidate combinations from both inverse folding models enhanced germline antibody fitness, with some combinations even exceeding the fitness of mature antibodies. Although a high weighted mutation score does not always correlate with HF, AiCE_{multi} effectively identified HF mutation combinations with high confidence. This suggests that AiCE_{multi} provides a computationally efficient alternative to traditional experimental approaches for optimizing multi-mutations. This enhances the success rate of screening for multi-mutations with HF potential (Figure S2C). We extended this analysis to two other proteins, His3 (from yeast, catalyzing the sixth step in histidine biosynthesis) and pfluGFP2, and identified several multi-mutations with higher relative and absolute fitness (Figure 3C).^{45,46} Interestingly, the fitness of these multi-mutations, identified through AiCE_{multi}'s evolutionary-coupling analysis, did not show a strong correlation with weighted mutation scores or EC scores alone. This suggests that AiCE_{multi} leverages additional structural and evolutionary constraints to identify HF multi-mutations that may be overlooked by previous methods.²² SaProt also predicts multi-mutation variants, allowing for a direct performance comparison with AiCE_{multi}. To compare its performance with AiCE_{multi}, we performed a comparative analysis. We selected the top 20 SaProt-predicted mutations and observed prediction accuracies of 95% and 70% for fitness-enhancing mutations in His3 and pfluGFP2, respectively (Figure S2D). AiCE_{multi} had a comparable predictive performance to SaProt but was alone in identifying mutations capable of overcoming epistatic constraints. Interestingly, SaProt exhibited lower predictive accuracy across four antibody libraries, as demonstrated by the top 30 mutations identified by the model (Figure S2C), likely due to the underrepresentation of antibody structures in SaProt's training data.

In summary, we introduce AiCE_{multi}, a module designed for efficient nomination of multi-mutations. As a critical component of the AiCE framework, it consists of two primary phases (Figure 3D). Initially, AiCE_{single} identifies potentially beneficial mutation types. Then EC or LD scores are used to identify high-probability evolutionarily coupled positions. We propose that high EC scores serve as predictive markers for identifying structurally and functionally compatible multi-mutations, facilitating rational protein design. We conclude that AiCE_{multi} represents a state-of-the-art (SOTA) approach that enables the prediction of HF multi-mutations.

Use of AiCE to design a single-stranded DNA adenine deaminase with a narrower deamination window

To assess the ability of AiCE to address complex protein engineering challenges, we used it to optimize genome-editing-associated proteins. Genome editing technologies, exemplified by the CRISPR-Cas system and its derivatives, offer unprecedented opportunities to study complex biological processes and address the root causes of genetic diseases.⁴⁷ Improving their efficiency and specificity through protein engineering is essential for expanding their application.^{47,48} The flexibility and high efficiency of AiCE provide a powerful framework for rapid and precise engineering of these proteins.

Our first goal was to optimize TadA8e, a highly engineered single-stranded DNA (ssDNA) adenine deaminase that serves as the core component of widely used adenine base editors (ABEs) due to its high A-to-G editing specificity and efficiency (STAR Methods).⁴⁹ Despite its efficacy, TadA8e has a relatively broad editing window, which leads to unintended bystander editing.^{49,50} Additionally, its editing efficiency at certain target sites is limited, and it has proved difficult to remedy using conventional techniques. Using the cryoelectron microscopy (cryo-EM) structure of TadA8e (PDB: 6VPC) as a template, we enlisted the help of AiCE_{single} with three different inverse folding models alongside six other strategies and generated 122 single mutations (Table S2). We introduced these mutations into ABEs and tested them at three endogenous target sites in HEK293T cells (Figure S3A). HF mutations were defined as those exhibiting at least a 10% increase in editing efficiency compared with the wild-type TadA8e enzyme. We identified 13 such mutations, six of which were in structurally flexible regions (Figure 4A). Of these, 11 were nominated by AiCE_{single}, with the top variant showing a ~47% increase in editing efficiency. AiCE_{single}-ProteinMPNN had the highest prediction accuracy (35%), outperforming other AI models, including ProMEP, which was previously effective for TadA1.2 (a precursor of TadA8e).¹⁷

To explore multi-mutations, we used AiCE_{multi}-ProteinMPNN to nominate 10 multi-mutants (AiCE_{multi} mutants, Table S2). As a control, we used BLOSUM62 matrices to generate nine multi-mutants (BLOSUM62 mutants, Table S2). Six HF AiCE_{multi} variants, containing up to 4 substitutions, and five HF BLOSUM62 variants were identified (Figures 4B and S3B). Ablation experiments revealed that the BLOSUM62 mutants had significantly lower deamination activities than the AiCE_{multi} mutants ($p < 0.0001$, Figures 4C and S3B). Furthermore, the latter exhibited fewer negative epistatic effects: the AiCE_{multi} and BLOSUM62 mutants partially or fully avoided epistatic effects in 6 and 4 cases, respectively, highlighting AiCE_{multi}'s ability to identify effective combinatorial mutation sites (Figure S3B). Since AiCE_{multi} selects mutations based on EC scores, we further introduced low EC score controls (Low_{multi}) by randomly designing five additional variants, including two Low_{multi} mutants. Deep sequencing confirmed two efficient mutations, both from AiCE_{multi} mutants (Figure 4D). The Tc22 variant (C83S/C86S) in HTT had the highest efficiency and could potentially be used to edit Huntington's disease, a disorder associated with repetitive sequence regions.⁵¹

To further refine the editing window and reduce bystander mutations, we performed a second evaluation. We randomly selected 18 HF mutants (13 single and 5 multiple) and evaluated

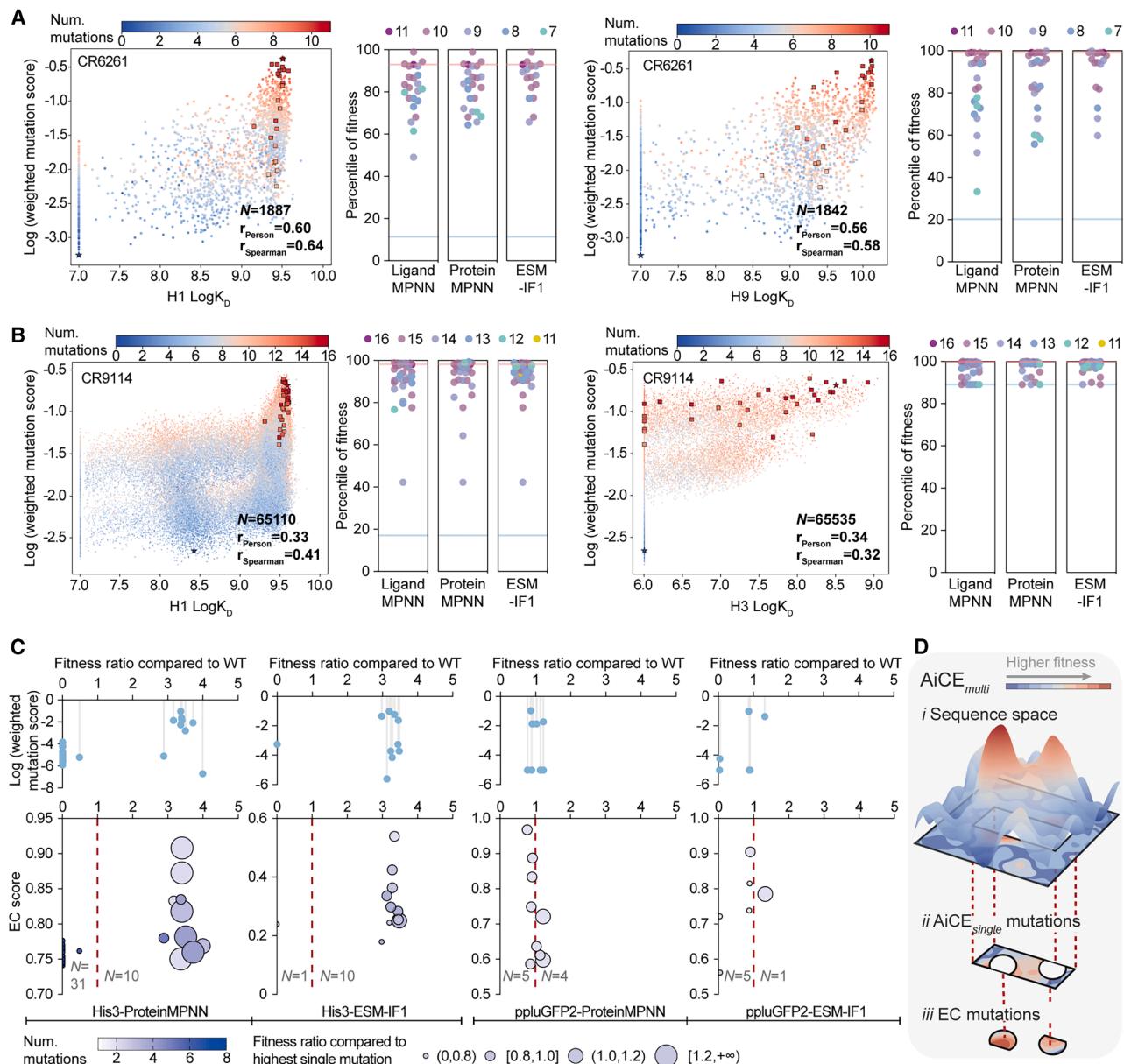


Figure 3. Evaluating the performance of AiCE for generating interpretable predictions of HF multi-mutations

(A and B) Predictive performance of AiCE_{multi} for HF multi-mutations in CR6261 and CR9114 antibodies. (A) Correlations between AiCE_{multi}-LigandMPNN predictions and experimentally determined dissociation constants, and ranking of predicted mutant fitness in libraries, for CR6261 binding to HA-H1 (left) and HA-H9 (right). (B) Analogous evaluation for CR9114 binding to HA-H1 (left) and HA-H3 (right). In scatterplots, rectangles represent candidate HF multi-mutations. Strip plots depict fitness percentile distributions across models. The vertical axis indicates the percentile of fitness in the mutation library, with each dot representing the number of multi-mutations in the mutant. Blue and red lines represent cumulative distributions of germline and mature antibodies, respectively.

(C) Predictive performance of AiCE_{multi} for high-fitness (HF) multi-mutations in His3 and ppluGFP2. Bubble color reflects the number of multi-mutations, and bubble size reflects relative fitness.

(D) Conceptual schematic of the AiCE_{multi} principle.

See also Figure S2.

their activities within the editing window at six additional target sites (Figures 4E and S3C). Despite target-dependent variations, we identified two improved variants, T1 (E1M) and Tc1 (A11V/G27A), with enhanced efficiency and precision. T1 displayed an improvement of up to ~70% in editing efficiency within the

target window, and the more precise Tc1 variant, which has two substitutions, may stabilize the N terminus, thereby reducing bystander editing (Figure S3D).

We next used Tc1 to construct enABE8e and performed a comprehensive evaluation of the latter alongside ABE9, one of

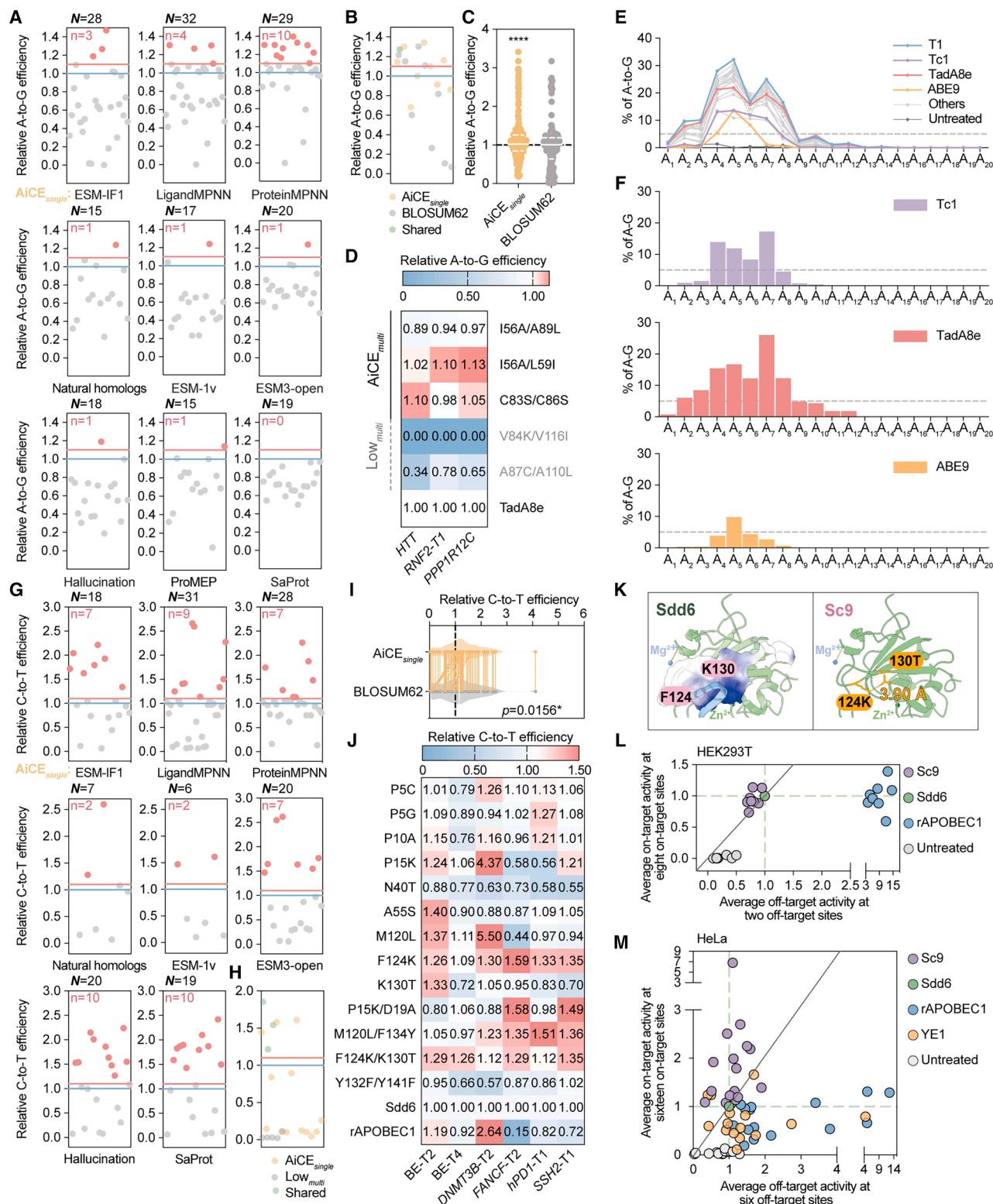


Figure 4. Proof of concept for customizing function-specific single-stranded DNA deaminases using AiCE

(A) Optimization of TadA8e variants using different approaches. Average relative editing efficiencies of TadA8e variants at three endogenous sites in HEK293T cells, normalized to wild-type TadA8e (set to 1). High-fitness (HF) mutations are marked in red, with their counts (n) indicated. Total number of variants (N) generated by each method is shown. Dots represent individual variants, and the blue line marks wild-type efficiency.

(legend continued on next page)

the most precise ABEs, across 24 target sites spanning eight genes in HEK293T, HeLa, K562, and U2OS cells (Figure 4F; Data S1).⁵² Overall, editing performance was consistent across cell lines, with the following editing windows: ABE8e (wild type), A₂-A₁₀ (9 bp window); ABE9, A₄-A₆ (3 bp window); and enABE8e, A₄-A₇ (4 bp window). The enABE8e variant exhibited a significantly narrower editing window than ABE8e while maintaining comparable or superior efficiency at more than half of the sites tested. Although enABE8e's editing window was only 1 bp wider than ABE9, it significantly outperformed ABE9 in editing efficiency. With its robust editing efficiency and significantly narrower editing window, enABE8e represents one of the most precise ABEs available.

Unconventional engineering can increase the specificity of ssDNA cytosine deaminases

Cytosine base editors (CBEs), built with ssDNA cytosine deaminases, enable C-to-T conversions.⁵³ Among the existing cytosine deaminases, rAPOBEC1 is the most widely used for its high editing efficiency. However, it suffers from poor specificity and often induces unintended off-target effects across the genome, limiting the safety of its use.^{54,55} Although high-fidelity variants such as rAPOBEC1-YE1 and Sdd6 have been developed, they remain limited in specificity and efficiency and require further optimization.^{56,57} We selected Sdd6 of the SCP1.201 family as the starting point for improving both activity and specificity. Following a strategy similar to TadA8e optimization, we used the AlphaFold3-predicted Sdd6-ssDNA structure to nominate 114 single and 21 multi-mutations (Table S2).⁵⁸ These mutations were incorporated into base editors and underwent preliminary testing for deamination activity at two endogenous targets in HEK293T cells (Figure S4A). Deep sequencing identified 48 single mutations (23 in the flexible regions) and six multi-mutations as HF mutations, none of which originated from Low_{multi}-based nominations (Figures 4G, 4H, and S4B; Table S2). The highest increase in editing efficiency (1.7-fold) was observed in a mutant nominated by AiCE_{single}-Li-

gandMPNN. HF multi-mutation variants exhibited editing efficiency increases from 22% to 85%, with Sc16 (a five-mutation combination) showing a 54% improvement. Although many Sdd6 multi-mutants were nominated by both AiCE_{multi} and BLOSUM62, ablation experiments showed that AiCE_{single}-nominated mutations yielded slightly higher deamination activity ($p = 0.0156$, Figures 4I and S4B; Table S2). Further analysis showed that, within the limited control dataset, three variants partially and three fully avoided epistatic constraints (Figure S4B; Table S2).

