Electronic Supplementary Information

Optimized TAL effector nucleases (TALENs) for use in treatment of sickle cell disease

Ning Sun, Jing Liang, Zhanar Abil and Huimin Zhao

Experimental

Construction of TALEN expression vectors

The DNA sequence encoding the nonspecific DNA cleavage domain of FokI restriction endonuclease was PCR-amplified from pST1374 (plasmid 13426; Addgene, Cambridge, MA). The PCR product was inserted into pET28a(+) vector (EMD, Madison, WI) through *Nde*I and *Sal*I recognition sites to yield plasmid pET28a-G4S-FokI. The DNA sequence encoding the central repeat DNA binding domain of TALE AvrXa10 was isolated from 0.7% agarose gel after digestion of the plasmid pAS202 (kindly provided by Dr. Frank F. White from Kansas State University, Manhattan, KS) using restriction enzymes *Stu*I and *Aat*II. The DNA fragment after gel extraction was cloned into the *Stu*I and *Aat*II sites of plasmid pET28a-G4S-FokI to make pET28a-TALEN-Xa10. The DNA sequences encoding the N-terminal extensions of the AvrXa10 protein with varying length were PCR-amplified from pAS202. The PCR products were digested with *Nde*I and *Stu*I and cloned into the large fragment of the pET28a-TALEN-Xa10 vector digested with the same enzymes. The DNA sequences encoding the C-terminal extensions of the AvrXa10 protein with varying length were PCR-amplified using the same template and cloned into the *Aat*II and *Afl*II sites of the plasmid pET28a-TALEN-Xa10. The gene encoding TALEN-Xa10 with different configurations were subcloned into pRS415 plasmid through the *Nde*I and *Sal*I sites in between the GAL10 promoter and the ADH2 terminator. The expression of TALENs in *Saccharomyces cerevisiae* was induced by adding 2% galactose in the cultural medium.

The DNA sequences encoding the custom-designed TALE central repeat domains targeting the specific sequence within the HBB gene locus were synthesized by DNA2.0, Inc (Menlo Park, CA). The synthesized DNA sequences were codon optimized to increase the protein expression level and reduce the repetitiveness among amino acid repeats in the DNA binding domain to facilitate DNA cloning. The genes of custom-designed TALENs with different N- and C-terminal extensions were first cloned into the pET28a(+) vector using the same strategy as described above. Next, the TALEN genes along with a FLAG tag sequence and a SV40 nucleus localization signal added to the N-terminus were subcloned into the pCMV5 mammalian expression vector[1] through the *Kpn*I and *Sal*I sites. The expression of these constructs are under the control of the cytomegalovirus (CMV) promoter.

References

1. S. Andersson, D. L. Davis, H. Dahlback, H. Jornvall and D. W. Russell, *J Biol Chem*, 1989, **264**, 8222-8229.

Fig. S1. Activity of TALEN-HBB-L variants as a homodimer in a mammanlian gene targeting system. (A) Sequence of the HBB-L palindromic site with two HBB-L sites (underlined) in a tail-to-tail orientation separated by a 15 bp spacer. (B) Representative flow cytometry plots of gene targeting. GFP-positive cells were quantitated in gate as depicted. Autofluorescence was measured using phycoerythrin-fluorescence channel. Hela cells with an EGFP reporter gene integrated into the genome were co-transfected with the expression vectors of TALEN-HBB-L variants and an EGFP donor as the DNA repair template. Empty vector pCMV5 served as negative control. The values represent the average (± standard deviation) of gene targeting rates from at least three independent experiments.

