Flow cytometric analysis of DNA binding and cleavage by cell surface-displayed homing endonucleases

Petra Volná[1,2], Jordan Jarjour[2,3], Sarah Baxter[4], Steve R. Roffler[5],
Raymond J. Monnat Jr[4], Barry L. Stoddard[6] and Andrew M. Scharenberg[1,2,3,*]

[1]Department of Pediatrics, University of Washington, Box 359300-CW, Seattle WA 98195, USA, [2]Children's Hospital, & Regional Medical Center, 307 Westlake Ave N Suite 300, Seattle WA 98109, USA, [3]Department of Immunology, University of Washington, 1959 Pacific Street NE, Seattle WA 98195, USA, [4]Department of Pathology, University of Washington, Seattle WA 98195, USA, [5]Institute of Biochemical Sciences, Academia Sinica, 128 Yen-chiu-yuan Rd., sec. 2, Tapei, Taiwan, [6]Division of Basic Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave N, A3-025, Seattle WA 98109, USA

ABSTRACT

LAGLIDADG homing endonucleases (LHEs) cleave 18–24 bp DNA sequences and are promising enzymes for applications requiring sequence-specific DNA cleavage amongst genome-sized DNA backgrounds. Here, we report a method for cell surface display of LHEs, which facilitates analysis of their DNA binding and cleavage properties by flow cytometry. Cells expressing surface LHEs can be stained with fluorescently conjugated double-stranded oligonucleotides (dsOligos) containing their respective target sequences. The signal is absolutely sequence specific and undetectable with dsOligos carrying single base-pair substitutions. LHE–dsOligo interactions facilitate rapid enrichment and viable recovery of rare LHE expressing cells by both fluorescence-activated cell sorting (FACS) and magnetic cell sorting (MACS). Additionally, dsOligos conjugated with unique fluorophores at opposite termini can be tethered to the cell surface and used to detect DNA cleavage. Recapitulation of DNA binding and cleavage by surface-displayed LHEs provides a high-throughput approach to library screening that should facilitate rapid identification and analysis of enzymes with novel sequence specificities.

INTRODUCTION

Homing endonucleases of the LAGLIDADG family (LHEs) form homodimers or pseudosymmetric monomers that generally recognize DNA sequences 18–24 bp in length (1). Their molecular structures are built around two conserved alpha helices that contain a LAGLIDADG consensus sequence, which forms the center of the interface between enzyme subunits or domains (2). The final acidic residues from the LAGLIDADG helix form part of each domain's active site that cleaves one strand of the double-stranded DNA target sequence. The DNA-binding interface of each domain is made up of a four-stranded antiparallel beta-sheet that is supported by a series of framework alpha helices, which form the core of the domain. Unlike restriction endonucleases which form densely packed and highly saturated DNA–protein interfaces, the DNA-binding interface of LHEs make fewer hydrogen bonds per target sequence base pair (3). These structural properties account for the ability of LHEs to withstand moderate variability in target sequence recognition (4–7), a characteristic which has been essential in maintaining their genetic mobility and horizontal proliferation (8) and which make LHEs ideal substrates for engineering altered DNA-binding interfaces with novel endonucleolytic specificities (9–13). The combination of high target sequence specificity and adaptable DNA-binding interfaces make LHEs attractive tools for genome engineering applications, which require the introduction of a double-stranded break at a precise genomic location (13–16).

Since only a limited number of native LHEs have been identified (1), attempts to use existing LHEs as scaffolds for creating novel enzymes able to target desired sequences have been widespread (4,5,17–20). Recent efforts have employed a variety of approaches that individually (5,20) or successively (12) utilize computational redesign with substrate cleavage screening

or two-plasmid selection systems (21,22). While these methodologies have shown promise, they are limited in their screening throughput because they require the generation of combinatorial endonuclease mutant libraries and the variant endonucleases must be well tolerated by the host's genomic DNA. An additional limitation is that the intracellular cleavage system must be redesigned and generated for each sequence targeted for selection. We sought to develop a system where LHE proteins could be rapidly screened to identify and isolate variants with new DNA target specificities. Here, we demonstrate that LHEs can be expressed on the plasma membrane of a lymphocyte cell line by targeting the expression of an LHE-CD80 transmembrane fusion protein to the secretory pathway. Surface-expressed LHEs faithfully recapitulate the properties of the native enzymes in solution, as assessed by flow cytometric analysis of both the binding and cleavage of fluorescently conjugated dsOligos. Furthermore, sequence-specific LHE interactions with dsOligos in conditions, which prohibit substrate cleavage allow for their physical isolation by multiple cell separation methods. The rapid analysis of LHE–DNA interactions on the cell surface with concurrent sorting options should facilitate a significant acceleration in the isolation of novel endonuclease variants with unique DNA target specificities.

METHODS

Plasmid construction and generation of stable LHE expressing DT40 clones

Vectors-containing cDNA for both LHEs were PCR amplified using following primers: I-AniI For SfiI and I-AniI Rev SalI; H-DreI For SfiI and H-DreI Rev SalI and cloned into the pLHCX-phOx expression vector (23,24) by SfiI and SalI digestion to replace phOx coding sequence. To place the NeoR gene in frame in the I-AniI construct, the NeoR cDNA including the HSV polyA sequence was amplified using CD80-NeoR For and NeoR Rev ClaI, while the existing I-AniI-CD80 expression construct (including the 5′ SP and HA epitope) was amplified by primers SP For Hind3 and CD80-NeoR Rev. The entire fusion molecule was generated by fusion PCR as described previously (26), and subcloned back into the pLHCX plasmid by HindIII and ClaI digestion. Mutation of residues K21, T27 for I-AniI^m generation was achieved by site-directed mutagenesis (Stratagene QuikChange II, no. 200523-5) using I-AniI K21 T27 SDM For and I-AniI K21 T27 SDM Rev, and the L223 mutation arose by PCR error. For transfection of DT40 cells, 30 μg of linearized plasmid DNA was electroporated into 10^7 DT40 cells (IgM-negative where indicated) using a Gene Pulser XCell (Bio-Rad, Hercules, CA, USA) in a final volume of 400 μl of serum-free RPMI media employing the exponential protocol: 550 V, 25 μF, ∞ resistance with a 4 mm cuvette gap. After 24 h of culture in drug-free media, cells were plated by limiting dilution in media containing 2 mg/ml G418 (Invitrogen, Carlsbad, CA, USA, no. 11811-098) for 10–14 days. Wells containing single G418-resistant clones

were expanded and screened by flow cytometry for HA surface expression.

I-AniI For SfiI: GGCCCAGCCGGCCATGGGC AGCAGCCATCATCATC

I-AniI Rev SalI: GTCGACATAATTTGAAGGTAT TTTTATTTTTTCTG

H-DreI For SfiI: GGCCCAGCCGGCCATGCATA ATAATGAGAATGTT

H-DreI Rev SalI: GTCGACCGGGGACGATTTCTT TTTTTCACT

CD80-NeoR For: CAGACCGTCTTCCTTGGATCG GCCATTGAACAAG

NeoR Rev ClaI: ATCGATGAACAAACGACCCA ACACCCGTGCG

SP For Hind3: AAGCTTATGGAGACAGACAC ACTCCTGCTATGGG

CD80-NeoR Rev: CTTGTTCAATGGCCGATCCAA GGAAGACGGTCTG

I-AniI K21 T27 SDM For: CAGCATCACCAACAAG GGTAAGTACCTACAGTATGAGCTGGGTATCGAG

I-AniI K21 T27 SDM Rev: CTCGATACCCAGC TCATACTGTAGGTACTTACCCTTGTTGGTGATG CTG

Western blotting and glycosylation analysis by PNGase F treatment

Here, 7.5×10^6 cells of the indicated cell lines were washed once in ice-cold PBS containing 0.1% BSA and lysed for 30 min at 4°C in lysis buffer (25 mM Tris-Cl pH 7.4, 140 mM NaCl, 2 mM EDTA, 1% NP-40, 0.05% sodium deoxycholate, 0.005% SDS and protease inhibitors). The crude cell lysates were clarified by centrifugation and 50 μg of total protein from post-nuclear cell lysates were used for incubation with PNGase F (New England Biolabs, Beverly, MA, USA, no. P0704S) for 2 h according to manufacturer's guidelines. Samples were analyzed by western blotting using anti-HA (Cell Signaling Technology, Danvers, MA, USA, no. 2367) and anti-β-actin Ab (Sigma-Aldrich, St. Louis, MO, USA, no. A1978) followed by HRP-conjugated anti-mouse-IgG (Amersham Biosciences, Piscataway, NJ, USA, no. NA931V).

