

Engineering domain fusion chimeras from I-Onul family LAGLIDADG homing endonucleases

Sarah Baxter^{1,2,3}, Abigail R. Lambert³, Ryan Kuhar³, Jordan Jarjour⁴, Nadia Kulshina⁵, Fabio Parmeggiani⁶, Patrick Danaher⁶, Jacob Gano⁴, David Baker⁷, Barry L. Stoddard⁵ and Andrew M. Scharenberg^{1,3,*}

¹Department of Immunology, ²Medical Scientist Training Program, University of Washington, Seattle, WA, 98195, ³Center for Immunity and Immunotherapies, Seattle Children's Research Institute, Seattle, WA, 98101, ⁴Pregenen Inc., Seattle, WA, 98103, ⁵Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, 98109, ⁶Department of Biochemistry and ⁷Department of Biostatistics, University of Washington, Seattle, WA, 98195, USA

Received March 16, 2012; Revised April 27, 2012; Accepted May 5, 2012

ABSTRACT

Although engineered LAGLIDADG homing endonucleases (LHEs) are finding increasing applications in biotechnology, their generation remains a challenging, industrial-scale process. As new single-chain LAGLIDADG nuclease scaffolds are identified, however, an alternative paradigm is emerging: identification of an LHE scaffold whose native cleavage site is a close match to a desired target sequence, followed by small-scale engineering to modestly refine recognition specificity. The application of this paradigm could be accelerated if methods were available for fusing N- and C-terminal domains from newly identified LHEs into chimeric enzymes with hybrid cleavage sites. Here we have analyzed the structural requirements for fusion of domains extracted from six single-chain I-Onul family LHEs, spanning 40–70% amino acid identity. Our analyses demonstrate that both the LAGLIDADG helical interface residues and the linker peptide composition have important effects on the stability and activity of chimeric enzymes. Using a simple domain fusion method in which linker peptide residues predicted to contact their respective domains are retained, and in which limited variation is introduced into the LAGLIDADG helix and nearby interface residues, catalytically active enzymes were recoverable for ~70% of domain chimeras. This method will be useful for

creating large numbers of chimeric LHEs for genome engineering applications.

INTRODUCTION

Rare-cleaving endonucleases are valuable tools for genome engineering, as they create double-strand breaks that become substrates for cell-intrinsic DNA repair pathways, enabling high efficiency sequence modification at or near their cleavage sites (1–4). Resolution of an endonuclease-induced DNA double-strand break through mutagenic non-homologous end joining (NHEJ) results in the generation of small insertions or deletions that can be exploited to disrupt a target gene's coding sequence (5,6). Alternatively, repair via the homologous recombination (HR) pathway with the codelivery of a rare-cleaving nuclease and a synthetic homologous repair template can achieve a variety of gene targeting outcomes (7–13).

Three platforms are available for generating customized rare-cleaving endonucleases for genome engineering: zinc-finger nucleases (ZFNs), TAL-effector nucleases (TALENs) and LAGLIDADG homing endonucleases (LHEs) (7,8,14–16). Whereas ZFNs and TALENs target a DNA hydrolysis reaction to a distinct target sequence by coupling the non-specific endonuclease domain from FokI with separate sequence-specific DNA-binding moieties, the hydrolytic active site of LHEs is integrated into their DNA-binding interface. The LHE protein family includes both homodimeric proteins, in which a single LAGLIDADG motif-containing subunit dimerizes to create a functional enzyme, and pseudo-symmetric

*To whom correspondence should be addressed. Tel: +1 206 987 7314; Fax: +1 206 987 7310; Email: andrewms@u.washington.edu
Present address:

Jacob Gano, Synthetic Biology Division, LS9 Inc, South San Francisco, CA 94080, USA.

monomers, where two structurally related domains, each possessing a single LAGLIDADG motif and similar folded topologies, are directly connected by a peptide linker. In the case of monomeric endonucleases, the N- and C-terminal protein domains (NTDs and CTDs) are individually responsible for recognition of the 5' and 3' half-sites of their corresponding DNA target sites.

Of the platforms listed above, LHEs offer several unique advantages for genome engineering. These include: (i) naturally high levels of specificity and a corresponding absence of genotoxicity observed when wild-type LHEs are expressed in a variety of cell types; (ii) small size, with a typical single-chain LHE open reading frame measuring 800–1000 bp; and (iii) a significant capacity for multiplexed use, as single-chain LHEs can function autonomously (17–19). While the importance of genomic-level specificity for therapeutic applications is obvious, the naturally small size of LHEs is also beneficial, as these compact enzymes are compatible with a wide range of both viral and non-viral vectorization strategies. Compatibility with viral vectors is particularly important for nuclease-based genome editing applications in primary cells where plasmid-based transfection approaches or the use of mRNA may be impractical (20). Similarly, as genome engineering strategies become increasingly complex, the ability of genome editing reagents to function autonomously becomes essential for applications where multiple genetic manipulations must be carried out simultaneously.

Although the unique properties of LHEs have driven their continued development as a genome editing platform, large-scale engineering of LHEs to cleave novel DNA target sequences remains challenging. Accumulating experience with both homodimeric and monomeric LHE scaffolds suggests that while engineering small changes to an enzyme's native cleavage target is generally well tolerated and can be readily achieved, the increased numbers of changes required for more radical alteration of specificity can exponentially increase both cost and effort, and often leads to less stable and less efficient enzymes (21,22). These challenges have significantly limited the widespread application of LHEs in genome engineering.

The identification of a large set of LHEs encompassing a wide range of target specificities would provide an alternative to the current paradigm. The availability of many diverse LHE scaffolds would allow a starting scaffold to be chosen with a recognition sequence closely matching the desired target, thus minimizing the engineering required to produce a high-quality, respecified enzyme. Although increasing numbers of novel single-chain LAGLIDADG nucleases have been identified from sequence databases, the total set of enzymes available as design scaffolds remains relatively limited. However, the structure of single-chain LHEs suggests that individual NTDs and CTDs from different parental single-chain LHEs could be fused into chimeric enzymes that cleave hybrid targets, as has been previously accomplished using the homodimeric LAGLIDADG enzyme I-CreI and the monomeric LAGLIDADG enzyme I-DmoI (23–25). An efficient structure-independent method for generation of

such chimeras would provide a rapid means to substantially expand the set of scaffolds available as starting points for redesign.

Here, we have systematically evaluated methods for creating functional enzymes by the fusion of individual NTDs and CTDs extracted from six members of a recently described group of pseudo-dimeric single-chain LHEs (26,27). Methods for choosing a linker peptide, introducing interface variation, and determining cleavage specificity across the central four (C4) base pairs of chimeric target sites were developed through analysis of fused domains from I-OnuI and I-LtrI, for which crystal structures are available. Insights from this work were incorporated into a structure-independent method for fusion of domain pairs. Using this approach, we were able to recover active chimeric enzymes from ~70% of attempted fusions. Taken together, our results suggest that a limited number of native single-chain LHEs enzymes can be expanded into a very large group of chimeric enzymes for use as design scaffolds, greatly facilitating the rapid generation of site specific nucleases for genome engineering applications.

MATERIALS AND METHODS

DNA constructs

The sequences of I-OnuI, I-LtrI, I-GpiI, I-GzeI, I-PanMI and I-SscMI were codon optimized for expression in both bacteria and yeast and synthesized by Genscript (Piscataway, NJ) into the pETCON vector (a hybrid of the pCTCON2 yeast surface expression vector with cloning sites from the pET vector series). This vector creates a fusion of the inserted protein sequence to the surface-expressed Aga2P yeast surface protein, and also incorporates an N-terminal hemagglutinin (HA) epitope tag and a C-terminal Myc epitope tag (used for fluorescent antibody staining). Individual NTDs and CTDs of the I-OnuI homologs were constructed by gene assembly PCR. Assembly primers (50–70 bp) were designed using the DNAWorks server (Helix Systems, <http://helixweb.nih.gov/dnaworks/>) and synthesized by Integrated DNA Technologies (IDT). For generation of libraries, randomized positions were introduced using assembly oligonucleotides with degenerate codons (NNS). The formulation used for synthesis of these randomized oligonucleotides was specified to be 'hand-mixed' by the manufacturer to ensure equal ratios of each nucleotide (Sigma and IDT). After transformation into yeast, the resulting library sizes were determined to consist of >10 million variants. Chimeras with the 'SGT' linker substitution were constructed by digestion of full-length enzyme with KpnI and either NdeI (for isolation of the NTD) or XhoI (CTD) (NEB). Digested fragments were purified using the Qiagen PCR Cleanup Kit (Qiagen), and combined in equimolar concentrations with a partner domain for ligation into the pETCON vector using T4 DNA ligase (NEB). Ligated DNA was transformed into chemically competent DH5 α cells and sequenced to isolate full-length clones; plasmid preparations of these clones were then transformed into yeast using the lithium

acetate protocol (28). See Supplementary Figure S1 for DNA and protein sequences used in all applications.

Modeling

Models of the Onu-Ltr and Ltr-Onu chimeras were created in Pymol (The PyMOL Molecular Graphics System, Version 1.5.0.1 Schrödinger, LLC.) by superposition of the I-OnuI (PDB 3QQY) and I-LtrI (PDB 3R7P) coordinates. The artificial helical linker tested with the Ltr-Onu chimera was originally designed for use in the wild-type I-OnuI structure. A short span of the linker is disordered in the I-OnuI crystal and, therefore, is missing from the deposited structure. The structure-building program Coot was used to model an ideal α helix across the missing portion of the I-OnuI structure (29). The length of the helix was trimmed to span the length of the gap (seven total residues), and amino acid sidechains were chosen to (i) encourage helix formation and (ii) pack against the I-OnuI surface (Lambert, A.R., unpublished data). Calculation of domain interface properties and energetics were performed using Rosetta (32–34).

Substrates for binding and cleavage assays

Biotinylated and fluorophore-conjugated double-stranded oligonucleotides (ds-oligos) were generated by PCR and purified from single-stranded contaminants by ExoI digestion (Fermentas) followed by size exclusion through a G-50 sephadex column (GE Healthcare). The final ds-oligos were analyzed by gel electrophoresis to be >98% pure. See Supplementary Figure S2 for oligonucleotide sequences used in all applications.

Yeast growth, transformation and plasmid recovery

Saccharomyces cerevisiae strain EBY100 was transformed using the lithium-acetate protocol described by Gietz and Schiestl (28). Yeast were grown in selective media (SC) with 2% glucose at 30°C overnight, followed by dilution and growth in SC+2% raffinose+0.1% glucose at 30°C for 12–20 h, to a density of 90–150 million cells/ml. Cells were then induced in SC+2% galactose for 2–3 h at 30°C, followed by 12–18 h at 20°C. Plasmids were isolated from yeast using the Zymoprep-II kit (Zymo Research). Plasmids were then chemically transformed into *Escherichia coli* DH10B (Invitrogen) for subsequent amplification and sequencing.