We randomly selected 13 HF candidates (9 single and 4 multi-mutants) for further evaluation at six additional target sites in HEK293T cells. Eleven of the mutants exhibited at least a 10% increase in deamination activity at one or more target sites (Figures 4J and S4C). Sc9 (F124K/K130T) was the most stable variant, increasing editing efficiency by 12%–35% at six target sites.

To enhance base-editing specificity, we aimed to develop a higher-fidelity deaminase essential for improving editing safety. Structures of Sdd6 mutant-ssDNA complexes predicted by AlphaFold3 unexpectedly revealed that Sc9 should also function as a high-fidelity variant. Specifically, it harbors mutations that ought to reduce the flexibility of the positively charged surface region, potentially reducing off-target deamination (Figure 4K). To evaluate Sc9 specificity, we performed an orthogonal R-loop assay in HEK293T cells.⁸ Sc9 showed a higher on-target/off-target ratio (referred to as “fidelity”) than wild-type Sdd6 at eight on-target and two off-target sites (Figures 4L and S4D), improving specificity by ~30% and increasing the on-target/off-target ratio by ~20%.

Given enSdd6-CBE's potential as a high-efficiency, high-fidelity CBE, we further characterized its deamination activity across different cell types. Compared with Sdd6, it exhibited editing efficiency increases of 19%, 84%, and 31% in HEK293T, HeLa, and U2OS cells, respectively. Compared with rAPOBEC1, it showed increases of 44%, 195%, and 127% (Figure S4G). Despite ~9% reduced deamination activity in K562 cells, its

(B) Prediction of multi-mutations using AiCE_{multi} and BLOSUM62 matrix. Average relative editing efficiencies were measured at three endogenous sites in HEK293T cells. Mutations occurred at the same positions but differed in substitution type. “Shared” variants (green) were identified by both methods. Other labeling is as in (A).

(C) Comparison of editing efficiencies between AiCE_{multi}- and BLOSUM62-predicted TadA8e multi-mutations and their constituent single mutations. Each dot represents editing efficiency of an individual mutation at three endogenous sites, measured in three independent biological replicates (each replicate plotted separately). Efficiencies were normalized to wild-type TadA8e (dashed line). p values from one-sided Mann-Whitney tests are shown: *** $p < 0.0001$.

(D) Relative editing efficiencies of three AiCE_{multi}-predicted multi-mutations, two randomly selected low-scoring multi-mutations (Low_{multi}), and wild-type TadA8e across three endogenous sites in HEK293T cells.

(E) Average editing efficiency across editing windows for 18 selected TadA8e variants (13 single, 5 multi-mutations), TadA8e, and ABE9 across six sites in HEK293T cells.

(F) Editing window profile of Tc1, TadA8e, and ABE9 across 24 target sites in HEK293T cells. The gray dashed line indicates a 5% editing efficiency.

(G) Optimization of Sdd6 variants using different approaches. Average relative editing efficiencies of Sdd6 variants at two endogenous sites. Labeling is as in (A).

(H) Prediction of multi-mutations using AiCE_{multi} and BLOSUM62 matrix. Low-scoring multi-mutations (Low_{multi}) were randomly selected. Other labeling is as in (B).

(I) Comparison of AiCE_{multi} and BLOSUM62 matrices for predicting Sdd6 mutations. Labeling is as in (C).

(J) Relative editing efficiencies of selected Sdd6 variants across six endogenous sites in HEK293T cells.

(K) Structural analysis of Sdd6 and Sc9. The AlphaFold3-predicted structures and electrostatic potential maps of Sdd6 and Sc9 proteins in complex with single-stranded DNA (ssDNA), magnesium ions (Mg^{2+}), and zinc ions (Zn^{2+}).

(L and M) Specificity assessment of Sc9 in HEK293T (L) and HeLa (M) cells using orthogonal R-loop assays, compared with Sdd6 (wild type), rAPOBEC1, and rAPOBEC1-YE1 (a high-fidelity variant) for off-target effects.

Data are the average of three independent biological replicates.

See also Figures S3 and S4.

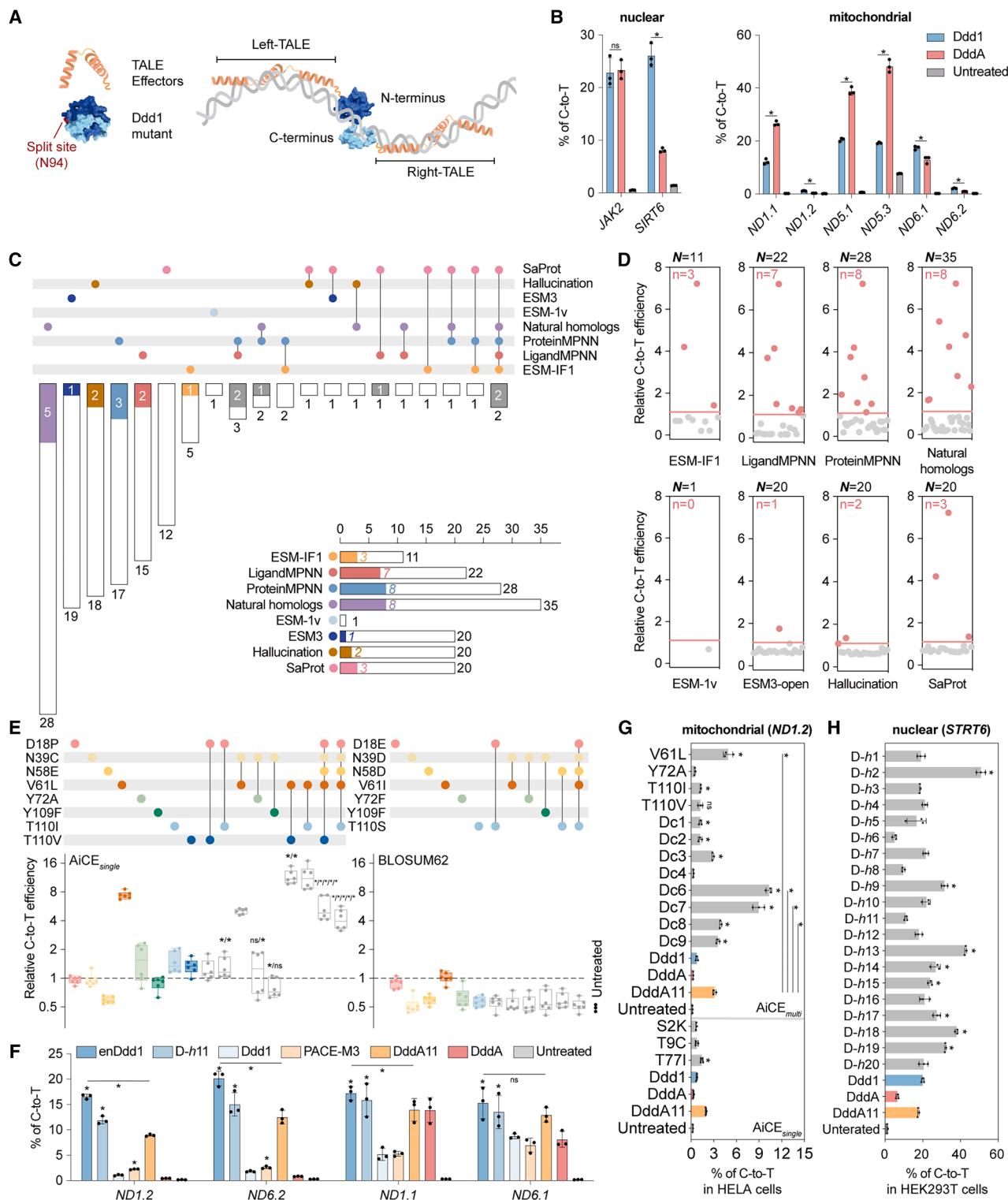


Figure 5. Multi-method-based evolution of the double-stranded DNA deaminase Ddd1 for environmental fitness

(A) Schematic of the DdcBE detection system using split Ddd1 fused to TALE effectors for C-to-T editing.
(B) Editing efficiencies of Ddd1 and DddA at two nuclear and six mitochondrial targets in HEK293T cells.

(legend continued on next page)

efficiency remained significantly higher than rAPOBEC1 (~160%). In terms of editing specificity, an on/off-target analysis of 16 on-off-target pairs at six off-target sites in HeLa cells showed that enCBE exhibited a significant specificity improvement. It improved specificity by ~28% and fidelity by ~132% compared with Sdd6 (Figures 4H, S4E, and S4F). Additionally, it exhibited a 51% improvement in specificity compared with rAPOBEC1-YE1 (Figures 4H, S4E, and S4F).

Strikingly, the two substitutions in Sc9, identified in a single round of rational design screening, would have been difficult to predict using traditional methods, as they lacked prior literature support and were located outside well-characterized catalytic or binding regions. This highlights the power of AI-informed methods such as AiCE that provide a new basis for rational protein design.

Improving the environmental adaptability of double-stranded DNA cytosine deaminases

Given the effectiveness of AiCE in custom mutagenesis, we applied it to engineer the double-stranded DNA deaminase Ddd1 to enhance its efficacy in mitochondrial DNA editing.⁵⁷ Double-stranded cytosine deaminases of the SCP1.201 family, which act on double-stranded DNA, can be fused with transcription activator-like effector (TALE) to form DdCBEs (double-stranded CBEs) (Figure 5A).⁵⁹ These enable mitochondrial DNA base editing, offering a potential strategy for correcting mutations linked to Leber's hereditary optic neuropathy and maternally inherited deafness.^{60,61} Ddd1 is more efficient than its homolog DddA in editing 5'-GC sequences, potentially expanding DdCBEs applications.⁵⁷ However, while Ddd1 exhibits efficient deamination activity in the nuclear environment, its editing efficiency in the mitochondrial environment is significantly lower (Figures 5B and S5B). Pronounced differences between mitochondrial and nuclear microenvironments make rational Ddd1 optimization challenging, but we hypothesized that AiCE could overcome this limitation. Hence, to further optimize Ddd1, particularly for editing 5'-GC sequences, we used the AlphaFold2-predicted structures of Ddd1 and the complex structure of the DddA homolog resolved by cryo-EM (PDB: 8E5E) as target backbones. Using multiple prediction methods, we identified 138 single mutants and 15 multi-mutants (Figure 5C; Table S2). These mutants were incorporated into individual TALE vectors (Figures 5A and S5A), and their deamination efficiencies assessed at two mitochondrial 5'-GC sequences in HEK293T cells. Target deep sequencing demonstrated that 20 single mutants (13 in flexible regions) and 8 multi-mutations (seven fully predicted by

AiCE_{multi}) were identified as HF mutations (Figures 5D and 5E). HF mutation prediction accuracies of AiCE_{single} using ESM-IF1, LigandMPNN, and ProteinMPNN were 27%, 32%, and 29%, respectively, while other methods ranged from 0% to 23%. To validate these predictions, we tested the top four sets of mutants with the highest prediction accuracies at two non-5'-GC target sites (Figures S5C and S5D). Notably, 21%–32% of these mutants exhibited higher deamination activities than DddA. The best of these mutants, D71 (V61L), had 0.9–6.5-fold increased deamination activity.

AiCE_{multi} outperformed BLOSUM62-based predictions in identifying the two most efficient multi-mutations: Dc6 (V61L/T110V) and Dc7 (V61L/T110I) (Figure 5E). Compared with BLOSUM62-derived mutants, AiCE_{multi}-nominated variants exhibited 6.9-fold higher deamination activity (Figure S5E). Both mutants share a high EC score mutation position, with substitutions nominated by AiCE_{single} and DddA11 (the most efficient DddA variant reported to date,⁶² which contains the PACE-M3 substitution) (STAR Methods). Dc6 and Dc7 performed deamination equally efficiently across different cell lines and mitochondrial targets (Figure S5F), demonstrating the predictive accuracy of AiCE and its compatibility with other protein engineering strategies. We designated Dc7 as enDdd1 and evaluated its deamination activity at four mitochondrial target sites in HEK293T cells. The results showed that enDdd1-DdCBE exhibited 1.8- to 14.3-fold higher activity than Ddd1-DdCBE and 1.2- to 1.9-fold higher activity than DddA11-DdCBE (Figures 5F, S5H, and S5I). In addition, enDdd1 edited 5'-GC sequences efficiently, achieving approximately 40% higher efficiency than DddA11 (Figure S5I), which had undergone multiple rounds of phage-assisted continuous evolution and contains six mutations (Figure S5G). This suggests that advantageous mutations obtained through AiCE are compatible with those derived by directed evolution, facilitating rapid functional iteration.

These HF mutants retained robust deamination activity across cell lines. To further validate their performance, we randomly selected 7 single mutants and 8 multi-mutants from the first screening round and evaluated their deamination activity at the ND1.2 target site in HeLa cells (Figure 5G). Of these mutants, 4 single mutants and 7 multi-mutants showed a significant increase in deamination activity. The 7 HF multi-mutants contained up to five amino acid substitutions, and four mutants had higher deamination activities than DddA11. We further tested AiCE's ability to enhance deamination activity in a nuclear environment. Among the 20 mutants generated using AiCE_{single}-ProteinMPNN, seven exhibited significantly elevated deamination activities in

(C) Computational strategies used to generate Ddd1 variants. Total number of tested variants for each method is indicated outside the bars, while the number of HF variants is shown inside the bars. Black lines connecting bars indicate overlaps between variants identified by different methods.

(D) Relative editing efficiencies of Ddd1 variants at two mitochondrial sites (ND1.2 and ND6.2) in HEK293 cells, normalized to wild-type Ddd1 (set to 1). Red line indicates 1.1 threshold. Other labeling is as in Figure 4A.

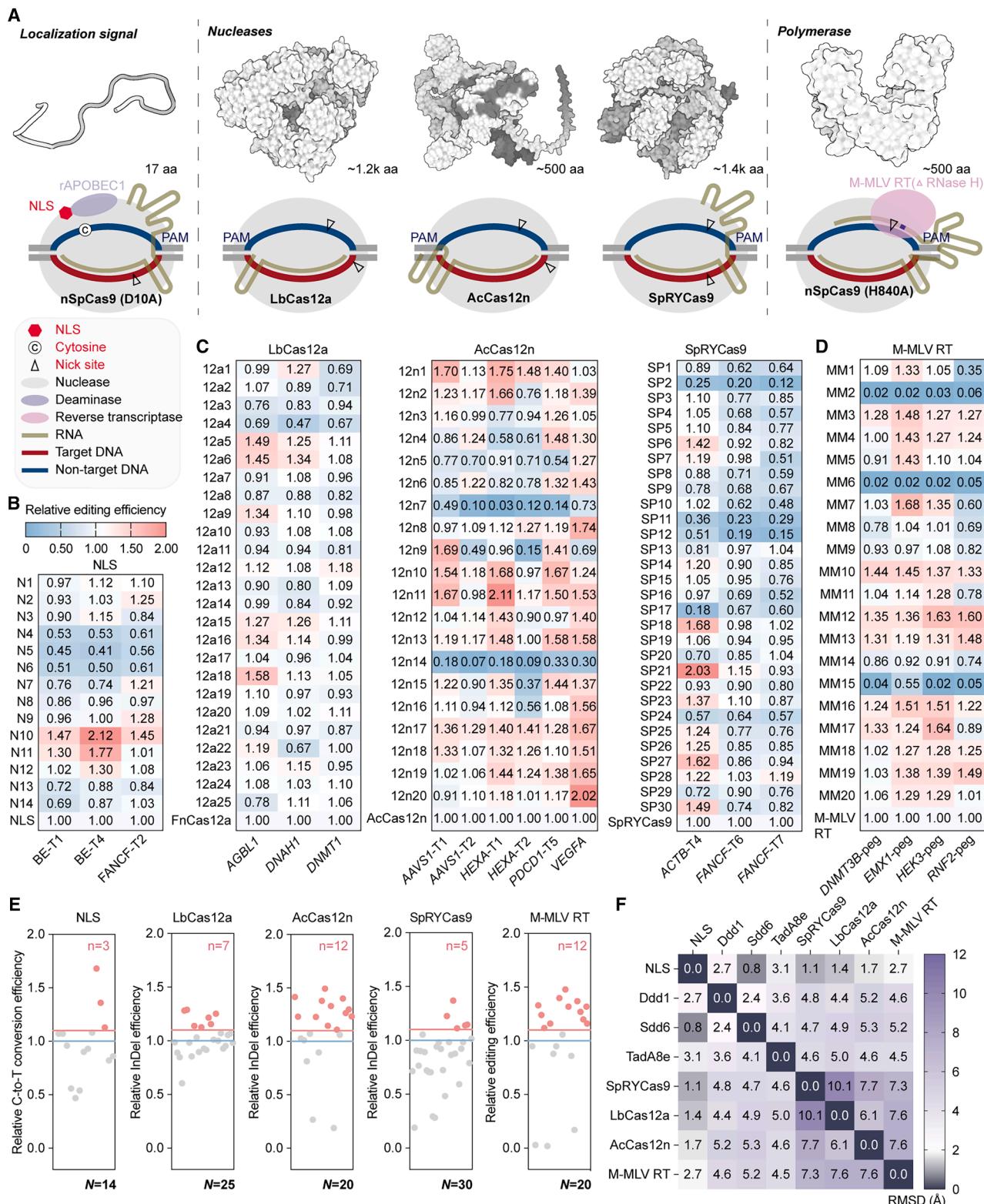
(E) Relative editing efficiencies of multi-mutations and the corresponding single mutations identified by AiCE_{multi} (left) and BLOSUM62 (right). The top panel shows the mutation types, with black lines connecting the single mutations that form each combinatorial variant.

