

ARTICLE

<https://doi.org/10.1038/s41467-019-10828-3>

OPEN

An anionic human protein mediates cationic liposome delivery of genome editing proteins into mammalian cells

Y. Bill Kim^{1,2,3,6}, Kevin T. Zhao^{1,2,3,6}, David B. Thompson^{4,5} & David R. Liu ^{1,2,3}

Delivery into mammalian cells remains a significant challenge for many applications of proteins as research tools and therapeutics. We recently reported that the fusion of cargo proteins to a supernegatively charged (-30)GFP enhances encapsulation by cationic lipids and delivery into mammalian cells. To discover polyanionic proteins with optimal delivery properties, we evaluate negatively charged natural human proteins for their ability to deliver proteins into cultured mammalian cells and human primary fibroblasts. Here we discover that ProT α , a small, widely expressed, intrinsically disordered human protein, enables up to -10-fold more efficient cationic lipid-mediated protein delivery compared to (-30)GFP. ProT α enables efficient delivery at low- to mid-nM concentrations of two unrelated genome editing proteins, Cre recombinase and zinc-finger nucleases, under conditions in which (-30)GFP fusion or cationic lipid alone does not result in substantial activity. ProT α may enable mammalian cell protein delivery applications when delivery potency is limiting.

¹Merkin Institute of Transformative Technologies in Healthcare, Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA. ²Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138, USA. ³Howard Hughes Medical Institute, Harvard University, Cambridge, MA 02138, USA. ⁴Department of Genetics, Harvard Medical School, Boston, MA 02115, USA. ⁵Wyss Institute for Biologically Inspired Engineering, Boston, MA 02115, USA. ⁶These authors contributed equally: Y. Bill Kim, Kevin T. Zhao. Correspondence and requests for materials should be addressed to D.R.L. (email: drliu@fas.harvard.edu)

Proteins including genome editing agents represent an increasing proportion of biomedical research tools and new human therapeutics¹. Due to their inability to spontaneously cross the lipid bilayer, however, current uses of protein tools and therapeutic agents mostly target extracellular components². Technologies to facilitate cytosolic access by proteins are critical to expanding the potential targets that can be accessed efficiently by exogenous proteins.

Researchers have developed many approaches to facilitate protein translocation into the cytosol. Positively charged cell-penetrating peptides, such as the HIV-1 transactivator of transcription (Tat) peptide^{3,4}, poly-arginine⁵, or superpositively charged proteins^{2,6,7}, allow a fused cargo protein to reach the cytosol via interaction with the highly anionic proteoglycans on the cell surface, followed by endocytosis⁸. However, escape from endosomes is generally inefficient, and unencapsulated proteins are susceptible to proteolytic degradation and neutralization by serum proteins and the extracellular matrix⁹.

Lipid nanoparticles can encapsulate proteins to prevent degradation and neutralization by antibodies¹⁰. While cationic lipids are routinely used to transfect polyanionic nucleic acids into mammalian cells, they have not been widely used for protein delivery, as proteins vary widely in their ability to be encapsulated by cationic lipids¹¹. To impart nucleic-acid-like properties onto proteins, we previously demonstrated that fusing an anionic protein, such as an engineered (−30)GFP with many surface-exposed negative charges, to a protein of interest enables its efficient encapsulation and delivery into mammalian cells¹².

While (−30)GFP enables cationic lipids to deliver a variety of fused cargo proteins¹², its discovery from protein engineering efforts unrelated to delivery¹³ suggests that more potent anionic proteins that mediate lipid-based protein delivery may exist. Moreover, the non-mammalian origin of (−30)GFP will likely result in immunogenicity, potentially compromising the safety or efficacy of using this protein *in vivo*¹⁴.

To address these limitations, we sought to discover an improved protein fusion partner for lipid-mediated protein delivery from the human proteome. We hypothesize that the human proteome, which contains many proteins with high theoretical net charge, likely includes some native highly anionic proteins that would mediate efficient lipid-based delivery of fused cargo proteins. Here we identify ProTα as a small (111-residue) supernegatively charged protein that enables potent delivery of fused cargo proteins into mammalian cells.

Results

Identification of highly anionic human proteins. We generated a candidate list of human proteins with >0.75 net theoretical negative charges per kDa from the UniProt protein database (Supplementary Fig. 1)¹⁵. To focus on proteins that are more likely to be easily expressed and purified as fusion partners with a wide range of potential cargo proteins, we narrowed down the list to proteins with published structures, known bacterial expression, and no annotated disulfide bonds. Based on these criteria, we identified 12 candidate proteins for further analysis (Table 1).

The genes encoding the candidate human proteins were cloned as N-terminal fusions to Cre recombinase, overexpressed in *E. coli*, and purified by affinity chromatography. Each purified fusion protein was encapsulated with Lipofectamine RNAiMAX, a commercially available lipid formulation we previously used to deliver anionic proteins and protein complexes¹². We used HeLa cells containing a genomically integrated DsRed gene preceded by a floxed transcriptional terminator (HeLa-DsRed)⁶ to screen the fusion proteins for Cre delivery activity. Functional Cre delivery catalyzes recombination to remove the terminator, resulting in

Table 1 Protein delivery using Lipofectamine RNAiMAX in HeLa-DsRed cell line

Protein fused to Cre	Concentration at which 50% of cells are recombined (nM)
ProTα	1.1 ± 0.18
Polyadenylate-binding protein-interacting protein 2	290 ± 60
Troponin C	57 ± 19
DPH3 homolog	36 ± 5
RNA Polymerase II Subunit F	110 ± 58
Multiple coagulation factor deficiency protein 2	140 ± 170
ADP-ribosylation factor-like protein 2-binding protein	20 ± 9
NF-kappa-B inhibitor alpha	63 ± 7
DNA damage-inducible transcript 3 protein	89 ± 5
Carbonic anhydrase VIII	84 ± 22
Protein S100-B	47 ± 60
Sirtuin-1	49 ± 18
(−30)GFP	11 ± 3.0
No fusion	73 ± 12

12 human proteins were chosen as candidates for lipid nanoparticle delivery. EC₅₀ is defined as the concentration at which 50 % of the cells contain red fluorescence

DsRed expression and red fluorescence. We defined Cre delivery potency as the protein concentration at which 50% of analyzed cells contain red fluorescence (EC₅₀). While unfused Cre with RNAiMAX lipid induced red fluorescence in our assay at an EC₅₀ of 73 ± 12 nM, fusion of (−30)GFP to Cre lowered the EC₅₀ with RNAiMAX to 11 ± 3.0 nM (Table 1), consistent with our previous findings¹².

