# Engineering domain fusion chimeras from I-OnuI family LAGLIDADG homing endonucleases for genome engineering applications

**Sarah Katherine Baxter**

**A dissertation**
**submitted in partial fulfillment of the**
**requirements for the degree of**

**Doctor of Philosophy**

**University of Washington**
**2012**

**Reading Committee:**
**Dr. Andrew Scharenberg, Chair**
**Dr. Barry Stoddard**
**Dr. David Baker**

**Program Authorized to Offer Degree:**
**Immunology**

# Table of Contents

# List of Figures

# List of Tables

University of Washington

## Abstract

Engineering domain fusion chimeras from I-OnuI family LAGLIDADG homing endonucleases for genome engineering applications

Sarah Katherine Baxter

Chair of the Supervisory Committee:
Dr. Andrew Scharenberg
Department of Immunology

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-OnuI family LHEs, spanning 40-70% amino acid homology. 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 approximately 70% of domain chimeras. This method will be useful for creating large numbers of chimeric LHEs for genome engineering applications.

# Chapter 1

# Introduction: Advancements in genome engineering

## *1.1 The promise of gene therapy*

Since its conception, the field of genome engineering – the process of altering the DNA of a target genome in a rational and controlled manner - has promised clinical benefits that would be difficult or impossible to achieve with standard therapeutic approaches. By manipulating the genomes of key cell populations, either through the removal or addition of genetic elements, gene therapy could harness a patient's own cells to combat disease: deleterious gene mutations could be corrected permanently, curing genetic diseases or replacing risk alleles that strongly predispose patients to disease.[1–4] Additional approaches could utilize exogenous cells engineered to limit immunogenicity in unrelated hosts, greatly increasing options for immunotherapeutics. Advances in the biochemical understanding of disease also present the potential to revolutionize therapy for non-genetic diseases such as cancer and AIDs using genome engineering.[5–8] For more than two decades, however, the implementation of gene therapy has proven difficult. The manipulation of the human genome by comparatively random insertions has proven unacceptably dangerous, with highly publicized carcinogenic adverse events.[9–19]

In 1970, before the advent of recombinant DNA tools, the first rudimentary gene-therapy trial was conducted using a papillomavirus to deliver a supposedly therapeutic gene to severely handicapped patients suffering from hyperargininemia, a nitrogen metabolism disorder. Unfortunately, the entire rationale of the trial was based on a mistaken assumption - that the virus expressed an arginase enzyme, which could correct the girls' genetic deficiency.[20] It was nearly twenty years before the first FDA-approved human gene therapy trials were conducted on

two young patients with Adenosine Deaminase-deficiency Severe Combined Immunodeficiency (ADA-SCID), a disease in which both T- and B-cells are impaired due to a defect in a single gene.[21] The resulting immunodeficiency is so severe that the disease, if untreated, is usually fatal within the first year of life. Although hematopoetic stem cell transplantation offers a cure for some patients, the lack of HLA-matched donors leaves many without the option of transplantation; moreover, those patients who are lucky enough to find a matching donor face the necessity of a lifetime of immunosuppression to control the ensuing graft-vs-host complications.[22–24] Gene therapy to correct SCID appeared simple in principle: the addition of a functional gene for adenosine deaminase would allow T- and B-cells to mature and function appropriately. The initial trial, delivering the gene by retroviral vectors, was a success by some measures; the patients regained some degree of function of their immune cells.

Enthusiasm for this heralded success, however, was soon tempered. Multiple patients, those from the original SCID trial as well as other trials involving both SCID and Chronic Granulomatous patients, were diagnosed with severe adverse events, resulting from integration of the gene proximal to a proto-oncogene.[25–27,13,14,28,18,19] In 1999, the first death of a patient from a gene therapy trial was reported, and investigation into the case uncovered widespread underreporting of adverse events within the field.[19,29] With non-specific integration, inserted genes are capable of integrating at sites that interrupt the coding sequence of important host genes, such as oncogenic suppressor genes, or at sites near genetic regulatory elements, leading to altered transcriptional regulation. The incidence of oncogenic side-effects was startlingly high in initial trials. One early trial resulted in five of its 20 subject subsequently being diagnosed with leukemia.[25]

The primary lesson from these studies, then, was two-fold: (1) gene-therapy can indeed be successful in treatment of certain disorders, but to do so safely, (2) the genome must be manipulated precisely and in a principled manner.

### *1.2 Rare-cleaving endonucleases*

One alternative solution to limit the oncogenic potential of gene therapy is the use of self-inactivating lentivirus and gamma-retrovirus-based vectors; these vectors lack promotor or enhancer sequences, thus limiting the risk of insertional mutagenesis leading to activation of proto-oncogenes. [30–36] The most promising alternative to classic gene therapy, however, is a targeted therapy approach, in which a genetic locus of interest in the host genome is directly altered. With the capacity to hit a given target, genes could be inserted into a selected "safe haven," or a host gene could be deleted or even corrected *in situ*. A number of potential delivery mechanisms have been proposed, included recombinases and transposons. One of the most promising modalities developed thus far for specific gene-targeting are rare-cleaving endonucleases.

Rare-cleaving endonucleases create double strand breaks that become substrates for cell-intrinsic DNA repair pathways, thus enabling high efficiency sequence modification at or near their cleavage sites.[37] Resolution of an endonuclease-induced DNA double-strand break through mutagenic non-homologous end joining (NHEJ) can result in the generation of small insertions or deletions that can be exploited to disrupt a target gene's coding sequence.[38–41] Alternatively, repair via the homologous recombination (HR) pathway with the co-delivery of a rare-cleaving nuclease and a synthetic homologous repair template can achieve a variety of gene targeting

outcomes, including *in-situ* "correction" of the target gene or gene insertion at a validated safe-harbor.[42–47]

Three platforms are currently available for generating customized rare-cleaving endonucleases for genome engineering: zinc finger nucleases (ZFNs), TAL-effector nucleases (TALENs), and LAGLIDADG homing endonucleases (LHEs).[44,45,48–52] 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. While zinc-finger and Tal-effector nucleases have the advantage of modular DNA-recognition sections, allowing relatively simple engineering to recognize target genes, there is some question as to whether these technologies will be appropriate for human therapeutics due to their potential limited specificity and apparent cellular toxicity.[53–58] In contrast, homing endonucleases (also called meganucleases) are naturally-occuring rare cleaving nucleases, with the hydrolytic active site integrated into their DNA binding interface.

Of the platforms listed above, homing endonucleases offer several unique advantages for genome engineering. These include: 1) naturally high levels of specificity and a corresponding absence of genotoxicity observed when wild type homing endonucleases are expressed in a variety of cell types; 2) small size, with a typical single-chain homing endonuclease open reading frame measuring 800-1000 base pairs; and 3) a significant capacity for multiplexed use, as single-chain homing endonucleases can function autonomously.[44,45,59–61] While the importance of genomic-level specificity for therapeutic applications is obvious, the naturally small size of homing endonucleases 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.[62–64] 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.

### *1.3 LAGLIDADG homing endonucleases*

LAGLIDADG homing endonucleases (LHEs), named for the two central helices composed of homologues to the canonical LAGLIDADG sequence, have been discovered in diverse unicellular eukaryotes, bacteria, and viruses.[13,14,25–28] LHE genes are able to exceed the normal vertical patterns of inheritance by engaging in a horizontal transposition process called 'homing'. The two prerequisites of horizontal propagation are: i) recipient genomic locations which can accept the insertion of an autonomous genetic element without negative phenotypic consequence; and ii) biochemical activity to achieve the transfer of genetic information into such a locus. LHE genes are embedded in and transpose to phenotypically neutral genomic locations, such as within self-splicing group I introns or as N- or C-terminal fusions with permissive recipient proteins.[65,66] Furthermore, LHE genes have devised a mechanism of transposition that relies on the nearly universal process of DNA repair by homologous recombination.[67] LHE genes encode for DNA endonucleases and embed themselves within a cleaved target site: they become self-propagating by creating DNA breaks and using their own coding sequences as repair templates for homology-driven DNA break resolution (figure 1).[68]

Because LHEs evolved specifically for genome-level single-site DNA recognition specificity and activity, recognizing target sequences 18-24 base-pairs in length, they show extremely low cellular toxicity and minimal off-target cleavage events.[62–64] The DNA target

sequences bound and cleaved by LHEs are approximately 22 base pairs in length. Similar to microRNAs or commonplace PCR primers, this length of DNA surpasses a hypothetical threshold of uniqueness within a genome, thereby enabling genome-specific operations. LHEs are one of two (the other being the recently discovered TAL effector proteins) natural protein scaffolds currently known which are able to recognize DNA with specificity properties at or above this 'genomic threshold'. Two decades of experimentation with natural LHEs has demonstrated that they can be expressed at high levels in orthogonal cell types, altering genomic loci without causing any overt toxicity, thereby confirming genomic-level DNA cleavage specificity.[50,69–73]

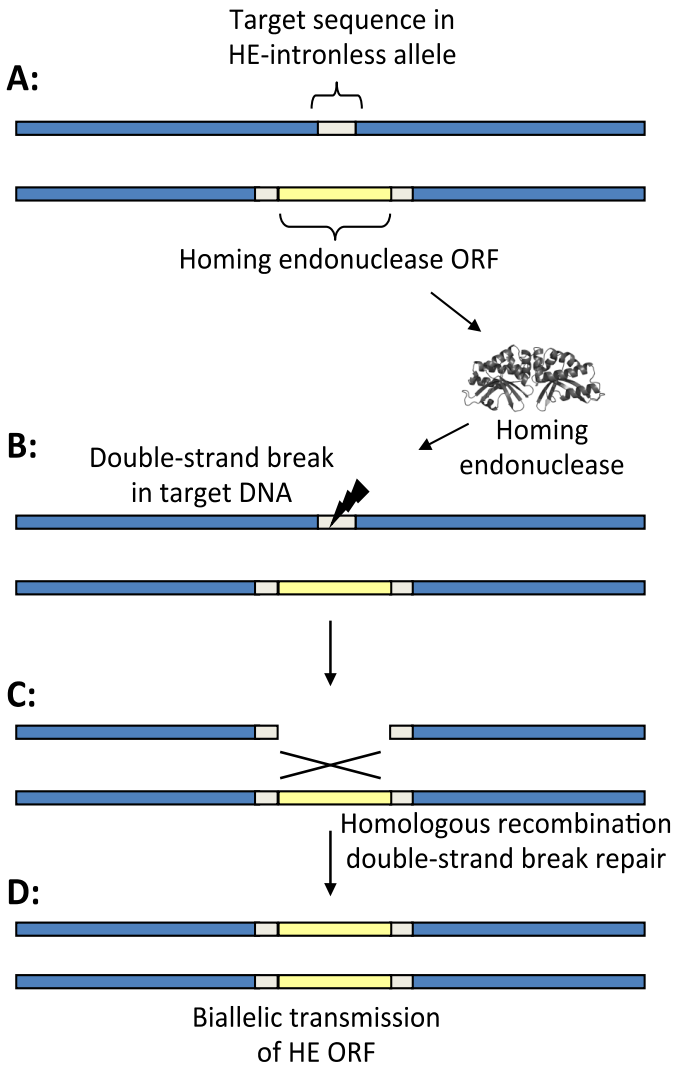



***Figure 1: Homing endonuclease lateral transfer.*** (a) In heterozygous hosts, the HE ORF is copied into the HE-intronless allele leading to biallelic transmission. This "homing" is accomplished by expression of the homing endonuclease by the coding allele; the HE recognizes DNA targets 18-24 nucleotides in length along the homologous, intron-less allele, and (b) introduces a double-strand break at the center of this target site. (c) If the induced DNA double-strand break is repaired by homologous recombination, the HE gene is laterally transferred, leading to (d) biallelic transmission of the HE ORF.

### *1.4 Engineering LAGLIDADG homing endonucleases*

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.[74] These challenges have significantly limited the widespread application of LHEs in genome engineering.

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 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-terminal 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.

This naturally occurring modularity suggests that domain chimerization might be utilized to generate an expanded set of starting scaffolds for engineering, thus effectively minimizing the engineering effort with respect to substrate re-specification. Native LHEs can be divided into artificial modules laterally across the entire protein-DNA interface, and libraries with randomizations in each module are screened to identify LHEs that recognize variant targets. N- and C-terminal domains from various parent enzymes could then be recombined to produce chimeric variants incorporating the alterations from the modular screening. Present results from

our collaborator, Jordan Jarjour, at Pregenen Inc, suggest that screening of a domain will yield approximately 200 enzymes with high activity towards their variant recognition sites (Personal communication, Jarjour). With our lab's recent discovery of up to 30 related enzymes related to the LHE I-OnuI (now designated the I-OnuI family)[66,75], combinatorial chimeras would yield nearly 900 potential starting scaffolds, and full modular screening of these 30 enzymes could thus yield 900*200*200= 32,000,000 enzymes with pre-defined specificities; for an 18 base pair recognition site, this would represent one enzyme for every 2000-base target sequences. Because there is some degree of tolerance to alternative bases in the central four nucleotides, this translates to nearly one cleaving enzyme for every 500 possible target sites. As any possible sequence could thus be modified by expressing two LHEs that bracket the desired modification site, this approach could lead to an "off the shelf" solution to support any desired targeted gene modification.

In summary, the identification of a large set of LHEs encompassing a wide range of target specificities provides an alternative to current techniques of deep engineering for broad respecification. The availability of many diverse LHE scaffolds allows 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, re-specified 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 N- and C-terminal domains (NTD and CTD) 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.[76,77] 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.

# Chapter 2

# Cell-surface display and interrogation of LHE stability, binding, and catalysis

## *2.1: Introduction*

Unlike endonucleases formed from zinc-finger and TAL binding domains, LHEs are not naturally modular; binding and cleavage properties are closely intertwined, and the alpha-helices forming the active site and beta-sheets straddling the DNA major-groove are entwined within a compact domain. LHEs have been difficult to engineer by computational redesign or by mutagenesis covering a small sequence space, due to the compact and interdependent nature of the binding and cleaving activities: while relatively small alternations in binding specificity have been achieved by targeted mutagenesis, larger-scale redesign remains difficult. Efforts to engineer these enzymes, therefore, depend on the ability to efficiently and rapidly mutate a large sequence space, and select active variants. Moreover, engineering has come to be viewed as a multi-step project, with final optimization steps often necessary to achieve a highly active variant.

A variety of engineering methodologies have been developed to utilize computational redesign for mutation-selection techniques.[66,75,78] These protocols employ reporters of high affinity DNA binding, or couple endonuclease activity either to the elimination of a reporter gene or to cell survival by elimination of a toxicity gene and concurrent recombination and reconstitution of a survival gene.[79–85]

Broad re-specification of LHEs is difficult using these techniques, however: the above-mentioned engineering protocols are limited by small molecular library sizes, low screening

throughput, low sensitivity or range of activity, few selection parameters, and poor downstream analytics. LHE interactions with the host genome are a confounding factor in the intracellular based selection methods, as the generation of variants with unknown specificity properties may lead to toxicity independent of the selection scheme. Moreover, directed evolution process, it is likely that broad re-specification would require lower-activity intermediates, which are difficult to detect or isolate in binary selection systems

We therefore developed a flow cytometry-based platform for efficient, semi-quantitative analysis of homing endonuclease binding and cleavage activities.[86] These assays can be performed to develop in-depth characterizations of LHE-DNA recognition properties. If carried out using a wide array of DNA substrate sequences, they can lead to a better understanding of correlations (or lack thereof) between DNA binding and cleavage and also how multiple simultaneous base pair substitutions influence the interaction. The system is fast to set up: following transformation with an episomal plasmid, yeast display on their cell surface a homing endonuclease of interest, and can be rapidly validated by antibody staining of a C-terminal Myc epitope to detect the presence of stable, full-length enzyme. Synthetically generated, fluorescently-labeled DNA substrates are then used to perform fast, multi-well scalable assays to uncover DNA binding and cleavage properties of interest to the investigator.

Using this novel assay, we are able to better determine (a) the in-depth patterns of their DNA recognition specificity, and (b) the degree to which this specificity can be redirected to target a broad range of non-native sequences relevant to the research and/or therapeutics communities.

We therefore developed an engineering platform with expanded capabilities, based on the premise that a flow-cytometry based assay would allow a number of advantages, including a

high screening rate and an expanded range of quantification of endonuclease activity. Additionally, a flow cytometry based assay allows the concurrent interrogation of multiple parameters of LHEs, including stability, and binding and cleavage of desired and off-target DNA sequences.

***Note: the flow-cytometry assay was developed in collaboration with Dr. Jordan Jarjour and Dr. Petra Volna. Surface display technology in B-cells was developed by Dr. Volna and Dr. Jarjour, and display technology in yeast was developed by Dr. Jarjour.*

### *2.2 Eukaryotic cell surface-display of LAGLIDADG homing endonucleases*

In order to efficiently evaluate the binding and catalytic properties of homing endonuclease variants, we developed a surface-display system of the endonucleases. Surface-display of LHEs allows the interrogation of LHE variants without exposure to the host genome, since transmembrane and secreted proteins are expressed through the secretory pathway and are not generally exposed to the nuclear or organellar genetic material in the cell.

In order to assess binding and catalytic profiles of homing endonucleases, it was necessary to develop surface-display of the proteins. Dr. Jordan Jarjour and Dr. Petra Volna developed an expression system to clonally display homing endonuclease variants on the cell surface of a lymphocyte cell line by targeting the expression of an LHE-CD80 transmembrane fusion protein to the secretory pathway. Transfection of the linearized constructs into DT40 cells resulted in the isolation of clonal lines with high levels of LHE surface expression.[78] Lymphocytes were used for initial surface-display technology, in order to utilize the B-cell intrinsic somatic hypermutation pathway to introduce significant mutations into the LHE ORF for directed

evolution experiments; the surface-expression system would lead to a clonally diversified population selectable in high-throughput flow-cytometry assays. Surface-expressed LHEs include an N-terminal HA-tag, allowing detection by fluorescent antibody staining.

In their native organisms, LHEs are expressed in the cytosol and targeted to DNA-containing organelles post-translationally. Developing a system of cell-surface display of LHEs requires both the cotranslational targeting to the secretory pathway and fusion to an appropriate transmembrane domain. [87,88] The strategy we chose was the one previously used to support surface display of antibody fragments.[89] The ORFs for I-AniI, an LHE isolated from Aspergillus nidulans[90], and H-DreI, an artificially engineered endonuclease consisting of domains from I-CreI (from Chlamydomonas reinhardtii) and I-DmoI (Desulfurococcus mobilis)[91] 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 2a). Clonal DT40 cell lines were created by transfection of the linearized constructs, and showed high levels of I-AniI and H-DreI surface expression, as detected by staining of an N-terminal HA tag (figure 2c). The inclusion of a neomyosin resistance gene allowed the isolation of multiple DT40 clones with stable surface expression (figure 2d).