(F) Editing efficiencies of enDdd1, D-h11, Ddd1 (wild type), PACE-M3, DddA11, and DddA at four mitochondrial sites in HEK293T cells.

(G and H) Editing efficiencies of Ddd1 variants, Ddd1 (wild type), DddA, and DddA11 at mitochondrial (G) and nuclear (H) sites in HeLa and HEK293T cells. The D-hs are a set of 20 mutants obtained by AiCE_{single}-ProteinMPNN with a parameter value of $\beta = 0.8$.

For (B) and (E)–(G), bars represent mean values \pm SD. p values from one-sided Mann-Whitney tests are shown: ^{ns} $p > 0.05$, $*p < 0.05$, $**p < 0.01$. Data are the average of three independent biological replicates.

See also Figure S5.



(legend on next page)

the nuclear environment, with up to a 1.6-fold improvement over wild-type Ddd1 (**Figures 5H** and **S5J**). These results demonstrate AiCE's ability to create double-stranded cytosine deaminase variants optimized for different environments.

AiCE_{single} facilitates the evolution of diverse complex proteins

Building on AiCE's success in deaminase engineering, we applied it to additional complex tasks to further improve genome-editing capabilities. We selected an NLS and four proteins as engineering targets (**Figure 6A**): (1) an NLS that facilitates nuclear import, potentially improving the nuclear localization and efficiency of genome-editing enzymes (**STAR Methods**). (2) LbCas12a and AcCas12n are Type V CRISPR effectors that differ substantially from Cas9 proteins in sequence, structure, and cleavage mechanisms.^{63,64} (3) SpRYCas9 is a type II CRISPR-Cas9 variant with increased tolerance for diverse PAM sequences, though its editing activity is suboptimal.⁶⁵ (4) Moloney murine leukemia virus (M-MLV) RT is an enzyme critical for prime editing, though its efficiency remains a limiting factor requiring protein engineering.^{66,67}

Using AiCE_{single}-ProteinMPNN, we designed mutations for each target and evaluated their genome-editing potential in HEK293T or HeLa cells (**Figure S6**). At least one mutation improved editing efficiency by $\geq 10\%$ at one or more target sites for NLS (8/14), LbCas12a (14/25), AcCas12n (18/20), SpRYCas9 (13/30), and M-MLV RT (14/20) (**Figures 6B–6D**). The average prediction accuracies across all target sites were 21%, 28%, 60%, 17%, and 60%, respectively, exceeding those of conventional protein engineering approaches (**Figure 6E**).^{1,2} These five proteins vary significantly in size (from tens to thousands of residues) and exhibit considerable structural heterogeneity (**Figure 6F**). Despite these differences, AiCE successfully improved all target proteins. In addition, AiCE reduces the experimental costs associated with protein engineering. Unlike traditional approaches that require extensive high-throughput screening, AiCE enables efficient protein evolution using only a limited number of tested variants (typically \sim tens per target).

DISCUSSION

The development of high-functioning mutants at low cost has long been a goal of protein engineering. The challenge arises from the multidimensional nature of protein structure and the complex relationship between sequence and function.¹ Dynamic proteins, such as enzymes, receptors, and channel proteins, are difficult to model using static structures, limiting the effectiveness of traditional approaches.¹³ Our findings demonstrate that AiCE, as an HF mutation prediction method, can identify HF mutations efficiently by sampling inverse-folding sequences.

Since AiCE does not require model training or transfer learning, it can rapidly identify and nominate single amino acid substitutions and double mutants for complex proteins such as SpCas9, requiring as little as 1.15 CPU hours. This enables rapid protein evolution without the need for high-throughput screening or specialized expertise in computational design methods.

AiCE operates on the assumption that inverse folding models inherently capture the natural dynamics governing the generation and distribution of protein sequences and structures. This implies that sampling inverse-folding sequences can predict HF mutations, a hypothesis supported by our experimental results across several protein families. While a concurrent study has conceptually validated this hypothesis for antibody evolution, AiCE extends this approach by integrating structural and evolutionary constraints to guide mutation nomination.²²

AiCE introduces several key innovations that enhance its predictive power and practical application. First, using a dataset nearly six times larger than those in previous studies, we found that mutations in flexible regions are more likely to increase fitness. This effect was not only observed for ProteinMPNN, which favors stabilizing mutations, but also for ESM-IF1. A likely explanation is that flexible regions experience fewer structural constraints, enabling them to accommodate mutations without disrupting overall functional stability.^{68–71} Building on this insight, we developed AiCE_{single}, a module that improves HF mutation prediction by incorporating structural-region-based filtering, achieving SOTA accuracy in predicting beneficial single amino acid substitutions while reducing experimental costs associated with protein engineering. Second, inspired by co-evolutionary theory, we developed AiCE_{multi}, a module for predicting multi-mutation variants, which we validated in deaminase engineering. Despite limited experimental data, our results suggest that identifying evolutionarily coupled functional sites can reduce negative epistatic effects, providing a simpler, more efficient, and interpretable alternative to combinatorial mutation models based on deep learning. Third, AiCE represents a paradigm shift in protein engineering, prioritizing the use of general-purpose large models over training specialized models from scratch. This approach is similar to that of BindCraft, which applies pre-trained AI models to design functional protein binders with high specificity.⁷² A consequence of these various innovations is that AiCE excels at engineering proteins related to genome editing, including protein-nucleic acid complexes and large proteins such as ABE8e and the CRISPR system.

Despite relying on generic protein models, AiCE generates HF mutations with sequences and distribution patterns that vary depending on the target protein backbone (**Figure 7A**). This adaptability is a key advantage of using generic protein models. Additionally, we observed that mutations nominated from inverse-folding sequences frequently appeared in unconventional

Figure 6. AiCE_{single}-based optimization of other functional protein domains

- (A) Schematic representation of the detection strategies used to evaluate the performance of AiCE_{single}-optimized variants across multiple proteins.
- (B–D) Optimization of variants for the NLS (B), nucleases (C), and RT (M-MLV RT) (D). Activities were normalized to the corresponding wild-type proteins (set to 1).
- (E) Distribution of relative editing efficiencies across five proteins. Labeling is as in **Figure 4A**.
- (F) Structural correlation analysis of eight AiCE-optimized proteins. Data represent root-mean-square deviation (RMSD) correlation values, indicating structural similarity across tested variants.

See also **Figure S6**.

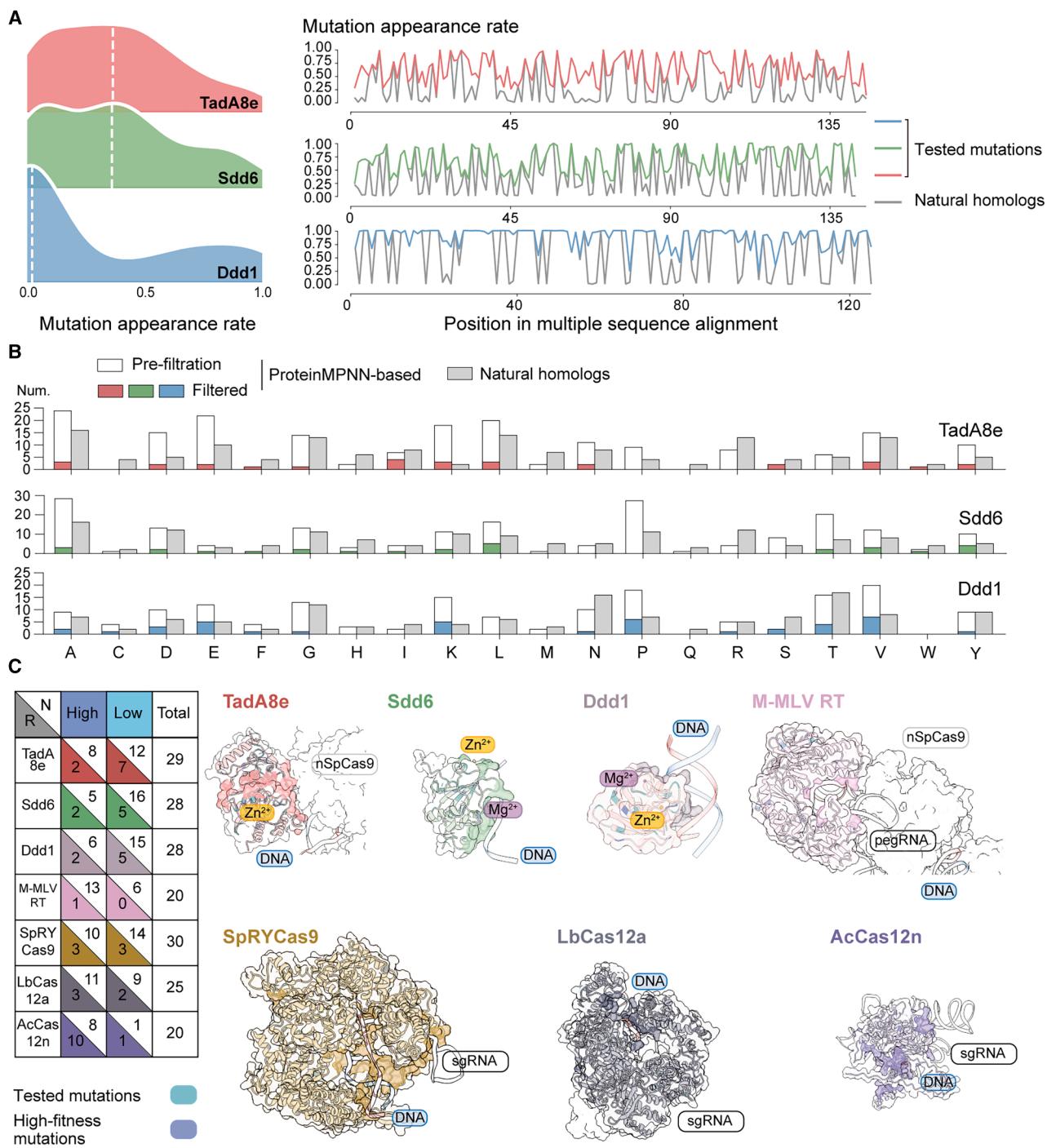


Figure 7. Analysis of inverse folding mutation generation patterns

(A) Distribution of ProteinMPNN-generated substitutions. Ridge plot illustrates the overall distribution of the incidence of ProteinMPNN-based mutations, with the dashed line indicating the median appearance rate. Line graph represents the prevalence of mutated versus wild-type amino acids at various positions.

(B) Amino acid substitution patterns across TadA8e, Sdd6, and Ddd1. Bar plots show the distributions of amino acid substitutions. Colored segments represent mutations selected by the AiCE strategy.

(C) Structural distribution of tested single mutations. Dark-colored regions in each structure represent hotspots identified by conventional rational design, while light blue identifies experimentally tested mutation sites, and dark blue highlights high-fitness (HF) mutations. The table summarizes the number of HF mutations (high), other mutations (low), and total experimental tested mutations (total) within rational design hotspot regions (R) and non-hotspot regions (N) across the seven proteins. The complex structures of TadA8e (PDB ID 6VPC) were resolved by cryo-EM, while others were predicted using AlphaFold3.

and seemingly counterintuitive forms, both in position and type. This effect was seen for both nominated mutations (regardless of structural constraints) and experimentally validated HF mutations (Figures 7B and 7C). It may explain AiCE's compatibility with other methods and its potential for structure-guided rational design (Figure 7C).

Due to its effectiveness, AiCE enabled enhancements to various genome-editing tools, particularly base editors, improving precision genome editing despite the substantial structural diversity of target proteins (Figure 6F). AiCE was also effective in enhancing specificity, editing windows, and environmental fitness across diverse evolutionary contexts. We have demonstrated the utility of this cross-domain protein evolution approach by benchmarking with extensive DMS libraries, modifying homotetrameric recombinases as part of a parallel project on manipulating large chromosomal fragments, and creating innovative cytosine and adenine deaminases, as well as a more effective enDdd1. We anticipate that the methods developed here will help accelerate the evolution of valuable proteins, particularly highly complex proteins and multi-subunit complexes.

Limitations of the study

Although the AiCE scheme enables rapid and efficient protein evolution, it has several limitations. First, it relies on sampling from inverse folding model outputs, which are inherently constrained by the accuracy of protein structure predictions. Some HF mutations identified in Sdd6 and AcCas12n were located in low-confidence regions of the predicted structures, highlighting this dependency. Future improvements may involve integrating multimodal reinforcement learning, experimentally validated active mutants, functional annotations, and additional biochemical constraints to enhance robustness. Second, AiCE prioritizes the most frequently occurring mutations without considering their retention under complex evolutionary pressures such as balancing selection, negative selection, and founder effects. Moderately frequent or rare mutations may harbor novel functional potential but could be overlooked by the current approach. Integrating additional principles of natural selection could improve AiCE's ability to identify rare but functionally important mutations. Third, AiCE considers the present distribution of amino acids, but incorporating amino acid substitution rates could improve its ability to model ancestral protein sequences and distribution patterns.

Although AiCE_{multi} was designed to mitigate negative epistasis by incorporating EC scores, its performance was not consistently superior to AiCE_{single}, likely due to limited combinatorial coverage and the additive model's inability to capture complex epistasis. These observations highlight the importance of developing more sophisticated approaches that can explicitly model higher-order interactions. Nonetheless, AiCE_{multi} markedly improved HF variant identification over random methods and enabled the development of three best-in-class base editors, demonstrating its practical utility and future potential.

From an application perspective, although AiCE has successfully engineered several genome-editing-associated proteins, the precise relationship between mutations, structure, and func-

tion remains incompletely understood. Future efforts involving molecular dynamics simulations or cryo-EM structural analyses may yield deeper mechanistic insights and help refine the AiCE framework. Finally, despite its broad applicability, occasional suboptimal performance likely reflects intrinsic differences in protein foldability, functional constraints, and engineering goals. Incorporating limited prior knowledge or iterative small-sample refinement may help overcome these challenges and further expand the utility of AiCE.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to the lead contact, Caixia Gao (cxgao@genetics.ac.cn).

Materials availability

All unique and stable reagents generated in this study are available from the lead contact with a completed materials transfer agreement.

Data and code availability

- All deep amplicon sequencing data have been deposited in the National Center for Biotechnology Information BioProject under accession code PRJNA1155632 and are publicly available as of the date of publication.
- All original code has been deposited at <https://github.com/ScorpioLea/AiCE> and is publicly available as of the date of publication.
- All additional data supporting the findings of this study are available from the corresponding author upon request.

ACKNOWLEDGMENTS

We thank Z. Zheng (Jingchu University of Technology) for his valuable advice on linkage disequilibrium analysis. We also extend our gratitude to X. Wang (Institute of Genetics and Developmental Biology, Chinese Academy of Sciences), P. Lu (Westlake University), K.T. Zhao (Qi-BioDesign), M. Jin (Northwest A&F University), and Y. Liu (National University of Singapore) for their helpful suggestions in revising the manuscript. This work was supported by the Agriculture Science and Technology Major Project, the National Natural Science Foundation of China (32388201), the National Key Research and Development Program (2022YFF1002802), the Beijing Municipal Science & Technology Commission (Z241100009024035), and the New Cornerstone Science Foundation.

AUTHOR CONTRIBUTIONS

Conceptualization: H.F., Y. Li, and C.G.; methodology: Y. Li and H.F.; formal analysis: Y. Li; investigation: H.F., Y. Liu, J.W., and A.C.; visualization: H.F., Y. Li, A.C., and Y. Liu; writing: C.G., Y. Li, and H.F.; supervision: C.G.

DECLARATION OF INTERESTS

The authors have submitted one patent application based on the results reported in this article. C.G. is a member of the *Cell* advisory board.

STAR METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODELS AND STUDY PARTICIPANT DETAILS
 - *E.coli* transfection
 - Mammalian cell lines and culture conditions
- METHOD DETAILS

- Generating inverse-folding sequences using generalized protein inverse-folding models
 - Use of AiCE_{single} for designing single substitution mutations
 - AiCE_{multi} for designing multi-mutations
 - Linkage disequilibrium analysis
 - Evolutionary coupling analysis
 - Multi-mutation nomination based on LD and EC score
 - Mutation selection based on natural homologs
 - Protein DMS libraries spanning multiple families as a benchmark for assessing single mutations
 - Mutation data from antibodies and other proteins as benchmarks for assessing multi-mutations
 - Selection of mutations for experimental testing
 - Identification of rational design hotspot regions for rational protein design
 - Protein structure prediction, comparison, and visualization
 - Plasmid construction
 - Mammalian cell transfection
 - DNA extraction
 - Amplicon deep sequencing and data analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS
 - Quantification
 - Statistical analysis
 - ADDITIONAL RESOURCES

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.cell.2025.06.014>.