A HBB-L palindromic site

5′-TGTAGCAACCTCAAACAGACACCATACTGATCTAGGACTAATGGTGTCTGTTTGAGGTTGCTAGT

3′-ACATCGTTGGAGTTTGTCTGTGGTATGACTAGATCCTGATTACCACAGACAAACTCCAACGATCA

B

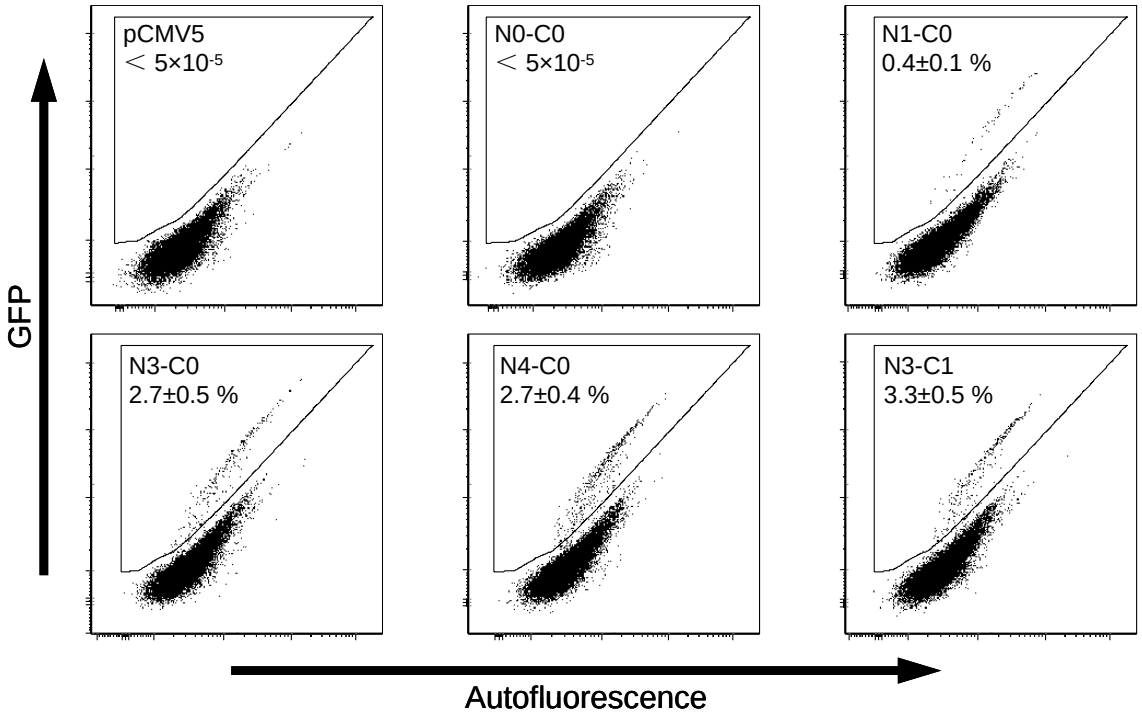

Fig. S2. Activity of TALEN-HBB-R variants as a homodimer in a mammanlian gene targeting system. (A) Sequence of the HBB-R palindromic site with two HBB-R sites (underlined) in a tail-to-tail orientation separated by a 15 bp spacer. (B) Representative flow cytometry plots of gene targeting. GFP-positive cells were quantitated in gate as depicted. Autofluorescence was measured using phycoerythrin-fluorescence channel. Hela cells with an EGFP reporter gene integrated into the genome were co-transfected with the expression vectors of TALEN-HBB-R variants and an EGFP donor as the DNA repair template. Empty vector pCMV5 served as negative control. The values represent the average (± standard deviation) of gene targeting rates at least three independent experiments.

A HBB-R palindromic site

5′-CCCTGTAACGGCAGACTTCTCCACAACTGATCTAGGACTATGTGGAGAAGTCTGCCGTTAGTGAA

3′-GGGACATTGCCGTCTGAAGAGGTGTTGACTAGATCCTGATACACCTCTTCAGACGGCAATCACTT

B

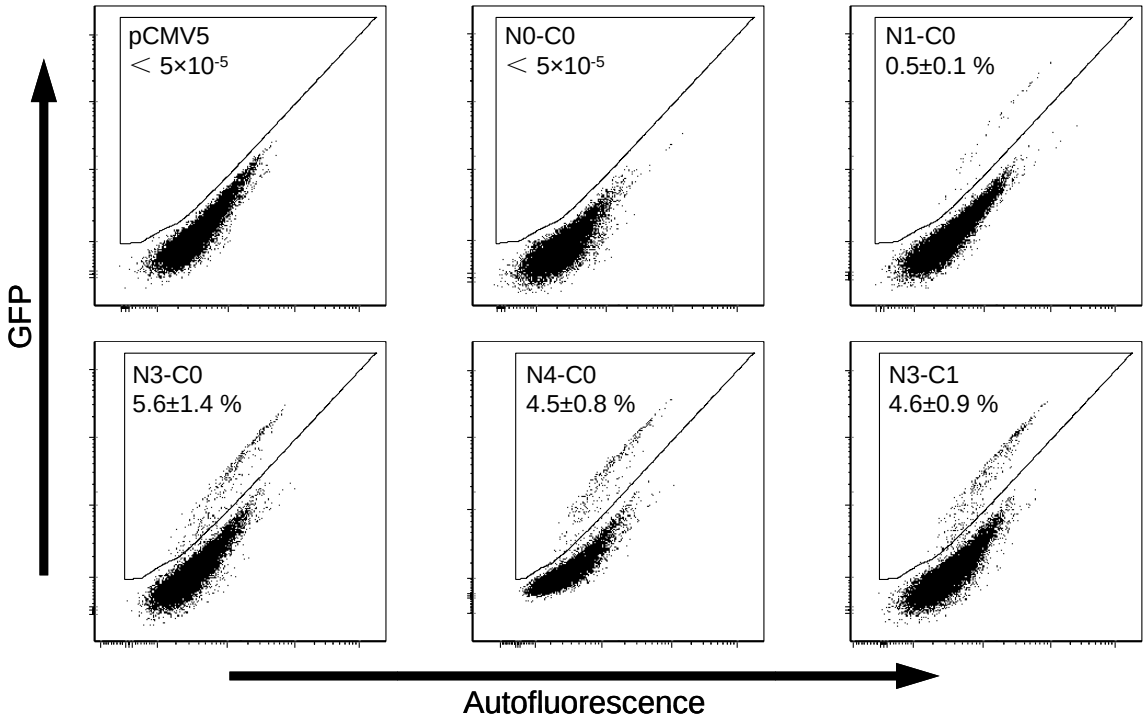

Fig. S3. Flow cytometric analysis of cellular γ-H2AX expression. (**A**) Representative flow cytometry histograms for etoposide treated cells with indicated concentrations. A gate was drawn (dashed line) to determine the percentage of γ-H2AX staining positive cells. (**B**) Representative flow cytometry histograms for cells transfected with nuclease-expression vectors as indicated.