Seven of the 12 anionic human proteins tested from the set of 12 candidates did not substantially improve the efficacy of protein delivery when fused to the N-terminus of Cre recombinase compared to that of Cre alone (Table 1). Their failure to enhance functional delivery of Cre may be due to an inability to support lipid encapsulation, or due to interference with recombinase activity when fused to Cre. Four human proteins showed compatibility with lipid complexation and enhanced Cre delivery compared to that of unfused Cre, but their delivery potency did not exceed that of (−30)GFP (Table 1). One protein, human prothymosin alpha (ProTα), a small, intrinsically disordered protein¹⁶ with 111 amino acids and a net theoretical charge of −44, greatly improved the potency of Cre protein delivery into HeLa-DsRed cells 10-fold compared to that of (−30)GFP and 73-fold compared to Cre alone, to an EC₅₀ of 1.1 ± 0.18 nM (Table 1 and Fig. 1).

ProTα is universally expressed in human tissues and contains multiple domains that are hypothesized to be modular in function¹⁶. Researchers have associated multiple biological functions with ProTα, including histone chaperone activity implicated in nucleosome exchange and extracellular immune modulation¹⁶. A stretch of highly anionic residues in the central part of ProTα was demonstrated to interact with cationic regions of histones¹⁷.

To verify the highly potent lipid encapsulation and delivery capabilities of ProTα, we delivered ProTα-Cre using lipid nanoparticles into BSR-tdTomato and HEK293-RFP cells, two additional human cell lines containing genomically integrated RFP reporters based on BSR and HEK293² cells. Consistent with the delivery improvement observed in HeLa-DsRed cells, ProTα-Cre also resulted in a large, ~10-fold improvement in delivery potency compared to (−30)GFP-Cre in BSR-tdTomato and HEK293-RFP cells (Fig. 1 and Supplementary Fig. 2). We visually

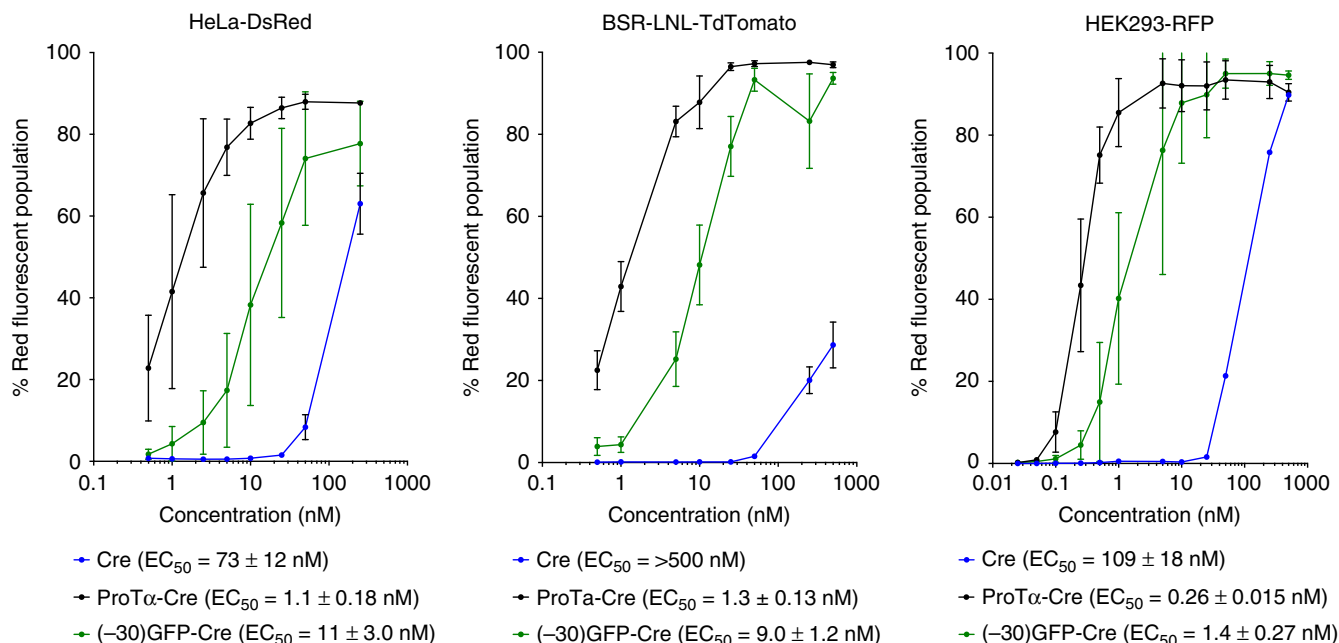


Fig. 1 ProT α enables efficient cationic lipid-mediated protein delivery into HeLa-DsRed, BSR-tdTomato, and HEK293-RFP cell lines. EC_{50} is defined as the protein concentration at which 50% of analyzed cells contain red fluorescence. In all cell lines tested, ProT α delivers Cre at low nM concentrations, -10-fold more potent than (-30)GFP and -100-fold more efficient than unmodified Cre. Values and error bars represent the mean and standard deviation of three independent biological replicates performed on different days. EC_{50} values were determined by nonlinear fit to the Hill equation, and are followed by the standard errors of the fit. Source data are available in the Source Data file

confirmed Cre-mediated fluorescent activation via microscopy to be consistent with these quantified changes in delivery potency (Supplementary Fig. 3). We did not observe any apparent cellular toxicity during ProT α -mediated delivery (Supplementary Fig. 4).