Flow cytometry expression, binding and cleavage assays

Expression, binding and cleavage activity of the yeast surface-expressed LHEs was quantified using flow-cytometry-based assays modified from the published protocol by Jarjour *et al.* (2009) (30). Briefly, expression was measured by incubating $0.25\text{--}0.5 \times 10^6$ induced yeast cells per sample in 100 μ l yeast staining buffer (YSB) [10 mM HEPES, 10 mM NaCl, 180 mM KCl, 5 mM CaCl₂, 0.1% galactose, 0.2% BSA, pH 7.5], containing biotin-conjugated anti-Myc antibody (ICL). Cells were incubated for 1–2 h at 4°C, washed with an excess of buffer, and then counter-stained with streptavidin-allophycocyanin (APC) for 1 h at 4°C. Binding activity

of surface-expressed LHEs was determined by incubating 0.5–50 nM fluorophore-labeled ds-oligo with $\sim 2\text{--}5 \times 10^5$ cells/sample in 100 μ l YSB (yielding an estimated 100 pM enzyme concentration, assuming $10^4\text{--}10^5$ molecules per yeast surface), supplemented with 5 mM calcium. Yeast were incubated for 2 h at 4°C to achieve equilibrium, washed and stained with fluorescein isothiocyanate (FITC)-conjugated anti-Myc antibody (ICL Labs). Cleavage activity of the surface-expressed LHEs was quantified using Jarjour *et al.*'s on-cell cleavage assay: $2.5\text{--}5 \times 10^5$ cells were stained with biotinylated anti-HA antibody (Covance) in YSB, washed and then stained with pre-conjugated streptavidin-PE (5 nM):biotin-ds-oligo-A647 (50 nM) in YSB supplemented with additional KCl to a final concentration of 580 mM (high-salt YSB). The high salt condition prevents binding of the ds-oligo by the expressed LHE, thus encouraging correct formation of the desired antibody-mediated tethering. Cells were washed and transferred to oligo cleavage buffer (OCB) [150 mM KCl, 10 mM NaCl, 10 mM HEPES, 0.5 mg/ml BSA, pH 8.25], with 5 mM MgCl₂ (for catalytic activity) or CaCl₂ (for binding without cleavage). These samples were incubated at 37°C for 15 min–1 h, and then washed with the high-salt YSB to release cleaved DNA. Cells were then incubated with FITC-conjugated anti-Myc antibody to determine concentration of enzyme on the yeast surface, as described above. Samples were run on a BD LSRII™ cytometer (BD Biosciences) or sorted using a BD FACSARIAII, and data was analyzed with FloJo software (Tree Star, Inc.).

In vitro cleavage assay

The *in vitro* cleavage assay was performed as described in Jarjour *et al.* (2009) (30). Briefly, 5–10 million induced yeast were incubated in 50 μ l YSB, as described above ($\sim 15\text{--}30$ nM enzyme), supplemented with 5 mM MgCl₂ or CaCl₂, 10 mM DTT (to release enzyme from the surface of yeast) and 20 nM Alexa 647-conjugated ds-oligo substrate, at 37°C for 15–60 min. Supernatants were run on a 15% non-denaturing polyacrylamide gel, and visualized using an Odyssey infrared imaging system (Li-Cor Biosciences).

In vitro HEK293T-cell culture assay

Open reading frames for I-LtrI, I-OnuI and Onu-Ltr were amplified by PCR and ligated into the CVL lentiviral backbone using the In-Fusion cloning system (Clontech Bioinformatics), for analysis in the Traffic Light Reporter (TLR) assay, as described in Certo *et al.* (1). Target sites for each enzyme were inserted into the TLR construct using standard molecular biology techniques. Lentivirus was produced as described previously (31). Briefly, HEK293T cells were transiently cotransfected with 6 μ g CVL-backbone TLR plasmids, 1.5 μ g pMD2G envelope plasmid (VSV-G) and 3 μ g psPAX2 for viral packaging. Cells were incubated in 10 ml DMEM without Phenol Red supplemented with 3–4% FBS and glutamine. Forty-eight hours post-transfection, viral supernatant was collected, filtered and stored at 4°C before being frozen at -80°C .

TLR cell lines were created by transducing 0.2×10^6 HEK293T cells with 0.5, 1 and 2 μ l of their respective unconcentrated reporter lentivirus. Three days after transduction, cells with integrated reporters were selected by treatment with 1 μ g/ml puromycin for 5 days. The cultures with the lowest number of surviving cells (those initially receiving 0.5 μ l lentivirus) were chosen as the final cultures and sorted using a BD FACSAriaII to remove background mCherry fluorescence resulting from integration errors.

For each experiment, 0.1×10^6 HEK293T cells were seeded in a 24-well plate 24 h prior to transfection. Cells were transiently transfected with 0.5 μ g of HE-expression construct, with the addition of 0.5 μ g of eGFP repair template for gene targeting experiments, using X-tremeGENE 9 DNA transfection reagent using the recommended manufacturer protocols (Roche Applied Science). Twenty-four hours after transfection, cells were split into a 12-well plate. Cells were collected 72 h after transfection and analyzed on a BD LSRII™ for BFP, mCherry and GFP fluorescence. A total of 0.1×10^6 cells per well were acquired for analysis. FloJo software (TreeStar, Inc) was used to analyze the flow cytometry data.

Protein expression and purification

Onu-Ltr and Ltr-Onu genes were subcloned into pET24b vectors with a stop codon preceding the C-terminal His-tag. Proteins were expressed in BL21 pLysS cells, and purified on a buffer gradient heparin column (50 mM to 1 M NaCl, 50 mM Tris, pH 8.0), followed by Super DEX gel filtration in 0.5 M NaCl, 50 mM Tris, pH 8.0 buffer. Proteins were concentrated and glycerol was added to 5% for storage. I-OnuI and I-LtrI were expressed and purified as described previously (27).

Circular dichroism melting curves

Circular dichroism (CD) thermal denaturation experiments were performed at 10 μ M protein concentration in 150 mM NaCl, 50 mM phosphate buffer. Measurements were made using a JASCO J-815 CD spectrometer with a Peltier thermostat. CD ellipticity at 220 nm was measured for samples in a 0.1-cm pathlength cell. The spectral bandwidth was 1.0 nm, and the response time was 8 s. Denaturation was performed over a 25°C to 96°C temperature range. The melting temperature was determined using JASCO software. Percent folded protein was determined using the formula $(X_{\text{obs}} - X_u)/(X_n - X_u) \times 100\%$, where X_n is the molecular ellipticity of the native protein, X_{obs} is the observed molecular ellipticity and X_u is the molecular ellipticity of fully denatured protein. X_n and X_u were determined by linear extrapolation of the folded and unfolded baselines to 25°C and 96°C, respectively.

RESULTS

Direct fusion of individual NTDs and CTDs extracted from I-OnuI and I-LtrI

With the goal of developing general principles for the fusion of NTDs and CTDs extracted from native single-chain LHEs, we began our studies by determining the

stability and catalytic properties of fusions of individual NTDs and CTDs derived from I-OnuI and I-LtrI: N'Onu-C'Ltr (Onu-Ltr) and N'Ltr-C'Onu (Ltr-Onu) (27). I-OnuI and I-LtrI were chosen for pilot studies because both of their crystal structures have been determined, thus offering the best opportunity to derive general insights into efficient generation of highly active chimeric enzymes. The structures of these two enzymes display remarkable homology, especially at the LAGLIDADG helices that form the primary interacting interface (Figure 1A). Furthermore, models of interface packing for Onu-Ltr and Ltr-Onu domain fusions, generated using ROSETTA macromolecular modeling, suggest that the I-OnuI NTD and I-LtrI CTD would be as energetically compatible with each other as they are with their native domain partners (e.g. Figure 1A and Table 1) (32–34).

To generate Onu-Ltr and Ltr-Onu chimeras, the open reading frames for the NTDs and CTDs of I-OnuI and I-LtrI were fused at the conserved residue P162 in I-OnuI (P160 in I-LtrI). To evaluate the behavior of the chimeric enzymes, we expressed them on the surface of yeast. In the yeast surface display method, an LHE is fused to the secreted Aga2P protein and expressed in the EBY100 *S. cerevisiae* strain under the control of a galactose-inducible promoter. Assuming comparable levels of transcription and translation, stability is generally correlated with surface expression in yeast, as unstable proteins are retained by the yeast secretory pathway, limiting their expression on the surface (35,36). We also used CD to measure the *in vitro* thermal stability of purified recombinant protein. Catalytic activity of the surface expressed enzymes can be assessed using a flow-cytometric on-cell cleavage assay, which measures the loss of a fluorophore due to cleavage of a labeled, double-stranded DNA target substrate which has been physically tethered to the surface-expressed enzyme (30,37).

As predicted by ROSETTA calculations of interface energetics, Onu-Ltr showed strong surface expression and was stable to 52°C, while Ltr-Onu showed significantly decreased surface expression and thermal stability in the CD assay (Figure 1B and C). Similarly, Onu-Ltr demonstrated cleavage activity against its putative DNA target comparable to that of its parental enzymes, while Ltr-Onu had reduced, albeit quite obvious, activity (Figure 1D and Supplementary Figure S2). The relative activities, as measured by the flow-cytometric cleavage assay, were further assessed by an *in vitro* cleavage assay (incorporating target binding efficiency) in which a non-tethered, fluorescently-labeled DNA target substrate is incubated with surface-released yeast enzyme and the resulting fragments visualized on a polyacrylamide gel. Using this *in vitro* assay, both Onu-Ltr and Ltr-Onu chimeras exhibited catalytic activity comparable to the on-cell yeast cleavage assay, and also demonstrated specificity for their predicted hybrid targets, as neither chimera cleaved the target sequences of the native I-OnuI or I-LtrI enzymes, nor did the chimeras cleave an unrelated target sequence (Figure 1E).