Flow cytometry

Standard antibody staining was done in PBS-containing 0.2% BSA using the following antibodies: mouse monoclonal anti-HA (Cell Signaling Technology, no. 2367) followed by PE-conjugated goat anti-mouse IgG1 (Southern Biotech, Birmingham, AL, USA, no. 1070-09S); FITC-conjugated anti-chicken IgM (Bethyl Laboratories Inc., Montgomery, TX, USA, no. A30-102F). Preparation of dsOligos and subsequent staining was performed as follows: complementary 5′-biotinylated and non-biotinylated DNA oligonucleotides (Figure 2) were annealed by incubation at 94°C for 5 min and allowed to cool slowly to room temperature, sterilized by ethanol precipitation and resuspended to a stock concentration of 1.6 μM. Cells were first incubated at 4°C for 30 min in our standard dsOligo blocking and staining buffer containing 135 mM NaCl, 5 mM KCl,

10 mM $CaCl_2$, 5.6 mM Glucose, 10 mM HEPES, 0.2% BSA and 1 µg/ml sonicated salmon sperm DNA, pH 7.4. Concurrent with this incubation, annealed dsOligos were complexed with SAv-PE (BD Biosciences, Palo Alto, CA, USA, no. 554061, Mw 300,000) at 1:1 molar ratio in the same buffer. The dsOligo-BT: SAv-PE complexes were used to stain the cells at a final concentration of 10–50 nM for 30–40 min at 4°C. Cells were washed twice with ice-cold buffer prior to analysis. Antibody and dsOligo-stained cells were analyzed by flow cytometry using the Beckton Dickenson FACSCalibur or LSRII instruments (BD Biosciences). 10 000–100 000 live cells were acquired per sample and the resulting raw data were processed using FlowJo software (FlowJo Ashland, OR, USA, LLC).

Fluorescence-activated cell sorting (FACS)

LHE expressing clones were mixed at the indicated ratios immediately prior to staining. The cells were stained using the above protocol with the indicated dsOligo complexes (SAv-Q655 from Invitrogen, no. Q10121MP). The PE- or Q655-positive populations of live-gated doublet-excluded cells were sorted using the BD Aria cell sorter. Sorted populations were cultured for 5–7 days and labeled with either dsOligos or anti-IgM for flow cytometry analysis. The above process was iterated for subsequent rounds of enrichment.

Magnetic cell sorting (MACS)

Cells were mixed at the indicated ratios (~ 5–10×10^7 cells per sample) and labeled for 30 min at 4°C with 100 nM dsAni1 in the same buffer used for flow cytometry. After washing, the mixed population was incubated with 20–50 µl SAv-coated magnetic beads (Miltenyi Biotec, Auburn, CA, USA, no. 130-048-101) in a final volume of 0.5–1.0 ml for 20 min at 4°C. The samples were washed twice and resuspended at a concentration of 2×10^7 cells/ml prior to loading onto the AutoMACS cell separator. The 'posselds' double column separation program was run and the positive fraction was washed and placed immediately in culture. Cells were analyzed by staining separately with anti-IgM and dsAni1 as described above.

Flow cytometry assay for dsOligo cleavage

Complementary 5'-biotin and 5'-Alexa Fluor647 conjugated (Invitrogen) DNA oligonucleotides were annealed as described above. The buffer used for all steps of the cleavage assay contained 10 mM NaCl, 90 mM KCl, 10 mM HEPES, 5.6 mM Glucose, 0.2% BSA, 1 µg/ml salmon sperm DNA and pH 8.5. Approximately 1×10^6 cells were first incubated at 4°C with biotinylated mouse anti-HA Ab (Abcam, Cambridge, UK, no. AB27987-100) at a dilution 1:300 for 30–40 min. After washing, the cells were stained with 30–50 nM 647-dsOligo-BT: SAv-PE for 30 min on ice. For cleavage 10 mM $MgCl_2$ was added to the buffer and the reaction was carried out at 42°C for the designated time points. The cells were washed in Mg^{2+}-free buffer and analyzed by flow cytometry.

In-vitro LHE cleavage assay and fluorescence gel imaging

Reaction conditions were identical to those described in the flow cytometry cleavage assay except that 30 nM recombinant I-AniI was used in place of cells for the *in vitro* assays. For the *in vitro* assay with bead-complexed oligos, 647-dsOligo-BT: SAv-bead complexes were formed by incubating 50 nM dsOligo with 20 µl SAv-conjugated Dynabeads for 30 min at room temperature. The unbound 647-dsOligo-BT was removed by extensive washing in cleavage assay buffer, followed by incubation with 30 nM recombinant I-AniI for 1 h at 42°C. Oligonucleotide fragments were purified by phenol extraction followed by ethanol precipitation. The purified samples were resuspended in Ficoll-based loading buffer and resolved by PAGE. The gels were scanned using the Typhoon 9410 system (GE Healthcare, Piscataway, NJ, USA) with excitation by the 633 nM laser. Images were acquired with detector PMT voltages at both optimal (between 450 and 600 V) and maximal (between 700 and 850 V) settings to observe all fluorescent species. Images were processed with Adobe Photoshop using linear adjustments and all detectible bands in each lane are visible.

RESULTS

Expression of homing endonucleases on the plasma membrane surface

LHEs are normally expressed in the cytosol and targeted to DNA-containing organelles post-translationally. Cell surface display requires both the cotranslational targeting to the secretory pathway and fusion to an appropriate transmembrane domain. The strategy we chose was the one previously used to support surface display of antibody fragments (23,24). LHE genes were inserted between the coding sequences of the N-terminal murine immunoglobulin signal peptide (SP) and the transmembrane region of the murine CD80 molecule (Figure 1a). Two different LHE-coding sequences were integrated into the CMV promoter-driven surface expression constructs: I-AniI, an endonuclease encoded in the mitochondrial genome of *Aspergillus nidulans* (25); and H-DreI (Hybrid-Dmo/CreI, formerly called 'E-DreI'), an engineered endonuclease containing an N-terminal domain derived from I-DmoI LHE (*Desulfurococcus mobilis*) and a C-terminal domain derived from I-CreI (*Chlamydomonas reinhardtii*) (10). These constructs included a hemagglutinin (HA) epitope tag downstream of the SP to facilitate biochemical and flow cytometric detection. Transfection of the linearized constructs into DT40 cells resulted in the isolation of clonal lines with high levels of I-AniI and H-DreI surface expression (Figure 1c).

