Cytosine base editing systems with minimized off-target effect and molecular size

Ang Li
Kobe University
Hitoshi Mitsunobu
Kobe University
Shin Yoshioka
Kobe University
Suzuki Takahisa
Kobe University
Akihiko Kondo
Kobe University https://orcid.org/0000-0003-1527-5288
Keiji Nishida (keiji_nishida@people.kobe-u.ac.jp)
Kobe University https://orcid.org/0000-0002-9371-4971

Letter

Keywords: cytosine base editing systems, biological modeling, structure-based rational engineering

Posted Date: April 28th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-414230/v1

Version of Record: A version of this preprint was published at Nature Communications on August 8th, 2022. See the published version at https://doi.org/10.1038/s41467-022-32157-8.

Abstract

Structure-based rational engineering of the cytosine base editing system Target-AID was performed to minimize its off-target effect and molecular size. By intensive and careful truncation, DNA-binding domain of its deaminase PmCDA1 was eliminated and additional mutations were introduced to restore enzyme function. The resulting tCDA1EQ was effective in N-terminal fusion (AID-2S) or inlaid architecture (AID-3S) with Cas9, showing minimized gRNA-independent off-targets, as assessed in yeast and human cells. Combining with the smaller Cas9 ortholog system, the smallest cytosine base editing system was created that is within the size limit of AAV vector.

Introduction

The cytosine base editing is mediated by cytidine deaminase guided by a nuclease-deficient CRISPR system. At the target site, deamination of cytosine generates uracil, which eventually converts C•G base pair (bp) to a T•A bp without introducing a double DNA strand break[1] [2] . The originally developed cytosine base editing systems, Base editor (BE) [1] and Target-AID[2], respectively employ rAPOBEC and PmCDA1 as deaminases and efficiently introduce mutations within the editing windows of 12–16 bases and 16–20 bases upstream of the PAM (protospacer-adjacent motif).

Recent studies have raised concern that deaminase-mediated base editing systems can induce genome-wide SNV off-targets especially if overexpressed for long period of time[3] [4]. In contrast to CRISPR-Cas9-dependent off-targets which is based on mismatch annealing of guide RNAs, the SNV off-targets induced by base editing appears to be independent of the target sequence and is thought to be caused by non-specific, random deamination by the deaminase domain. BE systems have been well studied and the original BE3 and BE4 have been shown to induce both DNA and RNA off-targets[3–6] and several rAPOBEC1 mutant variants were then identified with reduced off-target effects[5,7,8]. Analytical methods have also been developed to evaluate off-target potential of the base editors. Genome-wide mutations can be thoroughly elucidated by whole genome sequencing (WGS). However, WGS is expensive, time-consuming, and has low throughput. In addition, it may not be sensitive enough for further comparative analysis of the improved base editors. Previous WGS-based studies had indicated that actively transcribed regions were prone to base editing off-target effect[3,9], because the R-loops which is formed by the exposure of single-stranded DNA by RNA transcription is a preferred substrate for the deaminases. To mimic such hot spots, localized R-loop was formed by using an orthogonal nuclease-defective CRISPR system which was co-transfected with base editing systems targeting another distant locus[7,8]. Deep sequencing of the R-loop region allowed for rapid and sensitive comparative evaluation of gRNA-independent off target potency. Alternatively, the rate of non-specific mutations can be monitored as the occurrence of drug resistant mutants in microbes such as yeast[2] and *E.coli*[10]. By conducting these studies simultaneously, potential biases can be compensated for each other[7].

For the treatment of genetic diseases, base editing is considered as a promising agent because it can install specific SNV without inducing DNA double-strand breaks or template DNAs. In vivo delivery is one of the major bottlenecks to achieve efficient and specific editing at the target tissue. Smaller molecular size is advantageous especially for in vivo delivery tools such as adeno-associated virus (AAV) vector. AAV vector is one of the promising delivery methods for gene therapy with greater safety and efficiency[11], although its DNA vector size limitation (4-5kb) hinders wider applications including base editing. In conventional genome editing, smaller CRISPR ortholog *Staphylococcus aureus* (Sa) Cas9 has led to the development of single AAV vector[12,13], but adding a deaminase domain has been difficult[14]. Instead of composing a single AAV vector, the base editing components could be split into two AAV vectors to circumvent the size limitation[15].