Although both chimeras exhibited detectable cleavage activity, the activity of Onu-Ltr appeared to be equivalent

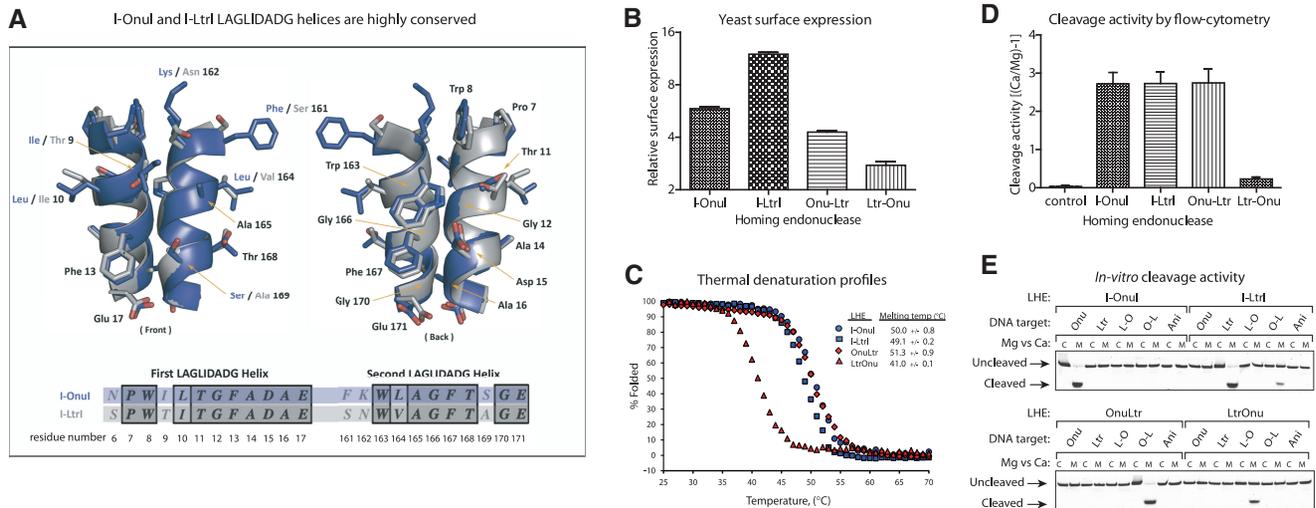


Figure 1. Comparison of Onu-Ltr and Ltr-Onu *in vitro* stability and activity. **(A)** Overlaid crystal structures of the LAGLIDADG helices of I-OnuI (blue) and I-LtrI (gray), which form the majority of the interface between the NTDs and CTDs. Residues are indicated by both name and corresponding sequence number. An alignment of the amino acid sequences illustrates their high level of conservation. **(B)** Comparison of enzyme expression levels in the flow-cytometric yeast surface display assay. Expression levels are quantified by intensity of fluorescent FITC signal (anti-Myc-FITC antibody bound to the C-terminal Myc epitope tag). Each bar represents the ratio of median FITC signal from ‘expressing’ versus ‘non-expressing’ cell populations. **(C)** Thermal denaturation was monitored by CD as an alternative measure for comparison of overall protein stability. See the ‘Materials and Methods’ section for details of data collection and calculations. **(D)** Comparison of DNA cleavage activity measured by the flow-cytometric yeast surface display assay. Activity is quantified by loss of A647 signal upon cleavage of a tethered, fluorescently-labeled DNA target substrate. Each bar represents the ratio of A647 signal from cells in the presence of calcium (no cleavage) to cells in the presence of magnesium (allows cleavage) minus one $[(Ca/Mg) - 1]$. A height of zero represents no detectable cleavage activity. The unrelated I-AniI DNA target site was used as the negative control. Cleavage reactions were incubated for 30 min at 37°C. Data represents five to seven separate experiments; in each individual experiment, all enzymes’ signals were normalized to the I-OnuI signal. **(E)** DNA cleavage activity measured by the *in vitro* gel cleavage assay. A647-labeled DNA target substrate was incubated with surface-released yeast protein in the presence of calcium (no cleavage) or magnesium (allows cleavage) and visualized on an acrylamide gel. Each homing endonuclease was assayed against the I-OnuI, I-LtrI, Ltr-Onu, Onu-Ltr, and I-AniI target sequences to detect any off-target activity. Cleavage reactions were incubated for 1 h at 37°C. (Compilation of three separate gels, run in parallel).

Table 1. Rosetta calculations for I-OnuI and I-LtrI chimeras

| NTD | CTD | $\Delta sasa$ | Δ_{tot} | Δ_{att} | Δ_{rep} | Δ_{hb} | Δ_{sol} | Δ_{dun} | Δ_{pair} | ΔdG |
|--------|--------|---------------|----------------|----------------|----------------|---------------|----------------|----------------|-----------------|-------------|
| I-OnuI | I-OnuI | 1755.1 | -37.8 | -204.7 | 33.3 | -11.6 | 102.8 | 40.3 | -2.7 | -18.5 |
| I-LtrI | I-LtrI | 1590.7 | -22.9 | -188.1 | 47.1 | -12.9 | 93.5 | 35.9 | -3.7 | -11.5 |
| I-OnuI | I-LtrI | 1703.9 | -36.8 | -213.9 | 42.5 | -15.0 | 111.5 | 36.8 | -3.8 | -18.4 |
| I-LtrI | I-OnuI | 1530.9 | 36.9 | -185.2 | 91.1 | -11.5 | 91.8 | 39.5 | -2.0 | 18.5 |

NTDs and CTDs are listed in the first two columns: wild-type I-OnuI and I-LtrI are shown in the first two rows. All values represent differences. ‘ $\Delta SASA$ ’ is the surface area that is shielded from solvent upon interaction of the two domains, expressed as angstroms squared. All other terms represent energetic differences between the domains when considered separately versus together, and are expressed in Rosetta units (RU). Therefore, these values measure the energetic contribution of the interface. ‘ Δ_{Tot} ’ is the total score. ‘ Δ_{Att} ’ is the attractive component of the Lennard-Jones potential, representing van der Waals interactions. ‘ Δ_{Rep} ’ is the repulsive component of the Lennard-Jones potential. ‘ Δ_{Hb} ’ is the contribution from hydrogen bonds, both from backbone and side chain atoms. ‘ Δ_{Sol} ’ is the solvation/desolvation term. ‘ Δ_{Dun} ’ is derived from the Dunbrack rotamer library and considers the frequency of side chain rotamers (32–34). ‘ Δ_{Pair} ’ refers to the interaction of full and partial charges. Other terms that contribute to a lesser extent to the total energy are not individually listed, but are included in ‘ Δ_{Tot} ’.

to that of its parental native enzymes. Since I-OnuI and I-LtrI both perform extremely well in cell-based assays, we compared the activity of Onu-Ltr to its wild-type parental enzymes using a recently developed *in vivo* system designed to simultaneously measure both NHEJ and HR resulting from endonuclease cleavage events in an integrated reporter cassette (1). The Onu-Ltr chimera, similar to native I-OnuI and I-LtrI, expressed efficiently in the reporter cells via transient transfection, as determined by expression of a BFP tag coupled to the

enzyme (Figure 2A). Onu-Ltr expression induced +3 frameshift mutations due to nonconservative end-joining (as measured by mCherry expression) in ~4% of cells; this rate was equivalent to disruption rates induced by I-LtrI against its native target and comparable to or slightly greater than that induced by I-OnuI (Figure 2B). The biological explanation for varying rates of HR and NHEJ observed among I-OnuI homologs is uncertain, and could include transcriptional timing, or the rate of enzyme release from cleaved DNA. Cleavage of the

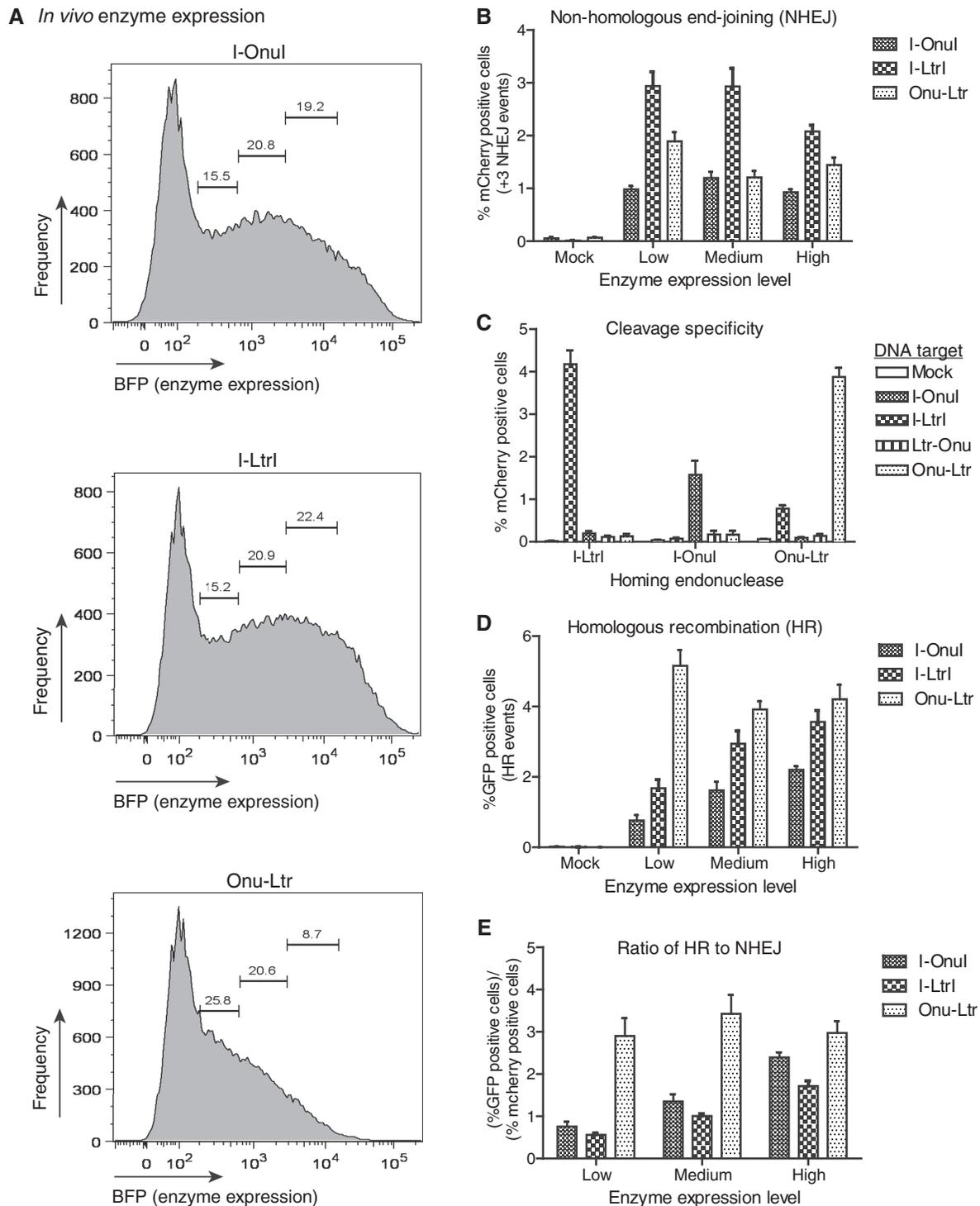


Figure 2. *In vivo* activity of Onu-Ltr compared to native enzymes. (A) Plasmids containing I-Onul, I-LtrI and Onu-Ltr with an N-terminal BFP tag were transfected into HEK293T cells containing the corresponding Traffic Light Reporter target plasmid. Enzyme expression is quantified by coexpressed BFP fluorescence. Low, medium and high levels of expressed enzyme are gated by BFP fluorescence to determine relative rates of NHEJ and HR. (B–E) *In vivo* activity of Onu-Ltr in the NHEJ vs HR Traffic Light Reporter assay. Cleavage of a target site plasmid can be repaired by either NHEJ or HR, and relative levels of each repair pathway can be simultaneously visualized using this reporter assay. (B) Mutagenic NHEJ events leading to +3 frameshifts are detected by mCherry fluorescence. mCherry-positive events represent ~33% of total mutagenic NHEJ events. (C) *In vivo* cleavage specificity. Mutagenic NHEJ (detected by mCherry fluorescence) was measured for each enzyme against the related I-Onul, I-LtrI, and chimeric DNA target sites. (D) Repair of a cleaved target site by the HR pathway (in the presence of a cotransfected GFP donor template) is detected by fluorescence of a correctly reconstituted GFP sequence. (E) Ratio of HR events (% GFP positive cells) to mutagenic NHEJ events resulting in +3 frameshifts (% mCherry positive cells).