Intracellularly expressed LHEs are not exposed to glycosyltransferase enzymes, however this is an important consideration when their expression is directed to the cell surface. Primary sequence analysis revealed that LHE fusion proteins do contain potential *N*-glycosylation motifs (*N-X-S/T* where $X \neq P$ or D). To evaluate their *N*-glycosylation status, we incubated lysates of LHE
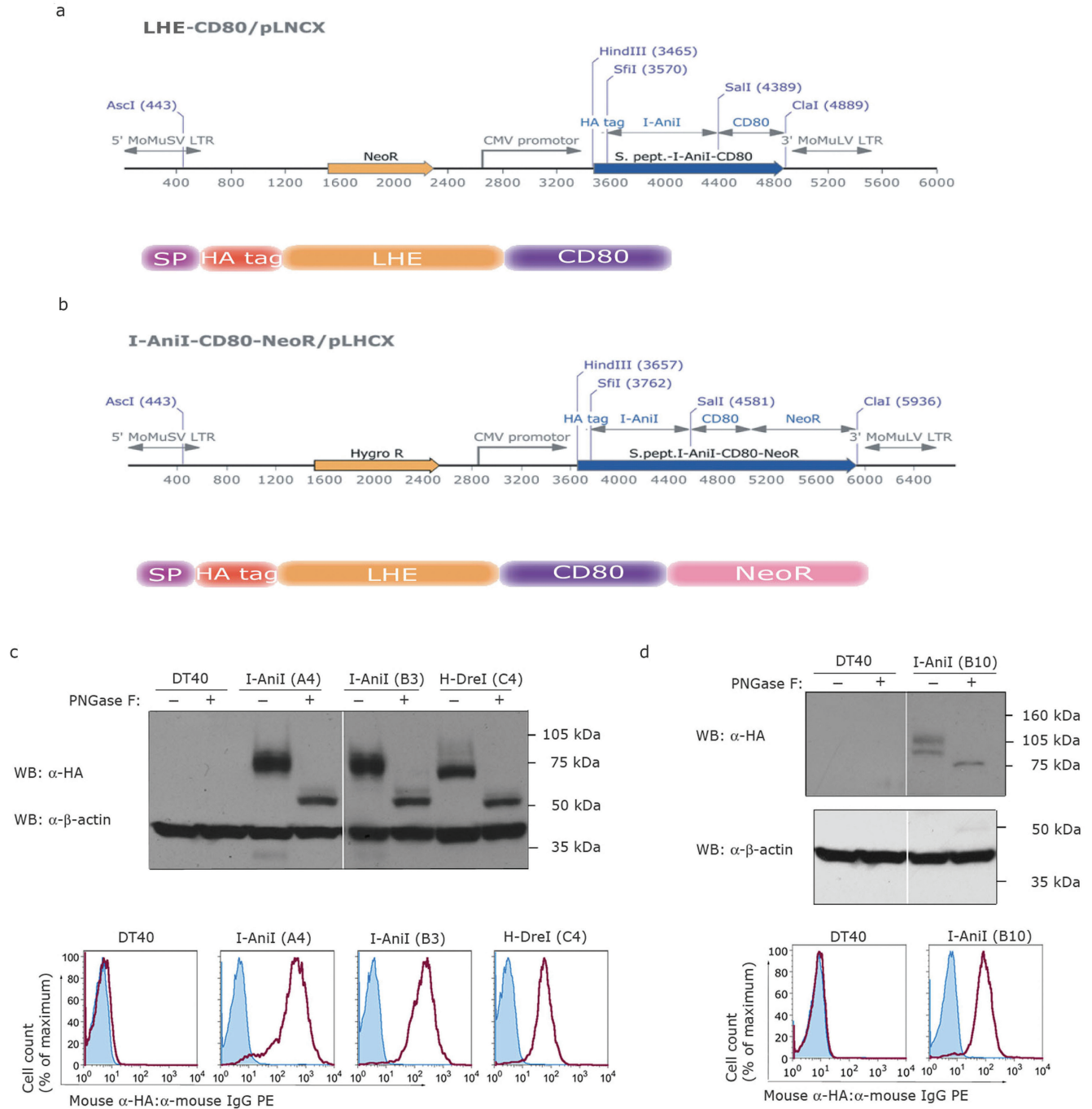
Figure 1. Vector schematics and validation of efficient LHE fusion protein expression in DT40 chicken B-cells. (**a**) LHE cDNAs were placed in-frame between a murine immunoglobulin-derived N-terminal signal peptide (SP) and the transmembrane spanning region of the murine CD80 molecule at the C-terminus. G418 resistance was conferred by a NeoR gene driven by an independent promoter. (**b**) The SP-HA-LHE-CD80 cassette was placed in-frame with the NeoR gene to allow coupled expression from a single promoter. Both constructs include an HA epitope tag at the N-terminus of the LHE and transcription is driven by the CMV promoter. (**c**) Western blot and flow cytometry analysis from clones expressing I-AniI (A4 and B3) and H-DreI (C4) and (**d**) from clone B10 expressing I-AniI as a fusion with C-terminal NeoR. Treatment with PNGase F is indicated above each lane. The corresponding clones were analyzed by flow cytometry for surface HA detection.

expressing cells with the enzyme peptide-*N*-glycosidase F (PNGaseF). The *N*-glycosylation status was estimated by observing changes in band mobility during electrophoresis, which demonstrated that PNGaseF-treated LHE fusion proteins migrated faster and with less variability compared with the untreated controls (Figure 1c and d). These results indicate that the membrane-anchored molecules were indeed *N*-glycosylated, consistent with their surface expression through the secretory pathway.
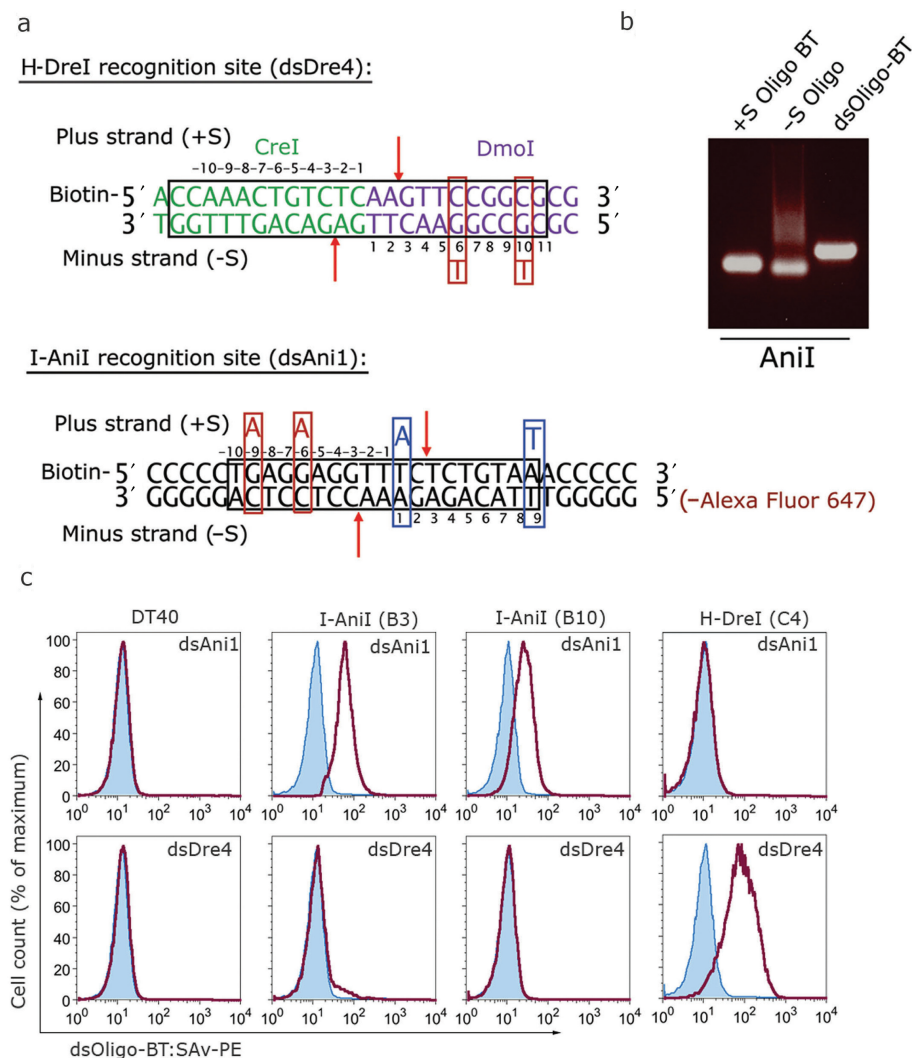
Figure 2. Fluorescently conjugated dsOligos bind cell surface LHEs in a manner, which is sequence specific and easily resolved by flow cytometry. (**a**) H-DreI is an engineered enzyme composed of domains derived from the I-CreI and I-DmoI LHEs. Its 23-bp recognition site (dsDre4, boxed) is therefore a complex of the natural target sequences bound by I-CreI (green) and I-DmoI (purple). The 19-bp I-AniI recognition site (dsAni1, boxed) was placed between stretches of five GC base pairs designed to enhance the formation and stability of the double-stranded complex. Single base-pair changes (dsDre4^{6T}, dsDre4^{10T}, dsAni1^{-6A} and dsAni1^{-9A}) are indicated by red boxes and the cleavage sites by red arrows. The alternative I-AniI target sequence (dsAni2) containing two base-pair changes are shown in blue boxes. Conjugations with biotin at the 5′ termini are depicted, and Alexa Fluor 647 conjugated oligonucleotides for dsAni1 and dsAni1^{-9A} were used in the flow cytometry cleavage assay. (**b**) Verification of efficient annealing of the complementary oligonucleotides run on a 3% agarose gel, with individual oligos (+S and −S) run as controls. (**c**) Flow cytometry analysis of clones stained with fluorescent dsOligos. Staining of I-AniI and H-DreI expressing clones in the presence of 10 mM Ca^{2+} are shown, with shaded and open histograms representing SAv-PE-only controls and dsOligo-BT: SAv-PE stained cells, respectively. The dsOligos used for each stain are indicated in the upper right corner of the histograms.