To complement the above Cre recombinase functional delivery experiments, we also quantified the delivery of fluorescent proteins by ProT α fusion. We expressed and purified mCherry, (-30)GFP-mCherry, and ProT α -mCherry, performed cationic lipid-mediated protein delivery into HEK293T cells across a range of protein concentrations, and analyzed the total amount of fluorescent protein in cells by flow cytometry. Compared to the delivery of mCherry alone, delivery of (-30)GFP-mCherry and ProT α -mCherry resulted in more potent delivery of mCherry (Supplementary Fig. 5). To test if ProT α -tagging allows protein delivery into primary cells, many of which suffer from poor liposome-mediated transfection efficiency¹⁸, we performed cationic lipid-mediated protein delivery using mCherry, (-30)GFP-mCherry, and ProT α -mCherry into human primary fibroblasts. Without any fusion, lipid-mediated mCherry protein delivery exceeded 10% only at concentrations ≥ 2.5 μ M (Supplementary Fig. 6). In contrast, lipid-mediated ProT α -mCherry and (-30)GFP-mCherry delivery was efficient at all concentrations tested (as low as 50 nM; see Supplementary Fig. 6). These results demonstrate that ProT α greatly facilitates cationic lipid-mediated protein delivery even in human primary cells that resist lipofection. In addition, these findings are consistent with a model in which both (-30)GFP and ProT α enable similarly efficient complexation and initial endocytosis into the cell, but ProT α allows proteins such as Cre to more efficiently reach its substrate (in the nucleus) in functional form than (-30)GFP.

Structure function analysis of ProT α . ProT α contains a stretch of aspartate and glutamate residues that provide a concentrated region of anionic charge¹⁶. To investigate if a simple stretch of acidic residues of similar negative charge magnitude is sufficient to confer potent delivery with cationic liposome delivery, we

generated two protein constructs that have the same theoretical net charge with those of (-30)GFP and ProT α , (-30 and -44, respectively) but with sequences containing only scrambled Glu and Asp residues: (-30)polyD/E and (-44)polyD/E (Fig. 2a). Importantly, Cre fusions to (-30)polyD/E or (-44)polyD/E resulted in much poorer protein delivery potencies ($EC_{50} = \sim 100$ nM and 21 ± 3.2 nM, respectively) than fusions with (-30)GFP or ProT α (Fig. 2b). These results show that additional features beyond net theoretical charge within ProT α strongly contribute to its ability to enable highly potent lipid-mediated protein delivery.

To identify other domains within ProT α responsible for cationic lipid-mediated protein delivery, we generated a series of truncated ProT α variants and repeated the Cre fusion protein delivery titrations in the presence of RNAiMAX lipid (Fig. 3a).

Deletion of the central anionic region almost completely abolished delivery enhancement by ProT α , consistent with the hypothesis that negative charges are critical for complexation with cationic lipids (Fig. 3b). Deletion of the N-terminal region, hypothesized to have receptor-binding properties¹⁶, also slightly increased (by 3-fold) the EC_{50} to 2.9 ± 0.39 nM, suggesting that ProT α might mediate liposome-cell membrane interaction via surface-exposed receptors, in addition to simple electrostatic charge attraction (Fig. 3b). The deletion of the C-terminal region resulted in an 8-fold impairment of delivery potency ($EC_{50} = 7.7 \pm 0.84$ nM, Fig. 4b). The C-terminus of ProT α contains a putative NLS¹⁶. To test the possibility that this NLS is responsible for loss of delivery potency by ProT α lacking the C-terminus, we cloned, expressed, and delivered Cre containing an SV40 NLS. Cre delivery potency did not benefit from the addition of this NLS, consistent with the known ability of Cre to localize to the nucleus spontaneously, and suggesting that the C-terminal domain augments ProT α -mediated protein delivery potency through mechanisms beyond simply providing an additional nuclear localization signal (Supplementary Fig. 7). Taken together, these results indicate that multiple domains within ProT α contribute to its unusual cationic lipid-mediated protein delivery potency.

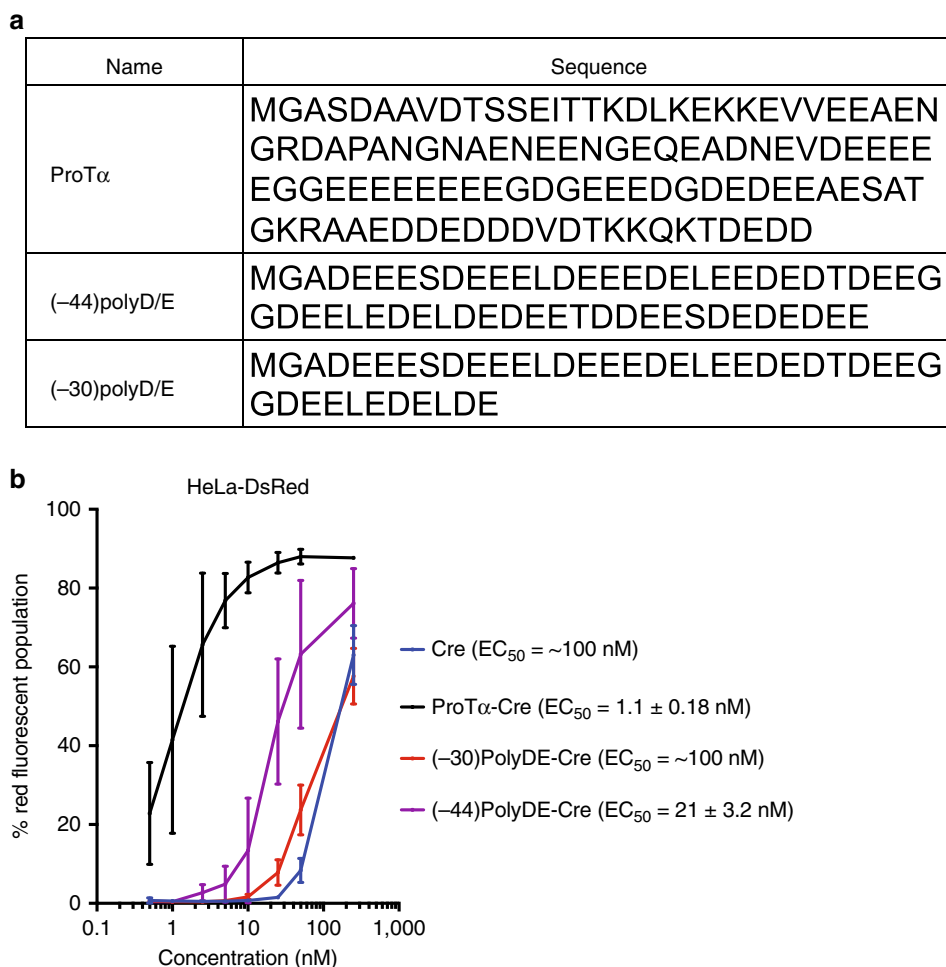


Fig. 2 ProT α delivery potency cannot be recapitulated by simple oligomers of acidic residues. **a** Sequences of ProT α and charged peptide analogs that contain -30 or -44 net theoretical charge for comparison. All proteins contain a Gly-Ala inserted after the initial Met to minimize expression level differences. **b** Despite sharing the same net theoretical charge as that of ProT α , (-44)PolyD/E-Cre does not promote potent protein delivery. EC_{50} is defined as the protein concentration at which 50% of analyzed HeLa-DsRed cells contain red fluorescence. Values and error bars represent the mean and standard deviation of three independent biological replicates performed on different days. EC_{50} values were determined by nonlinear fit to the Hill equation, and are followed by the standard errors of the fit. Source data are available in the Source Data file