I-LtrI target by Onu-Ltr was also visible at low rates (Figure 2C), corroborating the observation that I-LtrI has low-level catalytic activity against the Onu-Ltr DNA target in the *in vitro* gel cleavage assay (Figure 1F), and suggesting that the CTD of I-LtrI may allow for some degree of promiscuity in cleavage, even when incorporated within a domain fusion chimera. HR events induced by Onu-Ltr were increased ~2-fold over those induced by I-OnuI and/or I-LtrI, emphasizing the high level of performance achieved in the chimeric enzyme (Figure 2D and E).

From the collective results above, we conclude that the NTDs and CTDs from both I-OnuI and I-LtrI can be effectively fused into active chimeric enzymes. These results further suggest that domains extracted from other single-chain I-Onu family members possessing homology comparable to that of I-OnuI and I-LtrI (~40% identity and 65% similarity) might also be excellent substrates for fusion into active chimeric enzymes.

Benchmarking a general method for fusion of individual NTDs and CTDs

We sought to develop efficient general strategies for (i) extraction of individual domains from a parental, pseudo-dimeric single-chain LHE and (ii) fusion of these domains into active chimeric enzymes. To this end, we next analyzed three aspects of the single-chain LHE structure–function relationship in I-OnuI and I-LtrI and their domain fusion chimeras: (i) the extent to which the peptide linking the NTDs and CTDs contributes to individual NTD and CTD function; (ii) the influence of interactions at the domain interface on the stability and activity of chimeric enzymes and (iii) the extent to which the central 4 nt in the parental target sites are conserved in the target site of a chimera, given that indirect protein–DNA interactions dictate the often high specificity at these nucleotides in the native enzymes.

The linker peptide between domains contributes to enzyme stability and activity

Although the successful domain fusions of Onu-Ltr and Ltr-Onu suggest that domains extracted from other I-Onu family single-chain LHEs could be compatible, an important region of sequence divergence throughout the enzyme family corresponds to the linker peptide connecting the NTDs and CTDs. The linker peptide is highly divergent even between otherwise highly homologous enzymes in the I-Onu family, to the extent that there is no clear position within the linker for dividing and combining NTDs and CTDs (Figure 3A). Though previous studies have demonstrated considerable flexibility in linking NTDs and CTDs from homodimeric LAGLIDADG enzymes, the role of the inter-domain linker in the stability and enzymatic activity of single-chain LHEs has not been examined (39,40).

To understand to what extent the linker peptide might contribute to the successful fusion of domains, we generated a set of Ltr-Onu domain fusion chimeras with linker peptides of varying structure. The Ltr-Onu chimera was chosen as our linker test scaffold based on the

hypothesis that the moderate level of stability and activity observed for this chimera in our pilot studies would allow for optimal sensitivity in measuring changes in activity due to choice of the linker. For the purpose of constructing the linker test chimeras, the NTD extracted from I-LtrI terminated at position 148 of the I-LtrI sequence. The CTD extracted from I-OnuI began at the conserved proline (P162 in I-OnuI and 160 in I-LtrI) at the top of the C-terminal LAGLIDADG helix and continued through the end of the I-OnuI ORF. The set of linker variants evaluated included: the native I-LtrI linker, used in the initial fusion study; the native I-OnuI linker; a linker peptide designed with high α helical content for stability (to evaluate whether a generic, artificial peptide would be compatible with Onu and Ltr domain fusion); and two hybrid ‘1/2-and-1/2’ linkers, with residues derived from the linker peptides of both the NTDs and CTDs, connected by a tri-peptide bridge that replaces a section of the linker which is poorly conserved in the I-OnuI family (Figure 3A, B and Supplementary Figure S3).

The two hybrid linkers preserved residues in the connecting regions that interact with their own domains, as observed in the available crystal structures. Two different sets of bridging residues were tested: (i) an ‘NGN’ tri-residue bridge that was suggested by computational analysis to be compatible with both the I-OnuI and I-LtrI structures and (ii) an ‘SGT’ tri-residue bridge, based on its predicted flexibility and broad structural compatibility. Each of these linkers was incorporated into Ltr-Onu, replacing the residues that lie between P149 and P164 (Figure 3B). Ltr-Onu chimeras with these variant linker peptides were evaluated using the flow-cytometric yeast surface display assay (37). All linker variants were stably expressed on the surface of yeast. Interestingly, the variant including the full native I-LtrI-derived linker (the original gene-synthesized version of Ltr-Onu direct fusion) exhibited significant catalytic activity, while that including a full native I-OnuI-derived linker was completely inactive, demonstrating that linker peptide composition can indeed have an important influence on single-chain LHE function.

Similarly, although the helical linker preserves full enzymatic activity of native I-OnuI (data not shown), and marginally increases the stability of Ltr-Onu, it is not able to support catalytic activity in the Ltr-Onu context (Figure 3C–E). Ltr-Onu variants incorporating the hybrid ‘1/2-and-1/2’ linkers showed stability and activity equivalent to the Ltr-Onu chimera incorporating the native I-LtrI linker peptide, which included only three residues from the C-terminal helix (Figure 3C–E). To further evaluate the ‘SGT’ hybrid linker approach for use in larger scale domain fusion experiments, we used it to generate new versions of I-OnuI and Onu-Ltr, and compared catalytic activity to that of the wild-type I-OnuI and Onu-Ltr direct fusion chimera, respectively. Incorporation of the ‘SGT’ tri-residue bridge did not alter the stability or catalytic activity of native I-OnuI, and resulted in only a slight change in activity of Onu-Ltr, visible in the *in vitro* gel cleavage assay (Figure 3F–H).

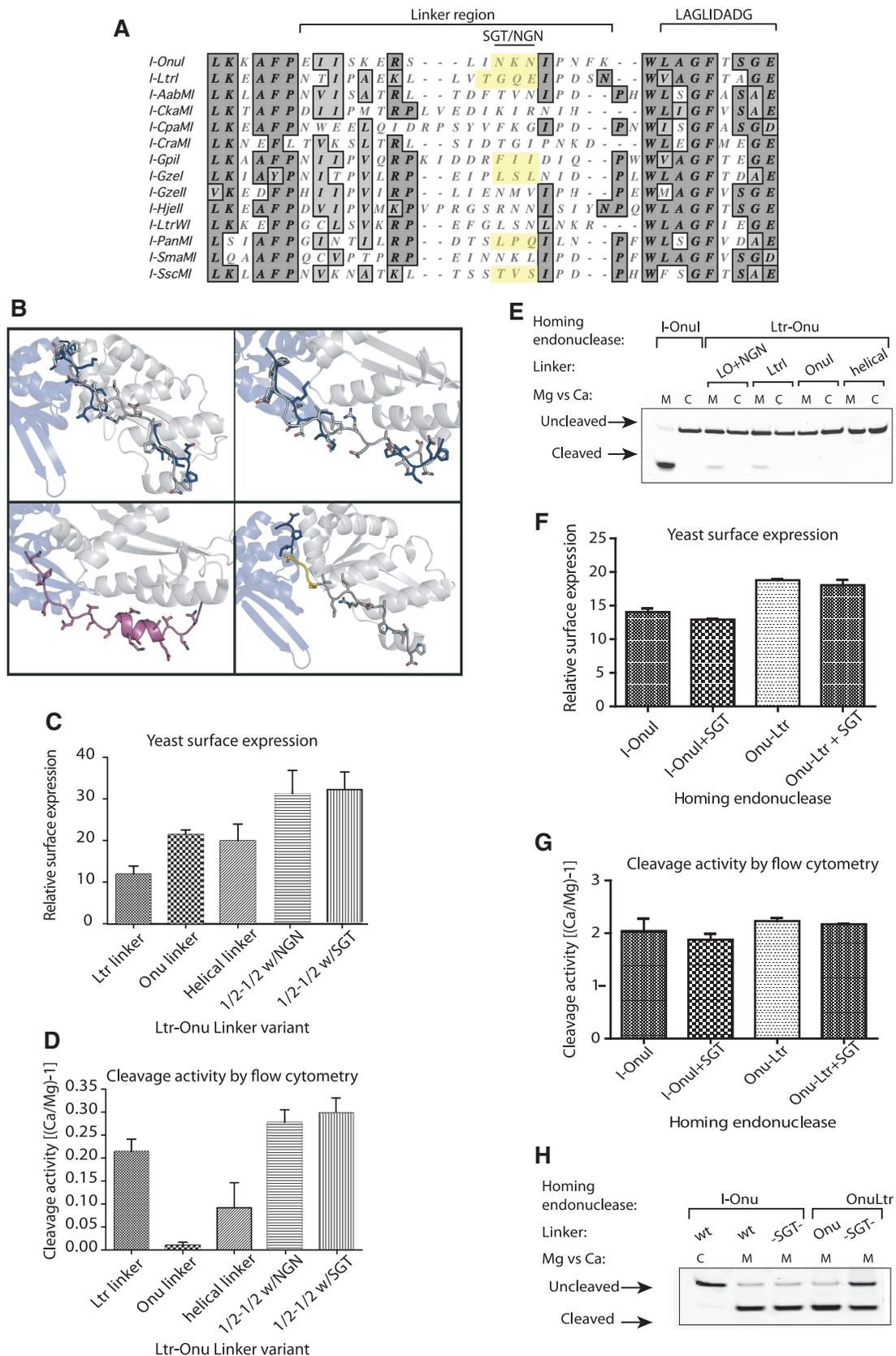


Figure 3. Linker peptide variations. Analysis of Ltr-Onu expression and activity with various linker strategies. (A) Alignment of the highly variable linker peptide sequences from 14 characterized I-OnuI homologs. Brackets indicate the linker peptide sequence and the position of the second LAGLIDADG helix. Residues replaced by the ‘SGT’ or ‘NGN’ flexible linkers are highlighted in yellow. (B) Models of the Ltr-Onu chimera illustrating the various linkers tested. Top left: Superposition of native I-OnuI (blue) and I-LtrI (gray) linkers, top view. A small portion of the

(continued)

Taken together, these data demonstrate that interactions between the linker and the NTD have an important influence on activity but not stability, consistent with the concept that the linker peptide may subtly influence the relative position of the two domains. In designing chimeras, our data suggest that the majority of the linker should be derived from the NTD. The data also emphasize the importance of accounting for the influence of the linker during the development of a general strategy aimed at producing fusions between NTDs and CTDs extracted from single-chain LHEs.