As one intended application of surface expressed LHEs is identification of desired LHE variants from large libraries generated by random or targeted mutagenesis, a tight linkage between surface LHE expression and a selection marker is desirable as a means to enrich for variants which are efficiently expressed. For this purpose, we evaluated a strategy involving fusion of a neomycin resistance (NeoR) gene in frame with the C-terminus of the CD80 transmembrane domain (Figure 1b) (26), such that the NeoR activity is positioned on the cytosolic face of vesicles and the plasma membrane after expression. Transfection of LHE-CD80-NeoR constructs and

application of neomycin selection allowed the isolation of multiple DT40 clones with stable surface expression of HA immunoreactivity from a single promoter (Figure 1d).

Surface-expressed LHEs are efficiently labeled with fluorescently conjugated dsOligos and detected by flow cytometry

We next tested the ability of the surface LHEs to bind annealed oligonucleotides representing their natural target specificities using flow cytometry. HEs are enzymatically active in the presence of Mg^{2+} ions, which are present in

the active site (27). When Mg^{2+} ions are replaced with Ca^{2+} ions, LHEs retain their DNA-binding properties, while the cleavage of target DNA sequence is abolished (27,28). We therefore used a buffer containing 10 mM Ca^{2+} for cell-surface staining of LHE expressing clones using fluorescently labeled dsOligos. In an effort to minimize the effects of variations in dissociation kinetics of different LHEs, we used a single-step staining protocol with pre-formed complexes of biotinylated dsOligos (dsOligo-BT, Figure 2b) with phycoerythrin-conjugated streptavidin (SAv-PE). Since streptavidin contains four high affinity biotin-binding subunits, we created complexes (dsOligo-BT: SAv-PE) at a 1:1 molar ratio to maximize the fluorescent signal per target sequence. Staining I-AniI and H-DreI expressing clones with dsOligos of their respective natural target sequences generated clearly labeled populations despite their apparent *N*-glycosylation (Figure 2c). This analysis indicates that glycosylation does not confound surface analysis of these particular LHEs.

To rule out the possibility that our expression and detection system leads to degenerate DNA substrate recognition, we stained I-AniI and H-DreI expressing clones with dsOligos containing modifications to their respective target sequences. As expected, we observed no detectable staining when dsAni1 or dsDre4 were used to stain non-corresponding LHE expressing clones (Figure. 2c). To achieve a precise characterization of staining specificity, we designed dsOligos bearing single base-pair differences from the known target sequence (dsAni1^{-9A}, dsAni1^{-6A}, dsDre4^{6T}, dsDre4^{10T}, Figure 2a). These substitutions were chosen to interrupt direct contacts within the I-AniI and H-DreI DNA–protein interfaces (10,25). Remarkably, these single base-pair changes resulted in little or no detectable staining above non-specific background levels (Figure 3), consistent with the predicted destabilization of the binding interactions with their respective LHEs. Conversely, we have generated NeoR-linked clones with mutant I-AniI enzymes (generally denoted as I-AniI^m) expressed stably on the cell surface (Figure 4b). Two I-AniI^m clones were used in our experiments and were predicted to have either core structural changes or designed to have lost specific contacts at the DNA-binding interface. Although we have not validated the structural consequences of these mutations, the failure of the mutant enzymes to bind dsAni1 indicates that structural alterations, which do not inhibit LHE expression have DNA-binding consequences that are resolvable by our approach. We further extended our analysis to a unique target sequence variation against which wild-type I-AniI is known to maintain its cleavage activity (dsAni2, unpublished data, Figure 2a). This second I-AniI target sequence readily stained clones expressing I-AniI, further supporting the correlation of dsOligo-based interrogation of LHEs on the cell surface with biochemical cleavage data (Figure 3a, bottom panels). These data show that surface-expressed LHEs reliably discriminate closely related dsOligo sequences in a manner which both parallels their reported target sequence cleavage specificities and is sensitive to
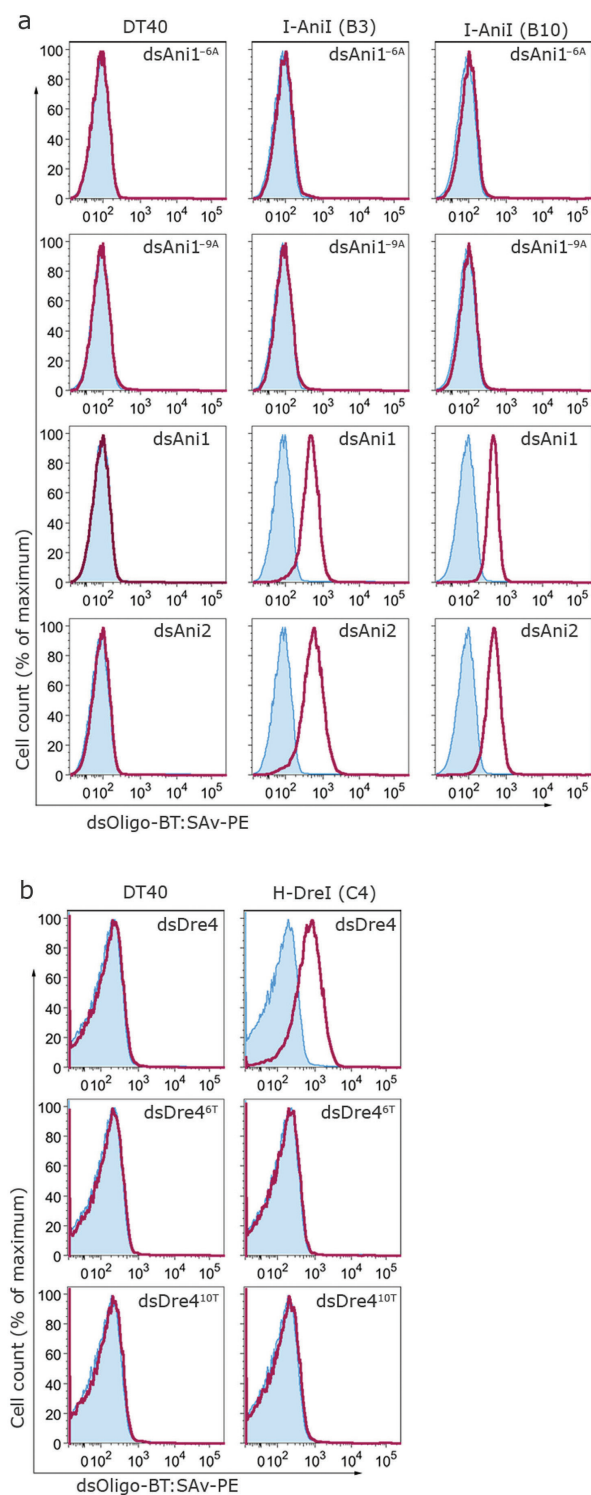