Next we probed potential mechanisms of ProT α -mediated protein delivery compared to those of canonical cationic liposomes, which are thought to enter cells through clathrin-dependent endocytosis^{19,20}. We applied several inhibitors of various pathways involved in endocytosis and micropinocytosis to observe possible effects on delivery, including sodium azide, which depletes cellular ATP;²¹ ethylisopropyl amiloride (EIPA), a micropinocytosis inhibitor;²² Dynasore, a dynamin inhibitor;²³ methyl- β -cyclodextrin (MBCD), a cholesterol-depleting molecule that prevents lipid raft formation;²⁴ chlorpromazine (CPZ), a clathrin-dependent endocytosis inhibitor²⁵, and wortmannin (WTM), a PI3-kinase inhibitor that prevents clathrin-dependent endocytosis²⁶. After pre-incubating each of these inhibitors with HEK293 cells containing the RFP Cre reporter, we combined either (-30)GFP-Cre or ProT α -Cre with RNAiMAX lipid and treated cells with the resulting complex for 4 h. Excess liposomes and inhibitors were removed using several washes with buffer containing heparin, and the cells were further incubated for 2 days.

Flow cytometry analysis showed that, in HEK293-RFP cells, MBCD almost completely negated cytosolic access, while CPZ and WTM caused mild inhibition of delivery (Fig. 4). These findings are consistent with those of a previous study²⁷ that

examined the mechanism of functional siRNA delivery using cationic lipids and implicated direct liposome fusion, which is strongly inhibited by MBCD, rather than clathrin-mediated endocytosis followed by endosomal escape, as the major source of functional cargo in the cytoplasm. As endocytosis inhibitors are known to widely differ in their degree of delivery inhibition across different cell types²⁵, further studies are needed to conclusively implicate a mechanism for cationic nanoparticle protein delivery. Moreover, (-30)GFP-Cre and ProT α -Cre were inhibited by the same molecules to nearly identical degrees, suggesting that the large delivery potency increase enabled by ProT α is predominantly due to its apparent ability to engage cationic lipids at ~ 10 -fold lower concentrations than (-30)GFP (Fig. 4). Minimal cell toxicity was observed for the inhibitors and doses used, other than moderate toxicity from sodium azide (Supplementary Fig. 8).

Delivery of zinc-finger nucleases. Next we tested the ability of ProT α to enhance the lipid-mediated delivery of zinc-finger nucleases (ZFNs), chimeric genome editing proteins composed of a modular DNA-binding zinc-finger domains and a heterodimeric FokI nuclease²⁸. Two ZFNs (left and right) when targeted to adjacent half-sites of a genomic locus will bind together,

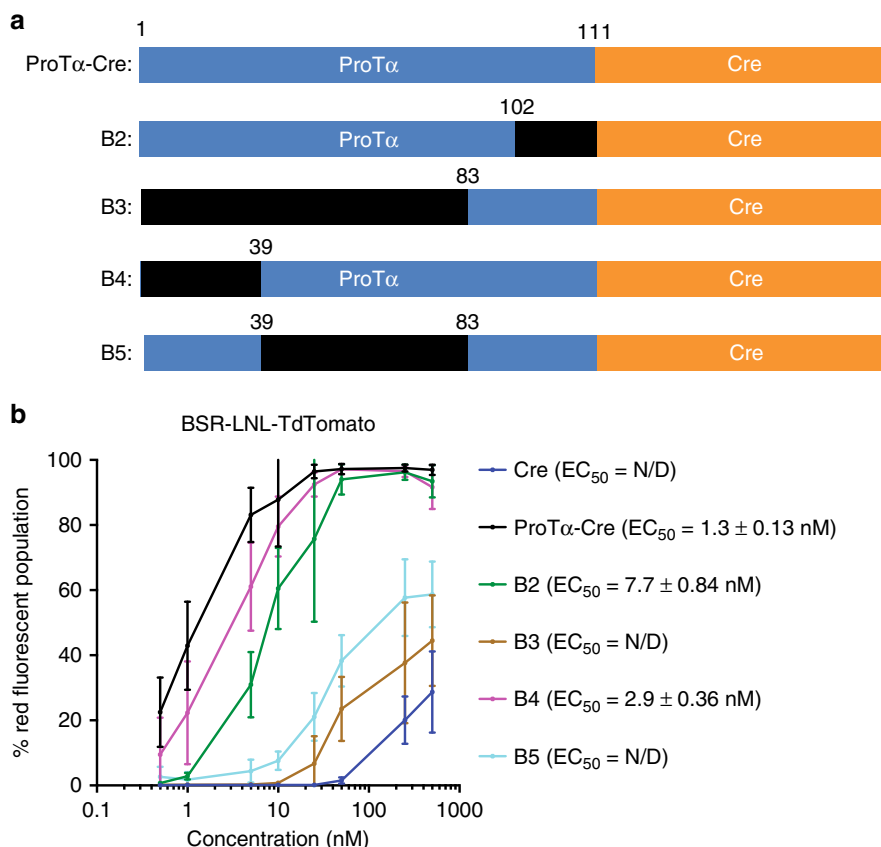


Fig. 3 Truncation of ProTα. **a** Primary structure of several truncated variants of ProTα-Cre. Deleted regions are shown in black. **b** Lipid-mediated delivery efficiency of Cre fused to each truncated ProTα variant into the BSR-TdTomato cell line. Removal of the central region containing stretches of acidic residues (B3, B5) results in a >100-fold reduction in protein delivery potency. Deletion of either the N-terminus or the C-terminus of ProTα (B4, B2) also reduces delivery efficiency, suggesting that the termini also play a role in the delivery process. EC_{50} is defined as the protein concentration at which 50% of analyzed cells contain red fluorescence. Values and error bars represent the mean and standard deviation of three independent biological replicates performed on different days. EC_{50} values were determined by nonlinear fit to the Hill equation, and are followed by the standard errors of the fit. Source data are available in the Source Data file

enabling their fused FokI nuclease domains to dimerize and induce a double stranded DNA cut, initiating end-joining processes that result in indels at the target locus²⁸. ZFNs are promising research tools and therapeutics, and are in multiple clinical trials for the treatment of diseases including HIV infection²⁹.