In this study, we addressed structure-based rational engineering of the cytosine base editing system Target -AID to minimize its off-target effects and molecular size.

Results

Elimination of DNA-binding region and restoration of deaminase activity of PmCDA1

DNA deaminases have an intrinsic affinity for DNA and cause nonspecific deamination. The structure of hAID, a human homolog of PmCDA1, has revealed its complex formation with double-stranded DNA in a region distinct from the catalytic core[16] (Fig. 1a). Based on the amino acid alignment of hAID and PmCDA1, the potential DNA-binding moieties for PmCDA1 were located to residues 21–27 and 172–192 of the total 208 amino acids (a.a.) length of the protein (Fig. 1a). To delete the predicted DNA-binding region, we first made a series of truncations from the C-terminal end (1-201, 1-197, 1-190, 1-183, 1-179, 1-176, 1-161) and tested their base editing activity in yeast *Saccharomyces cerevisiae* (BY4741) cells (Supplementary Fig. S1). Although the previous report had shown that 47 a.a. truncated PmCDA1 (1-161) still exhibited comparable editing efficiency to that of full-length PmCDA1 (1-208) in yeasts[17], our versions which were fused to the C-terminus of nCas9 and devoid of uracil DNA glycosylase inhibitor (UGI) showed a gradual decrease of their activity as truncation proceeded, as we assumed that the UGI-fusion was too effective in yeast and masked evaluation of the net activity of the deaminase. Next, we performed a series of truncations from the N -terminus of the 1-161 version by fusing to the N-terminus of nCas9. The N-terminus truncations of CDA1 (1-161) first showed further decreased activity, which was then recovered as truncation proceeded to 21 and 28 a.a (Supplementary Fig. S2). The predicted structure of the protein indicated that simultaneous truncation of the N- and C- terminus minimizes cross-section and gives a smoother protein surface with less exposure of hydrophobic residues (Supplementary Fig. S2). Further truncation to CDA1(30-150), which was predicted to be the smallest one with minimum exposure of the hydrophobic surface while retaining its enzymatic core domain intact (Fig. 1b, Supplementary Fig. S2), showed recovered activity (Supplementary Fig. S2). These results suggest that the changes in their editing activity were attributed to the conformational stability of the protein. To further improve its activity, we introduced a series of mutations to the hydrophobic residues that were

exposed after the truncation. Six mutations were tested in the first round and W122E was found to significantly gain activity to CDA1(30-150) (Supplementary Fig. S3). Additional seven mutations were tested in combination with W122E to find W133R/Q with further improvement of the activity (Supplementary Fig. S3). CDA1(30-150) containing W122E and W133Q was termed as tCDA1EQ hereafter.

As the engineered deaminase supposedly has less affinity to DNA by itself and might be less stable than the original PmCDA1, nCas9-fusion architecture may have a greater impact on its base editing property. Other than fusing to the nCas9 termini, the deaminase can be inlaid in the middle by splitting nCas9 polypeptide and fusing both termini of the protein to the split site. Structurally, 1054 a.a. position in RuvC domain of Cas9 is on the protein surface with flexibility and close to the non-target DNA strand[18] which is subject to deamination. While N-terminally fused tCDA1EQ showed varying editing efficiency among target sites assessed by CAN1 assay[2], the inlaid version showed consistent editing efficiencies comparable to that of the original Target-AID (Fig. 1d, Supplementary Fig. S4).