DNA-distal LAGLIDADG motif residues contribute to chimera stability

Although the LAGLIDADG helices are highly conserved within the I-OnuI family, the active sites of LHEs depend on the precise orientation of the two domains and their respective LAGLIDADG helices. The moderate stability and activity profile of the Ltr-Onu chimera suggested that the hybrid interface was slightly suboptimal, and that introduction of variation within both the LAGLIDADG helix and nearby interface residues might allow for enhanced recovery of stable, active enzymes from domain fusion experiments. We therefore used chimera models (based off the I-OnuI and I-LtrI structure) and sequence alignments of I-OnuI family LHEs to predict residues that would be most likely detrimental for packing and stability of Ltr-Onu. This analysis identified four residues at the DNA-distal end of the LAGLIDADG helices, along with two residues in two side loops (Figure 4A). These residues show extreme diversity within the I-OnuI family, and the residues at the distal ends of the LAGLIDADG helices have been previously targeted for engineering LHE dimeric interfaces (41).

To experimentally evaluate the importance of these residues, we created an Ltr-Onu library, from the original Ltr-Onu fusion enzyme with an I-LtrI-derived linker, of over 20 million variants by fully randomizing the six chosen residues, and analyzed the library using yeast surface display. Approximately 2% of the library yielded stable, high surface expressing enzymes. These yeast were sorted, expanded, reinduced and resorted for variants with detectable cleavage activity using the flow-cytometric cleavage assay (39,42). The top 1–2% of cleaving variants were selected and reanalyzed by yeast surface display. This analysis revealed a selected population with markedly improved surface expression, along with significantly increased catalytic activity compared

to the original direct-fusion chimera, although the resulting cleavage activity did not reach the level of either parental enzyme (Figure 4B).

Sequence analysis of the recovered population demonstrated strong patterns of residue selection in the sorted Ltr-Onu variants (Figure 4C and Supplementary Table S1). Three of the positions tested, including the two residues present on loops interacting with the opposite domain, showed conservative selection, with S6 in Ltr-Onu being strongly selected for both serine and threonine, T50 selected for serine, asparagine and threonine, and V154 selected for isoleucine and leucine. D161 and K163, immediately preceding the second LAGLIDADG helix, were primarily represented by I-OnuI residues, suggesting that these positions are most strongly influenced by adjacent residues within their own domain rather than interactions across the chimeric interface. Other positions showed more compelling and radical selections: at T9 in Ltr-Onu, a position in the interfacial region which is an isoleucine in I-OnuI and a threonine in I-LtrI, large aromatic residues were incorporated in a majority of solutions. Structural modeling suggests that the substitution of an aromatic at this position could allow more compact packing of the enzyme, thus accounting for the improved surface display properties. Interestingly, in alignments of I-OnuI homologs, this position is primarily held by large aromatics. The selection for a similar residue within the Ltr-Onu chimera, despite neither parental enzyme possessing an aromatic at the corresponding position, is consistent with the idea that incorporation of a large hydrophobic at this position may have a uniformly stabilizing effect on I-OnuI family domain interfaces.

Overall, these data suggest that incorporation of sequence variation into even a limited number of domain interface residues is adequate to allow the rapid isolation of domain fusion chimeras with improved performance.

Chimeras maintain predicted specificity at the 'C4' base pairs their target sequences

A potentially confounding factor in the analysis of a chimeric single-chain LHE is whether a simple bipartite DNA target site, composed of exactly half of each parental site, is consistently a valid substrate (Figure 5A). The four middlemost bases of a DNA target sequence cleaved by any type of LHE (designated the 'C4' base pairs) are typically not directly contacted by amino acid residues; rather, they appear to be read out indirectly through energetics related to the kinking and unwinding

Figure 3. Continued

I-OnuI linker is missing from the structure due to disorder in the crystal. Top right: Superposition of native I-OnuI (blue) and I-LtrI (gray) linkers, side view. Bottom left: Artificial helical linker (magenta), originally designed for use with wild-type I-OnuI. Bottom right: Half-and-half linker with 'SGT' residues highlighted yellow. (C) Comparison of expression levels in the flow-cytometric yeast surface display assay. Expression levels were quantified by intensity of fluorescent APC signal (antibody staining of a C-terminal Myc epitope tag). Each bar represents the ratio of median APC signal from the 'expressing' versus 'non-expressing' cell populations. (D) Comparison of DNA cleavage activity measured by the flow-cytometric yeast surface display assay. Activity is quantified by loss of A647 signal upon cleavage of a fluorescently-labeled DNA target substrate. Each bar represents the ratio of A647 signal from cells in the presence of calcium (no cleavage) to cells in the presence of magnesium (allows cleavage) minus one $[(Ca/Mg) - 1]$. A height of zero represents no detectable cleavage activity. Reactions were incubated at 37°C for 30 min. (E) Effect of linker variation on catalytic activity, as measured by the *in vitro* gel cleavage assay. A647-labeled target substrate was incubated (for 30 min at 37°C) with surface-released yeast protein in the presence of calcium (no cleavage) or magnesium (allows cleavage) and visualized on an acrylamide gel. (F–H) Comparison of the expression and cleavage activity of I-OnuI and Onu-Ltr with the 'SGT' linker, as measured by the flow-cytometric yeast surface display assay and the *in vitro* cleavage assay (as described above in parts C–E).

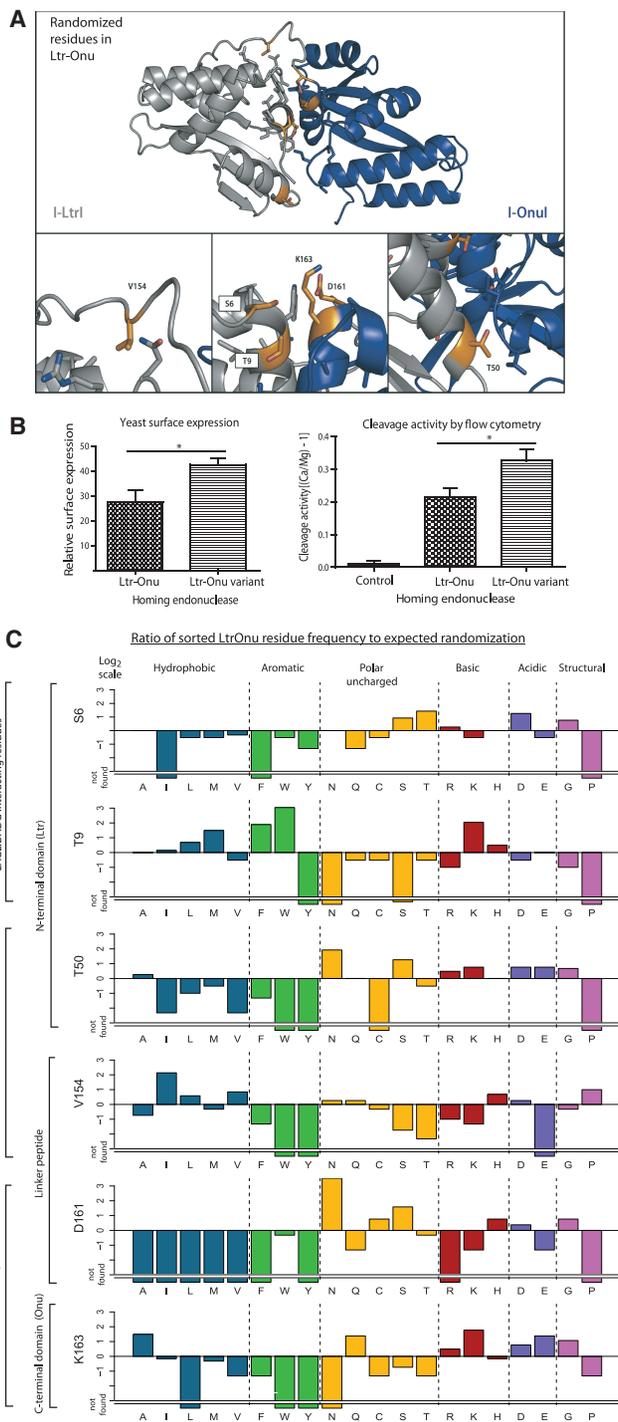


Figure 4. Variation of DNA-distal LAGLIDADG residues. Ltr-Onu variants were selected for increased expression and cleavage activity from a library with six randomized interface residues. (A) Randomized residues (colored orange) were chosen in three separate locations: the DNA-distal end of the LAGLIDADG helices (bottom middle), and in loops on either side of the central helices (bottom left and bottom right). (B) Left: Yeast surface expression is increased in the sorted Ltr-Onu library, as measured by FITC staining of the C-terminal Myc epitope tag. Each bar represents the ratio of median FITC signal from the ‘expressing’ versus ‘non-expressing’ cell populations. Right: Activity is quantified by loss of A647 signal upon cleavage of a fluorescently-labeled DNA target substrate. Each bar represents the ratio of A647 signal from cells in the presence of calcium (no cleavage) to cells in the presence of magnesium (allows cleavage) minus one

of target DNA observed in LHE/DNA structures (17). This is especially important given the limitations of engineering at the central 4 nt: with a large database of starting scaffolds, design for a given target would likely begin with a search for the scaffold with the closest identity to the desired sequence. Only after this search might the given chimeric scaffold be constructed. Therefore, understanding the extent to which the optimal C4 target of a chimeric enzyme diverges from those of its parental enzymes is essential to developing a general approach to generating domain fusion chimeras.

To evaluate whether C4 cleavage specificity is substantially altered in chimeric enzymes generated by domain fusion, we screened the activity of both parental enzymes and both domain fusion chimeras against panels of C4 targets; I-OnuI was screened against a subset of these targets, whereas I-LtrI, and the domain fusion chimeras were screened against all 256 (for these analyses, we used the sorted stabilized Ltr-Onu variant as it allowed increased sensitivity) (Figure 5B–D). This screen showed that the chimeric enzymes possess optimal or near-optimal activity against bipartite hybrid DNA targets (i.e. those consisting of exact fusions of 5′- and 3′-DNA half-sites from the original parental targets). In the case of Onu-Ltr, one other C4 target sequence—ATAA, differing in one nucleotide from the bisected ATAC—was cleaved with high efficiency. Four of the six targets showing moderate cleavage by Onu-Ltr differed from the optimal sites by only 1 bp (Figure 5C). Likewise, Ltr-Onu showed optimal catalytic activity against the bisected C4 variant ATTC, as well as two sequences—AATC and TTTC—varying by 1 bp. A majority (7/11) of the sequences against which Ltr-Onu displayed moderate activity also differed by only 1 nt (Figure 5D).