Figure 3. LHEs expressed on the cell surface reliably discriminate dsOligos containing single base-pair differences from their natural target sequences. (**a**) I-AniI and (**b**) H-DreI expressing clones were stained with dsOligo-BT: SAv-PE complexes containing the natural target sequences (dsAni1 and dsDre4) or containing single base-pair changes (dsAni1^{-6A} and dsAni1^{-9A}; dsDre4^{6T} and dsDre4^{10T}). Known target sequence degeneracy for I-AniI is also recapitulated by dsOligo staining and analysis by flow cytometry. The cells expressing I-AniI were efficiently stained with dsAni2 corresponding to an alternative I-AniI target sequence known to be cleaved with an efficiency that is similar to the natural target sequence.
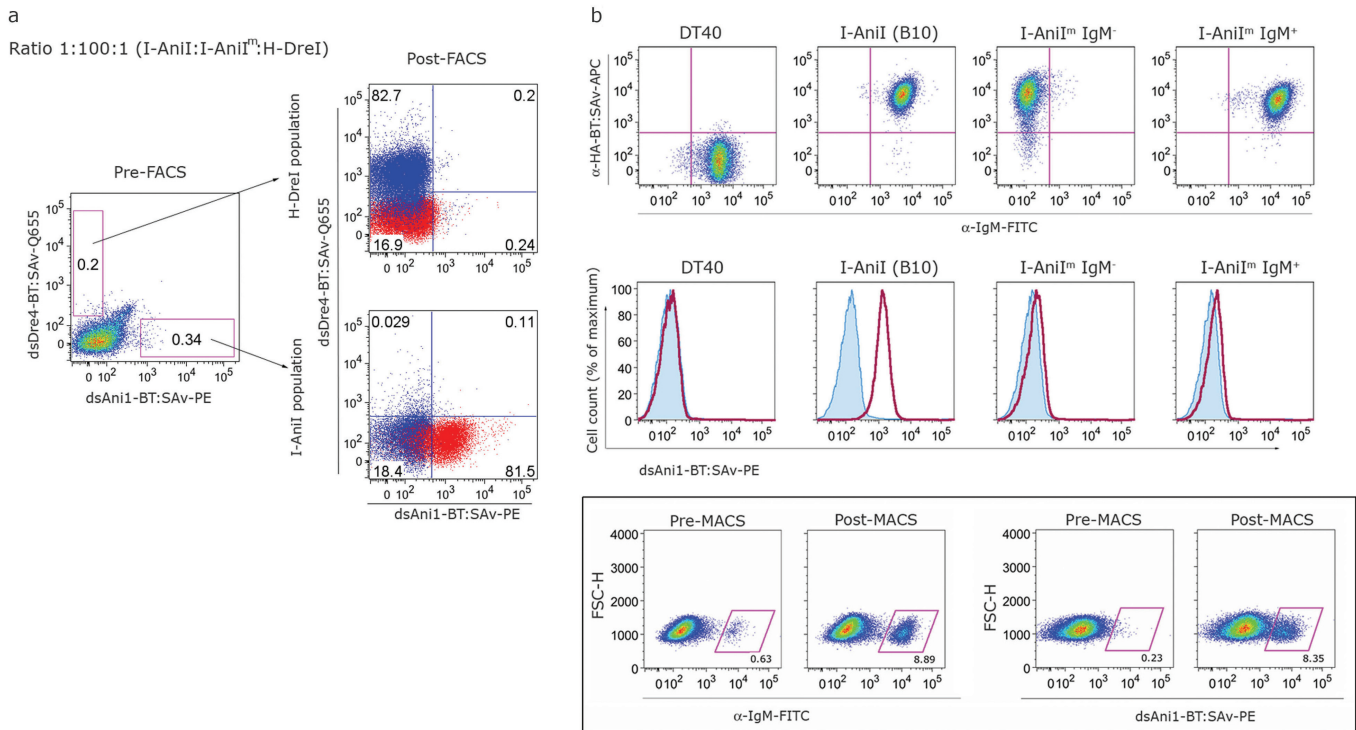
Figure 4. Fluorescent and magnetic strategies facilitate sequence specific sorting of cells expressing surface LHEs. (**a**) Three populations of cells expressing different LHEs (I-AniI, I-AniI^m and H-DreI) were mixed at a 1:100:1 ratio and double stained with dsAni1-BT: SAv-PE and dsDre4-BT: SAv-Q655, followed by FACS. The resulting sorted populations were cultured for 5–7 days prior to analysis and subsequent rounds of sorting. In post-sort analyses, cells stained with dsAni1 and dsDre4 are shown in red and blue, respectively. (**b**) Enrichment of low frequency dsOligo-binding cells by MACS. IgM-negative DT40 cells expressing I-AniI^m (top row, third panel) were used as a background population into which IgM-positive B10 cells were added at a frequency of 0.1%. IgM-positive I-AniI^m cells were included at 0.5% to control for potential background dsOligo binding caused by surface immunoglobulin expression, leading to a total of 0.6% IgM-positive cells in the input population, the majority of which do not stain with dsAni1. This mixed population was stained and sorted using AutoMACS (see Methods section for details). The positive fraction was grown out and analyzed for IgM expression. Staining with dsAni1 confirmed that the enriched IgM-positive population primarily expressed wild-type I-AniI.

mutations in the DNA binding and core regions of the enzyme.

Multi-parameter fluorescence-activated cell sorting (FACS) of cells labeled with dsOligos

We next evaluated whether our labeling method is suitable for sequence-dependent physical separation of LHE expressing cells by flow cytometry. We mixed three DT40 clones expressing different LHEs: clone B3 expressing I-AniI; clone C4 expressing H-DreI; and an I-AniI^m clone carrying a mutation proximal to the LAGLIDADG dimerization alpha-helix was utilized as the background population. The cells were mixed at a ratio of 1:100:1 for B3:I-AniI^m:C4 clones, respectively, and the mixed population was then stained with dsAni1-BT: SAv-PE and a quantum dot-conjugated dsDre4-BT: SAv-Q655. The dsAni1-specific and dsDre4-specific populations were isolated concurrently using fluorescence-activated cell sorting (FACS) and analyzed for their relative target specificities (Figure 4a). We achieved a significant enrichment of both I-AniI and H-DreI positive populations to 80% after the first round of sorting, and essentially no cross-contamination of the purified I-AniI or H-DreI populations was detected. The

capacity of dsOligo-dependent cell sorting was further explored by assessing the enrichment of low frequency I-AniI expressing cells from a background of I-AniI^m expressing cells, for which two iterative rounds of FACS sorting enriched an initial 0.01% population to 33% (Supplementary Figure 1). These data demonstrate that FACS sorting using fluorescently conjugated dsOligos is a highly effective method for the viable recovery of LHE expressing cells based on their DNA target specificity, and that rare clones with desired specificities may be isolated and enriched from large background populations.