Received: September 3, 2024

Revised: March 14, 2025

Accepted: June 10, 2025

Published: July 7, 2025

REFERENCES

1. Romero, P.A., and Arnold, F.H. (2009). Exploring protein fitness landscapes by directed evolution. *Nat. Rev. Mol. Cell Biol.* **10**, 866–876. <https://doi.org/10.1038/nrm2805>.
2. Kazlauskas, R.J., and Bornscheuer, U.T. (2009). Finding better protein engineering strategies. *Nat. Chem. Biol.* **5**, 526–529. <https://doi.org/10.1038/nchembio0809-526>.
3. Shen, X., Song, S., Li, C., and Zhang, J. (2022). Synonymous mutations in representative yeast genes are mostly strongly non-neutral. *Nature* **606**, 725–731. <https://doi.org/10.1038/s41586-022-04823-w>.
4. Hino, T., Omura, S.N., Nakagawa, R., Togashi, T., Takeda, S.N., Hiramoto, T., Tasaka, S., Hirano, H., Tokuyama, T., Uosaki, H., et al. (2023). An As-Cas12f-based compact genome-editing tool derived by deep mutational scanning and structural analysis. *Cell* **186**, 4920–4935.e23. <https://doi.org/10.1016/j.cell.2023.08.031>.
5. Bloom, J.D., Silberg, J.J., Wilke, C.O., Drummond, D.A., Adam, C., and Arnold, F.H. (2005). Thermodynamic prediction of protein neutrality. *Proc. Natl. Acad. Sci. USA* **102**, 606–611. <https://doi.org/10.1073/pnas.0406744102>.
6. Guo, H.H., Choe, J., and Loeb, L.A. (2004). Protein tolerance to random amino acid change. *Proc. Natl. Acad. Sci. USA* **101**, 9205–9210. <https://doi.org/10.1073/pnas.0403255101>.
7. 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. <https://doi.org/10.1038/s41587-022-01532-7>.
8. Jin, S., Fei, H., Zhu, Z., Luo, Y., Liu, J., Gao, S., Zhang, F., Chen, Y.-H., Wang, Y., and Gao, C. (2020). Rationally designed APOBEC3B cytosine base editors with improved specificitye726. *Mol. Cell* **79**, 728–740.e6. <https://doi.org/10.1016/j.molcel.2020.07.005>.
9. Poelwijk, F.J., Kiviet, D.J., Weinreich, D.M., and Tans, S.J. (2007). Empirical fitness landscapes reveal accessible evolutionary paths. *Nature* **445**, 383–386. <https://doi.org/10.1038/nature05451>.
10. Tabebordbar, M., Lagerborg, K.A., Stanton, A., King, E.M., Ye, S., Tellez, L., Krunnfusz, A., Tavakoli, S., Widrick, J.J., Messemer, K.A., et al. (2021). Directed evolution of a family of AAV capsid variants enabling potent muscle-directed gene delivery across species. *Cell* **184**, 4919–4938.e22. <https://doi.org/10.1016/j.cell.2021.08.028>.
11. Arnold, F.H. (1996). Directed evolution: creating biocatalysts for the future. *Chem. Eng. Sci.* **51**, 5091–5102. [https://doi.org/10.1016/S0009-2509\(96\)00288-6](https://doi.org/10.1016/S0009-2509(96)00288-6).
12. 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 A-T to G-C in genomic DNA without DNA cleavage. *Nature* **551**, 464–471. <https://doi.org/10.1038/nature24644>.
13. Lovelock, S.L., Crawshaw, R., Basler, S., Levy, C., Baker, D., Hilvert, D., and Green, A.P. (2022). The road to fully programmable protein catalysis. *Nature* **606**, 49–58. <https://doi.org/10.1038/s41586-022-04456-z>.
14. He, Y., Zhou, X., Chang, C., Chen, G., Liu, W., Li, G., Fan, X., Sun, M., Miao, C., Huang, Q., et al. (2024). Protein language models-assisted optimization of a uracil-N-glycosylase variant enables programmable T-to-G and T-to-C base editing. *Mol. Cell* **84**, 1257–1270.e6. <https://doi.org/10.1016/j.molcel.2024.01.021>.
15. Jiang, K., Yan, Z., Di Bernardo, M., Sgrizzi, S.R., Villiger, L., Kayabolen, A., Kim, B., Carscadden, J.K., Hiraizumi, M., and Nishimasu, H. (2024) Rapid In Silico Directed Evolution by a Protein Language Model with EVOLVEpro. *Science*, eadr6006.
16. Yang, J., Lal, R.G., Bowden, J.C., Astudillo, R., Hameedi, M.A., Kaur, S., Hill, M., Yue, Y., and Arnold, F.H. (2025). Active learning-assisted directed evolution. *Nat. Commun.* **16**, 714. <https://doi.org/10.1038/s41467-025-55987-8>.
17. Cheng, P., Mao, C., Tang, J., Yang, S., Cheng, Y., Wang, W., Gu, Q., Han, W., Chen, H., Li, S., et al. (2024). Zero-shot prediction of mutation effects with multimodal deep representation learning guides protein engineering. *Cell Res.* **34**, 630–647. <https://doi.org/10.1038/s41422-024-00989-2>.
18. Su, J., Han, C., Zhou, Y., Shan, J., Zhou, X., and Yuan, F. (2023). Saprot: Protein language modeling with structure-aware vocabulary. *bioRxiv*, 2001.560349.
19. Hayes, T., Rao, R., Akin, H., Sofroniew, N.J., Oktay, D., Lin, Z., Verkuil, R., Tran, V.Q., Deaton, J., and Wiggett, M. (2025). Simulating 500 Million Years of Evolution with a Language Model. *Science* **387**, 850–858.
20. Hsu, C., Verkuil, R., Liu, J., Lin, Z., Hie, B., Sercu, T., Lerer, A., and Rives, A. (2022). Learning inverse folding from millions of predicted structures. In *Proceedings of the 39th International Conference on Machine Learning*, **162**, pp. 8946–8970.
21. Dauparas, J., Anishchenko, I., Bennett, N., Bai, H., Ragotte, R.J., Milles, L.F., Wicky, B.I.M., Courbet, A., de Haas, R.J., Bethel, N., et al. (2022). Robust deep learning-based protein sequence design using ProteinMPNN. *Science* **378**, 49–56. <https://doi.org/10.1126/science.add2187>.
22. Shanker, V.R., Bruun, T.U.J., Hie, B.L., and Kim, P.S. (2024). Unsupervised evolution of protein and antibody complexes with a structure-informed language model. *Science* **385**, 46–53. <https://doi.org/10.1126/science.adk8946>.
23. Nishimasu, H., Ran, F.A., Hsu, P.D., Konermann, S., Shehata, S.I., Dohmae, N., Ishitani, R., Zhang, F., and Nureki, O. (2014). Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**, 935–949. <https://doi.org/10.1016/j.cell.2014.02.001>.

24. 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. <https://doi.org/10.1126/science.abb1390>.
25. Bozüyüük, K.A.J., Präve, L., Kegler, C., Schenk, L., Kaiser, S., Schelhas, C., Shi, Y.-N., Kuttelochner, W., Schreiber, M., Kandler, J., et al. (2024). Evolution-inspired engineering of nonribosomal peptide synthetases. *Science* 383, eadg4320. <https://doi.org/10.1126/science.adg4320>.
26. Lockless, S.W., and Ranganathan, R. (1999). Evolutionarily conserved pathways of energetic connectivity in protein families. *Science* 286, 295–299. <https://doi.org/10.1126/science.286.5438.295>.
27. Mabesoone, M.F.J., Leopold-Messer, S., Minas, H.A., Chepkirui, C., Chawengrum, P., Reiter, S., Meoded, R.A., Wolf, S., Genz, F., Magnus, N., et al. (2024). Evolution-guided engineering of trans-acyltransferase polyketide synthases. *Science* 383, 1312–1317. <https://doi.org/10.1126/science.adj7621>.
28. De Juan, D., Pazos, F., and Valencia, A. (2013). Emerging methods in protein co-evolution. *Nat. Rev. Genet.* 14, 249–261. <https://doi.org/10.1038/nrg3414>.
29. Dauparas, J., Lee, G.R., Pecoraro, R., An, L., Anishchenko, I., Glasscock, C., and Baker, D. (2025). Atomic context-conditioned protein sequence design using LigandMPNN. *Nat. Methods* 22, 717–723. <https://doi.org/10.1038/s41592-025-02626-1>.
30. Cummings, H.S., and Hershey, J.W. (1994). Translation initiation factor IF1 is essential for cell viability in *Escherichia coli*. *J. Bacteriol.* 176, 198–205. <https://doi.org/10.1128/jb.176.1.198-205.1994>.
31. Verger, A., Perdomo, J., and Crossley, M. (2003). Modification with SUMO: a role in transcriptional regulation. *EMBO Rep.* 4, 137–142. <https://doi.org/10.1038/sj.emboj.embor738>.
32. Richter, K., and Buchner, J. (2001). Hsp90: chaperoning signal transduction. *J. Cell. Physiol.* 188, 281–290. <https://doi.org/10.1002/jcp.1131>.
33. Zilfou, J.T., and Lowe, S.W. (2009). Tumor suppressive functions of p53. *Cold Spring Harbor Perspect. Biol.* 1, a001883. <https://doi.org/10.1101/cshperspect.a001883>.
34. Doudna, J.A., and Charpentier, E. (2014). Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* 346, 1258096. <https://doi.org/10.1126/science.1258096>.
35. Karplus, M., and Kuriyan, J. (2005). Molecular dynamics and protein function. *Proc. Natl. Acad. Sci. USA* 102, 6679–6685. <https://doi.org/10.1073/pnas.0408930102>.
36. Starr, T.N., Greaney, A.J., Hilton, S.K., Ellis, D., Crawford, K.H., Dingens, A.S., Navarro, M.J., Bowen, J.E., Tortorici, M.A., and Walls, A.C. (2020). Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. *Cell* 182, 1295–1310.
37. Rockah-Shmuel, L., Tóth-Petróczy, Á., and Tawfik, D.S. (2015). Systematic mapping of protein mutational space by prolonged drift reveals the deleterious effects of seemingly neutral mutations. *PLOS Comp. Biol.* 11, e1004421. <https://doi.org/10.1371/journal.pcbi.1004421>.
38. Stiffler, M.A., Hekstra, D.R., and Ranganathan, R. (2015). Evolvability as a function of purifying selection in TEM-1 β -lactamase. *Cell* 160, 882–892. <https://doi.org/10.1016/j.cell.2015.01.035>.
39. Wu, T., Liu, C., Zou, S., Lyu, R., Yang, B., Yan, H., Zhao, M., and Tang, W. (2023). An engineered hypercompact CRISPR-Cas12f system with boosted gene-editing activity. *Nat. Chem. Biol.* 19, 1384–1393. <https://doi.org/10.1038/s41589-023-01380-9>.
40. Brennan, L., Andreev, A., Cohen, O., Pantel, S., Kamburov, A., Cacchiarelli, D., Persky, N.S., Zhu, C., Bagul, M., Goetz, E.M., et al. (2016). Phenotypic characterization of a comprehensive set of MAPK1/ERK2 missense mutants. *Cell Rep.* 17, 1171–1183. <https://doi.org/10.1016/j.celrep.2016.09.061>.
41. Starita, L.M., Young, D.L., Islam, M., Kitzman, J.O., Gullingsrud, J., Hause, R.J., Fowler, D.M., Parvin, J.D., Shendure, J., and Fields, S. (2015). Massively parallel functional analysis of BRCA1 RING domain variants. *Genetics* 200, 413–422. <https://doi.org/10.1534/genetics.115.175802>.
42. Lyons, D.M., Zou, Z., Xu, H., and Zhang, J. (2020). Idiosyncratic epistasis creates universals in mutational effects and evolutionary trajectories. *Nat. Ecol. Evol.* 4, 1685–1693. <https://doi.org/10.1038/s41559-020-01286-y>.
43. Phillips, P.C. (2008). Epistasis—the essential role of gene interactions in the structure and evolution of genetic systems. *Nat. Rev. Genet.* 9, 855–867. <https://doi.org/10.1038/nrg2452>.
44. Phillips, A.M., Lawrence, K.R., Moulana, A., Dupic, T., Chang, J., Johnson, M.S., Cvijovic, I., Mora, T., Walczak, A.M., and Desai, M.M. (2021). Binding affinity landscapes constrain the evolution of broadly neutralizing anti-influenza antibodies. *eLife* 10, e71393. <https://doi.org/10.7554/elife.71393>.
45. Pokusaeva, V.O., Usmanova, D.R., Putintseva, E.V., Espinar, L., Sarkisyan, K.S., Mishin, A.S., Bogatyreva, N.S., Ivankov, D.N., Akopyan, A. V., Avvakumov, S.Y., et al. (2019). An experimental assay of the interactions of amino acids from orthologous sequences shaping a complex fitness landscape. *PLOS Genet.* 15, e1008079. <https://doi.org/10.1371/journal.pgen.1008079>.
46. Gonzalez Somermeyer, L., Fleiss, A., Mishin, A.S., Bozhanova, N.G., Igolkina, A.A., Meiler, J., Alaball Pujol, M.-E., Putintseva, E.V., Sarkisyan, K.S., and Kondrashov, F.A. (2022). Heterogeneity of the GFP fitness landscape and data-driven protein design. *eLife* 11, e75842. <https://doi.org/10.7554/elife.75842>.
47. Pacesa, M., Pelea, O., and Jinek, M. (2024). Past, present, and future of CRISPR genome editing technologies. *Cell* 187, 1076–1100. <https://doi.org/10.1016/j.cell.2024.01.042>.
48. Li, B., Sun, C., Li, J., and Gao, C. (2024). Targeted genome-modification tools and their advanced applications in crop breeding. *Nat. Rev. Genet.* 25, 603–622. <https://doi.org/10.1038/s41576-024-00720-2>.
49. 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. <https://doi.org/10.1038/s41587-020-0453-z>.
50. Rees, H.A., and Liu, D.R. (2018). Base editing: precision chemistry on the genome and transcriptome of living cells. *Nat. Rev. Genet.* 19, 770–788. <https://doi.org/10.1038/s41576-018-0059-1>.
51. Caspi, A., Sugden, K., Moffitt, T.E., Taylor, A., Craig, I.W., Harrington, H., McClay, J., Mill, J., Martin, J., Braithwaite, A., et al. (2003). Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. *Science* 301, 386–389. <https://doi.org/10.1126/science.1083968>.
52. Chen, L., Zhang, S., Xue, N., Hong, M., Zhang, X., Zhang, D., Yang, J., Bai, S., Huang, Y., Meng, H., et al. (2023). Engineering a precise adenine base editor with minimal bystander editing. *Nat. Chem. Biol.* 19, 101–110. <https://doi.org/10.1038/s41589-022-01163-8>.
53. 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. <https://doi.org/10.1038/nature17946>.
54. Jin, S., Zong, Y., Gao, Q., Zhu, Z., Wang, Y., Qin, P., Liang, C., Wang, D., Qiu, J.-L., Zhang, F., et al. (2019). Cytosine, but not adenine, base editors induce genome-wide off-target mutations in rice. *Science* 364, 292–295. <https://doi.org/10.1126/science.aaw7166>.
55. Zuo, E., Sun, Y., Wei, W., Yuan, T., Ying, W., Sun, H., Yuan, L., Steinmetz, L.M., Li, Y., and Yang, H. (2019). Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos. *Science* 364, 289–292. <https://doi.org/10.1126/science.aav9973>.
56. Doman, J.L., Raguram, A., Newby, G.A., and Liu, D.R. (2020). Evaluation and minimization of Cas9-independent off-target DNA editing by cytosine base editors. *Nat. Biotechnol.* 38, 620–628. <https://doi.org/10.1038/s41587-020-0414-6>.