To assess non-specific, gRNA-independent off-target effects, we performed a measurement of the occurrence of thialysine-resistance mutants (LYP1 assay)[2] for the engineered versions fused with UGI. Both N-terminal and inlaid tCDA1EQ versions showed significant decreases (5~79 fold) in the mutant occurrences compared to the original Target-AID (Fig. 2a), indicating that their gRNA-independent off-target effects were greatly reduced. We named these N-terminal and inlaid tCDA1EQ versions as AID-2S (Small and Specific) and AID-3S (Small, Specific and Superior), respectively.

Evaluation of AID-2S and AID-3S in mammalian cells

Next, we evaluated the editing efficiency and window of AID-2S and AID-3S in human HEK293T cells and compared them with existing improved cytosine base editors YE1, YE2, and R33A+K34A that were reported to exhibit reduced off-target effects[7]. The well-studied four on-target sites (HEK2, HEK3, RNF2, and VEGFA) were edited by plasmid DNA vector transfection and analyzed by amplicon deep sequencing. Target-AID, AID-2S, and YE1 showed consistent high efficiency for all four target sites tested. AID-3S and YE2 showed middle-to-high efficiencies dependent on target sites. R33A+K34A showed poor efficiency at HEK3 target site (Fig. 1e, Supplementary Fig. S5). Averaged editing window width for AID-2S is narrower than Target-AID and comparable to YE1 and YE2 (Fig. 1f).

The gRNA-independent off-target effects were assessed by using orthogonal SaCas9 R-loop assay[7] in HEK293T cells (Fig. 2b). SaCas9 off-targets site 1~6 were selected following the previous studies[7] and an additional site 7 (VEGFA locus) was chosen as its C-rich context may provide higher sensitivity to deamination by CBEs. Target-AID showed detectable off-target editing at all seven sites (Fig. 2c), while AID2S showed no detectable off-target occurrence in the site 1, 3 and significantly reduced off-target editing at site 2, 5, 6, 7, which was comparable to YE2 and R33A+K34A. YE1 showed rather higher off-target editing at site 6 and 7. AID-3S showed the lowest, hardly detectable off-target editing across all seven sites tested. This may be attributed to the inlaid architecture which sterically limits the access of

the enzyme beyond Cas9-binding DNA strand, in addition to the eliminated DNA affinity. On average, AID-2S and -3S respectively exhibited approximately 4.5-folds and 13.7-folds reduction of R-loop off-target editing compared to the original Target-AID while maintaining appreciable on-target editing efficiency (Fig. 2c, 2d). Combined with yeast LYP1 assay, these consistently support that the genome wide, gRNA-independent off-target effect is greatly mitigated in AID-2S and -3S. We also looked into the gRNA-dependent off-target effect by deep-sequencing of the 6 reported sites (HEK2_OF1, 2; VEGFA_OF1, 2, 3, 4) [19,20] (Fig. 2e, Supplementary Fig. S6). AID-2S and AID-3S, together with YE2 and R33A+K34A showed substantially reduced off-target editing at all the sites analyzed.

Minimization of cytosine base editing system

The engineered PmCDA1 (tCDA1EQ) is substantially smaller (121 a.a.) in size compared to the wild-type (208 a.a.). Smaller molecular size as a genome editing component is advantageous especially for in vivo delivery tools such as AAV vector, which limits the size of DNA length to 4-5kb. Even using the small ortholog SaCas9 system addition of base editing components apparently exceeds the size limit (Fig. 1g). To develop SaAID-3S that contain all necessary base editing components in a loadable size for a single AAV vector, tCDA1EQ was inlaid into 615-616 a.a. position of nSaCas9 within the HNH domain facing to the polynucleotide-binding cleft. Small Scp1 promoter[21] and SpA terminator were also employed to compose a total length of 4036 bp plus 332 bp of gRNA expression cassette. For comparison, the conventional form of SaCas9 version of Target-AID (SaAID) was also developed, which contains full-length PmCDA1 with linker, UGI, CMV promoter, and SV40 terminator to compose a total length of 5220 bp without gRNA cassette. To normalize transfection efficiency that may vary depending on the vector size, the transfected cells were sorted by the fluorescent signal of iRFP670 expressed from the vector backbone. At the two target sites tested, both constructs showed comparable editing efficiency (Fig.1h) with differences in mutation window (Supplementary Fig. S8).