Dr. Jarjour extended this work to develop a yeast surface-display system, fusing the LHE to the secreted protein Aga2P.[86] Yeast surface display is a widely used platform to analyze and engineer intermolecular binding interactions. Importantly, efficient transformation protocols in yeast allow for the generation of large and complex molecular libraries: PCR-generated ORFs, including random or targeted mutations by standard protocols, are clonally inserted into a population of cells, thereby creating a library of variants up to $10^8$ in size, which can then undergo selection in order to identify desired variants.

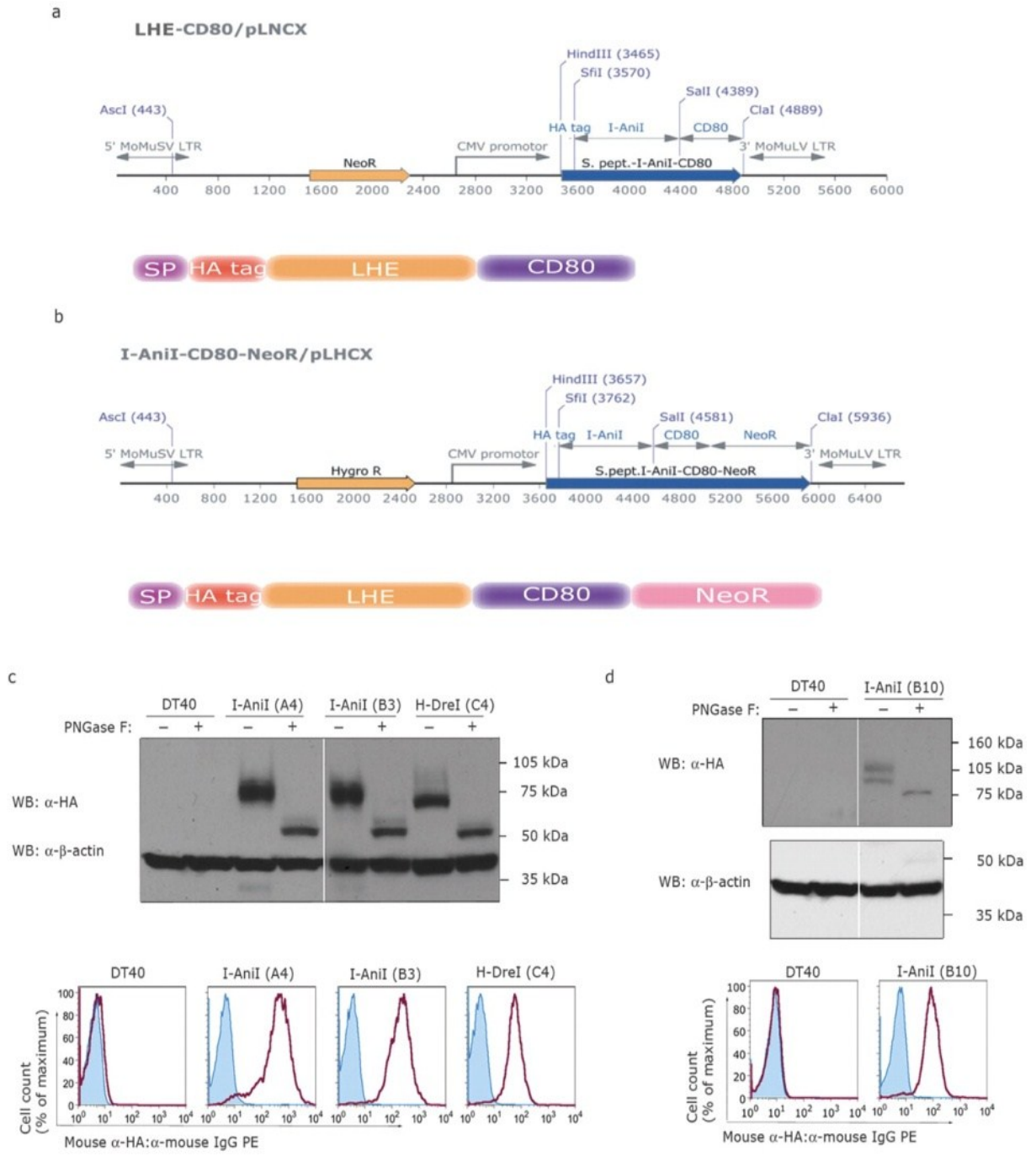



***Figure 2: Homing endonuclease cell-surface display.*** Vector schematics and validation of efficient LHE fusion protein expression in DT40 chicken B-cells. (a) LHE cDNAs were placed in-frame between a murine immunoglobulin-derived N-terminal signal peptide (SP) and the transmembrane spanning region 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 ***(cont.)***

Moreover, stability can be generally correlated with surface expression in yeast, with stable proteins typically exhibiting higher surface expression than unstable proteins. The eukaryotic translational machinery provides a stringent folding checkpoint, with unstable or poorly-folded proteins being degraded in the endoplasmic reticulum before export to the surface.[92,93] The use of the eukaryotic expression system as a tool and measure for stability has been reported by groups in varying fields and journals, and yeast surface-display has even been utilized to construct a stable T-cell receptor variant using yeast-surface display.[94]

### *2.3 Binding analysis*

One of the primary advantages to the surface-expressed homing endonuclease system is the capacity to assay activity against a wide array of target DNA substrates without the construction of new cell lines. Binding DNA target substrates are assembled by PCR with fluorophore-labeled primers (or biotinylated primers, incubated with streptavidin-conjugated fluorophores to create a pre-formed complex), allowing the direct detection of LHE binding to substrate in flow-cytometric analysis; this single-step staining also effectively minimizes the effects of variations in dissociation kinetics of different LHEs.

LHEs are enzymatically active in the presence of $Mg^{2+}$ ions, which are present in the active site.[95] To detect binding of the LHE to target oligonucleotides, the surface-expressing cells were

---

***Fig 2 cont*:** 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. (From Volna et al, 2007)

incubated iwth $Ca^{2+}$ ions, in the presence of which LHEs maintain their DNA-binding properties while target DNA cleavage is abolished.[96] Pre-formed complexes of biotinylated dsOligos with phycoerythrin-conjugated streptavidin (SAv-PE). Cells with surface-expressed I-AniI and H-DreI were stained with pre-formed complexes of biotinylated oligonucleotides with phycoerythrin-conjugated streptavidin (SAv-PE); expressing I-AniI and H-DreI showed specific binding of their respective natural target sequences (figure 3).

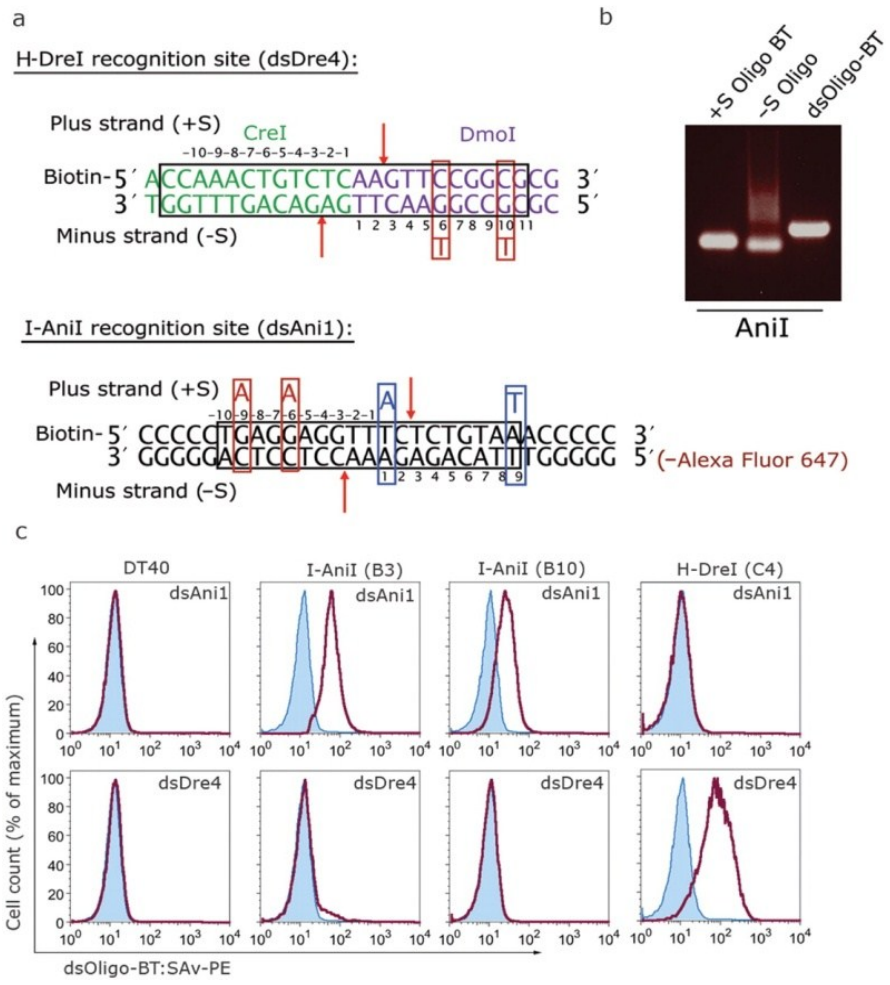


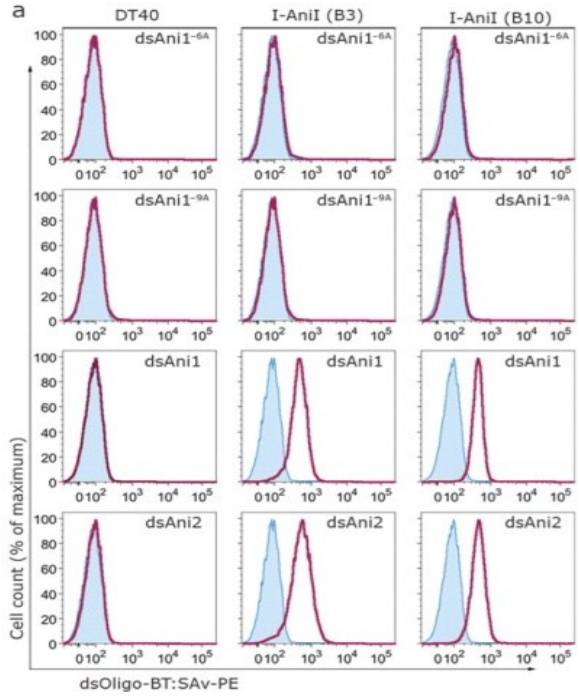

***Figure 3: Cell-surface expressed LHEs specifically bind target DNA.*** Fluorescently conjugated double-stranded oligonucleotide substrates bind cell surface LHEs in a manner, which is sequence specific and easily resolved by flow cytometry. (a) H-DreI is an engineered enzyme composed of domains derived 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 (dsDre46t, dsDre410t, 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 alexafluor 647-conjugated oligonucleotides for dsAni1 and dsAni1−9a were used in the flow cytometry cleavage assay. (b) Verification of efficient annealing of the complementary ***(continued)***

To determine the specificity of the binding assay, we stained I-AniI and H-DreI expressing clones with oligonucleotides containing modifications to their respective target sequences. No staining was observed when dsAni1 or dsDre4 were used to stain non-corresponding LHE expressing clones (Figure 3c). Likewise, staining was not detected when the LHEs were incubated with target sequences specifically designed with single base-pair alterations (dsAni1$^{-9A}$, dsAni1$^{-6A}$, dsDre4$^{6T}$, dsDre4$^{10T}$, Figure 3a) predicted by structural analysis to disrupt direct contacts within the DNA–protein interfaces (Figure 4).[90,91] In contrast, binding activity is maintained against a target sequence variation known to be efficiently cleaved (dsAni2, unpublished data, Fiugre 3a, Figure 4a, bottom panels). To demonstrate that the assay is sensitive not only to alterations in the target sequence but also to mutations in the structure of the LHE, I-AniI mutated enzymes (denoted as I-AniI$^{m}$) (Figure 4b), predicted to have either core structural changes or to have disrupted interactions at DNA-binding interface, were shown to have no binding activity against I-AniI target DNA. These data demonstrate the reliable discrimination of target DNA sequences by surface-expressed LHEs: binding analysis by fluorescent staining parallels known target sequence cleavage specificities, and is sensitive to mutations in the enzyme.

These binding studies were extended by Dr. Jarjour, using yeast-surface display (figure 5).[86] Substrate interactions were quantified using equilibrium ligand binding approximations

---

***figure 3 cont.:*** oligonucleotides run on a 3% agarose gel, with individual oligonucleotides (+s and −s) run as controls. (c) Flow cytometry analysis of clones stained with fluorescent double-strand oligonucleotides. 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. (from Volna P *et al*, 2007)



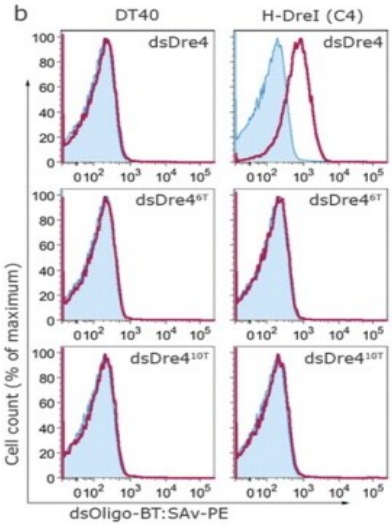




***Figure 4: Cell surface-expressed LHEs bind their target with high specificity.*** LHEs reliably discriminate dsOligos containing single base-pair differences from their natural target sequences. (a) I-AniI and (b) H-DreI expressing clones were stained with dsOligo-BT: SAv-PE complexes containing the natural target sequences (dsAni1 and dsDre4) or containing single base-pair changes (dsAni1−6A and dsAni1−9A; dsDre46T and dsDre410T). 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. (from Volná P *et al*., 2007)

analogous to those used in the exploration of protein–protein interactions in this surface display system.[97] As with the DT40 surface-display assay, samples were incubated to equilibrium with a given concentration of substrate and the amount of bound substrate was detected by a secondary staining step with streptavidin-PE. An antibody to the C-terminal Myc epitope was included to evaluate substrate-binding capacity as a function of surface expression (figure 5a). Staining with I-AniI's wild-type target site generated a signal that fit a standard one-site binding equation with an approximate dissociation constant ($K_d$) of 30 nM, which is consistent with the $K_d$ as measured by traditional biophysical methods.[98]

Equilibrium binding interactions were further evaluated for a panel of DNA target substrates comprising the three alternative bases at all target site positions (Figure 5c). This analysis

demonstrates that I-AniI discriminates with exclusive specificity at multiple positions of the corresponding (−) half-site, which corresponds with structural analysis showing specific hydrogen bond interactions in this region of the interface. In contrast with the high-specificity binding readout of the N-terminal domain, the I-AniI C-terminal domain displays a lack of binding specificity, implying that the majority of binding discrimination is accomplished by I-AniI's N-terminal domain, and highlighting the value of a quantitative assay for describing local binding specificity contributions in protein–DNA interactions.[86]

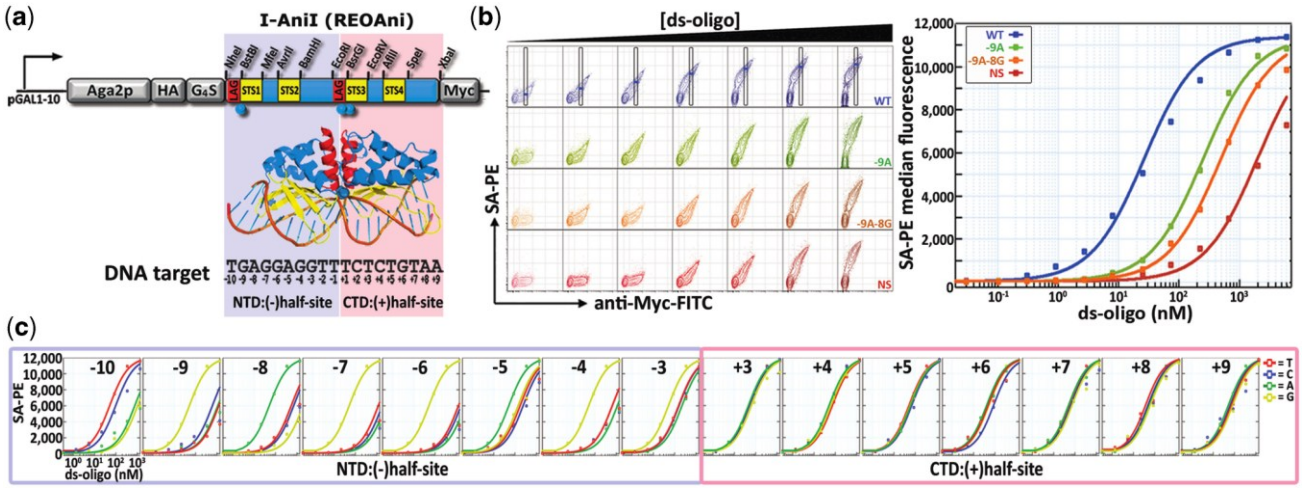



***Figure 5: Profiling protein–DNA binding specificity.*** (a) Schematic representation of the Aga2p-REOAni expression construct. Silent restriction sites unique within the vector and embedded adjacent to domain boundaries are indicated. A corresponding structural diagram (PDB file: 2QOJ) with emphasis on the motifs that comprise the N-terminal (STS1, STS2) and C-terminal (STS3, STS4) DNA-binding domains and annotated target sequence are shown below. NTD, N-terminal domain; CTD, C-terminal domain; LAG, LAGLIDADG helix; STS, strand-turn-strand; G4S, 3 × (Gly)4-Ser linker motif. (b) Flow cytometry contour plots for equilibrium titration staining with increasing concentrations of four ds-oligo substrates. For the wild-type target, Myc epitope normalization gates are shown with the ds-oligo:SA-PE median fluorescence intensity levels marked. Non-linear regression curves (see 'Materials and Methods' section) for binding affinity measurements using epitope-normalized fluorescence values from flow cytometry stains are shown in the adjacent panel. (c) Equilibrium binding curves from a representative experiment with target site substrate analogs bearing each single mutation across I-AniI's native target sequence. The curve representing I-AniI's interaction with its native target is reproduced on each plot for comparison. (from Jordan J *et al*, 2009)

### *2.4 Cleavage analysis*

To evaluate whether surface-expressed LHEs retain catalytic activity, we designed an assay to measure cleavage of specific target sequences. Target oligonucleotides are modified at the 5′ terminus with either Alexa Fluor 647 or biotin during synthesis and annealed to obtain dually conjugated double-stranded substrate (647-dsOligo-BT, figure 6a). The oligonucleotides are then conjugated to SAv-PE to obtain a bifluorescent 647-dsAni1-BT: SAv-PE staining reagent. Recombinant I-AniI was used to verify that fluorescently labeled substrate oligonucleotides, either alone or in complex with SAv-coated beads, are readily accessible and efficiently cleaved.