Magnetic cell sorting rapidly isolates LHE expressing cells labeled with biotin-conjugated dsOligos

As an alternative to FACS, we tested the utility of magnetic cell sorting (MACS) for isolation of low-frequency LHE expressing cells (Figure 4b). The principle advantage of MACS is its ability to process extremely large sample sizes in short time periods (screening rates greater than 10^5 cells per second were routinely used in our protocols), thereby providing a convenient mechanism to sample large libraries of LHE clones. We employed an IgM-negative background population expressing high levels of an I-AniI^m clone

containing a mutated DNA-binding interface that was designed to eliminate direct contacts with one side of the asymmetric wild-type target sequence. Consistent low-level staining with dsAni1 indicates that low-affinity interactions with the wild-type target sequence are retained (Figure 4b, middle panels). The IgM-positive B10 clone expressing wild-type I-AniI was added at a frequency of 0.1%. The use of IgM as a surrogate marker for wild-type I-AniI expression allows for more accurate discrimination of low-percentage populations after dsOligo-dependent sorting due to a higher signal to noise ratio compared with dsOligo staining. To control for potential low-affinity interactions of dsOligos with IgM on the cell surface we included IgM-positive cells expressing I-AniI^m in the initial sample at a frequency of ~0.5%. The mixed population was labeled with dsAni1-BT in the presence 10 mM Ca^{2+}, followed by incubation with SAv-coated magnetic beads. Binding and non-binding fractions were isolated using a double-column positive selection protocol on an AutoMACS cell sorter. Our initial experiments indicate that 0.1% starting populations can be consistently enriched to by two orders of magnitude after a single round of MACS with sample sizes as large as 10^8 cells, despite residual low-affinity interactions with the bulk of cells expressing a mutated enzyme. Importantly, the enriched IgM-positive population was entirely composed of dsAni1-binding cells expressing wild-type I-AniI and not the IgM-positive fraction expressing I-AniI^m (Figure 4b, lower panels). Importantly, these results establish that high-level expression of surface molecules with the potential for both spurious (IgM) and specific (I-AniI^m) low-affinity interactions with DNA substrates do not compromise the specificity of dsOligo-dependent enrichment by MACS.

Flow cytometry-based cleavage assay for surface-expressed LHEs

To evaluate whether surface LHEs retained sequence-specific endonuclease activity, we designed LHE target sequences with two distinct fluorophores at opposite termini. Each oligo was modified at its 5′ terminus with either Alexa Fluor 647 or biotin during synthesis and were annealed to obtain dually conjugated dsOligos (647-dsOligo-BT, Figure 2a) which were mixed with SAv-PE at a 1:1 molar ratio to obtain a bifluorescent 647-dsAni1-BT: SAv-PE staining reagent. Cells were first labeled with a biotin-conjugated anti-HA mono-clonal antibody (α-HA-BT) followed by the addition of pre-formed 647-dsAni1-BT: SAv-PE complexes which should contain an average of three remaining BT-binding sites per SAv tetramer. This staining protocol serves to tether the 647-dsAni1-BT: SAv-PE to the cell surface independent of any specific LHE–dsOligo interaction, yet still placing the dsOligo within the LHE's immediate environment (Figure 5a). We reasoned that if the tethered 647-dsOligo-BT can be cleaved by the surface LHE, the cells would lose the fluorescence signal contribution from Alexa Fluor 647 yet retain signal from the tightly bound bridging SAv-PE.

As both antibody binding and SAv: BT interactions are independent of divalent cation contribution, we used a Ca^{2+} and Mg^{2+}-free buffer to stain I-AniI expressing cells with α-HA-BT followed by 647-dsAni1-BT: SAv-PE. The cells were then spiked with 10 mM Mg^{2+} and placed at 42°C in order to restore optimal cleavage conditions (29) (without Mg^{2+} for control samples). Using bifluorescent dsAni1, we were able to readily assay sequence-specific endonuclease activity by clones expressing wild-type I-AniI by monitoring changes in the fluorescence signals from each fluorophore (Figure 5b). Time-course experiments were performed to observe the relative disappearance of Alexa Fluor 647 fluorescence, which indicated that the signal progressively decreased during the first 20 min of incubation (data not shown). Given our data demonstrating the strict sequence specificity of the surface-expressed LHE DNA-binding interaction, we utilized bifluorescent dsAni1^{-9A} as a stringent control for the specificity of the cleavage reaction. Consistent with the clear differences in the binding data for these dsOligos, we observed no relative fluorescent signal changes for dsAni1^{-9A} under optimal cleavage conditions, confirming that dsAni1^{-9A} was not cleaved by the surface LHEs. We calculated the PE:647 fluorescence ratios and their relative changes with each dsOligo species as an indicator of the relative substrate cleavage. This quantification clearly demonstrates a substantial increase in the PE:647 ratio only where the bifluorescent dsOligo matched the natural target sequence for I-AniI (Figure 5c). One possible interpretation of this result is that the sequence-specific reduction of the Alexa Fluor 647 signal was due to fluorophore quenching following LHE binding and not necessarily from cleavage and release of the fragment. We therefore verified the presence of the cleaved fragment in the supernatants of cleavage experiments (Figure 5d). Importantly, the cells used for the cleavage reactions were analyzed by flow cytometry to confirm specific loss of the Alexa Fluor 647 signal (as in Figure 5b). Control cleavage assays were performed *in vitro* using recombinant I-AniI to confirm that 647-dsAni1-BT alone or complexed with SAv-coated beads was readily accessible and efficiently cleaved by the purified enzyme. In both experiments, we identified co-migrating fluorescent fragments of smaller molecular weight compared to full-length double-stranded and residual single-stranded oligonucleotides. Smaller fragments were not detected in controls with dsAni1^{-9A} or where the cleavage reaction was performed in the absence of either Mg^{2+} or I-AniI.

