Supplementary Information, Data S1, Supplementary Text and Materials and Methods

Supplementary Text

To understand how different C-terminal tags affect the editing efficiency of Cas9, we investigated the conformations of two linkers (SRAD and GGSGP), which have drastically different effects on Cas9 (Figure 1B), using database analysis and molecular dynamics (MD) simulation [1, 2]. Through statistical analysis of all available structures containing the sequences of GGSGP and SRAD (see below, Materials and Methods), we found that GGSGP is structurally more flexible than SRAD, as indicated by the end-to-end distance (EED) for the first four residues (Supplementary information, Figure S2G). The same trend is reinforced in MD simulations, where the conformational spaces for both linkers together with the last three residues of Cas9 (GGD) were sufficiently sampled (Supplementary information, Figure S2H). Both analyses suggest that the GGSGP linker is capable of adopting various conformations, while the SRAD linker tends to take a locally bent structure, due to a stable electrostatic interaction between its Arg residue and the Asp residue (Supplementary information, Figure S2H).

Since the available Cas9 crystal structures are without any C-terminal tag [3-6], we modeled the structures of Cas9 I and Cas9 II by taking the representative conformations of SRAD and GGSGP from simulation trajectories and the structures of NLS and HA from structural databases (Supplementary information, Figure S2I, S2J and see below, Materials and Methods). According to the modeled structures

(Supplementary information, Figure S2I), the structurally flexible GGSGP linker allows the highly positively charged NLS sequence in Cas9 II to interact favorably with negatively charged nucleic acids, which likely reinforces the interaction between Cas9 and DNA or/and sgRNA and thus enhances the cleavage activity and specificity of Cas9 [3-6](Figure 1B). By contrast, the shorter and locally bent SRAD linker in Cas9 I does not provide sufficient flexibility to facilitate the interaction between the NLS tag and nucleic acids as in Cas9 II (Supplementary information, Figure S2J). Moreover, the presence of the HA sequence at the C-terminus of Cas9 II may further stabilize the formation of the Cas9/sgRNA/DNA ternary complex, resulting in further increase of the Cas9 activity (Figure 1B and Supplementary information, Figure S2I). This model is consistent with the predictions from the RBRIdent program [7], which suggests that the NLS sequence has a strong possibility to bind nucleic acids, while the HA sequence does not appear to interact with nucleic acids.

Materials and Methods

Worm Strains

We used the Bristol N2 strain as the wild type strain. Animals were maintained in standard nematode growth medium (NGM) agar plates at 20℃[8], except for the *ben-1*-editing experiments. For experiments targeting only the *ben-1* gene, all injected P0 animals and their F1 and F2 progeny were maintained on NGM plates containing 14 μM benomyl[9]. For gene editing experiments using the *ben-1*^{153} single guide RNA

(sgRNA) as the co-CRISPR or co-conversion marker[10, 11], only injected P0 animals were maintained on 14 μM benomyl plates.

Plasmid construction.

Cas9 I expression construct (Addgene plasmid #46168), Cas9 II expression construct (pDD162, Addgene plasmid #47549), and pCFJ104 ($P_{myo\text{-}3}$mCherry, Addgene plasmid #19328), which was used as a transgenic marker, were obtained from Addgene. To build the modified sgRNA$^{(F+E)}$ expression vector, PU6::*unc-119* sgRNA$^{(F+E)}$, a DNA fragment containing the *U6* promoter, 20 bp *unc-119* sgRNA guide sequence, the modified sgRNA$^{(F+E)}$ scaffold[12], and the down stream regulatory sequence of the *U6* gene was generated through multiple steps of PCR amplification and used to replace the corresponding DNA fragment in the PU6::*unc-119* sgRNA vector (Addgene plasmid #46169) through its EcoR I and Hind III sites[13]. To construct different sgRNA expression vectors, the 20 bp *unc-119* sgRNA guide sequence in the PU6::*unc-119* sgRNA vector or the PU6::*unc-119* sgRNA$^{(F+E)}$ vector was replaced with different sgRNA guide sequences as described previously[13]. sgRNAs targeting the *ben-1*^{22} and *ben-1*^{153} positions correspond to the sense strand of the *ben-1* gene, whereas sgRNAs targeting the *ben-1*^{1340} and *ben-1*^{1499} positions correspond to the anti-sense strand.