The majority of C4 wobble/promiscuity lay in the -1 and $+1$ positions, with the -2 and $+2$ positions more strictly conserved in accordance with the parent enzyme’s target sequence. The total rates of off-target cleavage agreed well with observations for the native I-LtrI (Figure 5B): I-LtrI effectively cleaves five C4 variants, including its native ATAC, and shows moderate activity against an additional seven variants. Analyses of I-OnuI against a smaller target set (Supplementary Figure S4) showed that it effectively cleaves four C4 variants, including its native ATTC, and shows low or moderate activity against an additional 11 variants. Overall, these data indicate that domain fusion chimeras are likely to maintain the high level of C4 specificity characteristic of their parental single-chain LHEs, and that the general usage of a predicted bipartite hybrid site to assess optimal cleavage activity of a domain fusion

Figure 4. Continued

[(Ca/Mg) – 1], as described in Figure 1. Asterisk indicates $P < 0.05$. (C) Approximately 150 clones from the sorted Ltr-Onu library were sequenced. Post-selection variation at each randomized position is represented by fold increase or decrease over the expected frequency (given complete randomization). Fold increase/decrease is presented along a log (2) axis. Residues that were not detected are scaled below the broken line. Selected residues are divided into groups with biochemically similar sidechains (hydrophobic, aromatic, polar uncharged, basic, acidic, structural).

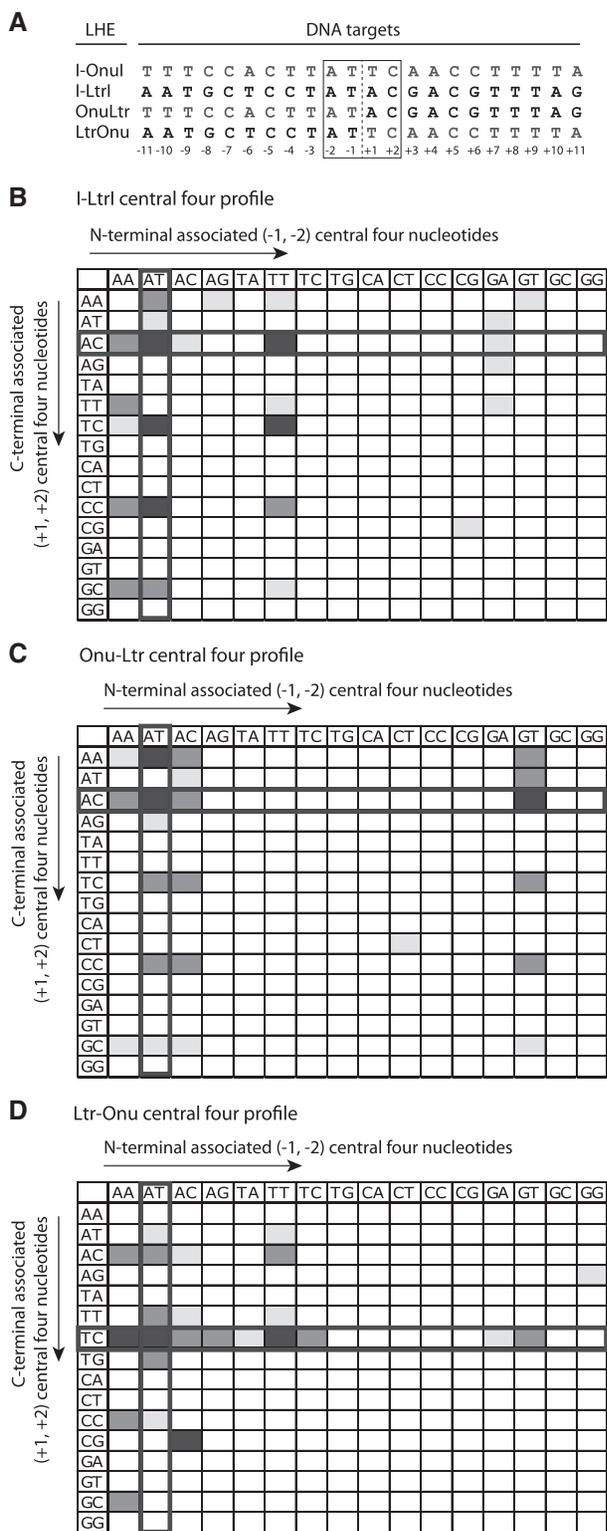


Figure 5. Specificity of I-LtrI, Onu-Ltr and Ltr-Onu chimeras against C4 target variants. (A) The 22-bp DNA recognition sequences for the Onu-Ltr and Ltr-Onu chimeras and their parental enzymes. The target sites are divided into a ‘minus half’ and a ‘plus half’, which are contacted by the NTDs and CTDs of the enzyme, respectively. The numbering scheme is indicated below the target sites, with the C4 base pairs boxed. (B) The catalytic activity of I-LtrI was analyzed in the flow-cytometric yeast surface display DNA cleavage assay against all potential 264 C4 target nucleotides. Nucleotides at positions -1 and -2 (interacting with the N terminal domain) are listed on the x-axis, with

chimera is reasonable. Moreover, the comprehensive nature of these C4 profiling experiments has uncovered a much higher degree of specificity within this region of the target site than has previously been identified.

Large-scale generation of chimeras by fusion of I-OnuI homolog domains

Large-scale generation of domain fusions with retention of full native interfaces

The data from our benchmarking studies of Onu-Ltr and Ltr-Onu led us to evaluate a general strategy for the structure-independent generation of domain fusion chimeras, in which NTDs and CTDs are extracted from parental I-Onu family single-chain LHEs in the following manner: NTDs are defined as starting six amino acids upstream from a conserved proline in the N-terminal LAGLIDADG helix, and ending eight residues upstream from a conserved tryptophan in the C-terminal LAGLIDADG helix; CTDs start five residues upstream from a conserved tryptophan in the C-terminal LAGLIDADG helix, and run through the end of the protein. A three residue ‘SGT’ bridge sequence with a KpnI restriction site is incorporated at the end of NTDs, and at the beginning of CTD. Using this approach, NTDs and CTDs can thus be rapidly extracted from their parental enzymes and fused into chimeric enzymes, singly or in combination, by digestion with the appropriate restriction enzymes, ligation into the yeast display vector pETCON and transformation into yeast.

To assess the potential of this approach for generating a greatly expanded set of novel LHE scaffolds for engineering, we generated domain fusions of all possible combinations of NTDs and CTDs extracted from I-OnuI, I-LtrI and four additional I-OnuI family homologs that have been identified and characterized in our lab, I-GpiI, I-GzeI, I-PanMI and I-SscMI (Supplementary Figure S5). These enzymes share ~40% amino acid sequence identity, with the exception of I-GzeI and I-PanMI, which share >70% sequence identity. Of the 36 enzymes made in total, 6 were reconstituted native enzymes with the ‘SGT’ tri-residue substitution in the linker peptide, and 30 were novel chimeras. Expression and binding of each chimeric enzyme was assessed by flow cytometry using yeast surface-displayed enzyme, and cleavage activity was determined by both the *in vitro* DNA cleavage assay (Figure 6A) and by flow cytometry. A summary of data for surface expression, binding and cleavage activity from this set of enzymes is shown in Figure 6E, left panel.

Importantly, all six reconstituted native enzymes exhibited surface expression and activity comparable to

Figure 5. Continued nucleotides at positions +1 and +2 (interacting with the CTD) listed on the y-axis. Boxes containing the native central 4nt for each domain are outlined in bold. Cleavage activity against each nucleotide combination is illustrated as a heat-map: white represents no measurable catalytic activity, with a Ca/Mg ratio of < 1.1 in the cleavage assay. Light grey, grey, and black represent low, medium and high levels of cleavage activity. (C–D) Catalytic activity of Onu-Ltr (C) and Ltr-Onu (D) against all 264 potential C4 target nucleotides.

their native forms (Figure 6A, E), validating our choice of location for division of the linker peptides in the native enzymes, and supporting the concept that an ‘SGT’ bridge incorporated into the linker is likely to be compatible with the vast majority of single-chain LHE enzymes.

The surface expression profiles for the domain fusion chimeras demonstrated that nearly half (14/30) were stable, well-folded enzymes. The percent of stable chimeras with measurable binding to their putative target was ~79% (11/14). Of these expressing and binding enzymes, cleavage was detectable in 81% (9/11) (Figure 6A, E, left panel and Supplementary Figure S6). A small number of chimeras with low expression showed some degree of cleavage, suggesting that the enzymes were minimally stable and therefore very weakly expressed by the yeast, but a minority were still able to fold appropriately and cleave their targets. Four chimeras were able to bind their putative targets, but showed no cleavage activity: N-terminal I-GpiI fused with C-terminal I-PanMI (Gpi-Pan), in particular, showed very strong binding but no cleavage activity. In order to verify that Gpi-Pan was not catalytically active against a slightly different target, we analyzed cleavage against the 16 C4 possibilities varying only 1 nt away from the predicted Gpi-Pan target. Gpi-Pan did not show cleavage against any of the alternative C4 targets (data not shown), indicating that it is unlikely that this chimera is able to form a catalytically competent complex despite a high-affinity interaction with the DNA substrate.

Large-scale generation of domain fusions with a uniform ‘common interface’

Two striking observations emerged from the above survey of simple domain fusions. First, domain-specific biases were prominent for the CTDs: the subset of CTDs extracted from I-GpiI, I-GzeI and I-SscMI were widely incompatible with domain fusion, resulting in enzymes with little to no activity; conversely, the subset of CTDs extracted from I-OnuI, I-LtrI and I-PanMI were widely compatible, resulting in several enzymes with near native levels of activity. Second, only 19% of chimeras demonstrated binding without catalytic activity, and likewise only 21% of stably expressed chimeras did not bind their putative target. Based on this, we hypothesized that the primary hurdle to successful domain fusion might lie in determining a compatible interface. The active site is functional in a majority of the stable proteins, suggesting a high degree of transferability of catalysis while maintaining catalytic specificity. Because inadequate interactions within the chimeric domain interface could be a primary destabilizing factor (despite high sequence conservation within the LAGLIDADG helices), we evaluated the use of a graftable ‘common interface.’ This approach has been previously attempted successfully via structure guided design for I-DmoI and I-CreI, despite the relative dissimilarity of those enzymes (38).