Discussion

In this study, we developed high-fidelity cytosine base editors through structural engineering of the deaminase PmCDA1 to remove the potential non-specific DNA binding moiety. Although the previous studies have explored a series of C-terminal truncations of PmCDA1 to show narrowed editing window and reduced genome-wide off-targets in yeasts[17], simple stepwise removal of the region led to a substantial reduction of its net deaminase activity when measured without UGI. As the UGI-fusion form is so effective in yeast that it may mask evaluation of the net activity of the deaminase, which may not be readily applicable in other organisms. We intentionally omit UGI for strict evaluation of the enzymatic activity in yeast, then add UGI in the following off-target assay and human cell applications. Based on the predicted structure, we deliberately truncated both N- and C- termini of PmCDA1 to cleanly eliminate the DNA-binding domain and to minimize the protein section. The amino acid substitutions to lessen hydrophobicity at the exposed surface further recovered the activity, probably due to improved protein stability or folding. The obtained tCDA1EQ version demonstrated comparable on-target activity and greatly reduced off-target effect, especially in the AID-3S inlaid form. Cas9-inlaid architectures for base

editing had been explored in CBE to expand editing window[18] and in ABE to reduce RNA off-target[14]. Consistently, AID-3S showed a wider editing window and further less R-loop off-target effect than AID-2S. As AID-2S also outperformed pre-existing base editors in on/off-target profile, its narrower editing window should be useful for precise on-target editing.

YE1 was initially developed to narrow editing window[22] then revealed to have reduced off-target effects for both DNA and RNA[23]. While YE2 and R33A+K34A elicited further lowered off-target effects, they performed less robust on-target editing[7]. In this study, YE1 showed relatively high off-target editing, especially at the R-loop site 6 and 7. This might be due to the motif preference of APOBEC1 to 5' TC motif[1,5], which was clearly observed at the both R-loop off-targets (Supplementary Fig. S6) . YE2 and R33A+K34A also showed the same trend but much less extent, probably due to weakened substrate binding capacity. In contrast to APOBEC1 and other APOBEC family proteins, PmCDA1 apparently did not show such strong motif preferences nor RNA editing[2,24,25]. Cas9-gRNA dependent off-target effect was also shown to be significantly reduced in AID-2S and -3S. Possibly, the DNA affinity provided by deaminase may cooperate with the off-target binding of Cas9-gRNA and subsequent editing.

Minimal off-target effects and robust on-target editing are expected to have a wide range of applications from plant and microbial breeding to clinical use. The AID-3S concept has also been demonstrated with the SaCas9 ortholog, providing the smallest base editing system of its size that can be carried in a single AAV vector, facilitating its application in safer gene therapy.

Declarations

Acknowledgments:

This work was supported by the New Energy and Industrial Technology Development Organization (NEDO). This work was also partly supported by the Japan Agency for Medical Research and Development (AMED), "Program on Open Innovation Platform with Enterprises, Research Institute and Academia (OPERA)"and Program for Building Regional Innovation Ecosystems from the Ministry of Education, Culture, Sports and Technology (MEXT) of Japan and the JSPS KAKENHI (grant number 26119710, 16K14654).

Data availability:

High-throughput sequencing and whole-genome sequencing data are deposited in the NCBI Sequence Read Archive (PRJNA721271). Nucleotide sequences of the DNA constructs developed in this study are provided in the Supplementary Sequences.

Author contributions:

A.L. conceived of and designed experiments with input from H.M., S.Y., T.K. and K.N.. A.L. performed plasmid construction, yeast assay, human cell assay and NGS with assistance from H.M. and S.Y.. A.L.

performed cell sorting with assistance from T.K. and S.Y.. K.N. and A.K. supervised the entire project. H.M., and K.N. wrote the manuscript and all authors edited the manuscript.