Cells with surface-expressed LHEs are labeled with a biotin-conjugated anti-HA monoclonal antibody (α-HA-BT), followed by the addition of pre-formed 647-dsAni1-BT: SAv-PE complexes, which contain an average of three remaining BT-binding sites per SAv tetramer. This staining protocol tethers the 647-dsAni1-BT: SAv-PE to the cell surface expressed LHEs independent of any specific LHE–DNA interaction, but still placing the target substrate within the LHE's immediate environment (figure 6a). Cleavage of the tethered 647-dsOligo-BT by the surface-expressed LHE leads to loss of the Alexa Fluor 647 signal, while fluorescent signal of the bridging SAv-PeE signal is retained.

Since both antibody binding and SAv: BT interactions are independent of divalent cation contribution, a $Ca^{2+}$ and $Mg^{2+}$-free buffer was used to stain LHE-expressing cells with α-HA-BT, followed by 647-dsAni1-BT: SAv-PE. The cells are then incubated with 5-10 mM $Mg^{2+}$ (which allows cleavage) or $Ca^{2+}$ (as a negative control), and placed at 37°C.[99] Using this fluorescent cell-surface cleavage assay, we were able to observe endonuclease activity of I-AniI against its target substrate (figure 6b). Consistent with the specificity demonstrated in binding experiments, the bi-fluorescent signal on surface-expressed I-AniI was unaltered when tethered to dsAni1$^{-9A}$,

confirming that I-AniI is not catalytically active against this altered target substrate. The ratio of PE:647 fluorescence was calculated to provide a quantitative measure of catalytic activity; this quantification demonstrates a substantial increase in I-AniI's PE:647 ratio with the native I-AniI target sequence (Figure 6c). To verify that the fluorophore loss in this cleavage assay occurs due to substrate cleavage and not from fluorophore quenching, the cleaved fragment was identified from the supernatants of cleavage experiments (Figure 6d). These results demonstrate that the surface-expressed LHEs retain their catalytic activity and high specificity.

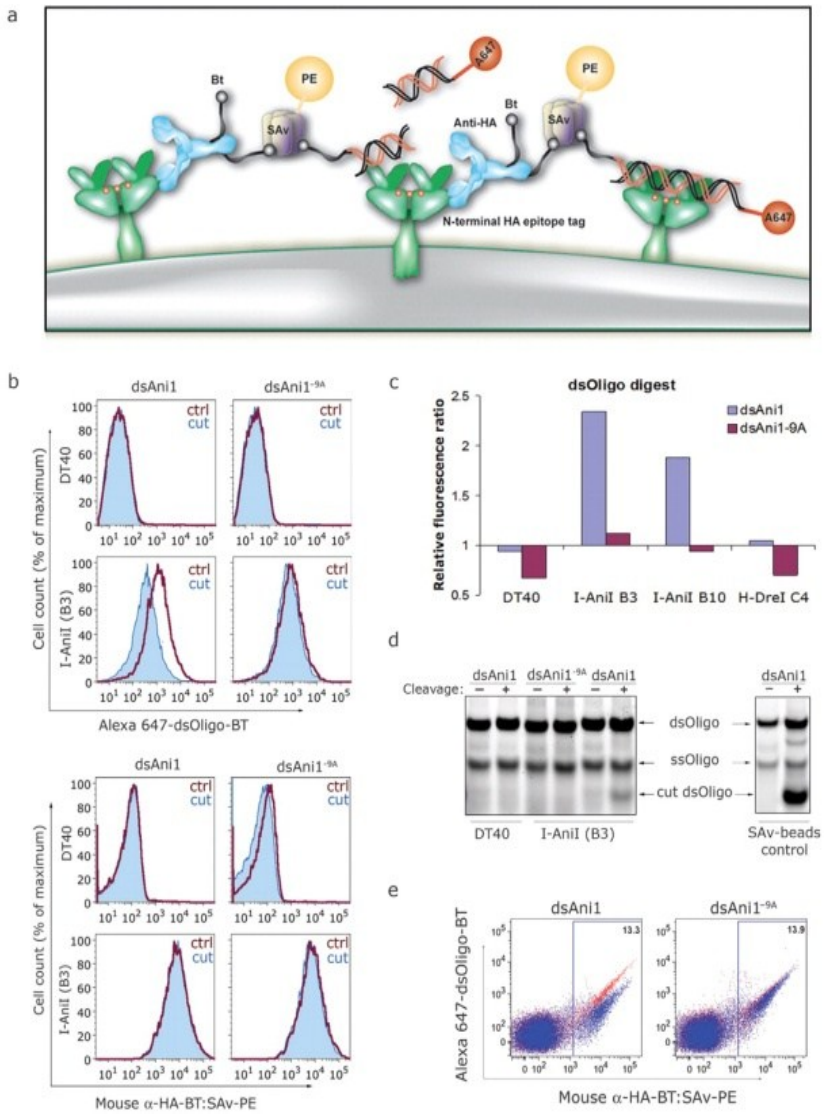



***Figure 6: LHE mediated cleavage of cell surface-tethered DNA substrates.*** (a) Schematic diagram for assaying surface LHE cleavage of α-HA-BT tethered dually fluorescent dsOligos and the release of Alexa Fluor 647 following addition of Mg2+ (red dots). (b) DT40 and B3 cells were stained with α-HA-BT 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) Mg2+ 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 ***(continued)***

As this assay was developed in part to enable FACS sorting of catalytically active LHE variants from large libraries, it was necessary to confirm that the tethered substrate was being cleaved by LHEs on the very cells to which they were tethered. To determine whether cleavage of substrate by adjacent cells would confound FACS sorting, I-AniI expressing (B3) cells were diluted with non-expressing DT40 cells at a 10:1 ratio, thus decreasing contacts between the LHE expressing cells. The decrease in Alexa Fluor 647 fluorescence was approximately equivalent to the loss measured in an undiluted I-AniI-positive population (Figure 6e). These data suggest that DNA substrates are primarily bound and digested by LHEs autonomously on the cell surface, and therefore FACS sorting in mutation/selection or directed evolution experiments should be effective.

### *2.5 Summary*

Engineering homing endonuclease variants to recognize and cleave alternative target sites is challenging with current screening technology.[64,81,84] Though powerful computational tools have been developed to increase our ability to understand protein-DNA binding interactions and

---

***Fig 6 cont:*** 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 Mg2+ (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) Mg2+ followed by flow cytometry analysis. (from Jordan J *et al,* 2009)

alter them accordingly, testing these variants for activity is a laborious and slow process.[76,77,100,101] In order to increase our capacity both to screen libraries of LHE variants for activity against a given target and to investigate the binding and cleavage profiles of a given LHE against a wide range of targets, we developed a high-throughput, flow cytometry based assay.

Using cell-surface display of homing endonuclease enzymes on DT40 and yeast cells, we are able to rapidly assess the binding and cleavage activity of an LHE against a PCR-generated target. This method has several useful properties. First, screening efficiency is significantly increased by abolishing the need to generate cell lines for every target site of interest, necessary for bacterial or yeast-mating protocols. Quantitative binding and cleavage profiles for novel LHE variants can also be rapidly assessed, which would otherwise require in vitro expression and purification followed by electrophoretic analysis of gel shifts and cleavage. Second, LHEs expressed on the cell surface are physically separated from the cell's genome, allowing for the possibility that iterative selections may produce lower specificity intermediates in the generation of a high-specificity enzyme; such intermediates could be toxic if the endonuclease were expressed intracellularly, with access to the host genomic DNA. Finally, this method offers significant flexibility and power in screening strategies. Current yeast transduction efficiencies can yield populations of up to $10^8$ mutated LHEs, which can be sorted in iterative flow-cytometry assays to isolate high performing variants. Selective gating could also allow for the concurrent enrichment of variants with desired specificities with the simultaneous exclusion of undesireable off-target sites, by performing the binding or cleavage assays with target substrate labeled with different flurophores; this sorting strategy would allow the isolation of especially high-specificity LHE variants (personal communication, J. Jarjour).

In summary, we have cell-surface display assay to facilitate the high-throughput screening of LHE cleavage and binding activity. The increased efficiency and throughput of the method should allow more efficient identification and analysis of enzymes with novel sequence specificities for use in a variety of genome engineering applications.

### *2.6 Materials and 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[87,88] 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[89], 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: GGCCCAGCCGGCCATGGGCAGCAGCCATCATCATC
I-AniI Rev SalI: GTCGACATAATTTGAAGGTATTTTTATTTTTTCTG
H-DreI For SfiI: GGCCCAGCCGGCCATGCATAATAATGAGAATGTT
H-DreI Rev SalI: GTCGACCGGGGACGATTTCTTTTTTTCACT
CD80-NeoR For: CAGACCGTCTTCCTTGGATCGGCCATTGAACAAG
NeoR Rev ClaI: ATCGATGAACAAACGACCCAACACCCGTGCG
SP For Hind3: AAGCTTATGGAGACAGACACACTCCTGCTATGGG
CD80-NeoR Rev: CTTGTTCAATGGCCGATCCAAGGAAGACGGTCTG
I-AniI K21 T27 SDM
For:CAGCATCACCAACAAGGGTAAGTACCTACAGTATGAGCTGGGTATCGAG
I-AniI K21 T27 SDM
Rev:CTCGATACCCAGCTCATACTGTAGGTACTTACCCTTGTTGGTGATGCTG

**Western blotting and glycosylation analysis by PNGase F treatment**

Here, $7.5 \times 10^6$ 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 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).

**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 $1 \times 10^6$ 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.

**Yeast growth, transformation, library construction and plasmid recovery**

*Saccharomyces cerevisiae* strain EBY100 was transformed using the lithium-acetate (LiAc) method.[102,103] For library construction, error-prone PCR was performed over the STS3/4 region of the I-AniI ORF using the GeneMorph-II Random Mutagenesis kit (Stratagene) according to the manufacturer's protocol. Library size for the STS3/4 library was determined by serial dilution to be $0.5 \times 10^6$ unique transformants. Mutation distribution and frequencies were verified by sequencing an unselected library and determined to be in the range of 0.5–1.0 mutations per kilobase with no major biasing of the type or positions of mutations. Yeast propagation was performed in the presence of 2% raffinose +0.1% glucose at 30°C for at least 12 h prior to induction. Cells were induced in 2% galactose for 2–3 h at 30°C followed by 18–26 h at 20°C. Plasmids were isolated from yeast populations using the Zymoprep-II kit (Zymo Research) and electroporated into *Escherichia coli* DH10B (Invitrogen) for amplification and/or sequencing. Sequencing was performed on 40–60 clones for a given selection output.

**Binding affinity calculation**

Epitope-normalized values for substrate binding were determined for each sample. This value was established as the median PE fluorescence in a 10% cell gate of FITC fluorescence. This gate position, which generated normally distributed PE fluorescence data from ~2000 cells per sample, was held constant across the entire experimental analysis. Median PE values were plotted versus ds-oligo concentration and the resulting distribution was fit using iterative least-squares modeling to the equation for equilibrium binding: $Y = [[B_{max}*L]/ [K_d + L]] + A$

In the VisualEnzymics (SoftZymics) module for IGOR Pro 6 (WaveMetrics), where [$L$] is the concentration of substrate and $Y$ is the measured median PE value for a given Myc-normalized sample. Non-linear regression analysis was applied using the Levenberg-Marquart (LM) Robust algorithm to isolate the maximum bound ligand, $B_{max}$ and the dissociation constant, $K_d$. Samples were weighted towards the lower ds-oligo concentrations which showed the least variation across samples. The constant term, $A$, was included to adjust for the median fluorescence value of unstained cells for a given cytometer's photomultiplier tube voltage setting.

# Chapter 3: Exploring chimerization of I-OnuI and I-LtrI

## *3.1: Introduction*

Phylogenetic analyses suggest that the LHEs, a dimeric and pseudodimeric family of enzymes with two interacting LAGLIDADG helices at their catalytic core, evolved through the fusions of unrelated domains.[62] The interfacial residues of the LAGLIDADG helices are very well conserved throughout the family, and superposition of known LHE structures suggests that domain fusions or dimerizations are possible; indeed, a limited number of domain fusions have been engineered using the well characterized LHEs I-CreI and I-DmoI.[91,104–106] The natural modularity suggests that domain chimerization could allow us to generate an expanded set of starting scaffolds for engineering, thus minimizing the engineering effort with respect to substrate re-specification. Using the highly conserved I-OnuI LHE homologues, a family of enzymes recently characterized in our lab, we began our studies of domain chimerization by investigating the stabilizing interactions within the LAGLIDADG interface, and defining the structural variation necessary to efficiently design stable LHE chimeras between any two LAGLIDADG domains.

To develop efficient general strategies for a) extraction of individual domains from a parental, pseudo-dimeric single-chain LHE, and b) fusion of these domains into active chimeric enzymes, we analyzed three aspects of the single-chain LHE structure/function relationship in I-OnuI and I-LtrI and their domain fusion chimeras: 1) the extent to which the peptide linking the N- and C-terminal domains contributes to individual NTD and CTD function; 2) the influence of interactions at the domain interface on the stability and activity of chimeric enzymes, and 3) the extent to which the central four nucleotides 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.

### *3.2: Direct fusion of individual N- and C-terminal domains 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 N- and C-terminal domains derived from I-OnuI and I-LtrI: N'Onu-C'Ltr (Onu-Ltr) and N'Ltr-C'Onu (Ltr-Onu).[66,75] 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 7a). Furthermore, models of interface packing for Onu-Ltr and Ltr-Onu domain fusions, generated using ROSETTA macromolecular modeling, suggest that the I-OnuI N-terminal domain and I-LtrI C-terminal domain would be as energetically compatible with each other as they are with their native domain partners (e.g. table 1). [107–109]

To generate Onu-Ltr and Ltr-Onu chimeras, the open reading frames for the N- and C-terminal domains 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 promotor. 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. [107–109] We also used circular

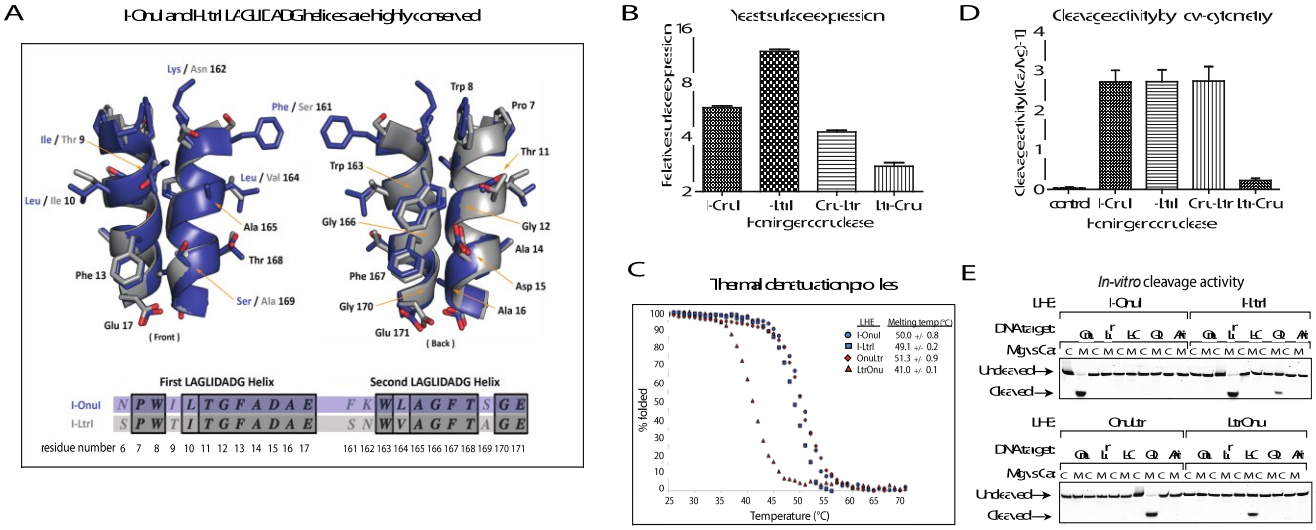



***Figure 7: Comparison of Onu-Ltr and Ltr-Onu in vitro stability and activity***. (a) Overlaid crystal structures of the LAGLIDADG helices of I-OnuI (blue) and I-LtrI (gray), which form the majority of the interface between the N- and C-terminal domains. Residues are indicated by both name and corresponding 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" vs. "non-expressing" cell populations. (c) Thermal denaturation was monitored by circular dichroism (CD) as an alternative measure for comparison of overall protein stability. See the 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 minutes at 37°C. Data represents 5-7 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 hour at 37°C. (Compilation of three separate gels, run in parallel).

dichroism (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.[92,93]

| NTD | CTD | Δsasa | Δtot | Δatt | Δrep | Δhb | Δsol | Δdun | Δpair | 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 |

***Table 1: Rosetta calculations for I-OnuI and I-LtrI chimeras.*** N- and C-terminal domains 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. "**Δ**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 vs. together, and are expressed in Rosetta units (RU). Therefore, these values measure the energetic contribution of the interface. "ΔTot" is the total score. "ΔAtt" is the attractive component of the Lennard-Jones potential, representing van der Waals interactions. "ΔRep" is the repulsive component of the Lennard-Jones potential. "ΔHb" is the contribution from hydrogen bonds, both from backbone and side chain atoms. "ΔSol" is the solvation/desolvation term. "ΔDun" is derived from the Dunbrack rotamer library and considers the frequency of side chain rotamers.(43) "ΔPair" 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 "ΔTot".