ZFNs have been shown to enter cells spontaneously at high concentrations under some conditions³⁰. We sought to test the ability of ProTα to mediate potent delivery of ZFNs with cationic lipids. We expressed both left and right ZFNs targeting the AAVS1 safe harbor site in the human genome fused with ProTα at the N-terminus (Fig. 5a). Before conducting protein delivery, we confirmed that ProTα did not affect the activity of ZFNs in DNA cleavage assays in vitro conducted using purified substrates (Supplementary Fig. 9a), and also did not affect in HEK293T editing levels following plasmid transfection of ZFN variants (Supplementary Fig. 9b). We optimized the concentration of protein and lipid for ZFN delivery into HEK293T cells (Supplementary Fig. 10).

Finally, we delivered ProTα-ZFN fusions or unmodified ZFNs complexed with RNAiMAX lipid into HEK293T cells in the presence of 10% serum and measured the resulting levels of AAVS1 target site genome editing. After 2 days, high-throughput sequencing (HTS) showed substantial ZFN-mediated indel formation at the target site only in cells treated with both pairs of ProTα-ZFN fusions and lipid in the mid-nM concentration

regime (Fig. 5b). In contrast, cells treated with ZFNs lacking ProTα complexed with RNAiMAX resulted in no significant levels of genome editing (Fig. 5b). Neither the ZFNs alone nor the ProTα-ZFN fusions resulted in substantial cytotoxicity at the concentrations tested when complexed with RNAiMAX (Supplementary Fig. 11). Previously demonstrated self-delivery of ZFNs requires serum-free media and μ M protein concentrations to generate moderate levels of indels, conditions that are prohibitive for some cell culture experiments and most in vivo applications³⁰. Our result shows that ProTα can mediate delivery of ZFNs in cells media containing serum to generate even higher levels of indels using sub- μ M concentrations of protein with a cationic lipid.

Discussion

In this study we screened high anionic human proteins to identify ProTα, a small, intrinsically disordered protein that mediates efficient liposome-mediated delivery of fused cargo proteins. ProTα enables potent delivery of both Cre recombinase and ZFNs at nM concentrations into human cells when combined with a simple, commercially available cationic lipid. To our knowledge, ProTα represents the most potent protein reported to date that enables delivery of fused proteins via cationic liposomes. As ProTα expression is known in all human tissues tested¹⁶, it may serve as a less immunogenic domain for protein delivery than other non-human alternatives such as (–30)GFP.

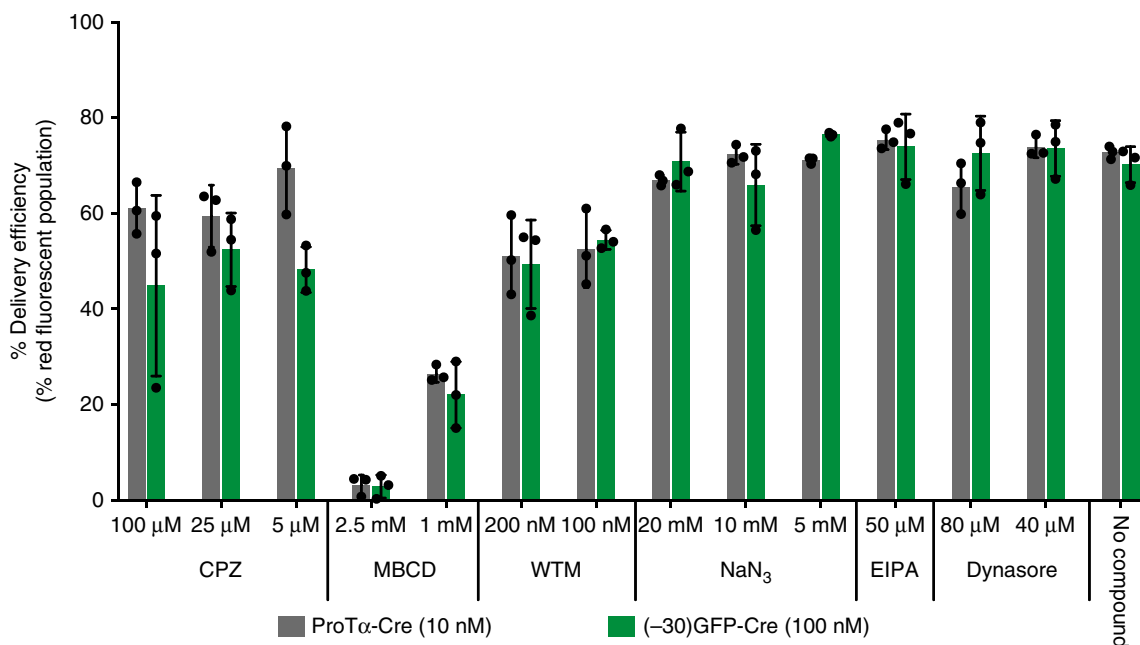


Fig. 4 Effect of various endocytosis inhibitors on ProTα- and (-30)GFP-mediated lipid nanoparticle delivery. Cationic lipid nanoparticles combined with either 10 nM ProTα-Cre or 100 nM (-30)GFP-Cre were incubated with HEK293-RFP cells that had been pre-treated with inhibitors that block various endocytosis pathways. After 2 day, delivery efficiency was measured via flow cytometry, and the percent of cells containing red fluorescence was quantified and normalized against a control sample lacking any inhibitor. Chlorpromazine (CPZ) and wortmannin (WTM) modestly decreased the delivery efficiency, whereas MBCD strongly blocked Cre recombination. Values and error bars represent the mean and standard deviation of three independent biological replicates performed on different days. Source data are available in the Source Data file