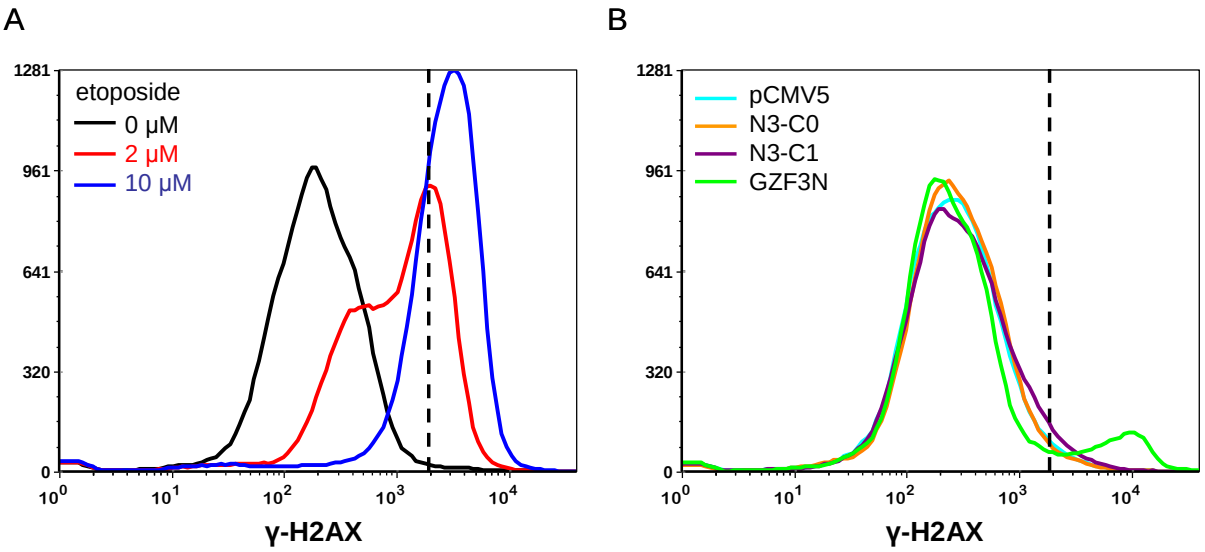

Table S1. Relative β-GAL activity of TALEN-AvrXa10 variants bearing N- and C-terminal extensions against target sites with the indicated spacers. The activity of I-CreI was normalized to 1. Values represent the average (± standard deviation) of at least three independent experiments. "EBE" indicates that the cleavage target contains just one monomer binding site.

	EBE	**6bp**	**8bp**	**10bp**	**12bp**	**14bp**	**15bp**	**16bp**	**24bp**	**32bp**
N0-C0	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
N1-C0	ND	ND	ND	1.7 ± 0.4	2.7 ± 0.7	2.3 ± 0.2	2.4 ± 0.4	2.6 ± 0.2	0.2 ± 0.0	ND
N2-C0	ND	ND	ND	2.1 ± 0.2	2.8 ± 0.1	2.5 ± 0.1	2.3 ± 0.6	3.2 ± 0.2	0.6 ± 0.1	0.2 ± 0.0
N3-C0	ND	ND	ND	2.7 ± 0.7	2.6 ± 0.7	2.2 ± 0.4	3.0 ± 0.4	3.1 ± 0.4	0.9 ± 0.2	0.3 ± 0.1
N4-C0	ND	ND	ND	2.4 ± 0.2	3.0 ± 0.1	1.9 ± 0.4	2.7 ± 0.3	2.2 ± 0.5	0.4 ± 0.2	0.2 ± 0.0
N3-C1	ND	ND	ND	ND	0.1 ± 0.0	1.6 ± 0.3	2.9 ± 0.7	3.1 ± 0.8	3.1 ± 0.5	3.1 ± 0.3
N3-C2	ND	ND	ND	ND	ND	0.5 ± 0.1	0.7 ± 0.5	2.1 ± 0.8	2.3 ± 0.5	1.9 ± 0.2
N3-C3	ND	ND	ND	ND	ND	ND	0.1 ± 0.0	0.8 ± 0.2	0.6 ± 0.1	0.3 ± 0.1
N3-C4	ND	ND	ND	ND	ND	ND	0.1 ± 0.0	1.5 ± 0.4	0.6 ± 0.0	0.4 ± 0.0
N3-C5	ND	ND	ND	ND	ND	0.1 ± 0.0	0.2 ± 0.0	0.5 ± 0.4	1.6 ± 0.1	1.5 ± 0.6

ND: no detectable activity