We next performed an experiment to confirm that the tethered dsOligos were being cleaved by LHEs on the very cells to which they were tethered. This is an important validation because cleavage caused by LHEs from adjacent cells might confound future attempts at FACS sorting by fluorescent signal loss following dsOligo cleavage. Using a mixed population of DT40 cells and I-AniI expressing (B3) cells at a 10:1 ratio where contacts between individual I-AniI expressing cells are decreased, we continued to observe sequence-specific reduction of Alexa Fluor 647 fluorescence to a similar extent as in a pure I-AniI-positive population (Figure 5e). We propose that individual dsOligos are primarily bound and digested
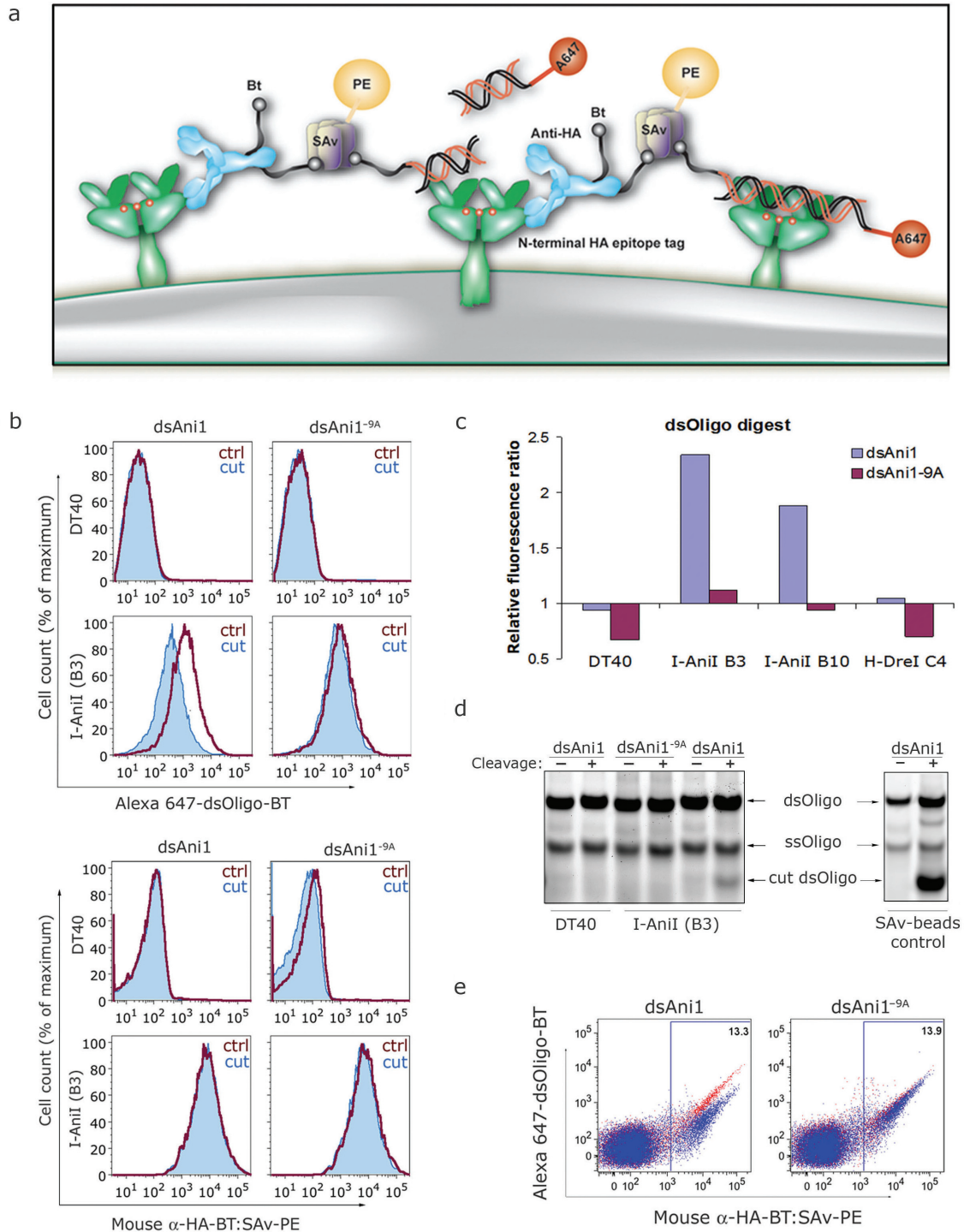
Figure 5. LHE mediated cleavage of cell surface-tethered dsOligo substrates conjugated with distinct fluorophores at opposite termini. (**a**) Schematic diagram for assaying surface LHE cleavage of α-HA-BT tethered dually fluorescent dsOligos and the release of Alexa Fluor 647 following addition of Mg^{2+} (red dots). (**b**) DT40 and B3 cells were stained with α-HA-BT followed by 647-dsOligo-BT: SAv-PE pre-formed complexes to tether the dsOligos to the surface LHE via the HA epitope. Cells with surface-tethered dsAni1 or dsAni1^{-9A} substrates were incubated at 42°C for 20 min with (filled histograms) or without (open histograms) Mg^{2+} and analyzed by flow cytometry. Although the fluorescence data was collected simultaneously, the fluorescence from Alexa Fluor 647 and PE are represented separately in the top and bottom panel sets, respectively, to demonstrate specific loss of the untethered fluorophore signal. (**c**) To quantify the extent of dsOligo cleavage by I-AniI, we calculated a ratio of the mean PE to Alexa Fluor 647 fluorescence intensities. Blue columns indicate changes in the PE:647 fluorescence ratio for dsAni1 cleavage whereas purple columns show relative ratio shifts for the dsAniI^{-9A} substrate. (**d**) DT40 cells and I-AniI expressing cells (B3) were stained as described in (b) and incubated at 42°C for 30 min in the presence (+) or absence (−) of Mg^{2+} (left panel). Also, 647-dsOligos-BT were bound to SAv-conjugated magnetic beads and incubated with recombinant I-AniI for 1 h at 42°C (right panel). DNA fragments were purified from supernatants and analyzed by PAGE followed by fluorescence imaging (see Methods section). (**e**) DT40 cells and I-AniI expressing cells (B3) were mixed at 10:1 ratio, labeled as described in (b) and incubated at 42°C for 20 min with (blue) or without (red) Mg^{2+} followed by flow cytometry analysis.

by LHEs autonomously on the cell surface. These results demonstrate that under optimal reaction conditions the surface-expressed LHEs are catalytically active and functionally recapitulate their highly sequence-specific nuclease activity.

DISCUSSION

We have generated a high-throughput flow cytometry based method for screening LAGLIDADG homing endonuclease DNA target specificity by expression and analysis of the normally intracellular endonucleases on the cell surface. Surface expression of LHEs followed by simple fluorescent staining protocols allowed us to verify cell surface display and DNA binding and cleavage specificity of LHEs with single cell resolution and high-throughput capacity. The precise recognition of their respective target sequence enabled multi-parameter FACS enrichment of specific LHE expressing cells. Low frequency clone enrichment by 3–4 orders of magnitude was accomplished from an initial 0.01% representation in a mixed population (Supplementary Figure 1). This was achieved with as little as two iterative selection rounds and we expect that this can be resolved for significantly lower frequencies and a greater number of fluorescent parameters with continued iteration of the process and/or with clone isolation by sorting into multi-well plates. Similarly, MACS achieved significant enrichment of I-AniI expressing cells from a low frequency mixed population following a single round of selection. Overall, the surface-expression approach is simple and appears to be readily applicable to multiple homing endonucleases of the LAGLIDADG family.

When compared to previously described methods used in selection experiments to identify LHEs with novel specificities, our approach has several useful properties. While transfection efficiencies are lower for mammalian cells, thus limiting transfected LHE library sizes, the potential exists to utilize somatic hypermutation to create and iteratively screen virtually limitless LHE libraries (30). In addition, LHEs expressed on the cell surface are physically separated from its host cell's genome. This allows for the possibility that iterative selections produce lower specificity endonucleases as intermediates to the generation of a high-specificity enzyme. Such intermediates would likely be toxic to a host organism if the endonuclease were expressed intracellularly where it would have access to host genomic DNA. Furthermore, this method allows rapid generation of quantitative information regarding binding and cleavage properties of novel LHE variants which otherwise would require *in vitro* expression and purification followed by electrophoretic analysis of gel shifts and cleavage. Finally, when combined with present generation flow cytometers, this approach offers significant flexibility in screening strategies. One may choose to screen a population of LHEs for multiple unique target sites in a single experiment. This approach would involve simultaneous staining with multiple dsOligos representing desired target sequences that are labeled with unique fluorophores. Alternatively, selective

population gating could enrich for desired specificities while simultaneously excluding undesired cross-reactivity as an approach to refining target sequence stringency. With the capacity to analyze dsOligo substrate cleavage, one could also potentially screen for both binding and catalytic activities in sequential flow cytometry assays to ensure that isolated binding properties do not compromise catalysis. Overall, the capacity to recapitulate DNA binding and cleavage by LHEs on the cell surface provides a high-throughput approach to LHE library screening that should facilitate the identification and analysis of enzymes with novel sequence specificities for use in a variety of genome engineering applications.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

Conflicts of interest statement. None declared.