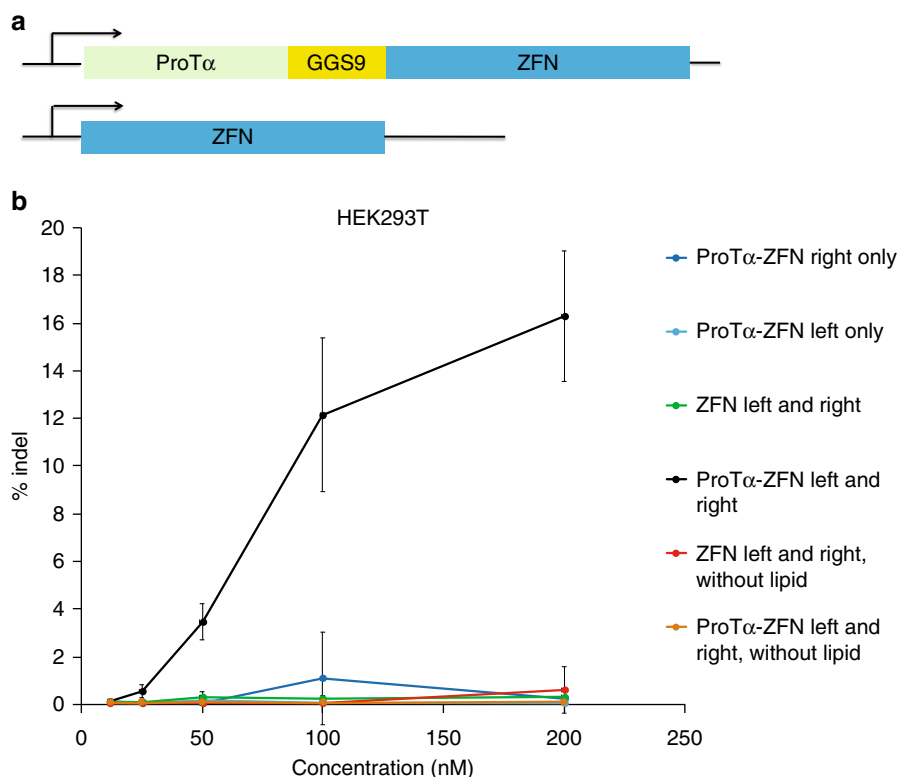


Fig. 5 Delivery of zinc-finger nucleases (ZFNs) using ProTα. **a** Structure of ProTα-fused ZFN versus ZFN alone. **b** ProTα-fused ZFNs or unfused ZFNs targeting the AAVS1 site in HEK293T cells were delivered using Lipofectamine RNAiMAX. ProTα enables efficient delivery of ZFNs and induce indels at a mid-nanomolar concentrations. Both left and right ZFN components, as well as lipid, are required for efficient indel generation. Values and error bars represent the mean and standard deviation of three independent biological replicates performed on different days. Source data are available in the Source Data file

Based on our previous work on the use of anionic proteins to mediate cationic lipid-based protein delivery^{12,31}, we anticipate that ProTa will be compatible with a variety of lipid reagents, although optimizing the dose and concentrations of both the lipid and protein before delivery, as shown above, maximizes delivery potency. Reagent dose optimization is especially important as proteins, even when fused to ProTa, will vary in their ability to be encapsulated into cationic liposomes, and will also require different degrees of delivery depending on their desired function within the cell. We also envision that ProTa may be particularly enabling when delivering proteins with adverse properties that preclude naked protein delivery via conventional cell-penetrating peptides, that are not tolerated by the cell at higher concentrations, or that are difficult to generate in quantities needed for less-potent delivery methods.

Methods

Cloning. PCR was performed using Q5 Hot Start High-Fidelity DNA Polymerase (New England BioLabs). Candidate human protein DNAs were purchased from IDT as gBlock Gene Fragments. Bacterial expression plasmids encoding human protein fused to Cre were made using USER-cloning (New England BioLabs). Truncation of ProTa was done using blunt-end ligation to delete regions of ProTa. Following PCR, KLD enzyme mix (New England BioLabs) was used to phosphorylate and circularize the PCR product before transformation into NEB10beta cells (New England BioLabs).

Protein expression. BL21 Star (DE3) chemically competent *E. coli* cells (ThermoFisher Scientific) were transformed with plasmids encoding the human proteins fused with Cre with a His₆ C-terminal purification tag. A single colony was grown overnight in 2×YT broth containing 50 µg/ml Carbenicillin at 37 °C. The cells were diluted 1:20 into 1 L of the same media and grown until OD₆₀₀ ~0.5. The cultures were incubated on ice for 60 min and protein expression was induced with 0.5 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG, GoldBio Sciences). Protein was expressed for 14–16 h with shaking at 16 °C. Cells were centrifuged at 20,000 × *g* for 20 min, and then resuspended in a high salt buffer (100 mM Tris(hydroxymethyl)aminomethane (Tris)-HCl, pH 8.0, 1 M NaCl, 20% glycerol, 5 mM Tris(2-carboxyethyl)phosphine (TCEP; GoldBio) with a protease inhibitor pellet (Roche). The cells were lysed using sonication and the supernatant was incubated with His-Pur nickel nitriloacetic acid (nickel-NTA) resin (ThermoFisher) with rotation at 4 °C for 30 min. The resin was washed with the high salt buffer before the protein was eluted with an elution buffer (high salt buffer supplemented with 200 mM imidazole). The eluent was purified on a 5 ml Hi-Trap Q (GE Healthcare) anion exchange column with an FPLC (AKTA Pure). The purified protein was quantified by a Pierce microplate BCA protein assay kit (Pierce Biotechnology) and snap-frozen in liquid nitrogen and stored at –80 °C until before use. ZFNs were purified as above with slight modification^{30,32}. Briefly, the ZFN proteins were induced at 22 °C for 4 h with 0.7 mM IPTG. Buffers contained additional 0.1 mM ZnCl₂ during Ni-column and FPLC purifications. Immediately after elution, the fractions were stored with 0.1 M L-Arg to improve protein stability. Amino acid sequences of proteins expressed in this study are provided in Supplementary Note 1.

In vitro DNA cleavage assay. DNA containing the AAVS1 locus was amplified from purified HEK293T genomic DNA using PCR. ~350 bp PCR product was purified using Minelute columns (Qiagen). Hundred nanogram of DNA substrate was incubated with 300 nM of ZFN or ProTa-ZFN pairs in Cutsmart buffer with 1 mM Arginine and 100 µM ZnCl₂ (New England Biolabs) at room temperature for 16 h. The cleavage product was detected by running the mixture on an agarose gel without further purification.