As predicted by ROSETTA calculations of interface energetics, Onu-Ltr showed strong surface expression and was stable to 52 degrees C, while Ltr-Onu showed significantly decreased surface expression and thermal stability in the circular dichroism assay (figure 7b and 7c). 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 7d

and table 1). 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 7e and table 2).

| Homing endonuclease | DNA target site |
|---|---|
| I-OnuI | TTTCCACTTAT/TCAACCTTTTA |
| I-LtrI | AATGCTCCTAT/ACGACGTTTAG |
| I-GpiI | TTTTCCTGTAT/ATGACTTAAAT |
| I-GzeI | GCCCCTCATAA/CCCGTATCAAG |
| I-PanMI | GCTCCTCATAA/TCCTTATCAAG |
| I-SscMI | AGGTACCCTTT/AAACCTATTAA |

***Table 2: DNA target sequences for native I-OnuI homologues.*** The "negative" half-sites (bound by the N-terminal domain) are shown on the left, and the "positive" half-sites (bound by the C-terminal domain) are shown on the right. Chimeric DNA target substrates were generated by matching the respective NTD-associated and CTD-associated half-sites.

Although both chimeras exhibited detectable cleavage activity, the activity of Onu-Ltr appeared to be equivalent 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 non-homologous end-joining (NHEJ) and homologous recombination (HR) resulting from endonuclease cleavage events in an integrated reporter cassette (figure 8).[38]

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 8a). Onu-Ltr expression induced +3 frameshift mutations due to nonconservative end-joining (as measured by mCherry expression) in approximately 4% of cells; this rate was equivalent to disruption rates induced by I-LtrI against its native target, and comparable to or

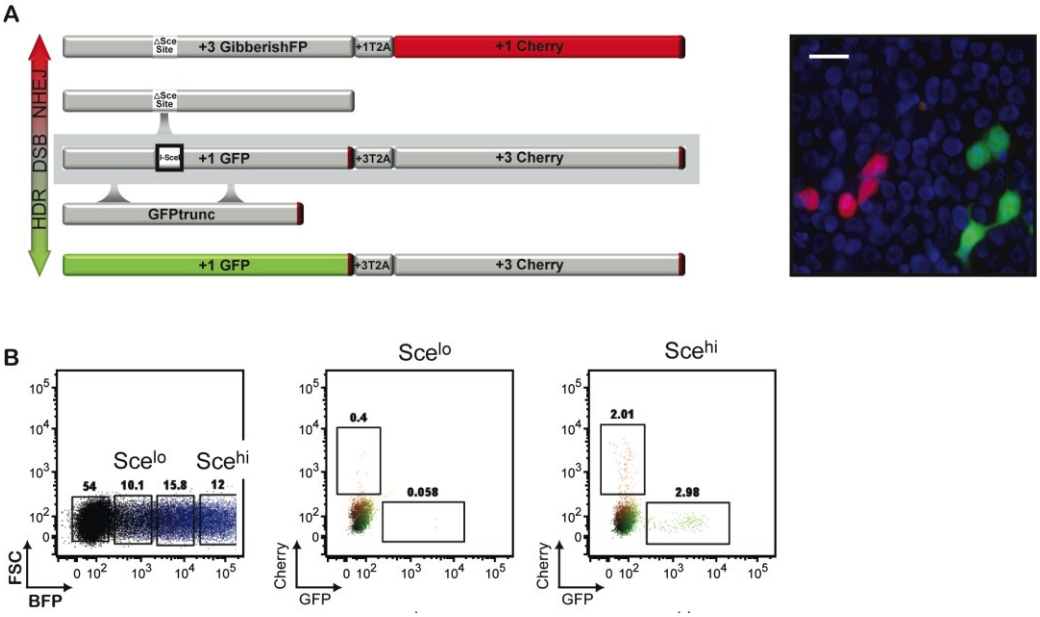



***Figure 8: The Traffic Light Reporter assay.*** This assay provides a quantitative measure of intracellular NHEJ and HR events induced by LHE expression. (a) The Traffic Light Reporter provides a fluorescent readout of repair events occurring at individual nuclease-induced DNA double-strand breaks. Left: the nave reporter (grey) is represented at center, and includes an inactive GFP with a target for the homing endonuclease I-SceI (black box) and an out of frame mCherry. Upon induction of the I-SceI homing endonuclease, cleavage occurs at the Sce site. This double strand break can be repaired by homologous recombination with an inactive GFP donor (bottom) to yield active GFP. Alternatively, the break can be repaired by non-homologous end joining: if this repair results in a mutagenic 2 base pair frame shift, an open reading frame is created which can read through the "GibberishFP," a T2A peptide, and the now in-frame mCherry (top). The T2A decouples mCherry and GibberishFP, resulting in a red fluorescent readout. Sequencing analysis confirmed that approximately 1/3 of mutagenic non-homologous end joining events result in an in-frame mCherry. Right: 293T cells with GFP and mCherry expression following I-SceI induction. (b) Expression of the LHE I-SceI induces mutagenic NHEJ and HR events in flow cytometric analysis. Top panel: Left: I-SceI expression is monitored by coexpression of BFP. Cells expressing low (center) and high (right) levels of I-SceI are gated to determine the levels of homologous recombination (GFP, x-axis) and non-homologous end joining (mCherry, y-axis). *(*from Certo et al, 2011)

slightly greater than that induced by I-OnuI (figure 9). The biological explanation for varying rates of HR and NHEJ observed among I-OnuI homologues is uncertain, and could include transcriptional timing, or the rate of enzyme release from cleaved DNA.

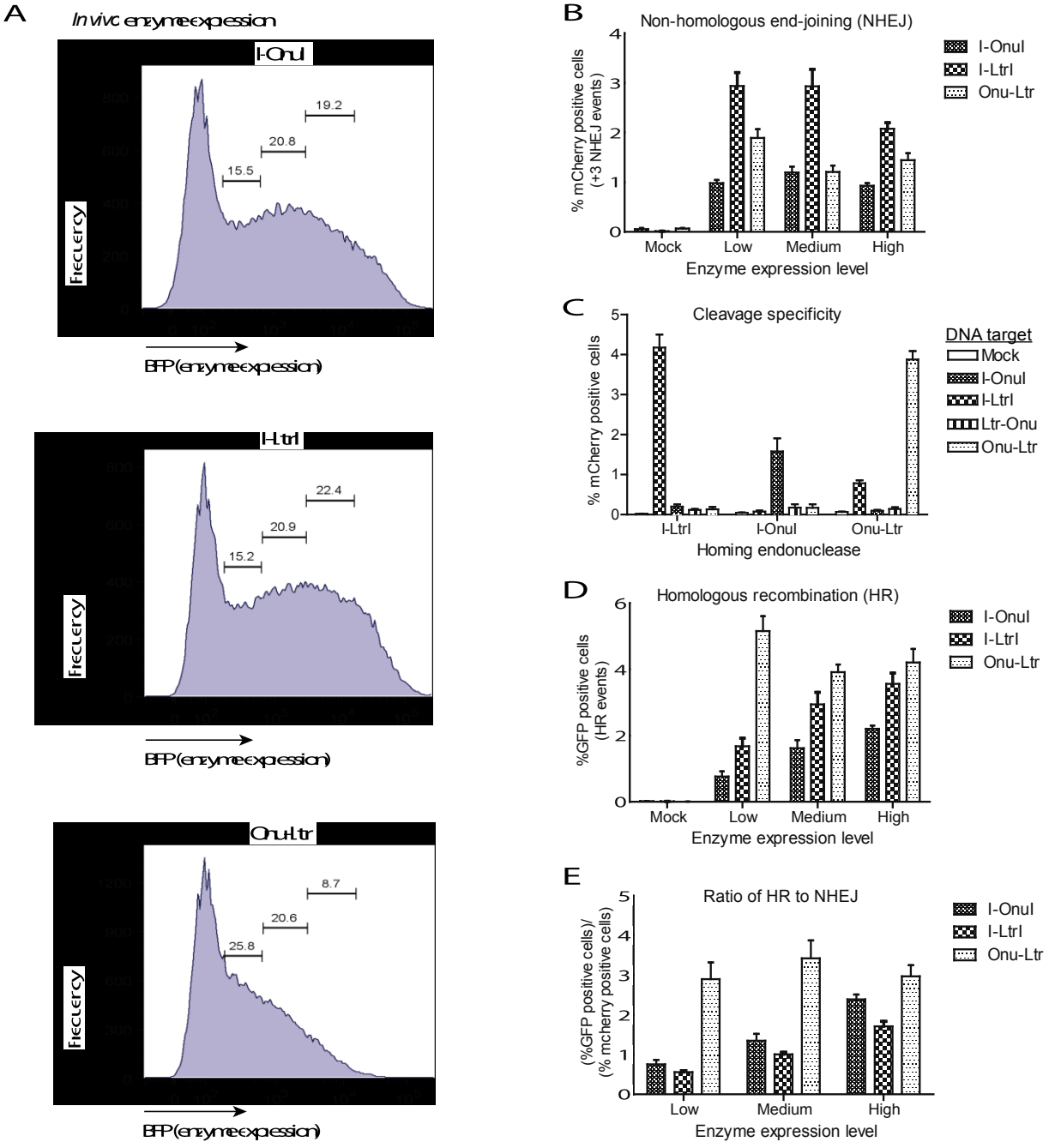



***Figure 9: In vivo activity of Onu-Ltr compared to native enzymes.*** (a) Plasmids containing I-OnuI, I-LtrI, and Onu-Ltr with an N-terminal BFP tag were transfected into HEK293T cells ***(continued)***

Cleavage of the I-LtrI target by Onu-Ltr was also visible at low rates (figure 9c), 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 7f), 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 approximately two-fold over those induced by I-OnuI and/or I-LtrI, emphasizing the high level of performance achieved in the chimeric enzyme (figure 9d, f).

From the collective results above, we conclude that the N- and C-terminal domains 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 (c.a. 40% identity and 65% similarity) might also be excellent substrates for fusion into active chimeric enzymes.

---

***Fig 9 cont:*** containing the corresponding Traffic Light Reporter target plasmid. Enzyme expression is quantified by co-expressed BFP fluorescence. Low, medium, and high levels of expressed enzyme are gated by BFP fluorescence to determine relative rates of non-homologous end-joining (NHEJ) and homologous recombination (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 roughly 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-OnuI, I-LtrI, and chimeric DNA target sites. (d) Repair of a cleaved target site by the HR pathway (in the presence of a co-transfected 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). (from Certo *et al,* 2011)

### *3.3: 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 N- and C- terminal domains. 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 N- and C-terminal domains (figure 10a). Though previous studies have demonstrated considerable flexibility in linking N- and C-terminal domains 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.[110]

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 alpha 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 N- and C-

terminal domains, connected by a tri-peptide bridge that replaces a section of the linker which is poorly conserved in the I-OnuI family (figure 10a-b, table 3).

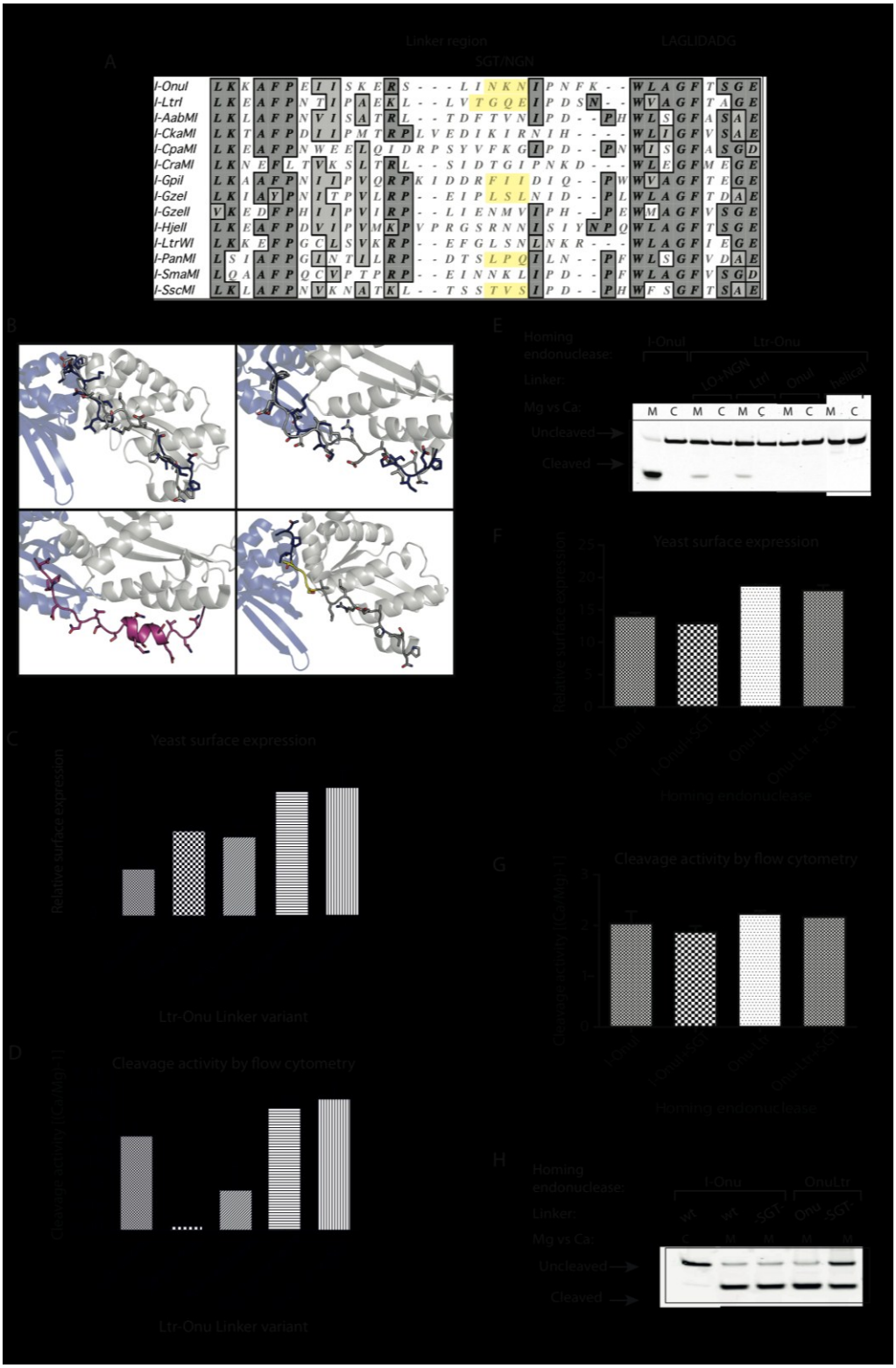



***Figure 10: Analysis of Ltr-Onu expression and activity with various linker strategies.*** (a) Alignment of the highly variable linker peptide sequences from 14 characterized I-OnuI homologues. Brackets indicate the linker peptide sequence and the position of the second LAGLIDADG helix. Residues replaced 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) ***(cont.)***

**Protein sequence:**

| Linker variants | Protein sequence |
|---|---|
| Ltr-Onu with Ltr-linker | LKEAFPNTI-PAEKLLV-TGQEIPDSKWLAGFT |
| Ltr-Onu with ½-and-½ + NGN | LKEAFPNTI-PAEKLLV--NGNIPNFKWLAGFT |
| Ltr-Onu with ½-and-½ + SGT | LKEAFPNTI-PAEKLLV--SGTIPNFKWLAGFT |
| Ltr-Onu with Onu-linker | LKEAFPENI-SKERSLI--NKNIPNFKWLAGFT |
| Ltr-Onu with helical-linker | LKEAFPNNIESASKNNSSVNGNIPNFKWLAGFT |

**DNA sequence:**

| Linker variants | DNA sequence |
|---|---|
| Ltr-Onu with Ltr-linker | TTGAAGGAAGCATTCCCAAACACTATCCCAGCTGAAAAGTTACTAGTTAC<br>TGGTCAAGAAATCCCAGACTCTAAATGGCTGGCTGGATTCACA |
| Ltr-Onu with ½-and-½ + NGN | TTGAAGGAAGCATTCCCAAACACTATCCCAGCTGAAAAGTTACTAGTTAA<br>CGGTAACATTCCGAATTTCAAATGGCTGGCTGGATTCACA |
| Ltr-Onu with ½-and-½ + SGT | TTGAAGGAAGCATTCCCAAACACTATCCCAGCTGAAAAGTTACTAGTTTC<br>GGGTACCATTCCGAATTTCAAATGGCTGGCTGGATTCACA |
| Ltr-Onu with Onu-linker | TTGAAGGAAGCATTCCCAGAGAACATTAGCAAAGAGCGCTCCCTTATCAA<br>TAAGAACATTCCGAATTTCAAATGGCTGGCTGGATTCACA |
| Ltr-Onu with helical-linker | TTGAAGGAAGCATTCCCAAACAACATCGAATCTGCTTCTAAGAACAACTC<br>TTCTGTTAACGGTAACATCCCAAATTTCAAATGGTTGGCTGGATTCACA |

***Table 3: DNA and protein sequences of linker variants tested in LtrOnu***

---

***Fig 10 cont:*** linkers, top view. A small portion of the 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 wildtype 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" vs. "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 minutes. (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 minutes 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).

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: 1) an 'NGN' tri-residue bridge that was suggested by computational analysis to be compatible with both the I-OnuI and I-LtrI structures, and 2) 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 10b). Ltr-Onu chimeras with these variant linker peptides were evaluated using the flow-cytometric yeast surface display assay.[39] 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 10c-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 10c-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 wildtype 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 (12f-h).