Competing financial interests:

Patents have been filed related to this work.

References

1. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).
2. Nishida, K. *et al.* Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science (80-.).* **353**, aaf8729 (2016).
3. Jin, S. *et al.* Cytosine, but not adenine, base editors induce genome-wide off-target mutations in rice. *Science (80-.).* **364**, (2019).
4. McGrath, E. *et al.* Targeting specificity of APOBEC-based cytosine base editor in human iPSCs determined by whole genome sequencing. *Nat. Commun.* **10**, (2019).
5. Grünewald, J. *et al.* Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors. *Nature* **569**, (2019).
6. Park, S. & Beal, P. A. Off-Target Editing by CRISPR-Guided DNA Base Editors. *Biochemistry* **58**, 3727–3734 (2019).
7. Doman, J. L., Raguram, A., Newby, G. A. & Liu, D. R. Evaluation and minimization of Cas9-independent off-target DNA editing by cytosine base editors. *Nat. Biotechnol.* **38**, (2020).
8. Yu, Y. *et al.* Cytosine base editors with minimized unguided DNA and RNA off-target events and high on-target activity. *Nat. Commun.* **11**, (2020).
9. Zuo, E. *et al.* Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos. *Science (80-.).* **364**, (2019).
10. Banno, S., Nishida, K., Arazoe, T., Mitsunobu, H. & Kondo, A. Deaminase-mediated multiplex genome editing in Escherichia coli. *Nat. Microbiol.* **3**, (2018).
11. Wang, D., Tai, P. W. L. & Gao, G. Adeno-associated virus vector as a platform for gene therapy delivery. *Nature Reviews Drug Discovery* **18**, (2019).
12. Nishimasu, H. *et al.* Crystal Structure of Staphylococcus aureus Cas9. *Cell* **162**, (2015).
13. Kumar, N. *et al.* The development of an AAV-based crispr sacas9 genome editing system that can be delivered to neurons in vivo and regulated via doxycycline and cre-recombinase. *Front. Mol. Neurosci.* **11**, (2018).
14. Nguyen Tran, M. T. *et al.* Engineering domain-inlaid SaCas9 adenine base editors with reduced RNA off-targets and increased on-target DNA editing. *Nat. Commun.* **11**, (2020).

15. Chen, Y. *et al.* Development of Highly Efficient Dual-AAV Split Adenosine Base Editor for In Vivo Gene Therapy. *Small Methods* **4**, (2020).
16. Qiao, Q. *et al.* AID Recognizes Structured DNA for Class Switch Recombination. *Mol. Cell* (2017). doi:10.1016/j.molcel.2017.06.034
17. Tan, J., Zhang, F., Karcher, D. & Bock, R. Expanding the genome-targeting scope and the site selectivity of high-precision base editors. *Nat. Commun.* **11**, (2020).
18. Wang, Y., Zhou, L., Liu, N. & Yao, S. BE-PIGS: A base-editing tool with deaminases inlaid into cas9 pi domain significantly expanded the editing scope. *Signal Transduction and Targeted Therapy* **4**, (2019).
19. Tsai, S. Q. *et al.* GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* **33**, (2015).
20. Fu, Y. *et al.* High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* **31**, (2013).
21. Juven-Gershon, T., Cheng, S. & Kadonaga, J. T. Rational design of a super core promoter that enhances gene expression. *Nat. Methods* **3**, (2006).
22. Kim, Y. B. *et al.* Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nat. Biotechnol.* (2017). doi:10.1038/nbt.3803
23. Zuo, E. *et al.* A rationally engineered cytosine base editor retains high on-target activity while reducing both DNA and RNA off-target effects. *Nat. Methods* **17**, (2020).
24. Lada, A. G. *et al.* Genome-Wide Mutation Avalanches Induced in Diploid Yeast Cells by a Base Analog or an APOBEC Deaminase. *PLoS Genet.* **9**, e1003736 (2013).
25. Grünewald, J. *et al.* CRISPR DNA base editors with reduced RNA off-target and self-editing activities. *Nat. Biotechnol.* **37**, (2019).