57. Huang, J., Lin, Q., Fei, H., He, Z., Xu, H., Li, Y., Qu, K., Han, P., Gao, Q., Li, B., et al. (2023). Discovery of deaminase functions by structure-based protein clustering. *Cell* 186, 3182–3195.e14. <https://doi.org/10.1016/j.cell.2023.05.041>.
58. Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A.J., and Bambrick, J. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold. *Nature* 3, 1–3.
59. Mok, B.Y., de Moraes, M.H., Zeng, J., Bosch, D.E., Kotrys, A.V., Raguram, A., Hsu, F., Raday, M.C., Peterson, S.B., Mootha, V.K., et al. (2020). A bacterial cytidine deaminase toxin enables CRISPR-free mitochondrial base editing. *Nature* 583, 631–637. <https://doi.org/10.1038/s41586-020-2477-4>.
60. Shamsnjafabadi, H., MacLaren, R.E., and Cehajic-Kapetanovic, J. (2023). Current and future landscape in genetic therapies for Leber hereditary optic neuropathy. *Cells* 12, 2013. <https://doi.org/10.3390/cells12152013>.
61. Chen, C., and Guan, M.-X. (2023). Induced pluripotent stem cells: ex vivo models for human diseases due to mitochondrial DNA mutations. *J. Biomed. Sci.* 30, 82. <https://doi.org/10.1186/s12929-023-00967-7>.
62. Mok, B.Y., Kotrys, A.V., Raguram, A., Huang, T.P., Mootha, V.K., and Liu, D.R. (2022). CRISPR-free base editors with enhanced activity and expanded targeting scope in mitochondrial and nuclear DNA. *Nat. Biotechnol.* 40, 1378–1387. <https://doi.org/10.1038/s41587-022-01256-8>.
63. Chen, W., Ma, J., Wu, Z., Wang, Z., Zhang, H., Fu, W., Pan, D., Shi, J., and Ji, Q. (2023). Cas12n nucleases, early evolutionary intermediates of type V CRISPR, comprise a distinct family of miniature genome editors. *Mol. Cell* 83, 2768–2780.e6. <https://doi.org/10.1016/j.molcel.2023.06.014>.
64. Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., Slaymaker, I.M., Makarova, K.S., Essletzbichler, P., Volz, S.E., Joung, J., Van Der Oost, J., and Regev, A. (2015). Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell* 163, 759–771.
65. Walton, R.T., Christie, K.A., Whittaker, M.N., and Kleinstiver, B.P. (2020). Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science* 368, 290–296. <https://doi.org/10.1126/science.aba8853>.
66. Anzalone, A.V., Randolph, P.B., Davis, J.R., Sousa, A.A., Koblan, L.W., Levy, J.M., Chen, P.J., Wilson, C., Newby, G.A., Raguram, A., et al. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* 576, 149–157. <https://doi.org/10.1038/s41586-019-1711-4>.
67. Sun, C., Lei, Y., Li, B., Gao, Q., Li, Y., Cao, W., Yang, C., Li, H., Wang, Z., Li, Y., et al. (2024). Precise integration of large DNA sequences in plant genomes using PrimeRoot editors. *Nat. Biotechnol.* 42, 316–327. <https://doi.org/10.1038/s41587-023-01769-w>.
68. Tomatis, P.E., Fabiane, S.M., Simona, F., Carloni, P., Sutton, B.J., and Vila, A.J. (2008). Adaptive protein evolution grants organismal fitness by improving catalysis and flexibility. *Proc. Natl. Acad. Sci. USA* 105, 20605–20610. <https://doi.org/10.1073/pnas.0807989106>.
69. Hochmair, I. (2013). The importance of being flexible. *Nat. Med.* 19, 1240–1244. <https://doi.org/10.1038/nm.3341>.
70. Fields, P.A. (2001). Review: Protein function at thermal extremes: balancing stability and flexibility. *Comp. Biochem. Physiol. A Mol. Integr. Physiol.* 129, 417–431. [https://doi.org/10.1016/s1095-6433\(00\)00359-7](https://doi.org/10.1016/s1095-6433(00)00359-7).
71. Teilum, K., Olsen, J.G., and Kragelund, B.B. (2011). Protein stability, flexibility and function. *Biochim. Biophys. Acta* 1814, 969–976. <https://doi.org/10.1016/j.bbapap.2010.11.005>.
72. Pacesa, M., Nickel, L., Schellhaas, C., Schmidt, J., Pyatova, E., Kissling, L., Barendse, P., Choudhury, J., Kapoor, S., Alcaraz-Serna, A., et al. (2024). BindCraft: one-shot design of functional protein binders. Preprint at bioRxiv, 2024.09.30.615802. <https://doi.org/10.1101/2024.09.30.615802>.
73. Kabadi, A.M., Ousterout, D.G., Hilton, I.B., and Gersbach, C.A. (2014). Multiplex CRISPR/Cas9-based genome engineering from a single lentiviral vector. *Nucleic Acids Res.* 42, e147. <https://doi.org/10.1093/nar/gku749>.
74. Koblan, L.W., Doman, J.L., Wilson, C., Levy, J.M., Tay, T., Newby, G.A., Maianti, J.P., Raguram, A., and Liu, D.R. (2018). Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. *Nat. Biotechnol.* 36, 843–846. <https://doi.org/10.1038/nbt.4172>.
75. Liang, R., He, Z., Zhao, K.T., Zhu, H., Hu, J., Liu, G., Gao, Q., Liu, M., Zhang, R., Qiu, J.-L., et al. (2024). Prime editing using CRISPR-Cas12a and circular RNAs in human cells. *Nat. Biotechnol.* 42, 1867–1875. <https://doi.org/10.1038/s41587-023-02095-x>.
76. 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. <https://doi.org/10.1038/s41586-021-03819-2>.
77. Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28, 3150–3152. <https://doi.org/10.1093/bioinformatics/bts565>.
78. Magoć, T., and Salzberg, S.L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27, 2957–2963. <https://doi.org/10.1093/bioinformatics/btr507>.
79. Steinegger, M., Meier, M., Mirdita, M., Vöhringer, H., Haunsberger, S.J., and Söding, J. (2019). HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinformatics* 20, 473. <https://doi.org/10.1186/s12859-019-3019-7>.
80. Johnson, L.S., Eddy, S.R., and Portugaly, E. (2010). Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC Bioinformatics* 11, 431. <https://doi.org/10.1186/1471-2105-11-431>.
81. Edgar, R.C. (2022). Muscle5: High-accuracy alignment ensembles enable unbiased assessments of sequence homology and phylogeny. *Nat. Commun.* 13, 6968. <https://doi.org/10.1038/s41467-022-34630-w>.
82. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., De Bakker, P.I.W., Daly, M.J., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. <https://doi.org/10.1086/519795>.
83. Rivoire, O., Reynolds, K.A., and Ranganathan, R. (2016). Evolution-based functional decomposition of proteins. *PLOS Comp. Biol.* 12, e1004817. <https://doi.org/10.1371/journal.pcbi.1004817>.
84. Wang, J., Lisanza, S., Juergens, D., Tischer, D., Watson, J.L., Castro, K. M., Ragotte, R., Saragovi, A., Miles, L.F., Baek, M., et al. (2022). Scaffolding protein functional sites using deep learning. *Science* 377, 387–394. <https://doi.org/10.1126/science.abn2100>.
85. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D. M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comp. Chem.* 25, 1605–1612. <https://doi.org/10.1002/jcc.20084>.
86. Kabsch, W., and Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22, 2577–2637. <https://doi.org/10.1002/bip.360221211>.
87. Dreyfus, C., Laursen, N.S., Kwaks, T., Zuijdgeest, D., Khayat, R., Ekiert, D. C., Lee, J.H., Metlagel, Z., Bujny, M.V., Jongeneelen, M., et al. (2012). Highly conserved protective epitopes on influenza B viruses. *Science* 337, 1343–1348. <https://doi.org/10.1126/science.1222908>.
88. Li, C., Zhang, R., Meng, X., Chen, S., Zong, Y., Lu, C., Qiu, J.-L., Chen, Y.-H., Li, J., and Gao, C. (2020). Targeted, random mutagenesis of plant genes with dual cytosine and adenine base editors. *Nat. Biotechnol.* 38, 875–882. <https://doi.org/10.1038/s41587-019-0393-7>.
89. Vaisvila, R., Johnson, S.R., Yan, B., Dai, N., Bourkia, B.M., Chen, M., Corrêa, I.R., Yigit, E., and Sun, Z. (2024). Discovery of cytosine deaminases enables base-resolution methylome mapping using a single enzyme. *Mol. Cell* 84, 854–866.e7. <https://doi.org/10.1016/j.molcel.2024.01.027>.

90. Wang, L., Li, L., Ma, Y., Hu, H., Li, Q., Yang, Y., Liu, W., Yin, S., Li, W., Fu, B., et al. (2020). Reactivation of γ -globin expression through Cas9 or base editor to treat β -hemoglobinopathies. *Cell Res.* 30, 276–278. <https://doi.org/10.1038/s41422-019-0267-z>.
91. 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. <https://doi.org/10.1038/s41587-020-0561-9>.
92. Huang, T.P., Newby, G.A., and Liu, D.R. (2021). Precision genome editing using cytosine and adenine base editors in mammalian cells. *Nat. Protoc.* 16, 1089–1128. <https://doi.org/10.1038/s41596-020-00450-9>.
93. Arbab, M., Shen, M.W., Mok, B., Wilson, C., Matuszek, Ž., Cassa, C.A., and Liu, D.R. (2020). Determinants of base editing outcomes from target library analysis and machine learning. *Cell* 182, 463–480.e30. <https://doi.org/10.1016/j.cell.2020.05.037>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
FastT1 Competent Cells	Vazyme	Cat#C505-02
Chemicals, peptides, and recombinant proteins		
2 × Phanta Max Master Mix (Dye Plus)	Vazyme	Cat#P525-01
Ampicillin sodium	Macklin	Cat#A6265
DMEM	Gibco	Cat#10569044
Fetal Bovine Serum	Gibco	Cat#10091148
Kanamycin Sulfate	Solarbio	Cat#AK177
Lipofectamine 2000 transfection reagent	Invitrogen	Cat#11668019
Opti-MEM	Gibco	Cat#31985-070
PBS pH 7.4 basic	Gibco	Cat#C10010500BT
Trypan Blue stain 0.4%	Invitrogen	Cat#T10282
TrypLE Express	Gibco	Cat#12605-010
Tryptone	OXOID	Cat#LP0042B
Yeast extract	OXOID	Cat#LP0021B
Critical commercial assays		
GeneJET Gel Extraction Kit	Thermo Scientific	Cat#K0692
Mycoplasma Detection Kit	Transgen	Cat#FM311-01
Plasmid Miniprep System	Promega	Cat#A1222
Plasmid Plus Midi Kit	QIAGEN	Cat#12945
Triumfi Mouse Tissue Direct PCR Kit	Genesand	Cat#SD312
Uniclon One Step Seamless Cloning Kit	Genesand	Cat#SC612
Deposited data		
Deep amplicon sequencing data	This paper	Accession ID PRJNA1155632
Experimental models: Cell lines		
HEK293T	ATCC	Cat#CRL-3216
HeLa	ATCC	Cat#CRM-CCL-2
K562	ATCC	Cat#CCL-243
U2OS	ATCC	Cat#HTB-96
Oligonucleotides		
Primers used in this paper, see Table S3	This paper	N/A
Recombinant DNA		
p2T-CMV-miniSdd6-BE4max-BlastR	Huang et al. ⁵⁷	Addgene #204850
p2T-CMV-rAPOBEC1-BE4max-BlastR	Huang et al. ⁵⁷	N/A
phU6 vector	Kabadi et al. ⁷³	Addgene #53188
pCMV-nSaCas9	Huang et al. ⁵⁷	N/A
p2T-CMV-miniSdd6 mutations-BE4max-BlastR	This paper	N/A
p2T-CMV-TadA8e mutations-BE4max-BlastR	This paper	N/A
pCMV-TALE-L-JAK2-Ddd9-N	Huang et al. ⁵⁷	Addgene#204853
pCMV-TALE-R-JAK2-Ddd9-C	Huang et al. ⁵⁷	Addgene #204854
pCMV-TALE-L-Ddd1 mutations-N	This paper	N/A
pCMV-TALE-R-Ddd1 mutations-C	This paper	N/A
pCMV_BE4max	Koblan et al. ⁷⁴	Addgene #112093

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
LbCas12a-D156R	Liang et al. ⁷⁵	Addgene #213744
pCMV-PE2	Anzalone et al. ⁶⁶	Addgene #132775
Software and algorithms		
Adobe Illustrator	Adobe	https://www.adobe.com/
AiCE	This paper	https://github.com/ScorpioLea/AiCE
AlphaFold2	Jumper et al. ⁷⁶	https://deepmind.google/technologies/alphafold/
AlphaFold3	Abramson et al. ⁵⁸	https://alphafoldserver.com/
CD-HIT	Fu et al. ⁷⁷	https://sites.google.com/view/cd-hit/home
ESM3-open	Hayes et al. ¹⁹	https://github.com/evolutionaryscale/esm?tab=readme-ov-file#esm-3
ESM-IF1	Hsu et al. ²⁰	https://github.com/facebookresearch/esm/tree/main/examples/inverse_folding
FLASH	Magoč et al. ⁷⁸	https://www.ccb.umd.edu/software/flash
GraphPad Prism 10	GraphPad Software	https://www.graphpad.com/scientific-software/prism/
HH-suite3	Steinegger et al. ⁷⁹	https://github.com/soedinglab/hh-suite
JackHMMER	Johnson et al. ⁸⁰	http://hmmer.org/
LigandMPNN	Dauparas et al. ²⁹	https://github.com/dauparas/LigandMPNN
MUSCLE	Edgar et al. ⁸¹	https://www.drive5.com/muscle/
PLINK	Purcell et al. ⁸²	https://www.cog-genomics.org/plink/
ProteinMPNN	Dauparas et al. ²¹	https://github.com/dauparas/ProteinMPNN
pySCA	Rivoire et al. ⁸³	https://github.com/reynoldsk/pySCA
RFdesign	Wang et al. ⁸⁴	https://github.com/RosettaCommons/RFDesign
SaProt	Su et al. ¹⁸	https://github.com/westlake-repl/SaProt
UCSF Chimera	Pettersen et al. ⁸⁵	https://www.cgl.ucsf.edu/chimera/

EXPERIMENTAL MODELS AND STUDY PARTICIPANT DETAILS***E.coli* transfection**

FastT1 *E.coli* competent cells were used to amplify plasmid DNA. Transfected *E. coli* cells were cultured overnight at 37°C in Lysogeny Broth (LB) medium supplemented with 100 mg/mL ampicillin or 50 mg/mL kanamycin, as appropriate.

Mammalian cell lines and culture conditions

Human HEK293T cells (ATCC, CRL-3216) and HeLa cells (ATCC, CRM-CCL-2) were maintained in Dulbecco's Modified Eagle's Medium (DMEM, Gibco), K562 cells (ATCC, CCL-243) in Iscove's Modified Dulbecco's Medium (IMDM, Gibco), and U2OS cells (ATCC, HTB-96) in McCoy's 5A Medium (Gibco). All media were supplemented with 10% fetal bovine serum (FBS, Gibco) and 1% penicillin-streptomycin (Gibco). All cell lines were cultured at 37°C in a humidified 5% CO₂ incubator using 75 cm² cell culture flasks (NEST). They were tested for mycoplasma using a Mycoplasma Detection Kit (TransGen Biotech, CN).

METHOD DETAILS**Generating inverse-folding sequences using generalized protein inverse-folding models**

Protein inverse folding models are designed to generate sequences that are compatible with given backbone structure. In this study, we evaluated the performance of AiCE using three commonly used generalized inverse folding models: ESM-IF1, LigandMPNN, and ProteinMPNN.

ESM-IF1 uses a transformers architecture to represent the inverse folding problem as a conditional distribution problem on a protein sequence X of length n , given a structure Y , namely:

$$p(Y|X) = \prod_{i=1}^n p(Y_i|Y_{i-1}, \dots, Y_1; X)$$

When provided with the 3D coordinates of the backbone atoms of the main chain (N, C α , C), the amino acid sequence of Y can be reconstructed by autoregressive modeling. An efficient sequence design is achieved by sampling or maximizing conditional probabilities. The inverse folding model used was esm_if1_gvp4_t16_142M_UR50, which comprises 124 million parameters.

ProteinMPNN is based on a message-passing neural network that autoregressively predicts protein sequences using protein backbone features, including distances between C α -C α atoms, relative C α -C α -C α frame orientations, and backbone dihedral angles. The model has 1.7 million parameters.

LigandMPNN may be conceived as an extension of ProteinMPNN, incorporating ligand structural information. The model treats protein residues as nodes and introduces nearest neighbor edges based on C α -C α distances to define a sparse protein graph. The protein backbone geometry is encoded as graph edges by pairwise distances between N, C α , C, O, and virtual C β atoms. In addition, two protein-ligand encoder layers are used to encode protein-ligand interactions. The model has 2.6 million parameters.

For all three models, 10,000 target inverse-folding sequences were generated, with the output sequence diversity parameter set to 0.5 to facilitate comparison between different models. The input to these models is a list of coordinates for the N, C α , and C atoms, which constitute the protein backbone. For each structure, the coordinate list coords has a shape of $L \times 3 \times 3$, where L is the number of amino acids in the structure. Specifically, coords[i][0] contains the 3D coordinate for the N atom in amino acid i , coords[i][1] contains the 3D coordinate for the C α atom, and coords[i][2] contains the 3D coordinate for the C atom.

Use of AiCE_{single} for designing single substitution mutations

AiCE_{single}, a module within AiCE, is a method to nominate candidate substitutions using $f_{\max}(i)$ under structural constraints. First, we define the β score for global screening of $f_{\max}(i)$ at any position i of the protein. The set of single substitutions where $f_{\max}(i)$ is greater than β is defined as AiCE_1 , which is:

$$\text{AiCE}_1 = \{i | f_{\max}(i) \geq \beta\}$$

Next, we define the γ -score as a threshold for screening $f_{\max}(i)$ in the flexible regions of the protein. The set of single substitutions where $f_{\max}(i)$ in the flexible region exceeds γ is defined as $\text{AiCE}_{2_extended}$, which is:

$$\text{AiCE}_{2_extended} = \{i | \text{is_flexible}(i) = 1 \text{ and } f_{\max}(i) \geq \gamma\}$$

where $\text{is_flexible}(i) = 1$ indicates that the most frequently occurring mutation $x_{i\text{mut}}$ at position i is located in a flexible region of the protein to be modified. The DSSP algorithm⁸⁶ is used to predict the secondary structure of proteins, with flexible regions labeled as S (bend), T (turn), and other relevant regions.

The set of mutations obtained by AiCE_{single} is the union of AiCE_1 and $\text{AiCE}_{2_extended}$, which is:

$$\text{AiCE}_{\text{single}} = \text{AiCE}_1 \cup \text{AiCE}_{2_extended}$$

By leveraging the most frequently occurring mutations $x_{i\text{mut}}$ and $f_{\max}(i)$ from the inverse-folding sequence, along with the β and γ scores, AiCE_{single} identifies high-fitness (HF) mutations without relying on prior knowledge of protein fitness or manual annotation.

AiCE_{multi} for designing multi-mutations

AiCE_{multi} is a module within AiCE designed to nominate multi-mutations by exploiting the concept of evolutionary coupling. AiCE_{multi} aims to identify amino acid position combinations with high co-evolution rates, an approach that is successfully applied in structure prediction tools such as AlphaFold2.⁷⁶ To achieve this, AiCE_{multi} uses linkage disequilibrium analysis and evolutionary coupling analyzes to evaluate candidate positions at both the nucleotide and amino acid levels.