Cell culture. HeLa-DsRed⁶, BSR-LNL-TdTomato², HEK293-loxP-GFP-RFP (HEK293-RFP) (GenTarget), and HEK293T (ATCC) cells were cultured in Dulbecco's Modified Eagle's Medium plus GlutaMax (ThermoFisher Scientific) supplemented with 10% (v/v) FBS, at 37 °C with 5% CO₂. Primary human fibroblasts (ThermoFisher Scientific C0045C) were cultured in Medium 106 (ThermoFisher Scientific M106500) supplemented with Low Serum Growth Supplement (ThermoFisher Scientific S00310) at 37 °C with 5% CO₂.

Protein delivery assays. For Cre delivery assays, proteins fused to Cre were serially diluted in OptiMem (ThermoFisher Scientific) at 12.5 µL volume. Equal volume of OptiMem containing 1.5 µL of Lipofectamine RNAiMAX (ThermoFisher Scientific) was mixed with the protein and incubated at 25 °C before delivery into cells that had been seeded on a 48-well collagen-coated BioCoat plate (Corning) at ~70% confluency (250 µL final volume). After 3 day, the cells were trypsinized using TrypLE reagent (ThermoFisher Scientific), and resuspended in

culture media before being analyzed on the CytoFlex flow cytometer (Beckman Coulter). To facilitate quantitative comparison of delivery efficacy among human proteins, EC₅₀ values were calculated by fitting the data for each protein to the Hill equation for proteins that resulted in greater than 50% Cre recombination at the doses assayed in the experiment.

For ZFN experiments, equimolar amounts of 'left' and 'right' ZFNs was diluted to 12.5 µL in OptiMem and was complexed with 12.5 µL of OptiMem containing 3.5 µL of Lipofectamine RNAiMAX (ThermoFisher Scientific). The resulting complex was delivered to cells that had been seeded on a 48-well collagen-coated BioCoat plate (Corning) at ~70% confluency at the final volume of 100 µL per well. After 4 h, the cells were incubated with fresh media for further 48 h. Then, the cells were lysed and DNA was purified DNA was isolated using the Agencourt DNAdvance Genomic DNA Isolation Kit (Beckman Coulter) according to the manufacturer's instructions. AAVS1 site was amplified by PCR with flanking high-throughput sequencing primer pairs (Supplementary Table 1). Then, DNA was further amplified by PCR with primers containing Illumina sequencing adaptors. The products were gel-purified and quantified using KAPA Library Quantification Kit-Illumina (KAPA Biosystems). Samples were sequenced on an Illumina MiSeq following the manufacturer's protocol. Indels were quantified within the 30-base window surrounding the cleavage site among the high-quality reads (Q > 30) using a previously reported Matlab script³³.

For mCherry delivery assays into HEK293T and primary human fibroblasts, proteins fused to mCherry were serially diluted in OptiMem (ThermoFisher Scientific) at 12.5 µL volume. Equal volume of OptiMem containing 1.5 µL of Lipofectamine RNAiMAX (ThermoFisher Scientific) was mixed with the protein and incubated at 25 °C for 20 min before delivery into cells that had been seeded on a 48-well collagen-coated BioCoat plate (Corning) at 80% confluency (250 µL final volume). After 5 h, the cells were trypsinized using TrypLE reagent (ThermoFisher Scientific), and resuspended in culture media before being analyzed on the CytoFlex flow cytometer (Beckman Coulter).

Inhibitor assays. HEK293-RFP cells at ~70% confluency were incubated with one of the following endocytosis inhibitors for 1 h: NaN₃ (Sigma), 5-(N-ethyl-N-isopropyl)amiloride (EIPA, Santa Cruz Biotechnology), Dynasore (Abcam), Chlorpromazine (CPZ, Sigma), Methyl-β-cyclodextrin (MBCD, Sigma), Wortmannin (Sigma). Then, liposomes containing either (-30)GFP-Cre or ProTa-Cre were delivered and incubated at 37 °C with 5% CO₂ for 4 h. The cells were washed three times with PBS containing heparin (Stem Cell Technologies) at 20 µg/ml. The cells were further recovered for 2 day, and the cells were analyzed by the CytoFlex flow cytometer (Beckman Coulter).

Charge analog assays. To mimic the densely charged central region of ProTa, 'DEEEE' and 'DEEE' motifs were repeated until the appropriate amount of theoretical negative charge (–44 for (–44)poly D/E) was achieved. To enable the synthesis of the repeats as a gene fragment, the sequence was modified with serine, threonine, glycine, and leucine residues to eliminate sequence repetitiveness. (–30) poly D/E was created via C-terminal truncation of (–44)poly D/E to achieve a theoretical negative charge of –30. The charge variants were fused to the N-terminus of Cre via a (GGG)_n linker.

ZFN transfections. HEK293T cells were plated on a 48-well collagen-coated BioCoat plate (Corning) 1 day prior to experiment. At ~70% confluency, 500 ng of "left" and "right" CMV-ZFNs and CMV-ProTa-ZFNs (1 µg total DNA content) were transfected using 1.5 µL Lipofectamine 2000 (ThermoFisher Scientific) according to the manufacturer's protocol.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The source data underlying Figs. 1, 2a, b, 3a, b, 4, and 5b and Supplementary Figs. 1, 2, 4, 5, 6, 7, 8, 9a, b, 10, and 11 are provided as a Source Data file. High-throughput sequencing data generated during the current study have been deposited in the NCBI Sequence Read Archive repository under BioProject ID [PRJNA507860](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA507860). Other datasets generated during the current study are available from the corresponding author on reasonable request.

Received: 26 September 2018 Accepted: 23 May 2019

Published online: 02 July 2019

References

1. Pisal, D. S., Kosloski, M. P. & Balu-Iyer, S. V. Delivery of therapeutic proteins. *J. Pharm. Sci.* **99**, 2557–2575 (2010).