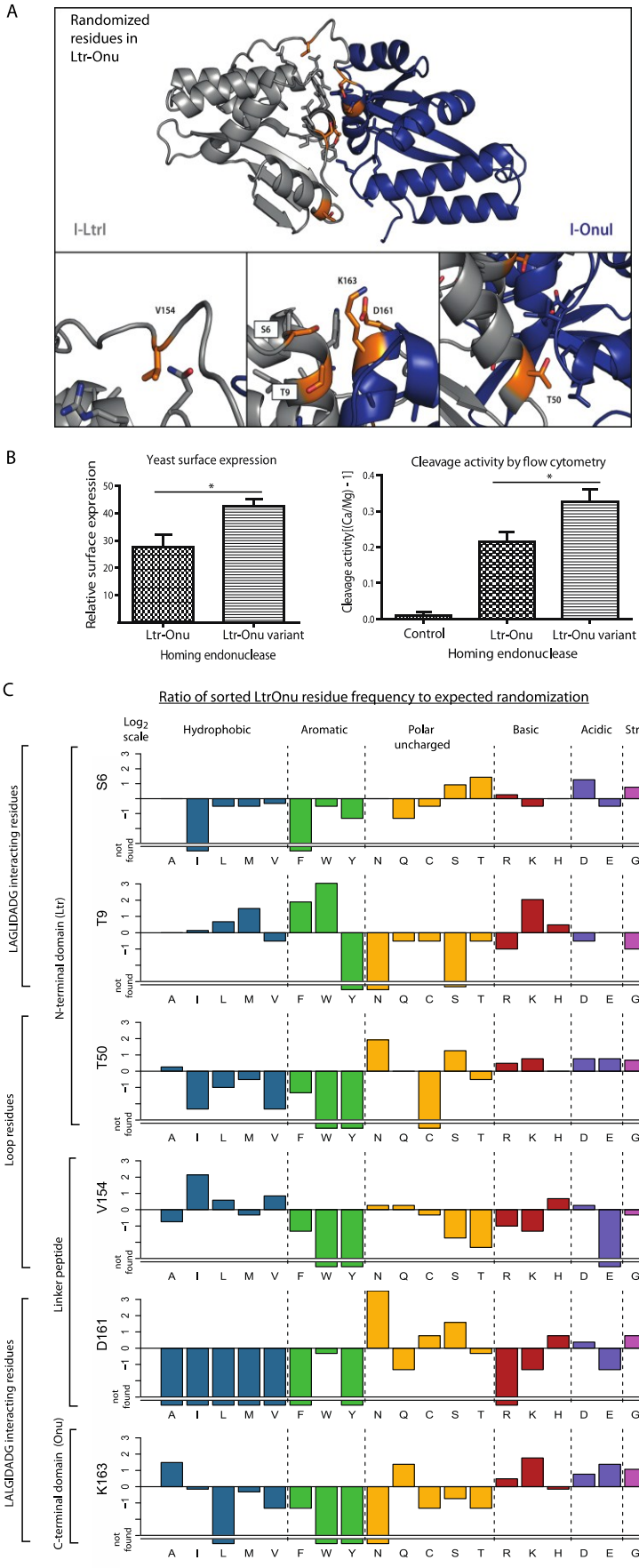
Taken together, these data demonstrate that interactions between the linker and the NTD have an important influence on activity but not stability, consistent with the concept that the linker peptide may subtly influence the relative position of the two domains. In designing chimeras, our data suggest 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 N- and C-terminal domains extracted from single-chain LHEs.

### *3.4: 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 4 residues at the DNA-distal end of the LAGLIDADG helices, along with 2 residues in two side loops (figure 11a). 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.[112]

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 expression. These yeast were sorted, expanded, re-induced, and re-sorted for variants with detectable cleavage activity using the flow-cytometric cleavage assay.[78,86] The top 1-2% of cleaving variants were selected and re-analyzed 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 11b).

Sequence analysis of the recovered population demonstrated strong patterns of residue selection in the sorted Ltr-Onu variants (figure 11c, table 4). 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 2$^{nd}$ 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 homologues, 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

***Figure 11: 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 3 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 ***(continued)***

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.

### *3.5: Chimeras maintain predicted specificity at the "central four" basepairs of 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 12 and table 2). The four middlemost bases of a DNA target sequence cleaved by any type of LHE (designated the 'central four' or 'C4' basepairs) 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 of target DNA observed in LHE/DNA structures.[62]

This is especially important given the limitations of engineering at the central four nucleotides: with a large database of starting scaffolds, design for a given target would likely

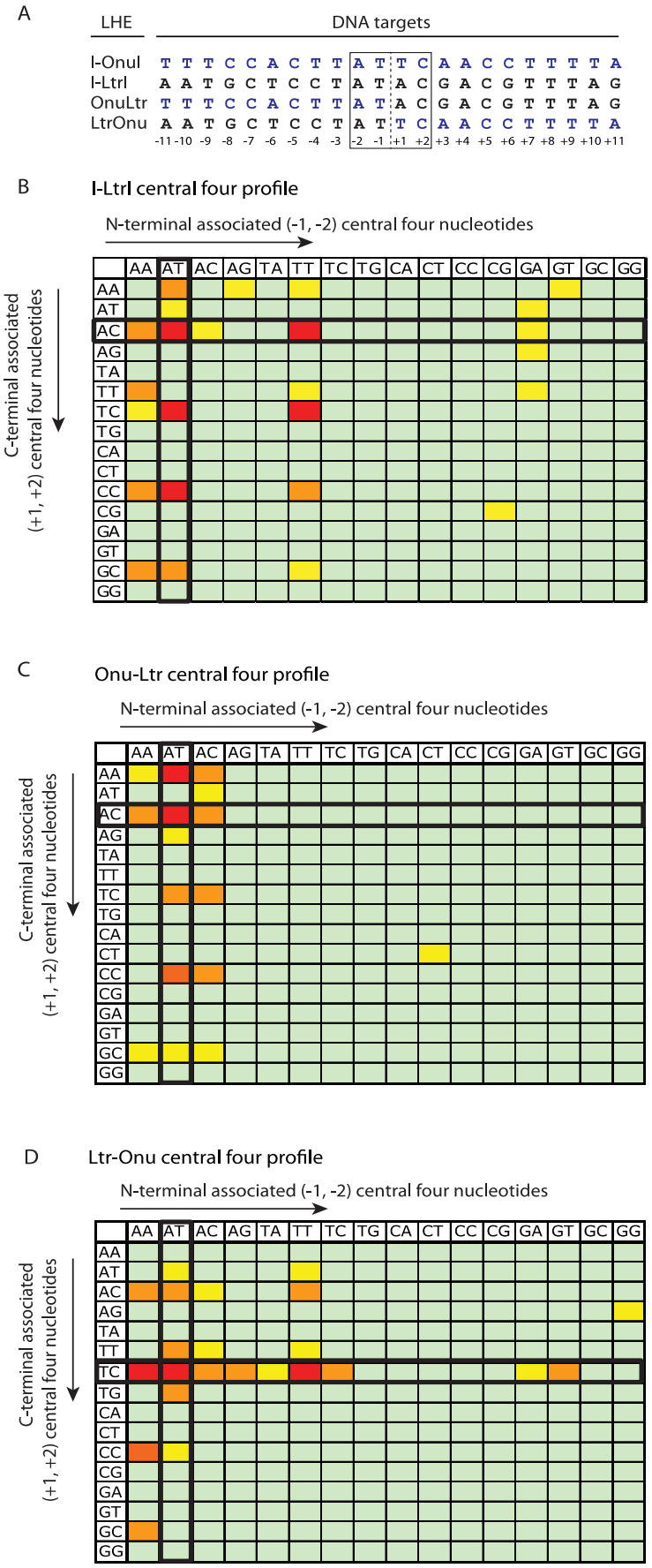
---

***fig 11 cont:*** C-terminal Myc epitope tag. Each bar represents the ratio of median FITC signal from the "expressing" vs. "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 [(Ca/Mg) – 1 ], as described in Figure 1. (*) = $p<0.05$. (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).

| Position: | S 6 | | T 9 | | T 50 | | V 154 | | D 161 | | K 163 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Amino acid | Freq | Fold change | Freq | Fold change | Freq | Fold change | Freq | Fold change | Freq | Fold change | Freq | Fold change |
| Ala | 6.4 | | 6.4 | | 7.5 | | 3.8 | | 0.0 | | 17.3 | |
| Ile | 0.0 | | 5.3 | | 1.1 | | 20.5 | | 0.0 | | 4.0 | |
| Leu | 6.4 | 0.6 | 14.9 | 1.2 | 4.3 | 0.5 | 14.1 | 1.7 | 0.0 | 0.0 | 0.0 | 0.8 |
| Met | 1.1 | | 4.3 | | 1.1 | | 1.3 | | 0.0 | | 1.3 | |
| Val | 5.3 | | 4.3 | | 1.1 | | 11.5 | | 0.0 | | 2.7 | |
| Phe | 0.0 | | 11.7 | | 1.1 | | 1.3 | | 0.0 | | 1.3 | |
| Trp | 1.1 | 0.3 | 12.8 | 2.9 | 0.0 | 0.1 | 0.0 | 0.2 | 1.3 | 0.2 | 0.0 | 0.2 |
| Tyr | 1.1 | | 0.0 | | 0.0 | | 0.0 | | 0.0 | | 0.0 | |
| Asn | 3.2 | | 0.0 | | 11.8 | | 3.8 | | 36.0 | | 0.0 | |
| Gln | 1.1 | | 2.1 | | 3.2 | | 3.8 | | 1.3 | | 8.0 | |
| Ser | 18.1 | 2.1 | 1.1 | 0.5 | 22.6 | 2.1 | 2.6 | 0.7 | 28.0 | 3.8 | 5.3 | 0.9 |
| Cys | 2.1 | | 2.1 | | 0.0 | | 2.6 | | 5.3 | | 1.3 | |
| Thr | 17.0 | | 4.3 | | 4.3 | | 1.3 | | 5.3 | | 2.7 | |
| Arg | 11.7 | | 4.3 | | 12.9 | | 5.1 | | 0.0 | | 13.3 | |
| Lys | 2.1 | 1.0 | 12.8 | 1.3 | 5.4 | 1.3 | 1.3 | 0.7 | 1.3 | 0.4 | 10.7 | 1.6 |
| His | 3.2 | | 4.3 | | 3.2 | | 5.1 | | 5.3 | | 2.7 | |
| Asp | 7.4 | 1.4 | 2.1 | 0.8 | 5.4 | 1.6 | 3.8 | 0.6 | 4.0 | 0.8 | 5.3 | 2.0 |
| Glu | 2.1 | | 3.2 | | 5.4 | | 0.0 | | 1.3 | | 8.0 | |
| Gly | 10.6 | 0.8 | 3.2 | 0.2 | 9.7 | 0.7 | 5.1 | 1.3 | 10.7 | 0.8 | 13.3 | 1.2 |
| Pro | 0.0 | | 0.0 | | 0.0 | | 12.8 | | 0.0 | | 2.7 | |

***Table 4: Selected DNA-distal LAGLIDADG residues.*** Amino acids were selected for six randomized positions in Ltr-Onu. As presented in Fig 11, Ltr-Onu variants were selected for increased expression and cleavage activity from a library with six randomized interface residues. Randomized libraries were created using PCR-based gene assembly, with residues S6, T9, T50, V154, D161, and K163 randomized to NNS (Integrated DNA Technologies). 150 clones were sequenced, with 72-95 sequences attained for each position. Total frequencies of each amino acid in the sorted population are shown in the first column for each position. Fold differences, calculated for biochemically similar amino acids, are shown in the second column; these fold differences were calculated by dividing the measured frequency of each group of residues by the expected frequencies given NNS randomization. (Biochemical groups, from top: hydrophobic, aromatic, polar uncharged, basic, acidic, structural).

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 a sorted stabilized Ltr-Onu variant as it allowed increased sensitivity) (figure 12). This screen showed that the chimeric enzymes

***Figure 12: Specificity of I-LtrI, Onu-Ltr and Ltr-Onu chimeras against central four target variants.*** (a) 22 base-pair 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 ***(continued)***

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 one base-pair (figure 12c). Likewise, Ltr-Onu showed optimal catalytic activity against the bisected C4 variant ATTC, as well as two sequences - AATC and TTTC - varying by one base pair. A majority (7/11) of the sequences against which Ltr-Onu displayed moderate activity also differed by only one nucleotide (figure 12d).

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 12b**)**: 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

---

***Fig 12 cont:*** by the N- and C-terminal domains of the enzyme, respectively. The numbering scheme is indicated below the target sites, with the central four basepairs boxed. (b) The catalytic activity of I-LtrI was analyzed in the flow-cytometric yeast surface display DNA cleavage assay against all potential 264 central four target nucleotides. Nucleotides at positions -1 and -2 (interacting with the N terminal domain) are listed on the x-axis, with nucleotides at positions +1 and +2 (interacting with the C-terminal domain) listed on the y-axis. Boxes containing the native central four nucleotides for each domain are outlined in bold. Cleavage activity against each nucleotide combination is illustrated as a heat-map: light green represents no measurable catalytic activity, with a Ca/Mg ratio of less than 1.1 in the cleavage assay. Yellow, orange, and red represent low, medium, and high levels of cleavage activity. (b-c) Catalytic activity of Onu-Ltr (b) and Ltr-Onu (c) against all 264 potential central four target nucleotides.

(figure 13) showed that it effectively cleaves four C4 variants, including its native ATTC, and shows low or moderate activity against an additional eleven 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 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.

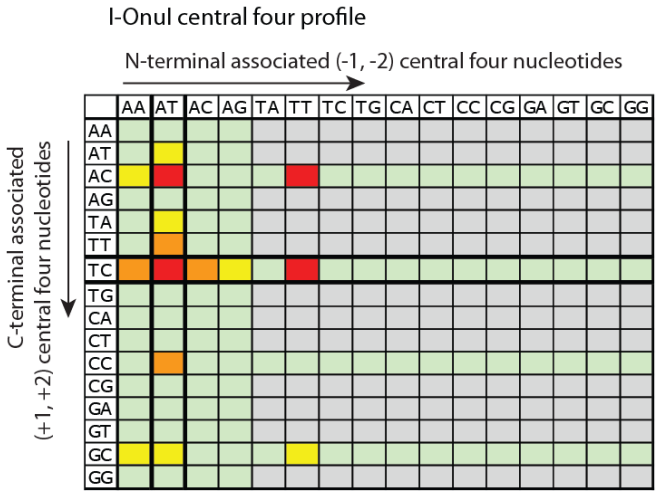



***Figure 13: I-OnuI cleavage activity.*** Cleavage was profiled against target variations at positions -2,-1, +1 and -1, +1, +2. Grey highlighted areas were not tested. Activity was quantified by the flow-cytometric cleavage assay, as described in Figure 12.

### *3.6: Summary*

In order to establish parameters for the extraction of N- and C-terminal domains 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 N- and C-terminal domains extracted from I-OnuI and I-LtrI.

A significant result emerging from our studies is that the linker peptide in single-chain LHEs forms not only important, predictable interactions with the N-terminal domain, but also functionally impacts the LAGLIDADG interface. We observed, for example, an unexpected decrease in catalytic activity in OnuLtr with the incorporation of an –SGT- bridge proximal to the helices and substitution of the Ltr linker residues proximal to the C-terminal domain (figure 10). 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.[49] 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 central four 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 "central four." These data demonstrate that I-OnuI family enzymes have remarkably tight central four 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 base-pairs, domain chimerization will allow for considerable expansion of potential target sites, as the central four nucleotides are not currently targeted for engineering due to their unpredictable biochemistry. Furthermore, the AT-rich nature of the central four targets that are typically cleavable by I-OnuI family enzymes suggests that the energetics of DNA unwinding in

the central four 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.

### *3.7: Materials and methods*

See Chapter 4.6: Materials and Methods

# Chapter 4: Large-scale generation of chimeras by fusion of I-OnuI homologue domains

## *4.1: Introduction*

Using a set of newly characterized enzymes from the I-OnuI family, we have investigated the stabilizing interactions within the pseudodimer interface, with the goal of understanding how to incorporate the structural variation necessary to capture active I-OnuI family chimeras. 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.[66,75] Methods for choosing a linker peptide, introducing interface variation, and determining cleavage specificity across the central four basepairs 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 approximately 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.

## *4.2: 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 N- and C-terminal domains 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, N-terminal and C-terminal domains 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 N-terminal and C-terminal domains extracted from I-OnuI, I-LtrI, and four additional I-OnuI family homologues that have been identified and characterized in our lab, I-GpiI, I-GzeI, I-PanMI, and I-SscMI (figure 14). These enzymes share approximately 40% amino acid sequence identity, with the exception of I-GzeI and I-PanMI, which share over 70% sequence identity. 36 enzymes were made in total, 6 of which were reconstituted native enzymes with the 'SGT' tri-residue substitution in the linker peptide, and 30 of which 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 15a) and by flow cytometry. A summary of data for surface expression, binding, and cleavage activity from this set of enzymes is shown in figure 15e, left panel.

Importantly, all six reconstituted native enzymes exhibited surface expression and activity comparable to their native forms (figure 15a, 15e), 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.

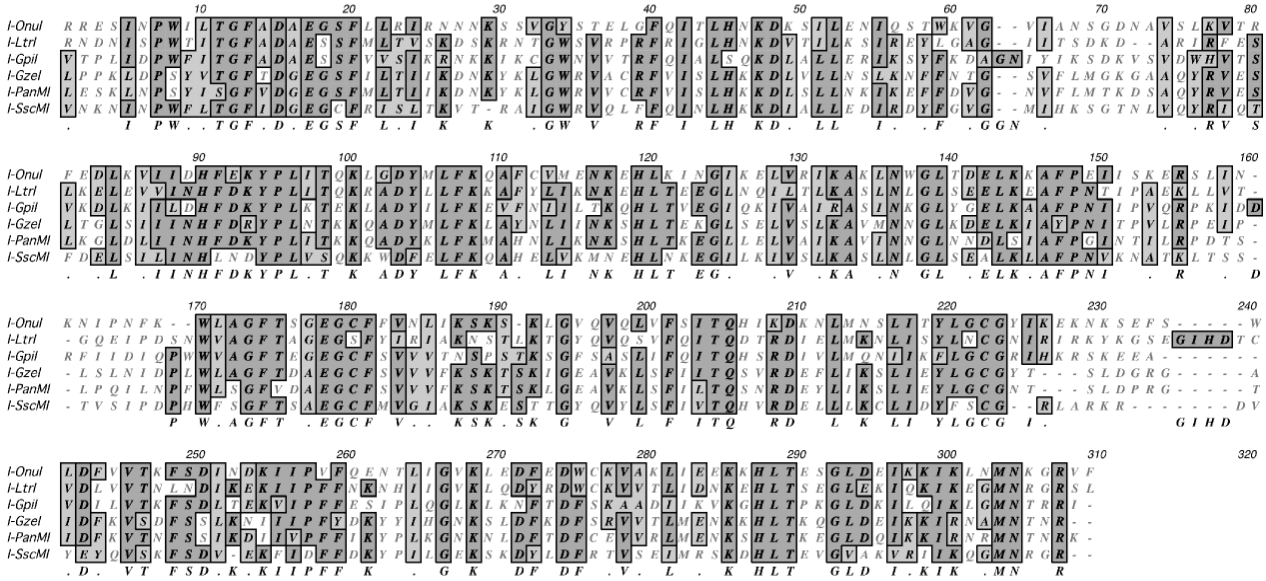



***Figure 14: Amino acid sequences for I-OnuI family LHEs.*** The homing endonucleases I-OnuI, I-LtrI, I-GpiI, I-GzeI, I-PanMI, and I-SscMI are aligned, with identities highlighted in dark grey and similarities in light grey. Our numbering scheme begins 9 residues prior to the canonical LAGLIDADG helix. LHEs are listed according to homology to I-OnuI, as also shown in **Figure 6.** Residue numbers shown above the alignment correspond to numbers referenced in the text.

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 approximately 79% (11/14). Of these expressing and binding enzymes, cleavage was detectable in 81% (9/11) (figure 15 and table 5). 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 central-four possibilities varying only one nucleotide away from the predicted Gpi-Pan target. Gpi-Pan did not show cleavage against any of the alternative central-four 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.