Linkage disequilibrium analysis

Amino acid sequences obtained from inverse folding were converted into pseudo-DNA sequences based on human optimal codon frequency. These pseudo-DNA sequences were then converted into VCF format files, with mutation information annotated using a custom script. Plink software (v1.90b6.21)⁸² was used to compute the LD score (r^2) for two-by-two positional linkage disequilibrium:

$$D = P_{AB} - P_A P_B$$

Here, P_A , P_B , and P_{AB} denote the population frequencies of allele A, allele B, and heterozygote AB, respectively. D is the coefficient of linkage disequilibrium. The linkage disequilibrium (LD) score r^2 between two positions is then calculated as:

$$r^2 = \frac{D^2}{P_A(1 - P_A)P_B(1 - P_B)}$$

Further mapping of DNA positions to corresponding amino acid positions is required in the analysis.

Evolutionary coupling analysis

We directly calculated the evolutionary coupling (EC) score between two amino acid positions using the Rivoire et al. method.⁸³ The joint frequency of amino acids *a* and *b* at positions *i* and *j* is denoted by C_{ij}^{ab} , and is defined as:

$$C_{ij}^{ab} = f_{ij}^{ab} - f_i^a f_j^b$$

where f_i^a and f_j^b are the frequencies of amino acids *a* and *b* at positions *i* and *j*, respectively.

To emphasize evolutionary conservation, we weight the covariance matrix:

$$EC = \widetilde{C}_{ij}^{ab} = \phi_i \phi_j C_{ij}^{ab}$$

Here, ϕ_i and ϕ_j are defined as:

$$\phi_i = \ln\left(\frac{f_{ai}(1 - q_a)}{(1 - f_{ai})q_a}\right)$$

$$\phi_j = \ln\left(\frac{f_{bj}(1 - q_b)}{(1 - f_{bj})q_b}\right)$$

where f_{ai} and f_{bj} are the frequencies of amino acids *a* and *b* at positions *i* and *j*, respectively, in the multiple sequence alignment. q_a and q_b are the background frequencies of amino acids *a* and *b*, respectively.

Multi-mutation nomination based on LD and EC score

$AiCE_{multi}$ defines amino acid position combinations with high LD scores and high EC scores as multi-mutations with potentially high fitness. We define a high LD score (r^2) as 0.5 and a high EC score as the top tenth of the global EC score distribution for an amino acid sequence. In this study, LD or EC scores (collectively referred to as $score_{\mathcal{M}}$ for a given set of mutant positions \mathcal{M}) were calculated as the global weighted score of all positions in the mutation combination:

$$score_{\mathcal{M}} = \frac{2}{|\mathcal{M}| \times (|\mathcal{M}| - 1)} \sum_{i < j, i, j \in \mathcal{M}} score(i, j)$$

where $|\mathcal{M}|$ denotes the total number of sites in the mutation set \mathcal{M} , and $score(i, j)$ represents the score between positions *i* and *j*. This calculation provides a measure of the evolutionary coupling strength of the mutation combination as a whole.

The final mutation position combination $AiCE_{multi}$ is defined as the union of the high r^2 combination ($AiCE_{LD}$) and the high EC score combination ($AiCE_{EC}$):

$$AiCE_{multi} = AiCE_{LD} \cup AiCE_{EC}$$

where:

$$AiCE_{LD} = \{\mathcal{M} | score_{\mathcal{M}-LD} \geq 0.5\}$$

$$AiCE_{EC} = \{\mathcal{M} | score_{\mathcal{M}-EC} \geq score_{percentile\ 90}\}$$

Here, $score_{\mathcal{M}-LD}$ and $score_{\mathcal{M}-EC}$ refer to the r^2 and EC score of the combination \mathcal{M} , respectively, and $score_{percentile\ 90}$ refers to the top 10th of the global score of the amino acid sequence.

As illustrated in Figure 3D, the amino acid type corresponding to the combination of the selected mutation positions should be the amino acid mutation x_{imut} with the highest appearance rate. By integrating $AiCE_{single}$ with evolutionary coupling analysis, HF multi-mutations can be identified without bias, a process we refer to as $AiCE_{multi}$. Furthermore, this nonlinear mapping of evolutionary coupling and mutation fitness enhances the interpretability of $AiCE_{multi}$.

Mutation selection based on natural homologs

In principle, the $AiCE_{single}$ mutation selection strategy using $f_{max}(i)$ is not limited to inverse-folding sequences. In this study, we also evaluated the preservation of HF mutations using natural homologs. The natural homologs were gathered from publicly accessible databases using two profile search schemes to ensure comprehensive candidate retrieval.

HH-suite3 (v3.3.0)⁷⁹ was used to identify homologs across a range of databases, including BFD (bfd_metaclust_clu_complete_id30_c90_final_seq_sorted_opt), dbCAN3 (dbCAN-fam-V9), pdb100 (accessed on May 03, 2022), and pdb70 (accessed on May 03, 2022), Pfam (35.0), SCOP40 (accessed on Mar 1, 2017), SCOP 70 (accessed on Mar 1, 2017), and UniRef30 (UniRef30_2023_02). Search parameters were set to ‘-e 0.0005 -B 1000 -maxfilt 50000’. These databases collectively comprised 2.3 terabytes of data.

Additionally, protein sequence files were obtained from non-redundant genome databases, including GTDB (Genome Taxonomy Database, accessed on March 23, 2022), MGnify (formerly known as EBI metagenomics, accessed in May 2022), and UniProt (accessed in May 2022). Encoded protein databases from microbial genomes and metagenomic sources were constructed, and homologs were iteratively searched using JackHMMER (v3.4)⁸⁰ with the following parameters: '-F1 0.0005 -F2 5e-05 -F3 5e-07 -incE 0.0001 -E 0.0001'. The database used for this search totaled 0.9 terabytes, with five iterative search rounds performed for both methods to maximize homolog retrieval.

Candidate homologs were initially filtered using CD-HIT (v4.8.1)⁷⁷ to remove redundant sequences with coverage and similarity exceeding 99%. Furthermore, sequences with coverage greater than 1.8 or less than 0.6 in comparison to the reference sequences were removed before multiple sequence alignment (MSA) processing. MUSCLE (v5.1)⁸¹ in '-super5' mode was used for rapid and precise multiple sequence alignment. Subsequently, a custom script was utilized to eliminate gap positions in the reference sequences, generating a multiple sequence alignment based on the natural homologs of the reference sequences. Ultimately, the candidate mutations were obtained through the AiCE_{single} strategy.

Protein DMS libraries spanning multiple families as a benchmark for assessing single mutations

Sixty protein DMS datasets, encompassing diverse families from a range of species, were systematically compiled to serve as benchmark for assessing the efficacy of AiCE_{single}. The datasets cover a diverse range of functional categories, including transcriptional regulation, oncogenes, and signaling (Table S1). Among them, 51 of the libraries correspond to 44 proteins with curated PDB structures predicted by AlphaFold2, with an average predicted local distance difference test (pLDDT) score of 84.7. Although the BRCA1 RING domain had a lower average pLDDT score of 48.3 (Table S1), it still exhibited satisfactory predictive performance. These libraries were used to evaluate the performance of mutation prediction across all models and methods, except LigandMPNN.

In addition, nine protein complex libraries were included. These complexes were also evaluated using the LigandMPNN model to assess its ability to generate HF mutations within the AiCE_{single} framework. The scarcity of certain human proteins or viral homolog stems from the limited availability of microbial sequence databases and the specificity of viral genomes. Our analysis incorporated four percentile thresholds (top 5th, 15th, 25th, and 50th), with particular emphasis on the top 5th percentile, which represents the optimal level of fitness.

Mutation data from antibodies and other proteins as benchmarks for assessing multi-mutations

Four antibody mutation datasets were initially used to assess the efficacy of three inverse folding models (ESM-IF1, LigandMPNN, and ProteinMPNN) as well as SaProt in generating multi-mutations. The input structural templates for these models were CR9114 in complex with H5 HA (PDB ID 4FQI)⁸⁷ and CR6261 with H1 HA (PDB ID 3GBN). These datasets, which exhaustively explore permutations from a single mutation up to 16 mutations, allow for a comprehensive assessment of AiCE_{multi}'s ability to predict combinations of HF mutations. To evaluate model performance, we analyzed the correlation between weighted mutation scores (log likelihood) for all mutations and the corresponding experimental binding measurements ($\text{Log } K_D$ for CR9114 and CR6261). The weighted mutation score, $\text{Log score}_{\mathcal{M}}$, for a mutation combination \mathcal{M} is defined as:

$$\text{Log score}_{\mathcal{M}} = \text{Log} \frac{1}{|\mathcal{M}|} \sum_{i \in \mathcal{M}} f_{\text{mut}}(i)$$

where $|\mathcal{M}|$ denotes the total number of sites in the mutation set \mathcal{M} , and $f_{\text{mut}}(i)$ represents the rate of appearance of mutations at site i . For simplicity, we represent the base 10 logarithm (\log_{10}) as Log.

The correlation between $\text{Log score}_{\mathcal{M}}$ and $\text{Log } K_D$ varied considerably across different datasets and models. LigandMPNN demonstrated the strongest correlation overall but performed poorly on the CR9114 dataset. This suggests a limited ability to predict combinations of high-fitness mutations based solely on weighted mutation scores. We developed the AiCE_{multi} method and demonstrated its effectiveness in guiding the selection of candidate multi-mutations with high success rates. AiCE_{multi} was tested on two new datasets: the yeast His3 protein and ppI GFP2, using input structures predicted by AlphaFold2.

Selection of mutations for experimental testing

Deaminases

As a first demonstration, we focused on engineering deaminases, the catalytic core of base editors, given their ability to mediate genome editing at single-nucleotide resolution.^{12,53} These enzymes have broad applications in mutagenesis, functional mutation screening, sequencing technology development, and therapeutics.^{48,88–90} Deaminase-based base editors can correct disease-causing mutations, which account for approximately 43% of human disease-associated variants (data from dbVar: <https://www.ncbi.nlm.nih.gov/dbvar/>, 08/2024), and introduce beneficial mutations.⁹¹ A number of base editing therapies have already reached the clinical trial stage, including VERVE-101 for lowering low-density lipoprotein cholesterol (LDL-C) and BEAM-101 for treating sickle cell disease and β-thalassemia. However, many deaminases remain suboptimal due to limited activity and frequent off-target or bystander edits.^{54,92} Addressing these issues through conventional engineering is labor-intensive and costly.^{57,92,93} We therefore applied AiCE's HF mutation identification prediction framework to efficiently generate and experimentally test a focused set of optimized deaminase variants.

The single-stranded DNA cytosine deaminase Sdd6, single-stranded DNA adenine deaminase TadA8e, and double-stranded DNA deaminase Ddd1 represent the deaminase components of the most commonly utilized cytosine base editors, adenine base editors, and mitochondrial base editors, respectively. The structures of Sdd6-ssDNA complex and Ddd1 were predicted by AlphaFold3 and AlphaFold2, respectively, whereas the structure of TadA8e was determined by cryo-EM (PDB ID 6VPC).

A range of computational methods was used to predict single mutants without preselection, including AiCE_{single}-ESM-IF1, AiCE_{single}-ProteinMPNN, AiCE_{single}-LigandMPNN, natural homologs, ESM-1v, ESM3, SaProt, and the hallucination-based design approach. AiCE_{single} consistently applied standardized AiCE filtering parameters. ESM-1v selected mutations following the methodology described by Brain et al., while ESM3 and hallucination-based design utilized masked approaches. ESM3 leveraged the publicly available ESM3-open model and hallucination-based design was derived from RFdesign's *Hallucination* design module.⁸⁴ In the TadA8e study, an additional 15 mutations were evaluated, sourced from ProMEP's prior work on TadA1.2. Across all methodologies, 20–30 mutations were randomly selected per model for experimental validation, resulting in the generation of 122, 113, and 141 single mutants in TadA8e, Sdd6, and Ddd1, respectively.

Six mutations were identified in DddA11: T26I, R38V, N39S, N67K, T77I, and T110I (aligned to the predicted Ddd1 structure). Among these, T26I, T77I, and T110I have been shown to enhance deamination efficiency in Ddd1 and are designated as PACE-M1, PACE-M2, and PACE-M3, respectively. To further assess the potential of these mutations, we conducted a combination test involving D-h11 (located at the N-terminus of Ddd1, split site N94) and PACE-M3 (located at the C-terminus of the deaminase).

AiCE_{multi} was used to generate multi-mutations, with the Low_{multi} series of mutants representing random combinations of mutations constrained by a low score_M. Given ProteinMPNN's favorable performance in previous evaluations and its computational efficiency, we used AiCE_{multi}-ProteinMPNN to design multi-mutants. Control mutants were included in ablation experiments, maintaining the same mutation sites as AiCE_{multi} but selecting mutation types using the BLOSUM62 matrix instead of AiCE_{single}. Additionally, the designed multi-mutations were deconstructed into single mutations to evaluate whether epistatic effect could be mitigated.

In total, we tested 131, 114, and 138 single mutants for the TadA8e, Sdd6, and Ddd1 proteins, along with 24, 21, and 15 multi-mutants.

Other complex tasks

Due to cost constraints, we limited our design to single mutants using AiCE_{single}-ProteinMPNN, which had demonstrated strong performance in previous evaluations. We designed 14, 25, 20, 30, and 20 mutants for NLS, LbCas12a, AcCas12n, SpRYCas9, and M-MLV RT, respectively. To evaluate the functional impact of engineered NLS variants on base editing, we incorporated a selected NLS mutation into a cytosine base editor (CBE) system containing rAPOBEC1. In parallel, nucleases were assessed for their indel efficiencies, and M-MLV RT variants were tested for their editing precision in a prime editing system.

Identification of rational design hotspot regions for rational protein design

Protein-protein interaction regions, protein-DNA interaction regions, and enzyme activity pockets are defined as hotspot regions for the rational design of deaminases. These regions were identified in complex structures derived from predicted or experimentally resolved data. Interaction regions were calculated as atoms situated within a 5 Å radius of DNA, Zn²⁺, Mg²⁺, ions, and other proteins.

Protein structure prediction, comparison, and visualization

All structures used for mutation design were predicted using AlphaFold2 (v2.3.0) with the default parameters. The predicted complexes in this study were obtained from the AlphaFold3 online server. Structural comparisons were performed using the US-align online server, and protein structure visualizations were generated using UCSF Chimera.⁸⁵

Plasmid construction

All protein components were codon-optimized for expression in human and mouse cells and synthesized commercially by GenScript. Plasmids encoding TALE-based base editors, including Ddd1, the TALE array, UGI (uracil glycosylase inhibitor), and a mitochondrial targeting signal, were codon-optimized and synthesized. The components were subsequently cloned into the vectors pCMV-TALE-L-JAK2-Ddd9-N (Addgene #204853) and pCMV-TALE-R-JAK2-Ddd9-C (Addgene #204854).⁵⁷ Vectors related to Sdd6 and TadA8e were constructed using p2T-CMV-miniSdd6-BE4max-BlastR (Addgene #204850) as the backbone.⁵⁷ The NLS sequence was introduced into pCMV_BE4max (Addgene #112093),⁷⁴ replacing the original N-terminal NLS of the encoded protein. Constructs for LbCas12a, AcCas12n, and SpRYCas9 were generated using the LbCas12a-D156R (Addgene #213744) backbone.⁷⁵ M-MLV reverse transcriptase (RT)-related vectors were derived from pCMV-PE2 (Addgene #132775) with the RNase H domain deleted.⁶⁶ All mutants were generated by fusion PCR and cloned using the Gibson assembly method.

All RNA expression elements were designed using the phU6 vector (Addgene #53188) as a template and assembled via circular PCR or Gibson assembly.⁷³ For circular, digested PCR products were transformed into *Fast-T1* competent cells (Vazyme). PCR amplification was performed using 2× Phanta Max Master Mix (Vazyme).

Mammalian cell transfection

Transfection was carried out 16–24 hours after seeding. In transfection experiments using the TALE-mediated base editing system, 0.4 µl of Lipofectamine 2000, 150 ng of the TALE-L vector, 150 ng of the TALE-R vector, and 10 ng of green fluorescent protein were co-incubated and transfected into the cells.

In transfection experiments involving the CRISPR-Cas-mediated base editing system, 0.4 µl of Lipofectamine 2000 (ThermoFisher Scientific), 300 ng of a vector containing a deaminase, 100 ng of an sgRNA expression vector, and 10 ng of Green Fluorescent Protein were co-incubated and transfected into the cells.

To examine off-target effects using the R-loop assay, four vectors were co-incubated and transfected with 0.4 µl of Lipofectamine 2000 (ThermoFisher Scientific). A total of 150 ng of the pCMV-deam-BE4max vector, 150 ng of the pCMV-nSaCas9 vector, and two corresponding sgRNA vectors (each at 50 ng), along with 10 ng of green fluorescent protein, were used.

DNA extraction

Following a 72-hour incubation period, the HEK293T cells were washed with PBS and then subject to genomic DNA extraction with Lysis Buffer and Proteinase K using a Triumfi Mouse Tissue Direct PCR Kit (Genesand).

Amplicon deep sequencing and data analysis

A series of primers with barcodes at their 5'-ends were designed for the purpose of amplifying targeted sequences. The amplicons were purified using a Thermo Scientific GeneJET Kit (Thermo Fisher Scientific) and quantified with a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific). Equal amounts of PCR product were pooled and sequenced commercially (GENEWIZ) using the NGS. For target sites that are difficult to amplify, an additional round of primers was designed for 500 bp amplification. The resulting product was diluted 20-fold, and then 1 µl of the diluted product was used as template for nested PCR using primers containing a barcode sequence. All the primers used are listed in a supplementary table ([Table S3](#)).