To generate expression vectors for Cas9 III, Cas9 IV and Cas9 VII, a DNA fragment encoding the C-terminal part of Cas9 I along with different C-terminal tags and the *ttb-2* 3'UTR sequence was PCR amplified and used to replace the

corresponding DNA fragment in the Cas9 I expression construct through its Kfl I and Hind III sites. To generate expression vectors for Cas9 V and Cas9 VI, a DNA fragment encoding the C-terminal part of Cas9 II along with different C-terminal tags was PCR amplified and used to replace the corresponding DNA fragment in the Cas9 II expression construct through its Kfl I and Spe I sites. To generate expression vectors for Cas9 VIII, Cas9 X and Cas9 XI, a DNA fragment encoding different N-terminal tags along with the N-terminal part of Cas9 II was PCR amplified and used to replace the corresponding DNA fragment in the Cas9 II expression construct through its Xba I and Msc I sites. To generate the expression vector for Cas9 IX, a DNA fragment encoding the C-terminal part of Cas9 II (without its C-terminal tag) was PCR amplified and used to replace the corresponding DNA fragment in the Cas9 VIII expression construct through its Kfl I and Spe I sites. To generate the expression vector for Cas9 XII, Cas9 XIII and Cas9 XIV, a DNA fragment encoding the C-terminal part of Cas9 II along with various C-terminal tags with different mutations in the HA motif was PCR amplified and used to replace the corresponding DNA fragment in the Cas9 II expression construct through its Kfl I and Spe I sites.

Selection of oligonucleotide repair templates

We normally synthesized oligonucleotide repair templates with 20-50 nt homology arm on both sides of the intended knock-in, which are sufficient to generate high percentages (>79%) of precise knock-ins at the *ben-1*^{153} position (Supplementary information, Figure S1K and S1L). Likewise, we were able to precisely insert a

FLAG tag at different target loci using oligonucleotide repair templates with 25-50 nt homology arm on both sides of the insertion site (Supplementary information, Figure S2C; Data not shown). Moreover, we found that the precise editing efficiency is inversely correlated with the distance between the PAM motif and the knock-in site (Supplementary information, Figure S1M and S1N), suggesting that it is beneficial to design a sgRNA that targets a sequence close to the knock-in site.

Microinjection

All plasmids were purified by QIAprep Spin Miniprep Kit (Qiagen). All repair oligonucleotides were purified by polyacrylamide gel (Invitrogen). Microinjection was performed as described using the standard protocol [14]. The injection mixture used in the *ben-1* gene editing experiments contains 50 ng/µl Cas9 expression vector, 45 ng/µl sgRNA expression vector, 5 ng/µl transgenic reporter pCFJ104, and 20 ng/µl repair oligonucleotide (Supplementary information, Table. S1A). For gene editing experiments using the co-CRISPR method, the *ben-1*^{153} sgRNA was also included in the injection mixture with a final concentration of 45 ng/µl.

Identification of precisely edited animals

ben-1 gene editing experiments [at the 22, 115, 127, 136, 153, 1340 and 1499 positions after the first nucleotide (designated as 1) of the translation initiation codon of the *ben-1* gene) were conducted as described previously with minor modifications [9, 15]. We took advantage of the feature that loss-of-function mutations generated at

the *ben-1* locus are dominant suppressors of the paralyzed or Uncoordinated (Unc) phenotype induced by the benomyl drug treatment to identify edited animals [9]. Briefly, all injected P0 animals and their F1 and F2 progeny were maintained on NGM plates with 14 μM benomyl (Sigma) at 20℃. Homozygous benomyl-resistant non-Unc F2 progeny of non-Unc F1 animals were selected and verified by DNA sequencing.

For gene editing experiments at the *drp-1, ced-9, fis-2* and *ubr-1* loci, the co-CRISPR or co-conversion method was used[10, 11]. The *ben-1*^{153} sgRNA expression vector was used as a co-CRISPR driver and was included in each injection mixture. Injected P0 animals were placed on the benomyl-containing NGM plates and kept at 20℃. 96 or 192 non-Unc F1 animals were cloned out to normal NGM plates and allowed to lay eggs for 1-2 days. The single F1 animals (or their F2 progeny) were lysed and used as PCR templates to screen for knock-in events occurring at the targeted gene loci. In *drp-1, fis-2* and *ubr-1* editing experiments, precise gene editing would create a BssH II site, a Hind III site, and a Hind III site, respectively. We thus identified knock-in animals through a combination of PCR and restriction fragment length polymorphism (RFLP) screens. Genomic DNA spanning each of the above three sgRNA target sites was PCR amplified (Supplementary information, Table S1B) and subjected to restriction enzyme digestion (NEB).