### *4.3: 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 C-terminal domains 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 C-terminal domains 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.[105]

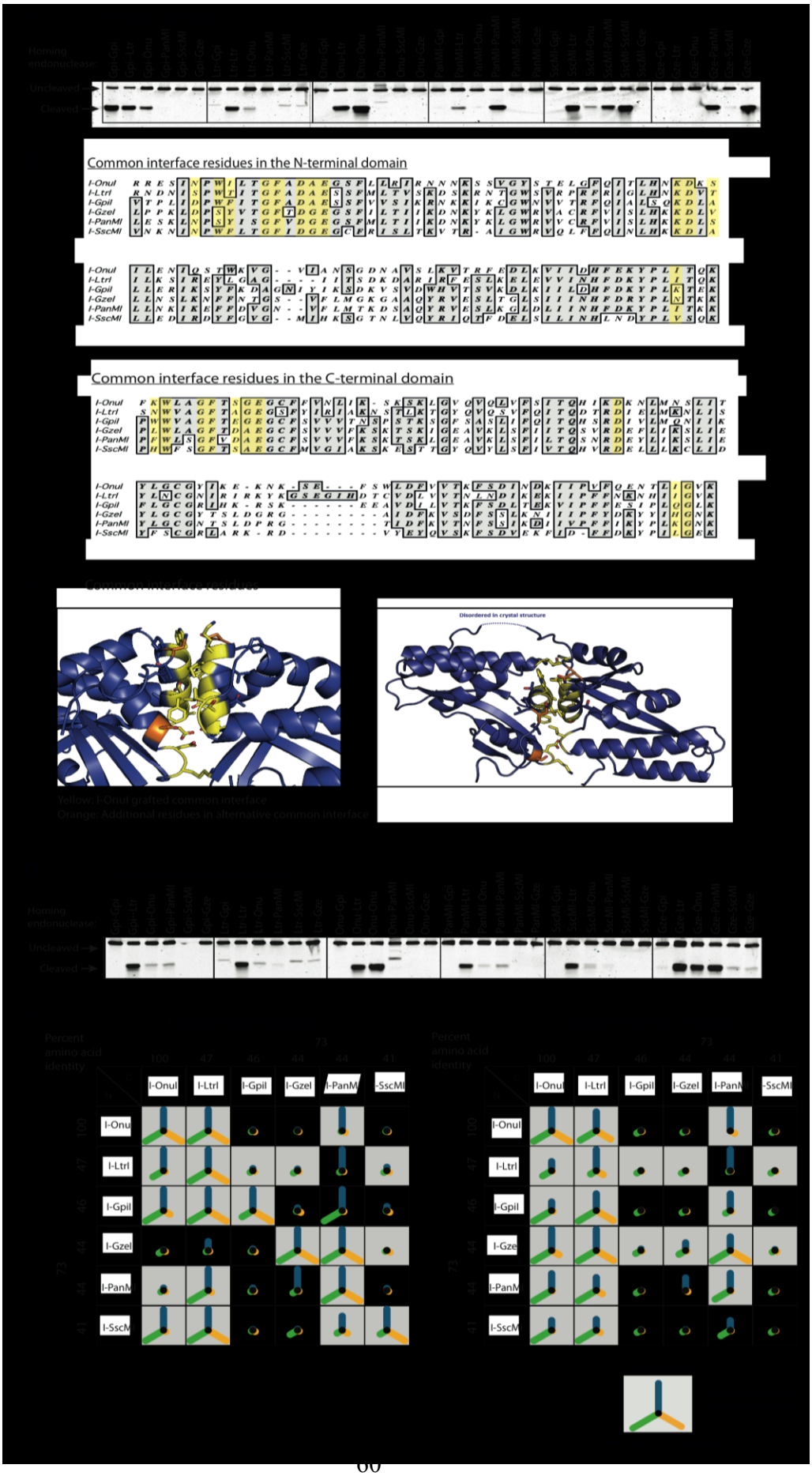



***Figure 15: Domain fusion chimera and common interface chimera screens.***

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 15b, 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

---

***Figure 15: Domain fusion chimera and common interface chimera screens***. N- and C-terminal domains from I-OnuI, I-LtrI, I-GpiI, I-GzeI, I-PanMI, and I-SscMI were combinatorially fused using the 'SGT' linker. Chimeras were generated with an interface composed of entirely native residues (fusion chimeras), and with a set of common interfacial residues originating from I-OnuI (common interface chimeras). (a) Catalytic activity of the fusion chimeras as measured by the *in-vitro* DNA cleavage assay. Enzymes are expressed on the surface of yeast, released with DTT, and incubated with A647-labeled DNA target substrate. Cleavage products are then visualized on an acrylamide gel. This figure is a compilation of 5 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. 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. N-terminal domains are listed along the vertical axis and C-terminal domains 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 vs. uncleaved target in the acrylamide gel. 50% cleavage of the DNA target substrate (after 1 hour 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.

Expression, fusion chimeras (ratio of median APC, expressing/non-expressing, +/- 95% confidence interval)

| N↓ C→ | I-GpiI | I-LtrI | I-OnuI | I-PanMI | I-SscMI | I-GzeI |
|---|---|---|---|---|---|---|
| I-GpiI | 28.07 (23.43, 32.71) | 103.6 (96.05, 111.14) | 46.78 (39.77, 53.78) | 49.53 (42.22, 56.85) | 6.55 (5.92, 7.17) | 5.61 (4.7, 6.53) |
| I-LtrI | 5.21 (5.07, 5.35) | 54.48 (28.77, 80.18) | 17.39 (16.84, 17.94) | 107.07 (92.52, 121.61) | 6.07 (5.7, 6.45) | 4.91 (4.67, 5.16) |
| I-OnuI | 4.3 (4.18, 4.43) | 42.65 (33.52, 51.78) | 17.48 (17.13, 17.82) | 45.18 (29.74, 60.63) | 4.42 (4.41, 4.43) | 4.24 (4.05, 4.44) |
| I-PanMI | 5.39 (4.9, 5.88) | 22.03 (21.31, 22.75) | 5.29 (5.21, 5.38) | 34.49 (26.18, 42.8) | 5.12 (4.98, 5.26) | 49.59 (47.97, 51.21) |
| I-SscI | 4.3 (3.95, 4.64) | 19.66 (17.99, 21.33) | 17.18 (7.65, 26.7) | 12.95 (11.88, 14.02) | 19.37 (17.34, 21.4) | 1.93 (1.92, 1.94) |
| I-GzeI | 4.94 (4.5, 5.38) | 9.68 (6.52, 12.84) | 4.06 (4.05, 4.08) | 33.38 (26.75, 40) | 3.81 (3.54, 4.08) | 17.79 (17.08, 18.51) |

Binding, fusion chimeras (ratio of median APC, expressing/non-expressing, +/- 95% confidence interval)

| N↓ C→ | I-GpiI | I-LtrI | I-OnuI | I-PanMI | I-SscMI | I-GzeI |
|---|---|---|---|---|---|---|
| I-GpiI | 2.36 (0.92, 3.81) | 43.85 (42.07, 45.62) | 15.76 (7.17, 24.35) | 4.99 (2.6, 7.37) | 1.04 (0.99, 1.09) | 1.1 (1.06, 1.14) |
| I-LtrI | 1.32 (0.89, 1.75) | 8.82 (8.74, 8.9) | 2.35 (2.16, 2.54) | 1.3 (1.28, 1.33) | 1.35 (1.23, 1.46) | 1.48 (1.47, 1.5) |
| I-OnuI | 1.15 (1.12, 1.18) | 13.62 (10.26, 16.98) | 3.31 (2.8, 3.81) | 1.12 (1.03, 1.2) | 1.08 (1.02, 1.14) | 1.04 (1.02, 1.06) |
| I-PanMI | 1.23 (1.19, 1.26) | 12.36 (11.04, 13.68) | 1.31 (1.03, 1.6) | 10.7 (4.23, 17.18) | 0.96 (0.1, 1.81) | 1.55 (1.47, 1.63) |
| I-SscI | 1.15 (1.14, 1.16) | 9.14 (8.69, 9.59) | 4.55 (0.75, 8.34) | 1.63 (1.27, 1.98) | 1.84 (1.74, 1.94) | 1.93 (1.92, 1.94) |
| I-GzeI | 1.18 (0.98, 1.38) | 1.22 (1.01, 1.44) | 1.54 (1.46, 1.61) | 24.17 (14.96, 33.37) | 1.16 (1.15, 1.17) | 12.22 (8.54, 15.9) |

Cleavage, fusion chimeras (ratio of median APC fluorescence, cleaved target substrate/uncleaved)

| N↓ C→ | I-GpiI | I-LtrI | I-OnuI | I-PanMI | I-SscMI | I-GzeI |
|---|---|---|---|---|---|---|
| I-GpiI | 6.105 | 2.928 | 0.356 | 0.093 | 0.154 | 0.178 |
| I-LtrI | 0.096 | 1.094 | 0.094 | 0.088 | 0.131 | 0.079 |
| I-OnuI | 0.097 | 3.362 | 5.800 | 0.164 | 0.077 | 0.077 |
| I-PanMI | 0.081 | 0.180 | 0.116 | 3.835 | 0.092 | 0.072 |
| I-SscI | 0.132 | 1.839 | 0.137 | 0.293 | 3.184 | 0.087 |
| I-GzeI | 0.124 | 0.078 | 0.101 | 4.450 | 0.174 | 5.905 |

Expression, common interface chimeras (ratio of median APC, expressing/non-expressing, +/- 95% confidence interval)

| N↓ C→ | I-GpiI | I-LtrI | I-OnuI | I-PanMI | I-SscMI | I-GzeI |
|---|---|---|---|---|---|---|
| I-GpiI | 4.65 (4.33, 4.96) | 20.15 (18.31, 21.99) | 12.6 (9.44, 15.76) | 20.02 (15.03, 25.01) | 4.61 (4.52, 4.7) | 4.72 (4.67, 4.77) |
| I-LtrI | 4.84 (4.46, 5.21) | 32.54 (28.74, 36.35) | 13.07 (11.61, 14.53) | 33.03 (31.45, 34.61) | 4.6 (4.51, 4.69) | 4.55 (4.47, 4.62) |
| I-OnuI | 4.94 (3.94, 5.94) | 22.18 (19.42, 24.95) | 24.1 (23.92, 24.29) | 29.26 (26.46, 32.06) | 4.58 (4.44, 4.73) | 4.7 (4.41, 4.99) |
| I-PanMI | 4.62 (4.58, 4.66) | 16.44 (12.56, 20.33) | 20.06 (16.99, 23.13) | 20.37 (16.97, 23.77) | 4.55 (4.38, 4.72) | 16.24 (13.95, 18.52) |
| I-SscI | 4.95 (4.29, 5.6) | 14.6 (11.88, 17.31) | 13.63 (10.89, 16.37) | 13.61 (11.69, 15.52) | 4.56 (4.27, 4.86) | 5.08 (5.03, 5.14) |
| I-GzeI | 5.91 (5.52, 6.3) | 25.64 (21.03, 30.25) | 26.48 (22.61, 30.35) | 17.2 (13.97, 20.43) | 5.46 (4.56, 6.37) | 11.74 (11.44, 12.03) |

Binding, common interface chimeras (ratio of median APC, expressing/non-expressing, +/- 95% confidence interval)

| N↓ C→ | I-GpiI | I-LtrI | I-OnuI | I-PanMI | I-SscMI | I-GzeI |
|---|---|---|---|---|---|---|
| I-GpiI | 1.93 (1.69, 2.17) | 4.94 (4.7, 5.18) | 8.39 (7.69, 9.08) | 2.08 (2.04, 2.13) | 1.84 (1.7, 1.98) | 1.87 (1.79, 1.95) |
| I-LtrI | 1.97 (1.71, 2.23) | 2.3 (2.22, 2.37) | 2.96 (2.77, 3.14) | 1.48 (1.45, 1.51) | 1.96 (1.75, 2.17) | 2 (1.8, 2.2) |
| I-OnuI | 2 (1.89, 2.11) | 4.38 (3.99, 4.77) | 4.76 (4.28, 5.25) | 2.18 (1.83, 2.54) | 1.72 (1.69, 1.76) | 1.69 (1.5, 1.89) |
| I-PanMI | 2.06 (1.94, 2.18) | 10.59 (8.52, 12.67) | 6.68 (6.22, 7.14) | 4.21 (4.09, 4.34) | 1.83 (1.77, 1.9) | 1.16 (1, 1.31) |
| I-SscI | 1.92 (1.43, 2.42) | 6.82 (6.65, 6.99) | 6.27 (6.19, 6.34) | 3.05 (3, 3.1) | 1.86 (1.51, 2.21) | 1.72 (1.61, 1.84) |
| I-GzeI | 2.2 (2.17, 2.23) | 38.54 (37.61, 39.46) | 27.68 (25.38, 29.98) | 15.9 (14.2, 17.6) | 1.68 (1.48, 1.88) | 2.49 (2.43, 2.55) |

Cleavage, common interface chimeras (ratio of median APC fluorescence, cleaved target substrate/uncleaved)

| N↓ C→ | I-GpiI | I-LtrI | I-OnuI | I-PanMI | I-SscMI | I-GzeI |
|---|---|---|---|---|---|---|
| I-GpiI | 0.031 | 0.937 | 0.087 | 0.080 | 0.000 | 0.072 |
| I-LtrI | 0.052 | 0.382 | 0.040 | 0.037 | 0.065 | 0.040 |
| I-OnuI | 0.060 | 0.797 | 3.793 | 0.246 | 0.057 | 0.046 |
| I-PanMI | 0.047 | 0.334 | 0.055 | 0.067 | 0.050 | 0.124 |
| I-SscI | 0.054 | 0.220 | 0.056 | 0.048 | 0.043 | 0.060 |
| I-GzeI | 0.062 | 0.950 | 0.416 | 2.122 | 0.110 | 0.071 |

***Table 5: Expression, binding and cleavage of I-OnuI fusion and common interface chimeras.*** N- and C-terminal domains from I-OnuI, I-LtrI, I-GpiI, I-GzeI, I-PanMI, and I-SscMI were combinatorially fused as described in figure 19**.** Data are represented as numerical values prior to normalization, including the 95% confidence interval for flow-cytometric assays. N-terminal domains are listed along the y-axis and C-terminal domains along the x-axis. Expression and binding are quantified according to the flow-cytometric assays, as described in the methods. Cleavage activity is quantified according to the *in-vitro* acrylamide gel cleavage assay, also as described in the methods.

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 15d, 15e, 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 15e, table 5). 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) (figure 16).

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 C-terminal domains 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 N-terminal domain of I-LtrI, for which these interface residues had originally been selected (figure 11, 16). 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: 1) the positioning of the N- and C-terminal domains to form stable, active chimeras is significantly influenced by the interfacial residues we identified; 2) 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 3) 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.

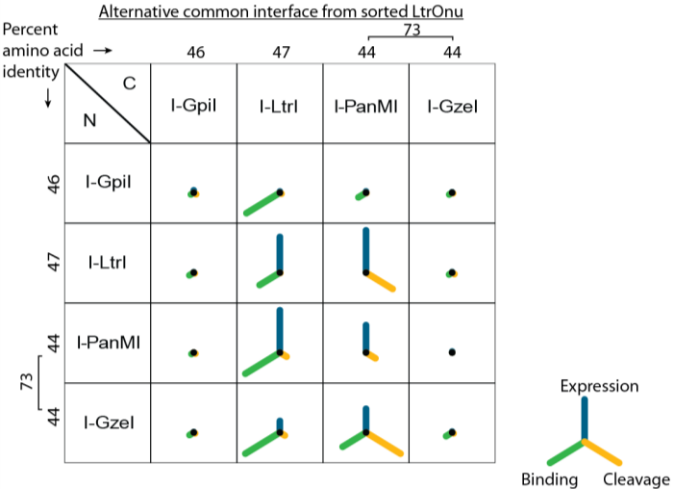



***Figure 16: Expression, binding, and cleavage activity of chimeras with an alternative common interface***
As presented in Figure 11, chimeras incorporating an "alternative common interface" were evaluated for expression, cleavage, and binding activity. The most highly selected residues from the Ltr-Onu sort (T6, W9, S50, I154, N161 and K163) were incorporated and/or substituted into the chimeras with the I-OnuI common interface. Expression, binding, and cleavage activity were quantified by flow-cytometric assays.

### *4.4: Alternative chimerization strategies*

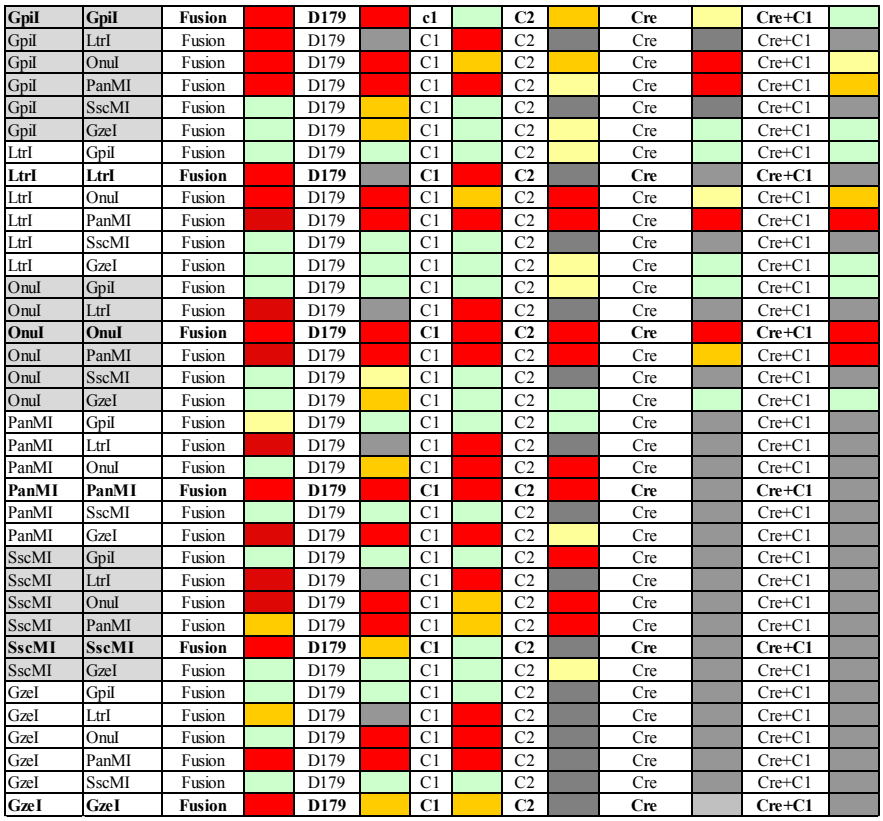
A host of alternative residue substitutions were evaluated using a subset of the above chimeras. These approaches primarily addressed the question of whether it is possible to form a more universal "common interface" that would stabilize a majority of I-OnuI family chimeras, or alternatively

whether the interface of these chimeras could be engineered to act more in accordance with homodimeric LHEs, which show more flexibility at their interface.