Figures

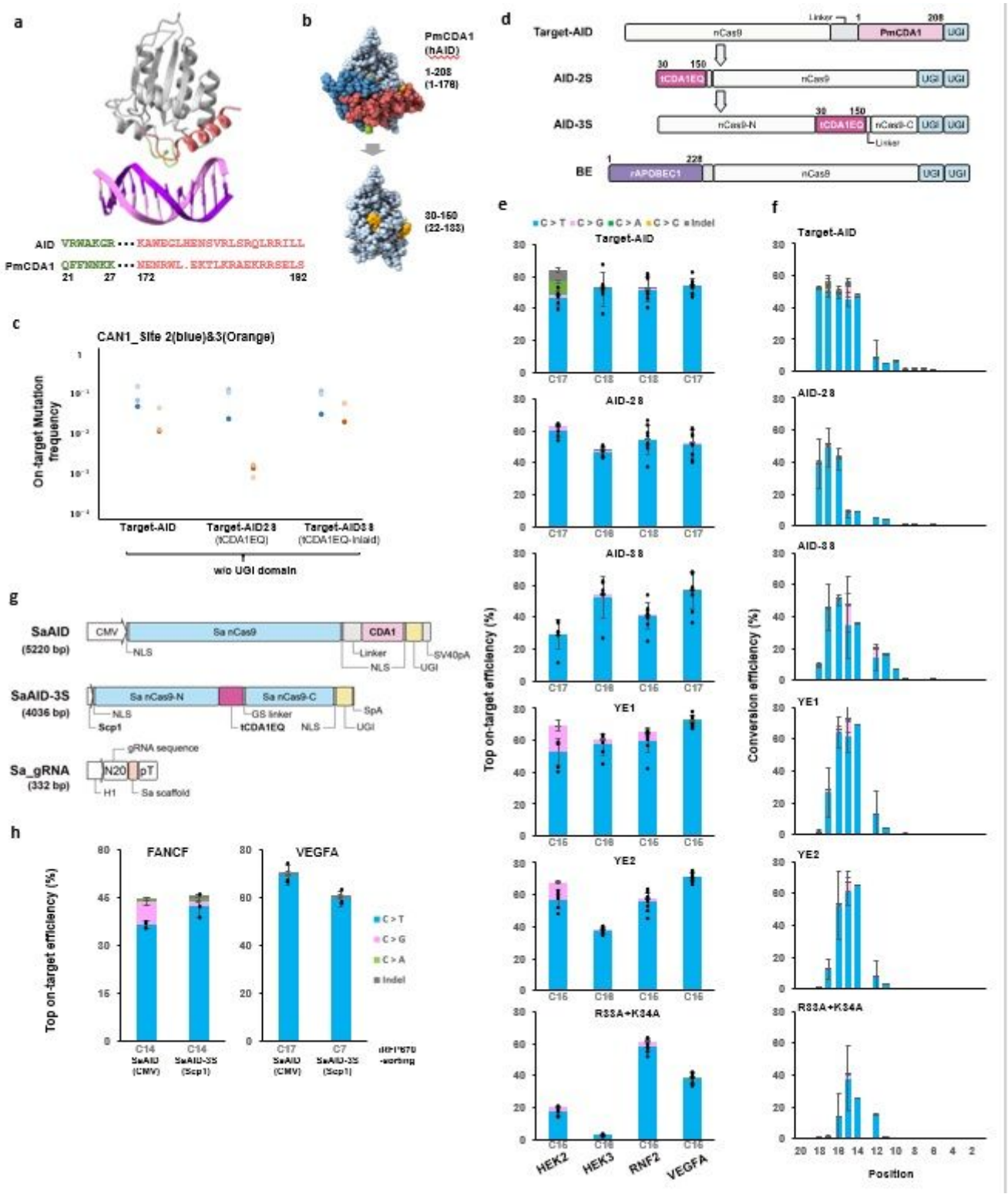

Figure 1

Rational engineering of smaller and specific Target-AID. a, Ribbon model of the structure of a complex of human AID with dsDNA. The non-catalytic dsDNA binding domain is shown in green (N-terminus) and red (C-terminus), of which amino acid sequences are aligned with that of PmCDA1 at the bottom. b, The predicted space-filling structure of PmCDA1 before and after engineering. In addition to the direct DNA-binding sites (green and red), segments shown in blue were trimmed to minimize the protein section. The

mutated amino acids (W122 and W139) are marked in yellow. c, On-target editing efficiencies for Target-AID, AID-2S, and AID-3S without UGI in yeast canavanine-resistance assay. CAN1-2 (blue dots) and CAN1-3 (orange dots) were selected as the target sites and the biological triplicates were plotted. d, Domain arrangements of CBE variants used in this study. BE architecture is common to YE1, YE2 and R33A+K34A, except for the point mutations in rAPOBEC1. e and f, On-target editing profiles of the CBE variants analyzed by deep sequencing at HEK2, HEK3, RNF2, and VEGFA sites in HEK293T. In e, the nucleotide positions (numbered from the PAM sequence side of the target sequence toward the 5’ side) with the highest C to T conversion frequencies for each target are shown. The mutation frequencies of each nucleotide position are also shown in Supplementary Fig 5. For f, the averaged editing windows for the four targets are shown. For e, f, and h, the mean scores (square bars) with standard deviations (error bars) are shown and each biological replicate is plotted with a dot if n<9. g, Domain architectures of SaAID and SaAID-3S. The gRNA expression cassette was combined into each effector plasmid. h, On-target editing frequencies of SaAID and SaAID-3S in HEK293T are shown as in e. To normalize transfection efficiencies, cells were sorted by the expression of iRFP670 from the plasmid backbone. The mutation frequencies of each nucleotide position are also shown in Supplementary Fig. 8.