In the *flag::ced-9* editing experiment, a PCR primer specific to the FLAG-encoding-DNA sequence and a primer specific to the *ced-9* promoter were used to identify *flag::ced-9* knock-in animals (Supplementary information, Table

S1B). In all cases, knock-ins were confirmed to be precise by DNA sequencing (Supplementary information, Figures S1C-S1G, S1J, S2A-S2F and S2M).

Extracting structures from the PDB database

We extracted the structures of the protein fragments containing the sequences of GGSGP, SRAD, PKKKRKV (NLS) and YPYDVPDYA (the HA tag) from the protein data bank (PDB) for structural analyses. To achieve this goal, the sequences of all PDB proteins were aligned using BLAST[16, 17], and the pisces server was used to remove redundant proteins that exhibit sequence identity of > 95%[18, 19]. Proteins that have 40 residues or less were also removed, unless otherwise mentioned. The conformations of the identified fragments were then extracted from the retained pdb structures. Specifically, 16, 52 and 5 chains were collected for the fragments of GGSGP, SRAD and the HA tag, respectively. Because all four pdb structures containing the NLS sequence have less than 40 residues, these four fragment structures were retained for further analysis, without any redundancy processing. In order to analyze the conformational difference between GGSGP and SRAD, we calculated the end-to-end distance (EED) for the first four residues, which are defined as the distance between the Cα atoms of the first and the fourth Gly residues in GGSGP or that between the Cα atoms of the Ser and Asp residues in SRAD.

MD Simulations

Since the limited structures available in the database analysis as described above may bias the conformational comparison, we subsequently conducted molecular dynamics (MD) simulations to further invesitigate the distinct conformational preferences for the GGSGP and SRAD fragments. In this analysis, the last three residues of Cas9 (GGD) were also included. Therefore, the final fragments adopted for simulation were GGDGGSGP and GGDSRAD, respectively, and their intial conformations were modeled based on two available crystal structures (PDB IDs of 1Y37 and 3KYI, respectively). Both fragments were blocked by chemical groups at the N- and C-termini to mimic continuous polypeptide chains, were solvated by the explicit TIP3P waters, and were then neutralized by 0.15 mol/L NaCl. All simulations were performed using the NAMD 2.9 package[20] with the CHARMM36 force field[2124]. The electrostatic potential was evaluated by the particle mesh Ewald (PME) method with the grid spacing of less than 1 Å. The cutoff of van der Waals energies was set at 12 Å with a smooth switch at 10 Å. The simulations were conducted under an NPT ensemble with the temperature and pressure held at 310 K and 1 atm, respectively, using the Langevin thermostat and Langevin barostat, respectively. Following 1000 steps of energy minimization, both systems were initially pre-equilibrated with gradually relaxed positional constraints (from 10 to 0.01 kcal/mol) applied on the Cα atoms in 4 ns. The productive simulations were then performed for 100 ns, with a time step of 2 fs. After the simulations, all structural snapshots during last 30 ns simulation were collected for structural analysis. We calculated the end-to-end distance (EED) for the last five residues, which is defined as the distance between the

Cα atoms of the first Gly residue and the fifth Pro residue in GGSGP or the distance between the Cα atom of the Ser residue in SRAD and the C atom of the C-terminal blocker of SRAD, which is equivalent to the Cα atom of the next residue in the continuous polypeptide.

Structural Modeling

Since the C-terminal residues (after Gly1366) were absent in the Cas9 crystal structure (PDB ID: 4OO8)[5], we tried to model the structure of the whole Cas9 protein, including its C-terminal tag, based on the fragment structures obtained from the simulations and the structural database analysis. Specifically, all structural snapshots from the MD simulations were analyzed by the R package and VMD 1.9.1[25, 26] and the representative structures for the highest probability point in histrogram were taken as the templates for the subsequent structural modeling of the fragments of GGDGGSGP and GGDSRAD. The structural templates for the NLS and the HA sequences were taken from the pdb structures (PDB IDs of 4WA1 and 1Q1S, respectively). The final structures were generated using the program MODELLER9v12 program[27] without any loop refining.

There are thus far four Cas9 crystal structures with sgRNA and target DNA[3-6]. To analyze how the C-terminal tags of Cas9 affect the protein function, we generated structural models for the whole Cas9 protein by aligning the backbone of the first residue of the modeled terminus (linker + NLS + HA) to the last residue (residue 1366) in the Cas9 crystal structure (PDB ID: 4OO8). In order to visually illustrate the same

effect on another Cas9 crystal structure (PDB ID: 4UN3), both protein and RNA in this complex structure were superposed to the above crystal structure (PDB ID: 4OO8) by aligning all Cα atoms in the protein. The two Cas9/sgRNA/DNA complexes were shown in Supplementary information, Figure S2I and S2J.