In Takeuchi 2010, a vital E->D mutation at the base of the second LAGLIDADG helix restored wild-type level activities in an engineered I-OnuI variant. Incorporating this substitution in the domain fusion chimeras proved successful specifically in chimeras that included an I-OnuI domain, for example by increasing expression in N-terminal I-PanMI with C'I-OnuI (PanMOnu), and yielding catalytically active N-terminal I-GzeI with C'I-OnuI (GzeOnu).[25,35] The majority of chimeras, however, were not improved with this active site mutation. The limited success of this substitution suggests that, while design solutions likely will translate in chimeras containing a domain from the parent engineered scaffold, many solutions on a given LHE scaffold will not necessarily translate to homologous scaffolds (figure 17).

A second generalized substitution method that we investigated stemmed from our previous selection of LtrOnu variants with increased stability by variation at the DNA-distal LAGLIDADG residues. To determine whether the substitutions selected for in the stabilized LtrOnu might be applicable in further I-OnuI family chimeras, we created a set of chimeras with substitutions of the most highly selected LtrOnu variant residues. The chimeras included a common interface of I-OnuI (as above), except with the substitution at the DNA-distal end of the LAGLIDADG helices of these selected residues. As with other techniques, this alternative common interface increased surface expression in some of the chimeras, and decreased expression in others. Catalytic activity was increased in a limited number of cases, such as N'I-LtrI-C'I-PanMI (LtrPanM). As above, the LtrPanM chimera that displayed increased activity includes the N-terminal domain of I-LtrI, on which these interfacial residues had been selected (albeit with a different C-terminal domain). Therefore this set of chimeras suggests that solutions found to stabilize a given LHE could contribute to its stability when paired with alternative domains. No set of interfacial residues, however, appears to be compatible with majority of chimeras.

## Surface expression




## Binding affinity

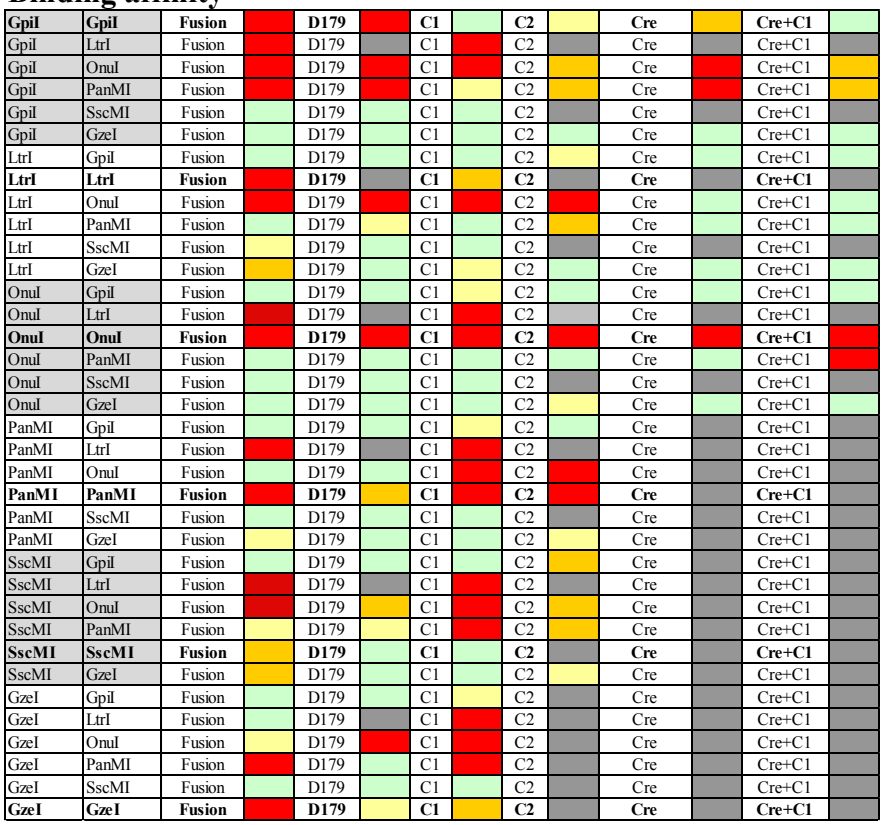



***Figure 17: Expression, binding, and cleavage heat-maps for alternative chimerization techniques***. N-terminal domain of each chimera is listed in the first column, the C-terminal domain in the second column. Left to right: E178: native E at the DNA-proximal end of LAGLIDADG (these are the relative activities of the domain-fusion chimeras shown in **figure 6**; D178: E to D mutation at the DNA-proximal end of the second LAGLIDADG helix, a mutation that reconstituted full activity in an engineered I-OnuI variant in Takeuchi, 2011; Onu (c1) common interface with Onu residues; LtrOnu (c2) common interface with residues from the sorted LtrOnu variant; Cre+fus: Onu common interface residues with water-coordinating residues grafted into the active sites from I-CreI; Cre+c1: I-OnuI common interface with water-coordinating residues from I-CreI. B: Binding heat-map. Binding was determined with 2-50nM target substrate, incubated with surface-expressed chimeras for 2-4 hour C: Cleavage heat map. Cleavage was determined using on-cell ***(continued)***

**Cleavage activity**

| GpiI | GpiI | Fusion | | D179 | | C1 | | C2 | | Cre+fus | | Cre+C1 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| GpiI | LtrI | Fusion | | D179 | | C1 | | C2 | | Cre+fus | | Cre+C1 | |
| GpiI | OnuI | Fusion | | D179 | | C1 | | C2 | | Cre+fus | | Cre+C1 | |
| GpiI | PanMI | Fusion | | D179 | | C1 | | C2 | | Cre+fus | | Cre+C1 | |
| GpiI | SscMI | Fusion | | D179 | | C1 | | C2 | | Cre+fus | | Cre+C1 | |
| GpiI | GzeI | Fusion | | D179 | | C1 | | C2 | | Cre+fus | | Cre+C1 | |
| LtrI | GpiI | Fusion | | D179 | | C1 | | C2 | | Cre+fus | | Cre+C1 | |
| **LtrI** | **LtrI** | **Fusion** | | **D179** | | **C1** | | **C2** | | **Cre+fus** | | **Cre+C1** | |
| LtrI | OnuI | Fusion | | D179 | | C1 | | C2 | | Cre+fus | | Cre+C1 | |
| LtrI | PanMI | Fusion | | D179 | | C1 | | C2 | | Cre+fus | | Cre+C1 | |
| LtrI | SscMI | Fusion | | D179 | | C1 | | C2 | | Cre+fus | | Cre+C1 | |
| LtrI | GzeI | Fusion | | D179 | | C1 | | C2 | | Cre+fus | | Cre+C1 | |
| OnuI | GpiI | Fusion | | D179 | | C1 | | C2 | | Cre+fus | | Cre+C1 | |
| OnuI | LtrI | Fusion | | D179 | | C1 | | C2 | | Cre+fus | | Cre+C1 | |
| **OnuI** | **OnuI** | **Fusion** | | **D179** | | **C1** | | **C2** | | **Cre+fus** | | **Cre+C1** | |
| OnuI | PanMI | Fusion | | D179 | | C1 | | C2 | | Cre+fus | | Cre+C1 | |
| OnuI | SscMI | Fusion | | D179 | | C1 | | C2 | | Cre+fus | | Cre+C1 | |
| OnuI | GzeI | Fusion | | D179 | | C1 | | C2 | | Cre+fus | | Cre+C1 | |
| PanMI | GpiI | Fusion | | D179 | | C1 | | C2 | | Cre+fus | | Cre+C1 | |
| PanMI | LtrI | Fusion | | D179 | | C1 | | C2 | | Cre+fus | | Cre+C1 | |
| PanMI | OnuI | Fusion | | D179 | | C1 | | C2 | | Cre+fus | | Cre+C1 | |
| **PanMI** | **PanMI** | **Fusion** | | **D179** | | **C1** | | **C2** | | **Cre+fus** | | **Cre+C1** | |
| PanMI | SscMI | Fusion | | D179 | | C1 | | C2 | | Cre+fus | | Cre+C1 | |
| PanMI | GzeI | Fusion | | D179 | | C1 | | C2 | | Cre+fus | | Cre+C1 | |
| SscMI | GpiI | Fusion | | D179 | | C1 | | C2 | | Cre+fus | | Cre+C1 | |
| SscMI | LtrI | Fusion | | D179 | | C1 | | C2 | | Cre+fus | | Cre+C1 | |
| SscMI | OnuI | Fusion | | D179 | | C1 | | C2 | | Cre+fus | | Cre+C1 | |
| SscMI | PanMI | Fusion | | D179 | | C1 | | C2 | | Cre+fus | | Cre+C1 | |
| **SscMI** | **SscMI** | **Fusion** | | **D179** | | **C1** | | **C2** | | **Cre+fus** | | **Cre+C1** | |
| SscMI | GzeI | Fusion | | D179 | | C1 | | C2 | | Cre+fus | | Cre+C1 | |
| GzeI | GpiI | Fusion | | D179 | | C1 | | C2 | | Cre+fus | | Cre+C1 | |
| GzeI | LtrI | Fusion | | D179 | | C1 | | C2 | | Cre+fus | | Cre+C1 | |
| GzeI | OnuI | Fusion | | D179 | | C1 | | C2 | | Cre+fus | | Cre+C1 | |
| GzeI | PanMI | Fusion | | D179 | | C1 | | C2 | | Cre+fus | | Cre+C1 | |
| GzeI | SscMI | Fusion | | D179 | | C1 | | C2 | | Cre+fus | | Cre+C1 | |
| **GzeI** | **GzeI** | **Fusion** | | **D179** | | **C1** | | **C2** | | **Cre+fus** | | **Cre+C1** | |

cleavage assay as well as flow-cytometry based on-cell cleavage assay with surface expressed enzyme. Cleavage was allowed to proceed 20 min – 1 hour at 37$^{o}$C.

We also evaluated the possibility that evolutionary drift in monomeric LHEs allowed divergence of active-site residues. In homodimeric LHEs, these residues act in concert to coordinate cleavage of the DNA target. In a monomeric LHE, the kinetics of strand cleavage are often reduced in cleavage of at least one strand; it is possible that mutual mutations have drifted the active site geometry within these monomeric enzymes. We therefore substituted into a panel of chimeras the active-site, water-coordinating residues from the homodimeric, highly-active I-CreI. These residue substitutions were tested in both fusion domain chimeras, in which the interfacial residues from the native enzymes are maintained, and in common-interface chimeras, with the grafted I-OnuI interfacial residues. This approach appeared to increase activity in a very small number of cases, such as LtrGze; the increase

in catalysis measured for LtrGze by the on-cell cleavage assay, however, was countered by an evident decrease in stability, as approximated by surface expression levels. The majority of chimeras showed decreased expression, binding, and activity with this substitution. While these water-coordinating active-site residues are effective in homodimeric I-CreI and engineered I-CreI variants, it is likely that their position has drifted significantly in monomeric LHEs to the point that they are no longer compatible. This is in accordance with the above strategies, which suggested that even within the highly-conserved, single-chain I-OnuI homologues, solutions selected for a given enzyme are primarily effective in chimeras containing a domain from the original enzyme.

### *4.5: Summary*

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.[75,113] Using insights from the fusion of N- and C-terminal domains of the well-characterized homologues I-OnuI and I-LtrI, 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 approximately 50% of cases, and that introduction of even limited variation into the interface residues allows for recovery of active enzymes from nearly 70% of domain fusion pairs.

As was shown in the studies of OnuLtr and LtrOnu chimeras, the linker peptide in single-chain LHEs forms important interactions with the N-terminal domain; the linker peptide appears to additionally interact with LAGLIDADG interface, altering the functionality of the enzyme with alterations introduced proximal to the helices. 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 15e, left panel), we observed a few examples where alteration of linker composition led to a decrease in activity, as was seen in OnuLtr.

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 over 70% identity, achieved only a 50% success rate (figure 18). 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 chimeras were achieved in 7/10 instances when an I-OnuI domain was used with the I-OnuI-derived common interface. Likewise, the incorporation of a vital E->D mutation previously documented for I-OnuI primarily increased stability and activity in chimeras containing and I-OnuI domain, and substitution of experimentally sorted residues on LtrOnu increased activity in a chimera containing an I-LtrI domain. While all of the generalized strategies served to increase activity in only a minority of the panel of chimeras, the solutions were effective in a number of chimeras containing a domain from the parent enzyme on which these "optimizations" were based.

Overall amino acid identity:

| | I-OnuI | I-LtrI | I-GpiI | I-GzeI | I-PanMI | I-SscMI |
|---|---|---|---|---|---|---|
| I-OnuI | 100% | 47% | 46% | 44% | 44% | 41% |
| I-LtrI | 47% | 100% | 43% | 44% | 44% | 44% |
| I-GpiI | 46% | 43% | 100% | 44% | 43% | 39% |
| I-GzeI | 44% | 44% | 44% | 100% | 73% | 42% |
| I-PanMI | 44% | 44% | 43% | 73% | 100% | 42% |
| I-SscMI | 41% | 44% | 39% | 42% | 42% | 100% |

LAGLIDADG helices amino acid identity:

| | I-OnuI | I-LtrI | I-GpiI | I-GzeI | I-PanMI | I-SscMI |
|---|---|---|---|---|---|---|
| I-OnuI | 100% | 68% | 68% | 50% | 56% | 68% |
| I-LtrI | 68% | 100% | 68% | 50% | 50% | 62% |
| I-GpiI | 68% | 68% | 100% | 56% | 50% | 62% |
| I-GzeI | 50% | 50% | 56% | 100% | 81% | 62% |
| I-PanMI | 56% | 50% | 50% | 81% | 100% | 68% |
| I-SscMI | 68% | 62% | 62% | 62% | 68% | 100% |

"Common interface" residues' amino acid identity:

| | I-OnuI | I-LtrI | I-GpiI | I-GzeI | I-PanMI | I-SscMI |
|---|---|---|---|---|---|---|
| I-OnuI | 100% | 73% | 65% | 52% | 65% | 65% |
| I-LtrI | 73% | 100% | 65% | 52% | 56% | 60% |
| I-GpiI | 65% | 65% | 100% | 56% | 52% | 65% |
| I-GzeI | 52% | 52% | 56% | 100% | 73% | 60% |
| I-PanMI | 65% | 56% | 52% | 73% | 100% | 65% |
| I-SscMI | 65% | 60% | 65% | 60% | 65% | 100% |

***Figure 18: Identity matrix of I-OnuI homologues.***
Comparative tables showing overall amino acid percent identity (top), as well as identity in only the LAGLIDADG helices (middle) and across all defined "common interface" residues (bottom).

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 N- and C-terminal domains 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 nearly 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 approximately 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.[91,106] 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.

### *4.6 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 N- and C-terminal domains of the I-OnuI homologues were constructed by gene assembly PCR. Assembly primers (50-70bp) 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 N-terminal domain) or XhoI (C-terminal domain) (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 alpha cells and sequenced to isolate full-length clones; plasmid preparations of these clones were then transformed into yeast using the lithium acetate protocol.[40] See below for DNA and protein sequences used in all applications.

**DNA and protein sequences of I-OnuI homologues used for chimerization**

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. The tri-residue -SGT- linker substitution is highlighted in bold and underlined; this substitution contains a KpnI restriction site which

allowed for efficient swapping of N- and C-terminal domains between homologues. Numbered residues in the manuscript refer to the sequence alignment shown in Figure 14.