2. Cronican, J. J. et al. Potent delivery of functional proteins into mammalian cells in vitro and in vivo using a supercharged protein. *ACS Chem. Biol.* **5**, 747–752 (2010).
3. Frankel, Alan D. & Pabo, Carl O. Cellular uptake of the Tat protein from human immunodeficiency virus. *Cell* **55**, 1189–1193 (1988).
4. Fawell, S. et al. Tat-mediated delivery of heterologous proteins into cells. *Proc. Natl Acad. Sci. USA* **91**, 664–668 (1994).
5. Fuchs, S. M. & Raines, R. T. Polyarginine as a multifunctional fusion tag. *Protein Sci.* **14**, 1538–1544 (2009).
6. Cronican, J. J. et al. A class of human proteins that deliver functional proteins into mammalian cells in vitro and in vivo. *Chem. Biol.* **18**, 833–838 (2011).
7. Thompson, D. B., Villaseñor, R., Dorr, B. M., Zerial, M. & Liu, D. R. Cellular uptake mechanisms and endosomal trafficking of supercharged proteins. *Chem. Biol.* **19**, 831–843 (2012).
8. Heitz, F., Morris, M. C. & Divita, G. Twenty years of cell-penetrating peptides: from molecular mechanisms to therapeutics: peptide-based drug delivery technology. *Br. J. Pharmacol.* **157**, 195–206 (2009).
9. Allen, T. M. & Cullis, P. R. Liposomal drug delivery systems: from concept to clinical applications. *Adv. Drug Deliv. Rev.* **65**, 36–48 (2013).
10. Judge, A. D., Bola, G., Lee, A. C. H. & MacLachlan, I. Design of noninflammatory synthetic siRNA mediating potent gene silencing in vivo. *Mol. Ther.* **13**, 494–505 (2006).
11. Colletier, J.-P., Chaize, B., Winterhalter, M. & Fournier, D. Protein encapsulation in liposomes: efficiency depends on interactions between protein and phospholipid bilayer. *BMC Biotechnol.* **2**, 9 (2002).
12. Zuris, J. A. et al. Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. *Nat. Biotechnol.* **33**, 73–80 (2015).
13. Lawrence, M. S., Phillips, K. J. & Liu, D. R. Supercharging proteins can impart unusual resilience. *J. Am. Chem. Soc.* **129**, 10110–10112 (2007).
14. Ansari, A. M. et al. Cellular GFP toxicity and immunogenicity: potential confounders in in vivo cell tracking experiments. *Stem Cell Rev. Rep.* **12**, 553–559 (2016).
15. The UniProt Consortium. UniProt: a hub for protein information. *Nucleic Acids Res.* **43**, D204–D212 (2015).
16. Piñeiro, A., Cordero, O. J. & Nogueira, M. Fifteen years of prothymosin alpha: contradictory past and new horizons. *Peptides* **21**, 1433–1446 (2000).
17. Borgia, A. et al. Extreme disorder in an ultrahigh-affinity protein complex. *Nature* **555**, 61–66 (2018).
18. Gresch, O. & Altrogge, L. Transfection of difficult-to-transfect primary mammalian cells. *Methods Mol Biol.* **801**, 65–74. (2012).
19. Felgner, J. H. et al. Enhanced gene delivery and mechanism studies with a novel series of cationic lipid formulations. *J. Biol. Chem.* **269**, 2550–2561 (1994).
20. Godbey, W. T., Wu, K. K. & Mikos, A. G. Tracking the intracellular path of poly(ethylenimine)/DNA complexes for gene delivery. *Proc. Natl. Acad. Sci.* **96**, 5177–5181 (1999).
21. Hoffmann, J. & Mendgen, K. Endocytosis and membrane turnover in the germ tube of *Uromyces fabae*. *Fungal Genet. Biol.* **24**, 77–85 (1998).
22. Koivusalo, M. et al. Amiloride inhibits macropinocytosis by lowering submembranous pH and preventing Rac1 and Cdc42 signaling. *J. Cell Biol.* **188**, 547–563 (2010).
23. Preta, G., Cronin, J. G. & Sheldon, I. M. Dynasore—not just a dynamin inhibitor. *Cell Commun. Signal.* **13**:24 (2015).
24. Rodal, S. K., Skretting, G., Garred, Ø. & Vilhardt, F. Extraction of cholesterol with methyl- α -cyclodextrin perturbs formation of clathrin-coated endocytic vesicles. *Mol. Biol. Cell* **10**, 14 (1999).
25. Vercauteren, D. et al. The use of inhibitors to study endocytic pathways of gene carriers: optimization and pitfalls. *Mol. Ther.* **18**, 561–569 (2010).
26. Posor, Y. et al. Spatiotemporal control of endocytosis by phosphatidylinositol-3,4-bisphosphate. *Nature* **499**, 233–237 (2013).
27. Lu, J. J., Langer, R. & Chen, J. A novel mechanism is involved in cationic lipid-mediated functional siRNA delivery. *Mol. Pharm.* **6**, 763–771 (2009).
28. Hockemeyer, D. et al. Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. *Nat. Biotechnol.* **27**, 851–857 (2009).
29. Tebas, P. et al. Gene editing of *CCR5* in autologous CD4 T cells of persons infected with HIV. *N. Engl. J. Med.* **370**, 901–910 (2014).
30. Liu, J., Gaj, T., Wallen, M. C. & Barbas, C. F. Improved cell-penetrating zinc-finger nuclease proteins for precision genome engineering. *Mol. Ther.* **4**, e232 (2015).
31. Wang, M. et al. Efficient delivery of genome-editing proteins using bioreducible lipid nanoparticles. *Proc. Natl Acad. Sci. USA* **113**, 2868–2873 (2016).
32. Carroll, D., Morton, J. J., Beumer, K. J. & Segal, D. J. Design, construction and in vitro testing of zinc finger nucleases. *Nat. Protoc.* **1**, 1329–1341 (2006).
33. 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).

Acknowledgements

This work was supported by NIH RM1 HG009490, R01 EB022376, U01 AI142756, and R35 GM118062, DARPA HR0011–17–2–0049, the Bill and Melinda Gates Foundation, and HHMI. Y.B.K. was supported by Natural Science and Engineering Research Council Post-graduate scholarship (NSERC-PGS). K.T.Z. was supported by the Harvard Chemical Biology Program NIH Training Grant T32 GM095450.

Author contributions

Y.B.K. and K.T.Z. designed the research, performed the experiments, analyzed data and wrote the manuscript. D.B.T. supervised the research and wrote the manuscript. D.R.L. designed and supervised the research and wrote the manuscript.

Additional information

Supplementary Information accompanies this paper at <https://doi.org/10.1038/s41467-019-10828-3>.

Competing interests: The authors have filed a patent application on protein delivery technologies. The patent applicant is the Broad Institute and Harvard University with Y. B.K., K.T.Z. and D.R.L. as the named inventors. This patent is in regards to the use of ProTa for protein delivery. D.R.L. is a co-founder and consultant of Editas Medicine, Beam Therapeutics, and Parwise Plants, companies that use genome editing. The remaining authors declare no competing interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

Peer review information: *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019