Sequences surrounding target regions were scanned for editing events using a Python script from previous algorithms. The cleaned sequencing data was first split according to the sequencing primers. The following approach was used: (1) *Data splitting and merging*: The pooled sequencing data was split into individual treatments based on sample-specific adapter sequences. The FLASH tool (v1.2.11)⁷⁸ was used to merge the forward and reverse reads of each treatment into a single read, with parameters set to '-m 5 -M 150' to ensure a minimum overlap of 5 bp and a maximum overlap of 150 bp for paired-end reads. (2) *Sequence comparison and SNP extraction*: The merged reads were compared to the wild-type reference sequence. The comparison strategy involved determining the sequence window for viewing editing events by obtaining seven base pairs before and after the reference sequence as index sequences. The index sequences were compared to the merged reads. If both end index sequences were present in the merged reads, the middle sequence was extracted. The extracted sequence was then compared with the window sequence to identify and extract SNP information. (3) *Counting of editing events*: The editing efficiency, mutation types and other relevant metrics of all base editing events were calculated from the information about the extracted SNPs.

QUANTIFICATION AND STATISTICAL ANALYSIS

Quantification

Datasets were assessed using GraphPad Prism 10 (GraphPad Software) and Python.

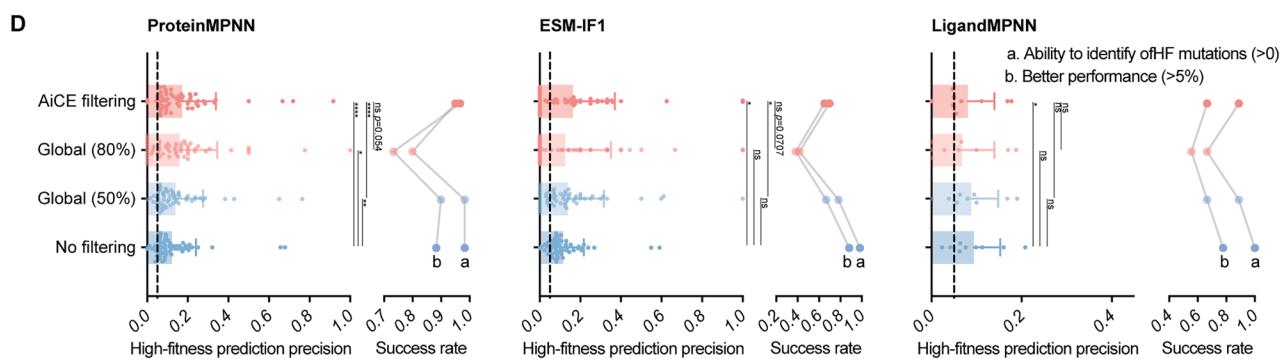
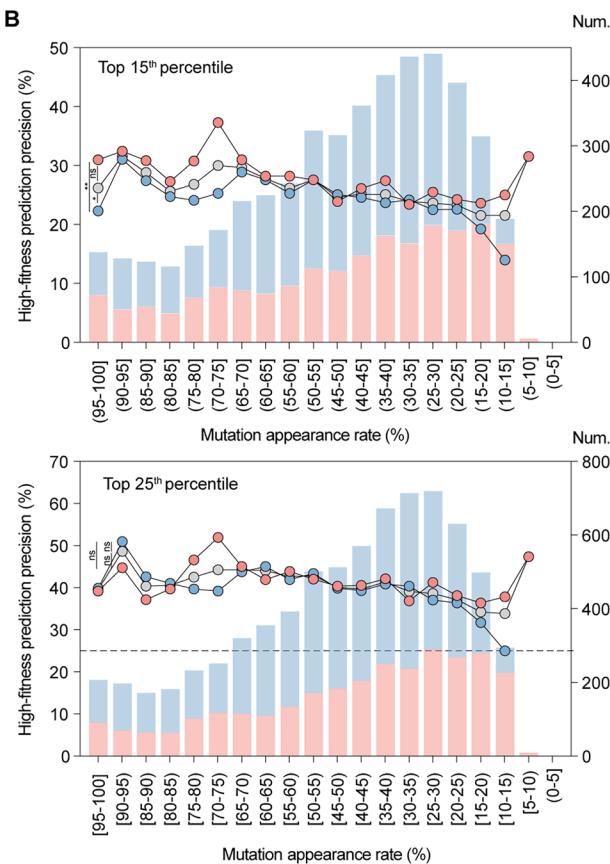
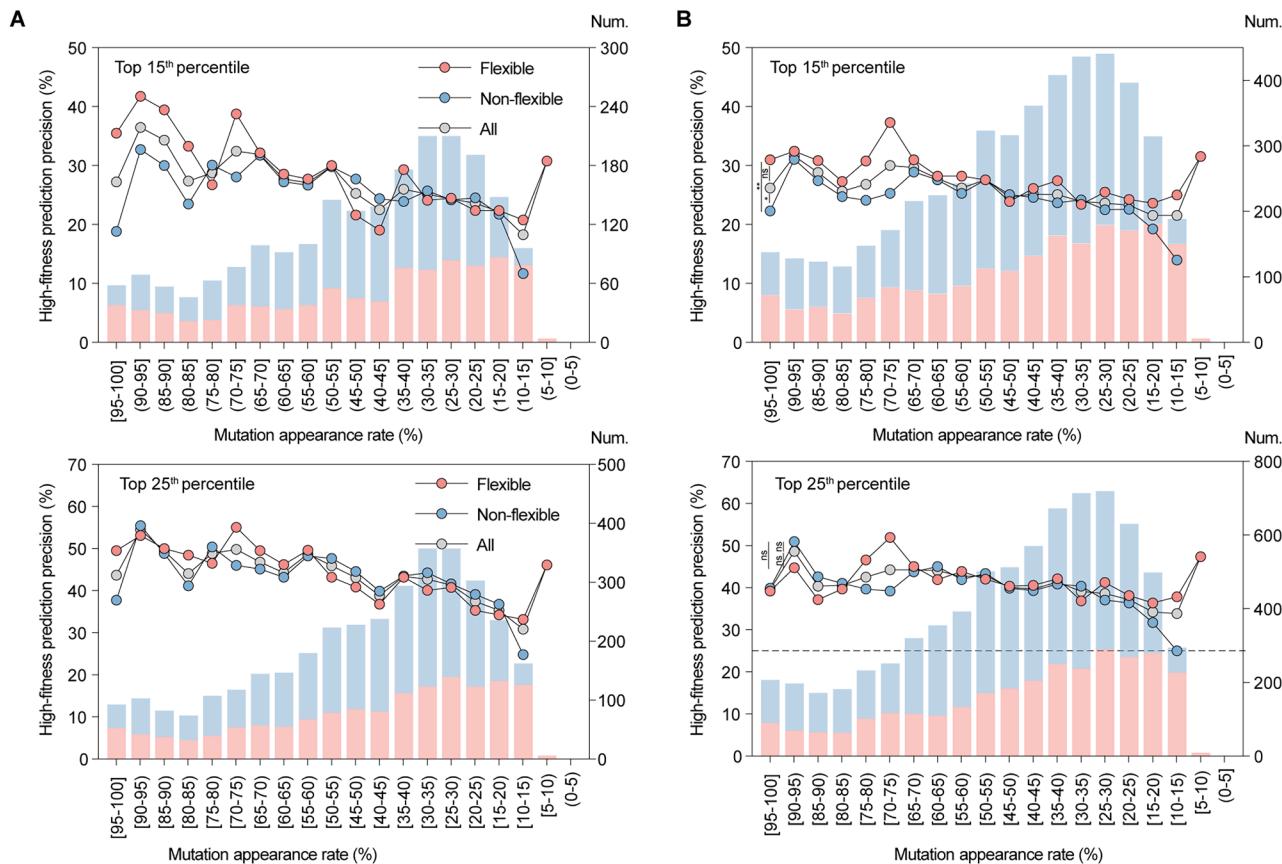
Statistical analysis

Statistical differences between treatments were tested using one-sided Mann-Whitney tests and Fisher's exact test, depending on the context. All numerical values are presented as means ± s.d.

ADDITIONAL RESOURCES

Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

Supplemental figures



(legend on next page)

- Figure S1.** AiCE_{single} under structural constraints can efficiently predict high-fitness mutations in various proteins, related to [Figure 2](#)
- (A) Predictive accuracy of 15th and 25th mutations at different mutation appearance rates across 31 DMS libraries. Labeling is as in [Figure 2C](#). *p* values from one-sided Mann-Whitney tests are shown: ⁿ*p* > 0.05, **p* < 0.05, ***p* < 0.01.
- (B) Same as (A), but for 60 DMS libraries.
- (C) Logistic regression analysis of factors influencing mutations predicted by different inverse folding models, based on data from 60 DMS libraries. *p* values from likelihood ratio tests are shown: **p* < 0.05, ***p* < 0.01, ****p* < 0.001, *****p* < 0.0001.
- (D) Comparison of different filtering strategies for predicting HF mutations using ProteinMPNN (left), ESM-IF1 (middle), and LigandMPNN (right). Labeling is as in [Figure 2F](#).

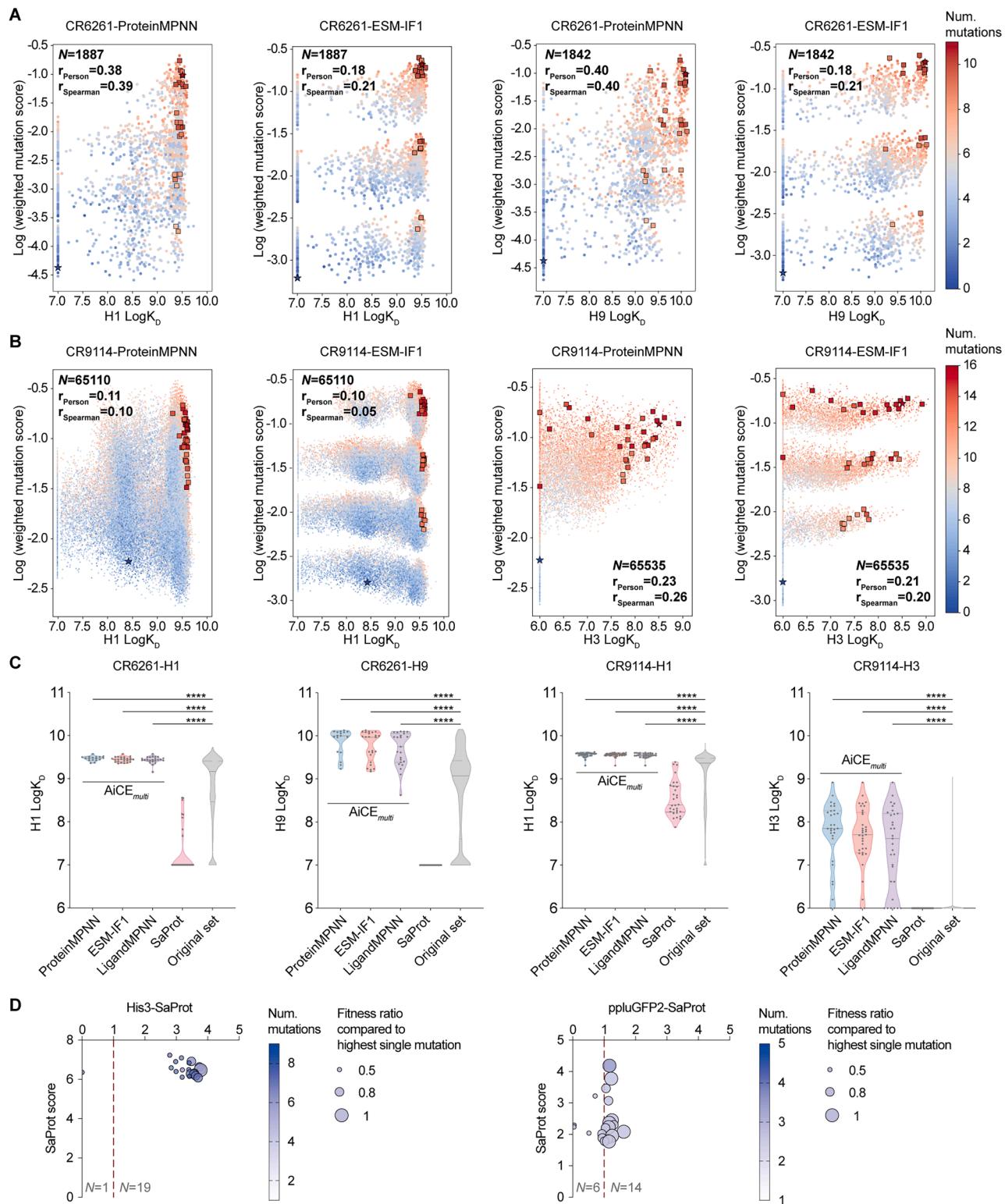


Figure S2. Testing the performance of AiCEcom's predictions of HF mutation combinations using different inverse folding models, related to Figure 3

(A and B) Predictive performance of AiCE_{multi} for HF multi-mutations in CR6261 (A) and CR9114 (B) antibodies using ProteinMPNN and ESM-IF1. Labeling is as in Figures 3A and 3B.

(legend continued on next page)

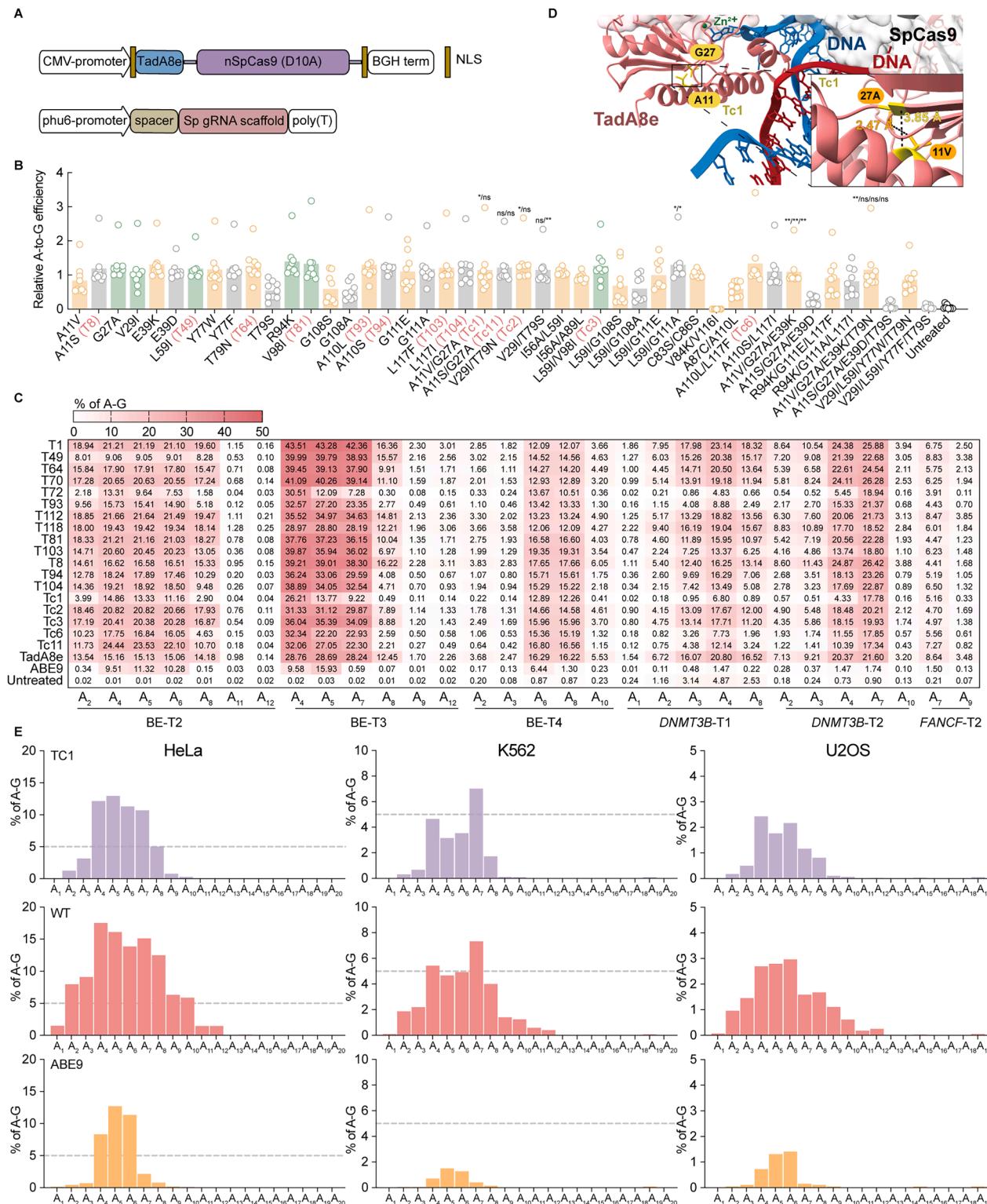


Figure S3. Design and testing of ABEs with narrower editing windows, related to Figure 4

(A) Schematic of the nCas9-based vector for ABEs.

(B) Relative editing efficiencies of TadA8e multi-mutants and their constituent mutations at three endogenous sites in HEK293T cells, normalized to wild-type TadA8e (set to 1). Each dot represents an independent replicate. Variants identified by AICE_{multi}, BLOSUM62, and both methods are labeled as orange, gray, and

(legend continued on next page)

green. Bars represent mean values. p values from one-sided Mann-Whitney tests for each of the HF multi-mutants that completely or partially overcame the epistatic effect with its constituent individual amino acid substitutions are shown: ${}^{ns}p > 0.05$, $*p < 0.05$.

(C) Average editing efficiency across editing windows for selected TadA8e variants across six sites in HEK293T cells. The PAM-proximal end is defined as position 1.