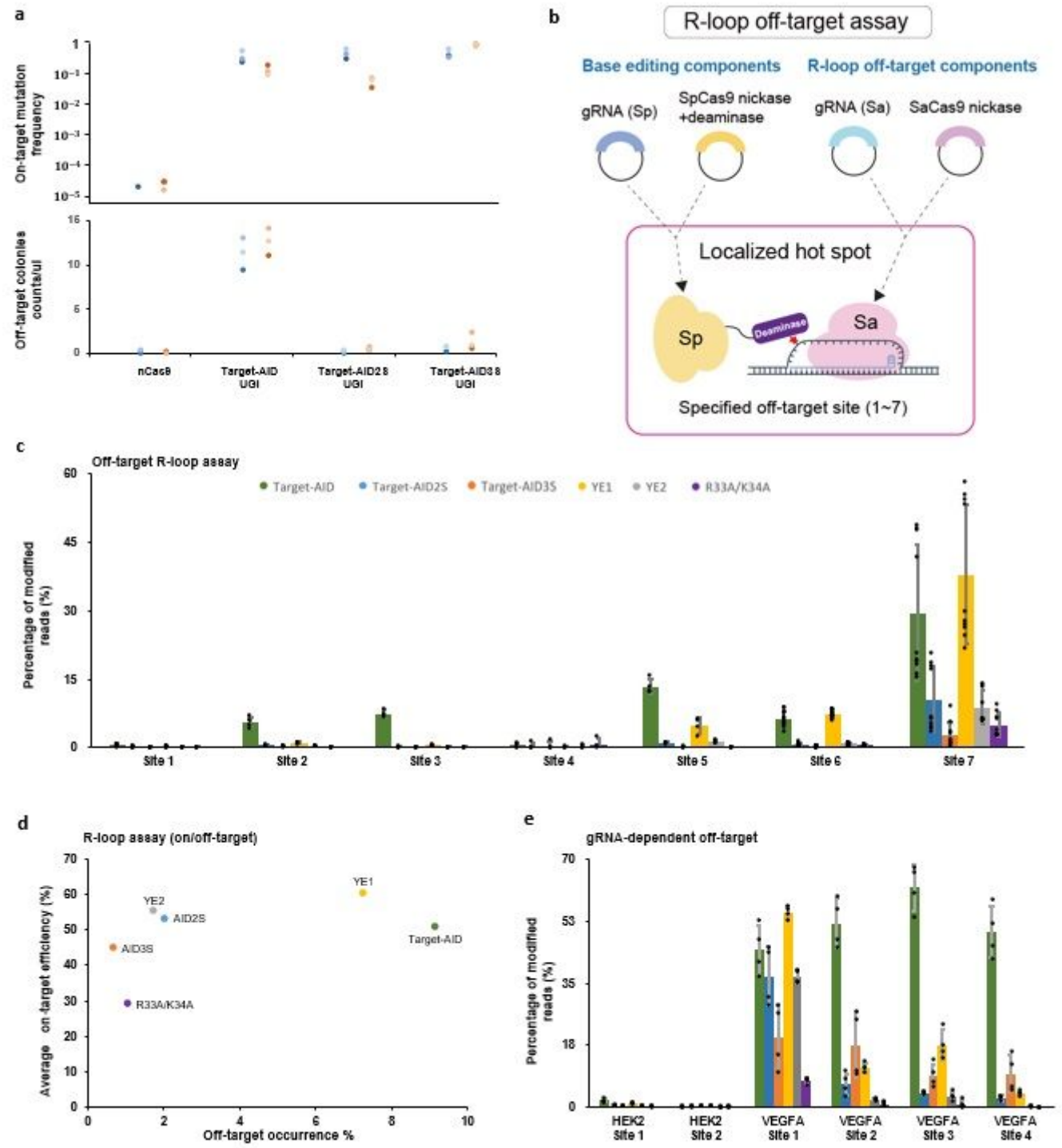

Figure 2

Off-target assessment of AID-2S, -3S and rAPOBEC1 base editors. a, Occurrences of on-target mutation (canavanine resistance) and off-target mutation (thialysine resistance) were measured after induction of each construct as indicated in yeast. Biological triplicates were plotted for CAN1-2 (blue dots) and CAN1-3 (orange dots) target sites.b, Schematic illustration of orthogonal R-loop off-target assessment. Local R-loop formed by nickase SaCas9 serves as a hot spot for off-target events by SpCas9-based base editors.

c, Seven off-target R-loop sites (1~7) were selected, co-transfected with one of on-target sites (HEK2, HEK3, RNF2, VEGFA) and analyzed by deep sequencing. Off-target frequencies are shown as the percentage of reads containing the mutation. Data sets of n=6, n=4, n=12 and n=10 for site 1, 2~5, 6 and 7, respectively, are plotted and their mean frequencies (square bars) and standard deviations (error bars) are shown. The mutation frequencies at each nucleotide position are also shown in Supplementary Fig. 6. d, On-target editing versus average off-target editing profile for all CBEs in this study. The y axis represents the mean on-target editing for the four on-target sites (HEK2, HEK3, RNF2, VEGFA) used in the R-loop assay, and the x axis represents the mean off-target editing (mean of modified reads in percentage) for the seven orthogonal R-loop sites. e, Evaluation of Cas9-dependent off-target effects. The two HEK2 off-target sites (1~2) and four VEGFA off-target sites (1~4) were analyzed by deep sequencing and shown as in e. The data sets maintain n=4. The mutation frequencies at each nucleotide position are also shown in Supplementary Fig. 7.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- AL0412Supplementarysequenceandtable.docx
- AL0412OLMethod.docx
- AL0412Supplementaryfigures.docx