For determination of an appropriate common interface, both inspection of structures and computational predictions were used to identify the interacting interfacial residues in I-OnuI and I-LtrI (Figure 6B and C). The designated residues from native I-OnuI were grafted

onto each chimera (keeping the ‘SGT’ linker), with sequence alignments used to predict the equivalent interfacial residues in I-GpiI, I-GzeI, I-PanMI and I-SscMI (designated as CI1, ‘common interface 1’). The Onu interface was chosen for grafting, as the structure of I-OnuI was available to us, allowing an unambiguous choice of interface residues, and because I-OnuI is the most well-characterized member of the family. Because the substitutions previously selected for stabilization of Ltr-Onu were predicted to be potentially more energetically favorable for the entire set of domain fusions, we also created a second set of common-interface chimeras including the residues selected for Ltr-Onu at the DNA-distal end of the LAGLIDADG helices (designated as CI2, ‘common interface 2’).

With the CI1 interface, half (15/30) of the chimeras stably expressed on the surface of yeast, with 80% (12/15) of the expressing chimeras showing binding of their putative target, and 92% (11/12) of these binding enzymes demonstrating catalytic activity (Figure 6D and E, right panel). The majority of enzymes previously demonstrating activity by simple domain fusion maintained some level of activity, and likewise many of the chimeras that were not previously stable or active remained so (Figure 6E, Supplementary Figure S6). For the CI2 interface, catalytic activity was increased in a limited number of cases, most impressively in N/I-LtrI-C/I-PanMI (Ltr-Pan) (Supplementary Figure S7).

Several important patterns become evident in the cleavage activities observed for ‘common-interface’ domain fusion chimeras. First, for domain fusions involving either the N- or CTDs of I-OnuI, in which interfacial residues were only substituted on the partner domain (since the common interface residues are derived from I-OnuI), an increased success rate was observed. Cleavage activity was substantially increased in N/I-PanMI with C/I-OnuI (Pan-Onu), and rescued in N/I-GzeI with C/I-OnuI (Gze-Onu). Similarly, the increased activity of the N/I-LtrI-C/I-PanMI (Ltr-Pan) chimeric enzyme is notable, as it includes the NTD of I-LtrI, for which these interface residues had originally been selected. Second, Gpi-Pan, which was stable and able to bind its putative target as a simple domain fusion chimera, gained partial catalytic activity with a grafted ‘common interface,’ suggesting that stable chimeras are promising candidates for further optimization with potentially only a limited number of changes. Finally, it was striking that catalytic activity of I-GpiI and I-SscMI were ablated by the swapping of interfacial residues, and activities of I-PanMI and I-GzeI were decreased. The significant changes in activity resulting from the substitution of interfacial residues emphasize three key points regarding both native and chimeric single chain LHEs: (i) the positioning of the NTDs and CTDs to form stable, active chimeras is significantly influenced by the interfacial residues we identified; (ii) despite their relatively high sequence identity and structural homology, the interfacial interactions are sufficiently diverged among native enzymes that introduction of variation within this residue set is required to consistently isolate stable and active chimeric enzymes; and

(iii) in forming domain fusion chimeras, choosing interfacial residues common to one or the other of the domains appears to increase the likelihood of forming stable and active chimeras.

DISCUSSION

Here, we have systematically explored the potential of domain fusion to expand the number of native pseudo-dimeric single-chain LHE scaffolds for genome engineering applications, focusing on the recently described I-OnuI family (26,43). To establish parameters for extraction of NTDs and CTDs from single-chain LHEs, and for development of a structure-independent method for generation of these domain fusion chimeras, we examined the structure/function relationships of chimeras generated by fusion of NTDs and CTDs extracted from I-OnuI and I-LtrI. Using insights from this work, we systematically generated domain fusion chimeras from I-OnuI, I-LtrI and four other I-OnuI family enzymes, and characterized their biochemical properties using yeast surface display. Our results suggest that simple direct fusion approaches can yield active enzymes in ~50% of cases, and that introduction of even limited variation into the interface residues allows for recovery of active enzymes from ~70% of domain fusion pairs.

A significant result emerging from our studies is that the linker peptide in single-chain LHEs forms not only important, predictable interactions with the NTD, but also functionally impacts the LAGLIDADG interface. Even when using a hybrid '1/2-and-1/2' approach, which was designed to conserve important linker interactions, and which preserved activity in all native enzymes (e.g. Figure 6E, left panel), we observed a few examples where alteration of linker composition led to a decrease in activity (e.g. incorporation of an 'SGT' bridge into Onu-Ltr, Figure 3E). Therefore, in contrast to the flexible parameters that may be used in designing linkers to create single-chain versions of the homodimeric enzyme I-CreI, it is evident that the linker peptides in single-chain enzymes have evolved to interact in a meaningful manner with the domains, as well as with the interfacial region (39). Linker composition must therefore be taken into account in LHE engineering, not only in the development of a strategy to generate chimeric enzymes, but also

potentially in both later stage optimization of a chimeric enzyme, as well as in the optimization of single-chain LHEs whose domains have been engineered separately and later recombined.

Our exploration of C4 cleavage specificity provides a comprehensive data set for the capacity of I-OnuI family enzymes to cleave targets with varying sequences at the middlemost base pairs, the 'C4.' These data demonstrate that I-OnuI family enzymes have remarkably tight C4 specificity, exhibiting significant cleavage activity towards only approximately 4–8 of 256 possible sequences in this region. This specificity is retained in domain fusion chimeras. As each domain appears to contribute to the specificity at these central basepairs, domain chimerization will allow for considerable expansion of potential target sites, as the C4 nt are not currently targeted for engineering due to their unpredictable biochemistry. Furthermore, the AT-rich nature of the C4 targets that are typically cleavable by I-OnuI family enzymes suggests that the energetics of DNA unwinding in the C4 region is an important influence on LAGLIDADG cleavage efficiency, and likely is of central importance to the biochemistry of cleavage within this class of enzymes.

Our survey of structure-independent domain fusions of six I-OnuI family LHEs revealed several patterns that may potentially be exploited to increase the chance of a successful domain fusion among domains from any of the I-OnuI family enzymes. One obvious pattern is that certain domains (e.g. NTD of I-LtrI or the CTDs of I-OnuI, I-LtrI and I-PanMI) proved extremely amenable to direct domain fusion, resulting in highly active chimeric enzymes for the majority of pairs, whereas other domains, (e.g. CTD of I-SscI) would not form active or even stable enzymes with any other domains. This effect was not related to the level of homology, as even chimeras of I-GzeI and I-PanMI, which share >70% identity, achieved only a 50% success rate (Supplementary Figure S8). Thus, choice of domain fusion pairs so as to include a promiscuous partner, and exclude non-promiscuous partners, is a simple method to increase the likelihood of an obtaining an active enzyme from a direct fusion. A second important pattern is that domain fusion success was increased when a 'common interface' between partners was introduced which was native to one of the partner domains. For example, domain fusion

Figure 6. Continued

A647-labeled DNA target substrate. Cleavage products are then visualized on an acrylamide gel. This figure is a compilation of five separate gels. (B) Amino acid sequence alignment of LHEs used in this study. Similarities and identities are highlighted in gray, and 'common interface' residues are highlighted in yellow. The first highlighted residue, N6 in I-OnuI, was grafted only in the alternative common interface, using the solutions from the Ltr-Onu variant sort (Figure 4). Threonine, the most highly selected residue at this position, was substituted. (C) I-OnuI structure with common interface residues colored yellow. The additional residues included in the alternative common interface are colored orange. (D) Catalytic activity of the common interface chimeras as measured by the *in vitro* DNA cleavage assay. (E) Vector graphs showing expression, binding and cleavage activity for all chimeras. NTDs are listed along the vertical axis and CTDs along the horizontal axis, and are organized by percent identity to I-OnuI. The blue line pointing upwards represents expression of the chimera on the surface of yeast. The green line pointing down left represents DNA-binding activity, measured by detection of fluorescently-labeled DNA target substrate bound to surface-expressed enzyme in the presence of calcium (allows for DNA binding, but not cleavage). The orange line pointing down right represents DNA cleavage activity, quantified from the *in vitro* cleavage assay (acrylamide gel). For expression and binding, the length of each line is proportional to the expression and binding of wild-type I-OnuI, holding I-OnuI as the maximum. For cleavage activity, the length of each line is determined as the ratio of cleaved versus uncleaved target in the acrylamide gel. 50% cleavage of the DNA target substrate (after 1 h at 37°C) is set as maximum activity, so chimeras cleaving 50% or more of their target are given a ratio of 1. Chimeras with any detectable level of cleavage activity, as determined by a visible cleaved target band in the gel, are highlighted with a grey background.

chimeras were achieved in 7/10 instances when an I-OnuI domain was used with the I-OnuI-derived common interface. This observation may be exploited in a general approach to domain fusion by introducing residue variation encompassing what is observed throughout the I-OnuI family, into the 'common interface' residue set for every fusion pair. With such an approach, our results suggest that small libraries could be screened with relatively minor efforts to identify domain fusions with high levels of activity for the vast majority of domain pairs.

From our studies, it is evident that domain fusion using NTDs and CTDs extracted from single-chain I-OnuI family enzymes is an efficient approach to generating highly active chimeric enzymes that specifically cleave hybrid target sites. With a simple domain fusion strategy, we achieved ~50% success in generation of active chimeras, and by introducing limited variation into the interface residues, we were able to attain catalytically active chimeras for ~70% of those attempted with relatively minor effort. Our results further suggest that introducing interface residue variation into each domain, followed by the generation of a small library of enzymes for each domain pair, would lead to recovery of highly active chimeric enzymes from the majority of domain fusion pairings. Significantly, the close correlation we observed between ROSETTA energetics calculations and the observed stability and cleavage properties of chimeric enzymes derived from I-Onu and I-LtrI supports previous work, in which structural analysis was used to create stable, active domain fusions from disparate LHEs (24,25). Structural analysis of multiple members of the I-OnuI family could thus facilitate choice of optimal domain partners for direct fusion, further reducing the cost and effort of generating active chimeric enzymes. With the expanding set of characterized LHEs, these methods promise to markedly expand the number of starting scaffolds for engineering, thus enabling broader use of LHEs in genome engineering applications.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Table 1 and Supplementary Figures 1–8.

FUNDING

National Institutes of Health (NIH) [RO1CA133832, RL1GM133833, 5RL1GM84433-04, and U19AI096111]. Funding for open access charge: NIH [RO1CA133832].

Conflict of interest statement. A.M.S. and B.L.S. are co-founders of and have stock in Pregenen Inc., a genome engineering company, which makes engineered homing endonuclease reagents, and have a pending patent application on the use of I-OnuI and I-LtrI crystal structures for generating variants that cleave new target sites. A.M.S. serves as Chief Scientific Officer for Collectis therapeutics, for which he receives salary and stock-based compensation. Collectis therapeutics is the therapeutics subsidiary of Collectis, which has a

proprietary homing endonuclease protein engineering platform. D.B. is also a co-founder and stockholder of Pregenen. J.J. is an employee at Pregenen, with stock and salary compensation. J.G. is a previous employee at Pregenen, with salary compensation.