(D) Cryo-EM structure of the TadA8e and SpCas9 in complex (PDB ID 6VPC) with single-stranded DNA (ssDNA) and magnesium ions (Mg^{2+}). Constituent mutations are highlighted in yellow within the structural diagrams. The structure illustrates the positioning of these substitutions, as well as their potential interactions in the case of Tc1.

(E) Editing window profile of Tc1, TadA8e, and ABE9 across 24 target sites within eight genes in HeLa (left), K562 (middle), and U2OS (right) cells. Data represent the average of three independent biological replicates.

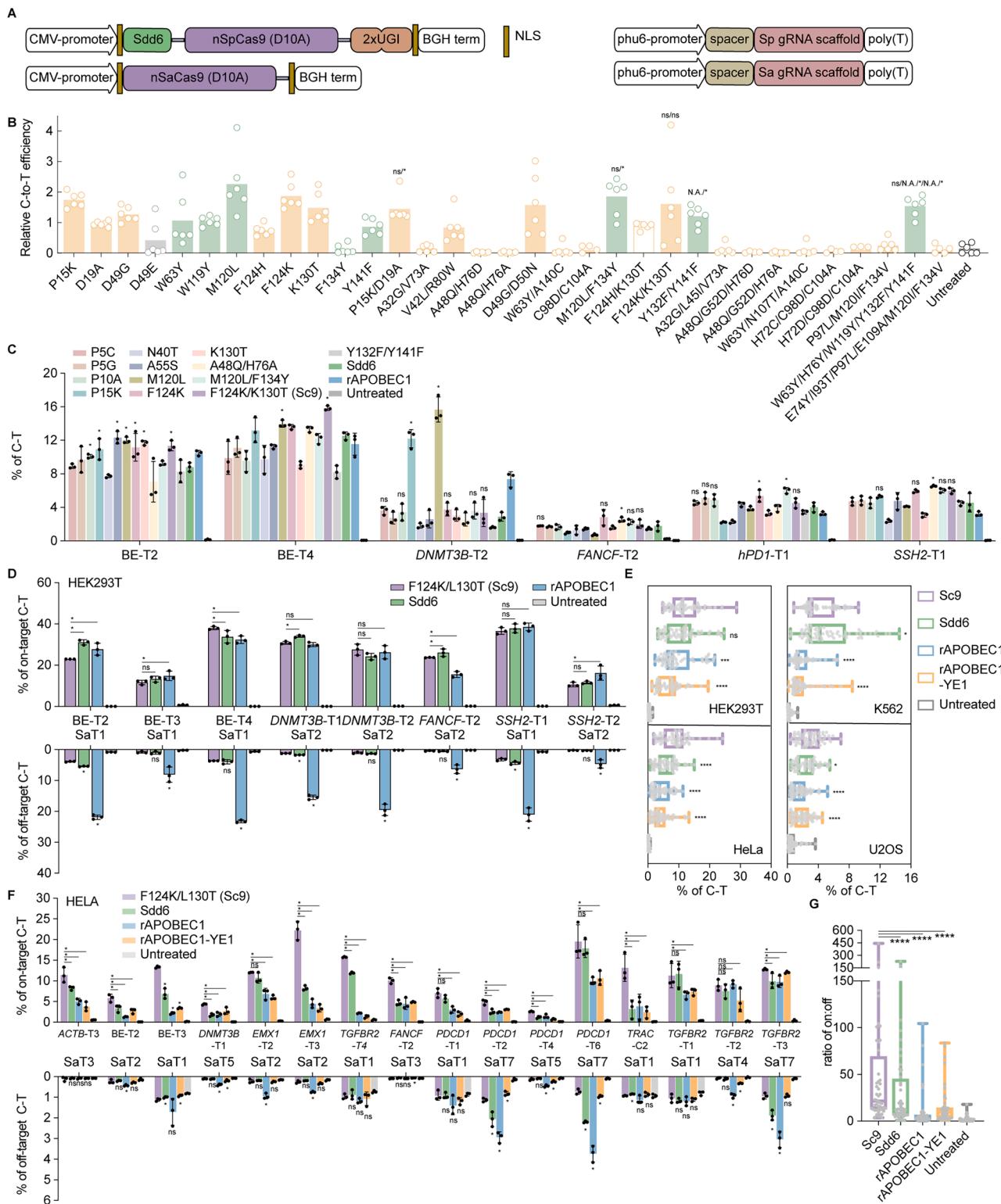


Figure S4. Design and testing of highly specific CBEs, related to Figure 4

(A) Schematic representation of CBE vectors. The constructs include an nSpCas9 (D10A)-based CBE and an nSaCas9 (D10A)-based system for the orthogonal R-loop assay, both containing a spacer-guide RNA scaffold.

(legend continued on next page)

-
- (B) Relative editing efficiencies of Sdd6 multi-mutants and their constituent mutations at two endogenous sites in HEK293T cells, normalized to wild-type TadA8e (set to 1). Labeling is as in [Figure S3B](#).
- (C) Editing efficiencies of selected Sdd6 variants, Sdd6, and rAPOBEC1 across six endogenous sites in HEK293T cells.
- (D) Specificity evaluation in HEK293T cells. Orthogonal R-loop assay measuring on-target efficiency across eight on-target sites (top) and sgRNA-independent off-target efficiency across two off-target sites (bottom) for Sc9, Sdd6 (wild type), and rAPOBEC1.
- (E) Editing efficiencies across multiple cell lines. Average editing efficiencies of Sc9, Sdd6, rAPOBEC1, and rAPOBEC1-YE1 across 24 endogenous sites in HEK293T, HeLa, K562, and U2OS cells.
- (F) Specificity evaluation in HeLa cells. Orthogonal R-loop assay measuring on-target editing efficiency across sixteen on-target sites (top) and sgRNA-independent off-target editing efficiency across six off-target sites (bottom) for Sc9, Sdd6 (wild type), rAPOBEC1, and rAPOBEC1-YE1.
- (G) On-target:off-target editing ratios for each base editor calculated from (E).
- For (C)–(G), bars represent mean values \pm SD. *p* values from one-sided Mann-Whitney tests are shown: ^{ns}*p* > 0.05, **p* < 0.05, ***p* < 0.001, and ****p* < 0.0001. Data represent the average of three independent biological replicates.

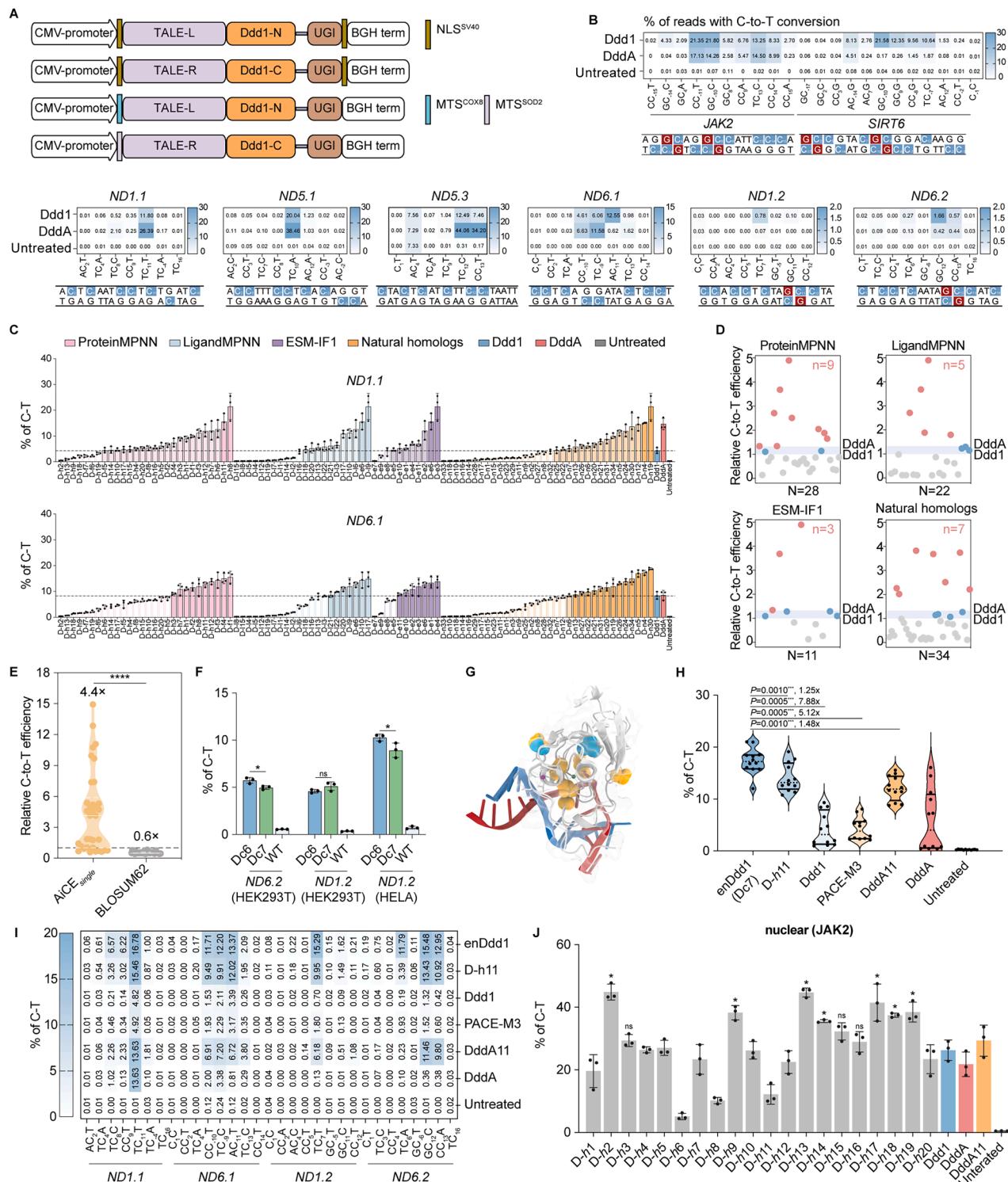


Figure S5. Engineering Ddd1 by different methods, related to Figure 5

(A) Schematic of the DdCBE vector construction. The DdCBE system is constructed using dual TALE effectors fused to Ddd1 variants for targeted C-to-T base editing.

(B) Editing efficiencies of Ddd1 and DddA at nuclear and mitochondrial targets. C-to-T conversion efficiencies were measured at eight endogenous target sites, including two nuclear targets (*JAK2* and *SIRT6*) and six mitochondrial sites. The 5'-GC sequence initiation point is indicated in red.

(legend continued on next page)

- (C) Editing efficiencies of Ddd1 variants obtained by different models across two mitochondrial sites (*ND1.1* and *ND6.1*) in HEK293T cells. The absence of color denotes variants with lower efficiency than both wild-type Ddd1 and DddA. Transparent colors indicate mutants with efficiency between Ddd1 and DddA, while solid colors represent mutants with higher efficiency than both Ddd1 and DddA. The dashed line represents wild-type Ddd1 efficiency.
- (D) Relative editing efficiencies of Ddd1 variants at four mitochondrial sites (*ND1.1*, *ND1.2*, *ND6.1*, and *ND6.2*) in HEK293T cells, normalized to wild-type Ddd1 (set to 1). Labeling is as in [Figure 5D](#). One variant from “natural homologs” is not shown because only two sites were tested.
- (E) Comparison of AiCE_{multi} and BLOSUM62 matrix for predicting Ddd1 mutations. Labeling is as in [Figure 4C](#).
- (F) Editing efficiencies of DC6, Dc7, Ddd1, and DddA at two mitochondrial sites (*ND1.2* and *ND6.2*) in HEK293T and HeLa cells.
- (G) Structural visualization of enDdd1 and DddA11 (DC7). Two substitutions in enDdd1 and DC6 are highlighted in blue, while the six substitutions in DddA11 are shown in yellow.
- (H) Comparison of enDdd1 and other variants in mitochondrial editing.
- (I) Heatmap illustrating the editing activities of Ddd1 variants within a window of endogenous sequences, corresponding to (F).
- (J) Editing efficiencies of Ddd1 variants, Ddd1 (wild type), DddA, and DddA11 at one nuclear target site (*SIRT6*) in HEK293T cells. Labeling is as in [Figure 5H](#). For (E), (F), (H), and (I), bars represent mean values \pm SD. *p* values from one-sided Mann-Whitney tests are shown: ^{ns}*p* > 0.05, **p* < 0.05, ***p* < 0.01, and ****p* < 0.0001. Data represent the average of three independent biological replicates.

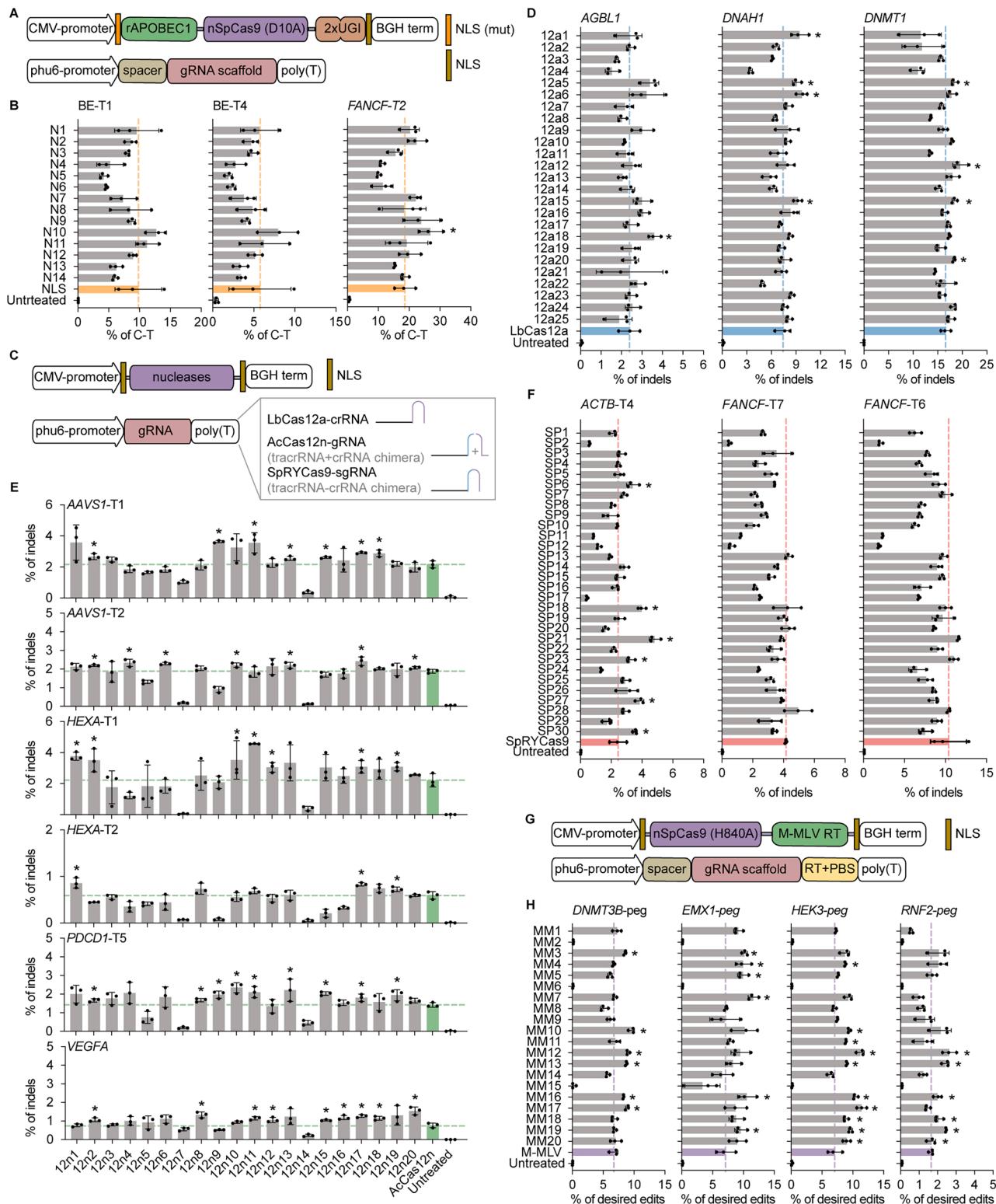


Figure S6. Engineering five new proteins using AiCEsingle, related to Figure 6

(A) Schematic representation of the NLS activity detection system. NLS variants were tested in a cytosine base editor (CBE) system, where the N-terminal region contained a mutant ^{BP}NLS, and the C-terminal region contained a wild-type ^{BP}NLS.

(B) Average efficiencies of NLS variants and wild-type NLS at three target sites in HeLa cells.

(legend continued on next page)

(C) Schematic of nuclease activity detection systems. Constructs for LbCas12a, AcCas12n, and SpRYCas9 include their respective RNA expression vectors for testing indel formation efficiency.

(D–F) Average indel efficiencies of LbCas12a variants (D), AcCas12n variants (E), and SpRYCas9 variants (F) and their wild-type counterparts at multiple target sites in HEK293T cells.

(G) Schematic of the M-MLV RT-based prime editing system. The system includes truncated M-MLV RT variants and RNA expression constructs for targeted editing.

(H) Average editing efficiencies of M-MLV RT variants and wild type at three target sites in HEK293T cells. Both variants and wild types are truncated proteins lacking the RNase H domain.

For (B), (D–F), and (H), bars represent mean values \pm SD. *p* values from one-sided Mann-Whitney tests are shown: **p* < 0.05. Data represent the average of three independent biological replicates.