Reference

1 Aneesur R. Correlations in the motion of atoms in liquid argon. *Phys Rev* 1964; **136**:405-411.

2 Loup V, Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Phys Rev* 1967; **159**:98-103.

3 Jinek M, Jiang F, Taylor DW*, et al.* Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* 2014; **343**:1247997.

4 Anders C, Niewoehner O, Duerst A, Jinek M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* 2014; **513**:569-573.

5 Nishimasu H, Ran FA, Hsu PD*, et al.* Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* 2014; **156**:935-949.

6 Jiang F, Zhou K, Ma L, Gressel S, Doudna JA. A Cas9-guide RNA complex preorganized for target DNA recognition. *Science* 2015; **348**:1477-1481.

7 Xiong D, Zeng J, Gong H. RBRIdent: an algorithm for improved identification of RNA-binding residues in proteins from primary sequences. *Proteins* 2015; **83**:1068-1077.

8 Brenner S. The genetics of *Caenorhabditis elegans*. *Genetics* 1974; **77**:71-94.

9 Driscoll M, Dean E, Reilly E, Bergholz E, Chalfie M. Genetic and molecular analysis of a *Caenorhabditis elegans* beta-tubulin that conveys benzimidazole sensitivity. *J Cell Biol* 1989; **109**:2993-3003.

10 Kim H, Ishidate T, Ghanta KS*, et al.* A Co-CRISPR strategy for efficient genome editing in *Caenorhabditis elegans*. *Genetics* 2014; **197**:1069-1080.

11 Arribere JA, Bell RT, Fu BXH*, et al.* Efficient marker-free recovery of custom genetic modifications with CRISPR/Cas9 in *Caenorhabditis elegans*. *Genetics* 2014; **198**:837-846.

12 Chen B, Gilbert LA, Cimini BA*, et al.* Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* 2013; **155**:1479-1491.

13 Friedland AE, Tzur YB, Esvelt KM*, et al.* Heritable genome editing in *C. elegans* via a CRISPR-Cas9 system. *Nat Methods* 2013; **10**:741-743.

14 Mello CC, Kramer JM, Stinchcomb D, Ambros V. Efficient gene transfer in *C.elegans*: extrachromosomal maintenance and integration of transforming

sequences. *Embo J* 1991; **10**:3959-3970.

15 Zhao P, Zhang Z, Ke H, Yue Y, Xue D. Oligonucleotide-based targeted gene editing in *C. elegans* via the CRISPR/Cas9 system. *Cell Res* 2014; **24**:247-250.

16 Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990; **215**:403-410.

17 Mount DW. Using the basic local alignment search tool (BLAST). *CSH Protoc* 2007; **2007**:pdb top17.

18 Wang G, Dunbrack RL Jr. PISCES: recent improvements to a PDB sequence culling server. *Nucleic Acids Res* 2005; **33**:W94-98.

19 Wang G, Dunbrack RL Jr. PISCES: a protein sequence culling server. *Bioinformatics* 2003; **19**:1589-1591.

20 Phillips JC, Braun R, Wang W, *et al.* Scalable molecular dynamics with NAMD. *J Comput Chem* 2005; **26**:1781-1802.

21 Vanommeslaeghe K, Hatcher E, Acharya C, *et al.* CHARMM general force field: a force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J Comput Chem* 2010; **31**:671-690.

22 MacKerell AD Jr., Feig M, Brooks CL, 3rd. Improved treatment of the protein backbone in empirical force fields. *J Am Chem Soc* 2004; **126**:698-699.

23 MacKerell AD, Bashford D, Bellott M, *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* 1998; **102**:3586-3616.

24 Best RB, Zhu X, Shim J, *et al.* Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backboneφ, ψ and side-chain χ_1 and χ_2 dihedral angles. *J Chem Theory Comput* 2012; **8**:3257-3273.

25 Hsin J, Arkhipov A, Yin Y, Stone JE, Schulten K. Using VMD: an introductory tutorial. *Curr Protoc Bioinformatics* 2008; **Chapter 5**:Unit 5.7.

26 Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graph* 1996; **14**:33-38, 27-38.

27 Eswar N, Webb B, Marti-Renom MA, *et al.* Comparative protein structure modeling using MODELLER. *Curr Protoc Protein Sci* 2007; **Chapter 2**:Unit 2.9.