REFERENCES

1. Chevalier,B.S. and Stoddard,B.L. (2001) Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. *Nucleic Acids Res.*, **29**, 3757–3774.
2. Heath,P.J., Stephens,K.M., Monnat,R.J.Jr and Stoddard,B.L. (1997) The structure of I-CreI, a group I intron-encoded homing endonuclease. *Nat. Struct. Biol.*, **4**, 468–476.
3. Galburt,E.A. and Stoddard,B.L. (2002) Catalytic mechanisms of restriction and homing endonucleases. *Biochemistry*, **41**, 13851–13860.
4. Jurica,M.S., Monnat,R.J.Jr and Stoddard,B.L. (1998) DNA recognition and cleavage by the LAGLIDADG homing endonuclease I-CreI. *Mol. Cell*, **2**, 469–476.
5. Chevalier,B., Turmel,M., Lemieux,C., Monnat,R.J.Jr and Stoddard,B.L. (2003) Flexible DNA target site recognition by divergent homing endonuclease isoschizomers I-CreI and I-MsoI. *J. Mol. Biol.*, **329**, 253–269.
6. Moure,C.M., Gimble,F.S. and Quiocho,F.A. (2003) The crystal structure of the gene targeting homing endonuclease I-SceI reveals the origins of its target site specificity. *J. Mol. Biol.*, **334**, 685–695.
7. Moure,C.M., Gimble,F.S. and Quiocho,F.A. (2002) Crystal structure of the intein homing endonuclease PI-SceI bound to its recognition sequence. *Nat. Struct. Biol.*, **9**, 764–770.
8. Burt,A. and Koufopanou,V. (2004) Homing endonuclease genes: the rise and fall and rise again of a selfish element. *Curr. Opin. Genet. Dev.*, **14**, 609–615.
9. Duan,X., Gimble,F.S. and Quiocho,F.A. (1997) Crystal structure of PI-SceI, a homing endonuclease with protein splicing activity. *Cell*, **89**, 555–564.

10. Chevalier,B.S., Kortemme,T., Chadsey,M.S., Baker,D., Monnat,R.J. and Stoddard,B.L. (2002) Design, activity, and structure of a highly specific artificial endonuclease. *Mol. Cell.*, **10**, 895–905.
11. Epinat,J.C., Arnould,S., Chames,P., Rochaix,P., Desfontaines,D., Puzin,C., Patin,A., Zanghellini,A., Paques,F. *et al.* (2003) A novel engineered meganuclease induces homologous recombination in yeast and mammalian cells. *Nucleic Acids Res.*, **31**, 2952–2962.
12. Arnould,S., Chames,P., Perez,C., Lacroix,E., Duclert,A., Epinat,J.C., Stricher,F., Petit,A.S., Patin,A. *et al.* (2006) Engineering of large numbers of highly specific homing endonucleases that induce recombination on novel DNA targets. *J. Mol. Biol.*, **355**, 443–458.
13. Steuer,S., Pingoud,V., Pingoud,A. and Wende,W. (2004) Chimeras of the homing endonuclease PI-SceI and the homologous Candida tropicalis intein: a study to explore the possibility of exchanging DNA-binding modules to obtain highly specific endonucleases with altered specificity. *Chembiochem*, **5**, 206–213.
14. Storici,F., Durham,C.L., Gordenin,D.A. and Resnick,M.A. (2003) Chromosomal site-specific double-strand breaks are efficiently targeted for repair by oligonucleotides in yeast. *Proc. Natl Acad. Sci. USA*, **100**, 14994–14999.
15. Tzfira,T., Frankman,L.R., Vaidya,M. and Citovsky,V. (2003) Site-specific integration of Agrobacterium tumefaciens T-DNA via double-stranded intermediates. *Plant Physiol.*, **133**, 1011–1023.
16. Miller,D.G., Petek,L.M. and Russell,D.W. (2003) Human gene targeting by adeno-associated virus vectors is enhanced by DNA double-strand breaks. *Mol. Cell Biol.*, **23**, 3550–3557.
17. Smith,J., Grizot,S., Arnould,S., Duclert,A., Epinat,J.C., Chames,P., Prieto,J., Redondo,P., Blanco,F.J. *et al.* (2006) A combinatorial approach to create artificial homing endonucleases cleaving chosen sequences. *Nucleic Acids Res.*, **34**, e149.
18. Sussman,D., Chadsey,M., Fauce,S., Engel,A., Bruett,A., Monnat,R.Jr, Stoddard,B.L. and Seligman,L.M. (2004) Isolation and characterization of new homing endonuclease specificities at individual target site positions. *J. Mol. Biol.*, **342**, 31–41.
19. Seligman,L.M., Chisholm,K.M., Chevalier,B.S., Chadsey,M.S., Edwards,S.T., Savage,J.H. and Veillet,A.L. (2002) Mutations altering the cleavage specificity of a homing endonuclease. *Nucleic Acids Res.*, **30**, 3870–3879.
20. Ashworth,J., Havranek,J.J., Duarte,C.M., Sussman,D., Monnat,R.J.Jr, Stoddard,B.L. and Baker,D. (2006) Computational redesign of endonuclease DNA binding and cleavage specificity. *Nature*, **441**, 656–659.
21. Doyon,J.B., Pattanayak,V., Meyer,C.B. and Liu,D.R. (2006) Directed evolution and substrate specificity profile of homing endonuclease I-SceI. *J. Am. Chem. Soc.*, **128**, 2477–2484.
22. Chames,P., Epinat,J.C., Guillier,S., Patin,A., Lacroix,E. and Paques,F. (2005) In vivo selection of engineered homing endonucleases using double-strand break induced homologous recombination. *Nucleic Acids Res.*, **33**, e178.
23. Chou,W.C., Liao,K.W., Lo,Y.C., Jiang,S.Y., Yeh,M.Y. and Roffler,S.R. (1999) Expression of chimeric monomer and dimer proteins on the plasma membrane of mammalian cells. *Biotechnol. Bioeng.*, **65**, 160–169.
24. Liao,K.W., Chou,W.C., Lo,Y.C. and Roffler,S.R. (2001) Design of transgenes for efficient expression of active chimeric proteins on mammalian cells. *Biotechnol. Bioeng.*, **73**, 313–323.
25. Bolduc,J.M., Spiegel,P.C., Chatterjee,P., Brady,K.L., Downing,M.E., Caprara,M.G., Waring,R.B. and Stoddard,B.L. (2003) Structural and biochemical analyses of DNA and RNA binding by a bifunctional homing endonuclease and group I intron splicing factor. *Genes Dev.*, **17**, 2875–2888.
26. Mohler,W.A. and Blau,H.M. (1994) Membrane-bound neomycin phosphotransferase confers drug-resistance in mammalian cells: a marker for high-efficiency targeting of genes encoding secreted and cell-surface proteins. *Somat. Cell Mol. Genet.*, **20**, 153–162.
27. Chevalier,B.S., Monnat,R.J.Jr and Stoddard,B.L. (2001) The homing endonuclease I-CreI uses three metals, one of which is shared between the two active sites. *Nat. Struct. Biol.*, **8**, 312–316.
28. Chevalier,B., Sussman,D., Otis,C., Noel,A.J., Turmel,M., Lemieux,C., Stephens,K., Monnat,R.J.Jr *et al.* (2004) Metal-dependent DNA cleavage mechanism of the I-CreI LAGLIDADG homing endonuclease. *Biochemistry*, **43**, 14015–14026.
29. Geese,W.J., Kwon,Y.K., Wen,X. and Waring,R.B. (2003) In vitro analysis of the relationship between endonuclease and maturase activities in the bi-functional group I intron-encoded protein, I-AniI. *Eur. J. Biochem.*, **270**, 1543–1554.
30. Wang,L., Jackson,W.C., Steinbach,P.A. and Tsien,R.Y. (2004) Evolution of new nonantibody proteins via iterative somatic hypermutation. *Proc. Natl Acad. Sci. USA*, **101**, 16745–16749.