REFERENCES

1. Certo, M.T., Ryu, B.Y., Annis, J.E., Garibov, M., Jarjour, J., Rawlings, D.J. and Scharenberg, A.M. (2011) Tracking genome engineering outcome at individual DNA breakpoints. *Nat. Methods*, **8**, 671–676.
2. Branzei, D. and Foiani, M. (2008) Regulation of DNA repair throughout the cell cycle. *Nat. Rev. Mol. Cell Biol.*, **9**, 297–308.
3. Ciccia, A. and Elledge, S.J. (2010) The DNA damage response: making it safe to play with knives. *Mol. Cell*, **40**, 179–204.
4. Karran, P. (2000) DNA double strand break repair in mammalian cells. *Curr. Opin. Genet. Dev.*, **10**, 144–150.
5. McCammon, J.M. and Amacher, S.L. (2010) Using zinc finger nucleases for efficient and heritable gene disruption in zebrafish. *Methods Mol. Biol.*, **649**, 281–298.
6. Doyon, Y., McCammon, J.M., Miller, J.C., Faraji, F., Ngo, C., Katibah, G.E., Amora, R., Hocking, T.D., Zhang, L., Rebar, E.J. et al. (2008) Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases. *Nat. Biotechnol.*, **26**, 702–708.
7. Bibikova, M., Beumer, K., Trautman, J.K. and Carroll, D. (2003) Enhancing gene targeting with designed zinc finger nucleases. *Science*, **300**, 764.
8. Bibikova, M., Carroll, D., Segal, D.J., Trautman, J.K., Smith, J., Kim, Y.G. and Chandrasegaran, S. (2001) Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. *Mol. Cell Biol.*, **21**, 289–297.
9. Pessach, I.M. and Notarangelo, L.D. (2011) Gene therapy for primary immunodeficiencies: looking ahead, toward gene correction. *J. Allergy Clin. Immunol.*, **127**, 1344–1350.
10. Cohen-Tannoudji, M., Robine, S., Choulika, A., Pinto, D., El Marjou, F., Babinet, C., Louvard, D. and Jaissier, F. (1998) I-SceI-induced gene replacement at a natural locus in embryonic stem cells. *Mol. Cell Biol.*, **18**, 1444–1448.
11. Gouble, A., Smith, J., Bruneau, S., Perez, C., Guyot, V., Cabaniols, J.P., Leduc, S., Fiette, L., Ave, P., Micheau, B. et al. (2006) Efficient in toto targeted recombination in mouse liver by meganuclease-induced double-strand break. *J. Gene Med.*, **8**, 616–622.
12. Heyer, W.-D., Ehmsen, K.T. and Liu, J. (2010) Regulation of homologous recombination in eukaryotes. *Annu. Rev. Genet.*, **44**, 113–139.
13. Kandavelou, K., Ramalingam, S., London, V., Mani, M., Wu, J., Alexeev, V., Civin, C.I. and Chandrasegaran, S. (2009) Targeted manipulation of mammalian genomes using designed zinc finger nucleases. *Biochem. Biophys. Res. Commun.*, **388**, 56–61.
14. Miller, J.C., Tan, S., Qiao, G., Barlow, K.A., Wang, J., Xia, D.F., Meng, X., Paschon, D.E., Leung, E., Hinkley, S.J. et al. (2011) A TALE nuclease architecture for efficient genome editing. *Nat. Biotechnol.*, **29**, 143–148.
15. Bobis-Wozowicz, S., Osiak, A., Rahman, S.H. and Cathomen, T. (2011) Targeted genome editing in pluripotent stem cells using zinc-finger nucleases. *Methods*, **53**, 339–346.
16. Mahfouz, M.M., Li, L., Shamimuzzaman, M., Wibowo, A., Fang, X. and Zhu, J.-K. (2011) De novo-engineered transcription activator-like effector (TALE) hybrid nuclease with novel DNA binding specificity creates double-strand breaks. *Proc. Natl Acad. Sci. USA*, **108**, 2623–2628.
17. Stoddard, B.L. (2005) Homing endonuclease structure and function. *Q. Rev. Biophys.*, **38**, 49–95.
18. Paques, F. and Duchateau, P. (2007) Meganucleases and DNA double-strand break-induced recombination: perspectives for gene therapy. *Curr. Gene Ther.*, **7**, 49–66.
19. Arnould, S., Chames, P., Perez, C., Lacroix, E., Duclert, A., Epinat, J.C., Stricher, F., Petit, A.S., Patin, A., Guillier, S. 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.
20. Urnov,F.D., Rebar,E.J., Holmes,M.C., Zhang,H.S. and Gregory,P.D. (2010) Genome editing with engineered zinc finger nucleases. *Nat. Rev. Genet.*, **11**, 636–646.
 21. Ashworth,J., Havranek,J.J., Duarte,C.M., Sussman,D., Monnat,R.J., Stoddard,B.L. and Baker,D. (2006) Computational redesign of endonuclease DNA binding and cleavage specificity. *Nature*, **441**, 656–659.
 22. Ashworth,J., Taylor,G.K., Havranek,J.J., Quadri,S.A., Stoddard,B.L. and Baker,D. (2010) Computational reprogramming of homing endonuclease specificity at multiple adjacent base pairs. *Nucleic Acids Res.*, **38**, 5601–5608.
 23. Grizot,S., Epinat,J.-C., Thomas,S., Duclert,A., Rolland,S., Pâques,F. and Duchateau,P. (2010) Generation of redesigned homing endonucleases comprising DNA-binding domains derived from two different scaffolds. *Nucleic Acids Res.*, **38**, 2006–2018.
 24. 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.
 25. Epinat,J.C., Arnould,S., Chames,P., Rochaix,P., Desfontaines,D., Puzin,C., Patin,A., Zanghellini,A., Paques,F. and Lacroix,E. (2003) A novel engineered meganuclease induces homologous recombination in yeast and mammalian cells. *Nucleic Acids Res.*, **31**, 2952–2962.
 26. Sethuraman,J., Majer,A., Friedrich,N.C., Edgell,D.R. and Hausner,G. (2009) Genes within genes: multiple LAGLIDADG homing endonucleases target the ribosomal protein S3 gene encoded within an rnl group I intron of *Ophiostoma* and related taxa. *Mol. Biol. Evol.*, **26**, 2299–2315.
 27. Takeuchi,R., Lambert,A.R., Mak,A.N.-S., Jacoby,K., Dickson,R.J., Gloor,G.B., Scharenberg,A.M., Edgell,D.R. and Stoddard,B.L. (2011) Tapping natural reservoirs of homing endonucleases for targeted gene modification. *Proc. Natl Acad. Sci. USA*, **108**, 13077–13082.
 28. Gietz,R.D. and Schiestl,R.H. (2007) High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.*, **2**, 31–34.
 29. Emsley,P. and Cowtan,K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2126–2132.
 30. Jarjour,J., West-Foyle,H., Certo,M.T., Hubert,C.G., Doyle,L., Getz,M.M., Stoddard,B.L. and Scharenberg,A.M. (2009) High-resolution profiling of homing endonuclease binding and catalytic specificity using yeast surface display. *Nucleic Acids Res.*, **37**, 6871–6880.
 31. Sather,B.D., Ryu,B.Y., Stirling,B.V., Garibov,M., Kerns,H.M., Humblet-Baron,S., Astrakhan,A. and Rawlings,D.J. (2011) Development of B-lineage predominant lentiviral vectors for use in genetic therapies for B cell disorders. *Mol. Ther.*, **19**, 515–525.
 32. Leaver-Fay,A., Tyka,M., Lewis,S.M., Lange,O.F., Thompson,J., Jacak,R., Kaufman,K., Renfrew,P.D., Smith,C.A., Sheffler,W. et al. (2011) ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Meth. Enzymol.*, **487**, 545–574.
 33. Rohl,C.A., Strauss,C.E.M., Misura,K.M.S. and Baker,D. (2004) Protein structure prediction using Rosetta. *Methods Enzymol.*, **383**, 66–93.
 34. Das,R. and Baker,D. (2008) Macromolecular modeling with Rosetta. *Ann. Rev. Biochem.*, **77**, 363–382.
 35. Hirsch,C., Gaus,R., Horn,S.C., Neuber,O. and Sommer,T. (2009) The ubiquitylation machinery of the endoplasmic reticulum. *Nature*, **458**, 453–460.
 36. Vembar,S.S. and Brodsky,J.L. (2008) One step at a time: endoplasmic reticulum-associated degradation. *Nat. Rev. Mol. Cell Biol.*, **9**, 944–957.
 37. Volná,P., Jarjour,J., Baxter,S., Roffler,S.R., Monnat,R.J., Stoddard,B.L. and Scharenberg,A.M. (2007) Flow cytometric analysis of DNA binding and cleavage by cell surface-displayed homing endonucleases. *Nucleic Acids Res.*, **35**, 2748–2758.
 38. Silva,G.H. and Belfort,M. (2004) Analysis of the LAGLIDADG interface of the monomeric homing endonuclease I-DmoI. *Nucleic Acids Res.*, **32**, 3156–3168.
 39. Li,H., Pellenz,S., Ulge,U., Stoddard,B.L. and Monnat,R.J. (2009) Generation of single-chain LAGLIDADG homing endonucleases from native homodimeric precursor proteins. *Nucleic Acids Res.*, **37**, 1650–1662.
 40. Silva,G.H., Belfort,M., Wende,W. and Pingoud,A. (2006) From monomeric to homodimeric endonucleases and back: engineering novel specificity of LAGLIDADG enzymes. *J. Mol. Biol.*, **361**, 744–754.
 41. Fajardo-Sanchez,E., Stricher,F., Pâques,F., Isalan,M. and Serrano,L. (2008) Computer design of obligate heterodimer meganucleases allows efficient cutting of custom DNA sequences. *Nucleic Acids Res.*, **36**, 2163–2173.
 42. Volna,P., Jarjour,J., Baxter,S., Roffler,S.R., Monnat,R.J., Stoddard,B.L. and Scharenberg,A.M. (2007) Flow cytometric analysis of DNA binding and cleavage by cell surface-displayed homing endonucleases. *Nucleic Acids Res.*, **35**, 2748–58.
 43. Takeuchi,R., Lambert,A.R., Mak,A.N.-S., Jacoby,K., Dickson,R.J., Gloor,G.B., Scharenberg,A.M., Edgell,D.R. and Stoddard,B.L. (2011) Tapping natural reservoirs of homing endonucleases for targeted gene modification. *Proc. Natl Acad. Sci. USA*, **108**, 13077–13082.