## I-OnuI

### DNA sequence:

TCTAGAAGAGAATCTATCAACCCATGGATCCTGACTGGTTTCGCTGACGCTGAAGGTTCTTTCCTGCTGAGAATCA
GAAACAACAACAAGTCTTCTGTCGGTTATTCTACTGAACTGGGTTTCCAGATCACTCTGCACAACAAGGACAAGTC
TATCCTGGAAAACATCCAGTCTACTTGGAAGGTCGGTGTCATCGCTAACTCTGGTGACAACGCTGTCTCTCTGAAG
GTCACTAGATTCGAAGACCTGAAGGTCATCATCGACCACTTCGAAAAGTATCCACTGATCACTCAGAAGCTGGGTG
ACTATATGCTGTTCAAGCAGGCTTTCTGTGTCATGGAAAACAAGGAACACCTGAAGATCAACGGTATCAAGGAAC
TGGTCAGAATCAAGGCTAAGCTGAACTGGGGTCTGACTGACGAACTGAAGAAGGCTTTCCCAGAAATCATCTCTA
AGGAAAGATCTCTGATCAACAAGAACATCCCAAACTTCAAGTGGCTGGCTGGTTTCACTTCTGGTGAAGGTTGTTT
CTTCGTCAACCTGATCAAGTCTAAGTCTAAGCTGGGTGTCCAGGTCCAGCTGGTCTTCTCTATCACTCAGCACATCA
AGGACAAGAACCTGATGAACTCTCTGATCACTTATCTGGGTTGTGGTTATATCAAGGAAAAGAACAAGTCTGAATT
CTCTTGGCTGGACTTCGTCGTCACTAAGTTCTCTGACATCAACGACAAGATCATCCCAGTCTTCCAGGAAAACACT
CTGATCGGTGTCAAGCTGGAAGACTTCGAAGACTGGTGTAAGGTCGCTAAGCTGATCGAAGAAAAGAAGCACCTG
ACTGAATCTGGTCTGGACGAAATCAAGAAGATCAAGCTGAACATGAACAAGGGTAGAGTCTTC

### Protein sequence:

SRRESINPWILTGFADAEGSFLLRIRNNNKSSVGYSTELGFQITLHNKDKSILENIQSTWKVGVIANSGDNAVSLKVTRFE
DLKVIIDHFEKYPLITQKLGDYMLFKQAFCVMENKEHLKINGIKELVRIKAKLNWGLTDELKKAFPEIISKERSLINKNIP
NFKWLAGFTSGEGCFFVNLIKSKSKLGVQVQLVFSITQHIKDKNLMNSLITYLGCGYIKEKNKSEFSWLDFVVTKFSDIN
DKIIPVFQENTLIGVKLEDFEDWCKVAKLIEEKKHLTESGLDEIKKIKLNMNKGRVF

## I-OnuI with SGT Linker (SGT Linker bold with underline)

### DNA sequence:

TCTAGAAGAGAATCTATCAACCCATGGATCCTGACTGGTTTCGCTGACGCTGAAGGTTCTTTCCTGCTGAGAATCA
GAAACAACAACAAGTCTTCTGTCGGTTATTCTACTGAACTGGGTTTCCAGATCACTCTGCACAACAAGGACAAGTC
TATCCTGGAAAACATCCAGTCTACTTGGAAGGTCGGTGTCATCGCTAACTCTGGTGACAACGCTGTCTCTCTGAAG
GTCACTAGATTCGAAGACCTGAAGGTCATCATCGACCACTTCGAAAAGTATCCACTGATCACTCAGAAGCTGGGTG
ACTATATGCTGTTCAAGCAGGCTTTCTGTGTCATGGAAAACAAGGAACACCTGAAGATCAACGGTATCAAGGAAC
TGGTCAGAATCAAGGCTAAGCTGAACTGGGGTCTGACTGACGAACTGAAGAAGGCTTTCCCAGAAATCATCTCTA
AGGAAAGATCTCTGATC**TCGGGTACC**ATCCCAAACTTCAAGTGGCTGGCTGGTTTCACTTCTGGTGAAGGTTGTTT
CTTCGTCAACCTGATCAAGTCTAAGTCTAAGCTGGGTGTCCAGGTCCAGCTGGTCTTCTCTATCACTCAGCACATCA
AGGACAAGAACCTGATGAACTCTCTGATCACTTATCTGGGTTGTGGTTATATCAAGGAAAAGAACAAGTCTGAATT
CTCTTGGCTGGACTTCGTCGTCACTAAGTTCTCTGACATCAACGACAAGATCATCCCAGTCTTCCAGGAAAACACT
CTGATCGGTGTCAAGCTGGAAGACTTCGAAGACTGGTGTAAGGTCGCTAAGCTGATCGAAGAAAAGAAGCACCTG
ACTGAATCTGGTCTGGACGAAATCAAGAAGATCAAGCTGAACATGAACAAGGGTAGAGTCTTC

### Protein sequence:

SRRESINPWILTGFADAEGSFLLRIRNNNKSSVGYSTELGFQITLHNKDKSILENIQSTWKVGVIANSGDNAVSLKVTRFE
DLKVIIDHFEKYPLITQKLGDYMLFKQAFCVMENKEHLKINGIKELVRIKAKLNWGLTDELKKAFPEIISKERSLI**SGT**IPN
FKWLAGFTSGEGCFFVNLIKSKSKLGVQVQLVFSITQHIKDKNLMNSLITYLGCGYIKEKNKSEFSWLDFVVTKFSDIND
KIIPVFQENTLIGVKLEDFEDWCKVAKLIEEKKHLTESGLDEIKKIKLNMNKGRVF

## I-LtrI

### DNA sequence:

TTCCCAGTTCAAGCTAGAAACGACAACATCTCTCCATGGACTATCACTGGTTTCGCTGACGCTGAATCTTCTTTCAT
GTTGACTGTTTCTAAGGACTCTAAGAGAAACACTGGTTGGTCTGTTAGACCAAGATTCAGAATCGGTTTGCACAAC
AAGGACGTGACTATCTTGAAGTCTATCAGAGAATACTTGGGCGCCGGTATCATCACTTCTGACAAGGACGCTAGAA
TCAGATTCGAATCTTTGAAGGAATTGGAAGTTGTTATCAACCACTTCGACAAGTACCCATTGATCACTCAAAAGAG
AGCTGACTACTTGTTGTTCAAGAAGGCTTTCTACTTAATTAAGAACAAGGAACACTTGACTGAAGAAGGTTTGAAC
CAAATCTTGACTTTGAAGGCTTCTTTGAACTTGGGTTTGTCTGAAGAATTGAAGGAAGCATTCCCAAACACTATCC
CAGCTGAAAAGTTACTAGTTACTGGTCAAGAAATCCCAGACTCTAACTGGGTTGCTGGTTTCACTGCTGGTGAAGG
TTCTTTCTACATCAGAATCGCTAAGAACTCTACTTTGAAGACTGGTTACCAAGTTCAATCTGTTTTCCAAATCACTC
AAGACACGCGTGACATCGAATTGATGAAGAACTTGATCTCTTACTTGAACTGTGGTAACATCAGAATCAGAAAGT
ACAAGGGTTCTGAAGGTATCCACGACACTTGTGTTGACTTGGTTGTTACTAACTTGAACGACATCAAGGAAAAGAT
CATCCCATTCTTCAACAAGAACCACATCATCGGTGTTAAGTTGCAAGACTACAGAGACTGGTGTAAGGTTGTTACT
TTGATCGACAACAAGGAACACTTGACTTCTGAAGGTTTGGAAAAGATCCAAAAGATCAAGGAAGGTATGAACAGA
GGTAGATCTTTG

### Protein sequence:

FPVQARNDNISPWTITGFADAESSFMLTVSKDSKRNTGWSVRPRFRIGLHNKDVTILKSIREYLGAGIITSDKDARIRFESL
KELEVVINHFDKYPLITQKRADYLLFKKAFYLIKNKEHLTEEGLNQILTLKASLNLGLSEELKEAFPNTIPAEKLLVTGQEI
PDSNWVAGFTAGEGSFYIRIAKNSTLKTGYQVQSVFQITQDTRDIELMKNLISYLNCGNIRIRKYKGSEGIIHDTCVDLVV
TNLNDIKEKIIPFFNKNHIIGVKLQDYRDWCKVVTLIDNKEHLTSEGLEKIQKIKEGMNRGRSL

## I-LtrI with SGT Linker (SGT Linker bold with underline)

### DNA sequence:

TTCCCAGTTCAAGCTAGAAACGACAACATCTCTCCATGGACTATCACTGGTTTCGCTGACGCTGAATCTTCTTTCAT
GTTGACTGTTTCTAAGGACTCTAAGAGAAACACTGGTTGGTCTGTTAGACCAAGATTCAGAATCGGTTTGCACAAC
AAGGACGTGACTATCTTGAAGTCTATCAGAGAATACTTGGGCGCCGGTATCATCACTTCTGACAAGGACGCTAGAA
TCAGATTCGAATCTTTGAAGGAATTGGAAGTTGTTATCAACCACTTCGACAAGTACCCATTGATCACTCAAAAGAG
AGCTGACTACTTGTTGTTCAAGAAGGCTTTCTACTTAATTAAGAACAAGGAACACTTGACTGAAGAAGGTTTGAAC
CAAATCTTGACTTTGAAGGCTTCTTTGAACTTGGGTTTGTCTGAAGAATTGAAGGAAGCATTCCCAAACACTATCC
CAGCTGAAAAGTTACTAGTT**TCGGGTACC**ATCCCAGACTCTAACTGGGTTGCTGGTTTCACTGCTGGTGAAGGTTC
TTTCTACATCAGAATCGCTAAGAACTCTACTTTGAAGACTGGTTACCAAGTTCAATCTGTTTTCCAAATCACTCAAG
ACACGCGTGACATCGAATTGATGAAGAACTTGATCTCTTACTTGAACTGTGGTAACATCAGAATCAGAAAGTACAA
GGGTTCTGAAGGTATCCACGACACTTGTGTTGACTTGGTTGTTACTAACTTGAACGACATCAAGGAAAAGATCATC
CCATTCTTCAACAAGAACCACATCATCGGTGTTAAGTTGCAAGACTACAGAGACTGGTGTAAGGTTGTTACTTTGA
TCGACAACAAGGAACACTTGACTTCTGAAGGTTTGGAAAAGATCCAAAAGATCAAGGAAGGTATGAACAGAGGTA
GATCTTTG

### Protein sequence:

FPVQARNDNISPWTITGFADAESSFMLTVSKDSKRNTGWSVRPRFRIGLHNKDVTILKSIREYLGAGIITSDKDARIRFESL
KELEVVINHFDKYPLITQKRADYLLFKKAFYLIKNKEHLTEEGLNQILTLKASLNLGLSEELKEAFPNTIPAEKLLV**SGT**IP
DSNWVAGFTAGEGSFYIRIAKNSTLKTGYQVQSVFQITQDTRDIELMKNLISYLNCGNIRIRKYKGSEGIIHDTCVDLVVT
NLNDIKEKIIPFFNKNHIIGVKLQDYRDWCKVVTLIDNKEHLTSEGLEKIQKIKEGMNRGRSL

## Onu-Ltr Chimera with N-terminal-derived Linker

### DNA sequence:

GCTACTGTTACTCCATTGATCGACCCATGGTTCATCACTGGTTTCGCTGACGCTGAATCTTCTTTCGTTGTTTCTATC
AAGAGAAACAAGAAGATCAAGTGTGGTTGGAACGTTGTTACTAGATTCCAAATCGCCTTAAGTCAAAAGGACTTG
GCTTTGTTGGAAAGAATCAAGTCTTACTTCAAGGACGCTGGTAACATCTACATCAAGTCTGACAAGGTTTCTGTTG
ACTGGCACGTTACTTCTGTTAAGGACTTGAAGATCATCCTTGATCACTTCGACAAGTACCCATTGAAGACTGAAAA
GTTGGCTGACTACATCTTGTTCAAGGAAGTTTTCAACATCATCTTGACTAAGCAACACTTGACTGTTGAAGGTATCC
AAAAGATCGTTGCTATCAGAGCTTCTATCAACAAGGGTTTGTACGGTGAATTGAAGGCTGCATTCCCAAACATCAT
CCCAGTTCAAAGGCCTAAGATCGACGACAGA**TCGGGTACC**GATATCCAACCATGGTGGGTTGCTGGTTTCACTGA
AGGTGAAGGTTGTTTCTCTGTTGTTGTTACTAACTCTCCATCTACTAAGTCTGGTTTCTCTGCTTCTTTGATCTTCCA
AATCACTCAACACTCTCGTGACATCGTTTTGATGCAAAACATCATCAAGTTCCTAGGTTGTGGTAGAATCCACAAG
AGATCTAAGGAAGAAGCTGTTGACATCTTGGTTACTAAGTTCTCTGACTTGACTGAAAAGGTTATCCCATTCTTCG
AATCTATCCCATTGCAAGGTTTGAAGTTGAAGAACTTCACTGACTTCTCTAAGGCTGCTGACATCATCAAGGTTAA
GGGTCACTTGACTCCAAAGGGTTTGGACAAGATCTTGCAAATCAAGTTGGGTATGAACACTAGAAGAATC

### Protein sequence:

SRRESINPWILTGFADAEGSFLLRIRNNNKSSVGYSTELGFQITLHNKDKSILENIQSTWKVGVIANSGDNAVSLKVTRFE
DLKVIIDHFEKYPLITQKLGDYMLFKQAFCVMENKEHLKINGIKELVRIKAKLNWGLTDELKKAFPEIISKERSLVTGQKI
PDSNWVAGFTAGEGSFYIRIAKNSTLKTGYQVQSVFQITQDTRDIELMKNLISYLNCGNIRIRKYKGSEGIHDTCVDLVV
TNLNDIKEKIIPFFNKNHIIGVKLQDYRDWCKVVTLIDNKEHLTSEGLEKIQKIKEGMNRGRSL

## Ltr-Onu Chimera with N-terminal-derived Linker

### DNA sequence:

TTCCCAGTTCAAGCTAGAAACGACAACATCTCTCCATGGACTATCACTGGTTTCGCTGACGCTGAATCTTCTTTCAT
GTTGACTGTTTCTAAGGACTCTAAGAGAAACACTGGTTGGTCTGTTAGACCAAGATTCAGAATCGGTTTGCACAAC
AAGGACGTGACTATCTTGAAGTCTATCAGAGAATACTTGGGCGCCGGTATCATCACTTCTGACAAGGACGCTAGAA
TCAGATTCGAATCTTTGAAGGAATTGGAAGTTGTTATCAACCACTTCGACAAGTACCCATTGATCACTCAAAAGAG
AGCTGACTACTTGTTGTTCAAGAAGGCTTTCTACTTAATTAAGAACAAGGAACACTTGACTGAAGAAGGTTTGAAC
CAAATCTTGACTTTGAAGGCTTCTTTGAACTTGGGTTTGTCTGAAGAATTGAAGGAAGCATTCCCAAACACTATCC
CAGCTGAAAAGTTACTAGTTACTGGTCAAGAAATCCCAGACTCTAAATGGTTGGCTGGATTCACATCTGGTGAAGG
TTGTTTCTTCGTTAATTTGATTAAATCTAAATCTAAATTGGGTGTTCAAGTTCAATTGGTTTTCTCTATTACTCAACA
TATTAAAGATAAAAATCTGATGAATTCTTTGATTACCTACCTAGGTTGTGGTTATATTAAAGAAAAAAATAAATCT
GAATTCTCTTGGTTGGATTTCGTGGTTACCAAATTCTCTGATATTAATGATAAAATTATTCCAGTTTTCCAAGAAAA
TACTTTGATTGGTGTTAAATTGGAAGATTTCGAAGATTGGTGTAAAGTTGCTAAATTGATTGAAGAAAAAAAACAT
TTGACTGAATCTGGTTTGGATGAAATTAAAAAAATTAAATTGAATATGAATAAAGGTAGAGTTTTC

### Protein sequence:

FPVQARNDNISPWTITGFADAESSFMLTVSKDSKRNTGWSVRPRFRIGLHNKDVTILKSIREYLGAGIITSDKDARIRFESL
KELEVVINHFDKYPLITQKRADYLLFKKAFYLIKNKEHLTEEGLNQILTLKASLNLGLSEELKEAFPNTIPAEKLLVTGQEI
PDSKWLAGFTSGEGCFFVNLIKSKSKLGVQVQLVFSITQHIKDKNLMNSLITYLGCGYIKEKNKSEFSWLDFVVTKFSDI
NDKIIPVFQENTLIGVKLEDFEDWCKVAKLIEEKKHLTESGLDEIKKIKLNMNKGRVF

**I-GpiI with SGT Linker** (SGT Linker bold with underline)

### DNA sequence:

GCTACTGTTACTCCATTGATCGACCCATGGTTCATCACTGGTTTCGCTGACGCTGAATCTTCTTTCGTTGTTTCTATC
AAGAGAAACAAGAAGATCAAGTGTGGTTGGAACGTTGTTACTAGATTCCAAATCGCCTTAAGTCAAAAGGACTTG
GCTTTGTTGGAAAGAATCAAGTCTTACTTCAAGGACGCTGGTAACATCTACATCAAGTCTGACAAGGTTTCTGTTG
ACTGGCACGTTACTTCTGTTAAGGACTTGAAGATCATCCTTGATCACTTCGACAAGTACCCATTGAAGACTGAAAA
GTTGGCTGACTACATCTTGTTCAAGGAAGTTTTCAACATCATCTTGACTAAGCAACACTTGACTGTTGAAGGTATCC
AAAAGATCGTTGCTATCAGAGCTTCTATCAACAAGGGTTTGTACGGTGAATTGAAGGCTGCATTCCCAAACATCAT
CCCAGTTCAAAGGCCTAAGATCGACGACAGA**TCGGGTACC**GATATCCAACCATGGTGGGTTGCTGGTTTCACTGA
AGGTGAAGGTTGTTTCTCTGTTGTTGTTACTAACTCTCCATCTACTAAGTCTGGTTTCTCTGCTTCTTTGATCTTCCA
AATCACTCAACACTCTCGTGACATCGTTTTGATGCAAAACATCATCAAGTTCCTAGGTTGTGGTAGAATCCACAAG
AGATCTAAGGAAGAAGCTGTTGACATCTTGGTTACTAAGTTCTCTGACTTGACTGAAAAGGTTATCCCATTCTTCG
AATCTATCCCATTGCAAGGTTTGAAGTTGAAGAACTTCACTGACTTCTCTAAGGCTGCTGACATCATCAAGGTTAA
GGGTCACTTGACTCCAAAGGGTTTGGACAAGATCTTGCAAATCAAGTTGGGTATGAACACTAGAAGAATC

### Protein sequence:

ATVTPLIDPWFITGFADAESSFVVSIKRNKKIKCGWNVVTRFQIALSQKDLALLERIKSYFKDAGNIYIKSDKVSVDWHV
TSVKDLKIILDHFDKYPLKTEKLADYILFKEVFNIILTKQHLTVEGIQKIVAIRASINKGLYGELKAAFPNIIPVQRPKIDDR
**SGT**DIQPWWVAGFTEGEGCFSVVVTNSPSTKSGFSASLIFQITQHSRDIVLMQNIIKFLGCGRIHKRSKEEAVDILVTKFS
DLTEKVIPFFESIPLQGLKLKNFTDFSKAADIIKVKGHLTPKGLDKILQIKLGMNTRRI

**I-PanMI with SGT Linker** (SGT Linker bold with underline)

### DNA sequence:

TCTACTTTGGAATCTAAGTTGAACCCATCTTACATCTCTGGTTTCGTCGACGGTGAAGGTTCTTTCATGTTGACTAT
CATCAAGGACAACAAGTACAAGTTGGGTTGGAGAGTTGTTTGTAGATTCGTTATCTCTTTGCACAAGAAGGACTTG
TCTTTGTTGAACAAGATCAAGGAATTTTTCGACGTCGGTAACGTTTTCTTGATGACTAAGGACTCTGCTCAATACAG
AGTTGAATCTTTGAAGGGTTTGGACTTGATCATCAACCACTTCGACAAGTACCCATTGATCACTAAGAAGCAAGCT
GACTACAAGTTGTTCAAGATGGCTCACAACTTAATTAAGAACAAGTCTCACTTGACTAAGGAAGGTTTGTTGGAAT
TGGTTGCTATCAAGGCTGTTATCAACAACGGTTTGAACAACGACTTGTCTATCGCTTTCCCAGGTATCAACACTATC
TTGAGGCCTGACACT**TCGGGTACC**ATCTTGAACCCATTCTGGTTGTCTGGTTTCGTTGACGCTGAAGGTTGTTTCTC
TGTTGTTGTTTTCAAGTCTAAGACTTCTAAGTTGGGTGAAGCTGTTAAGTTGTCTTTCATCTTGACTCAATCTAACA
GAGACGAATACTTGATCAAGTCTTTGATCGAATACCTAGGTTGTGGTAACACTTCTTTGGACCCAAGAGGTACTAT
CGACTTCAAGGTTACTAACTTCTCTTCTATCAAGGACATCATCGTTCCATTCTTCATCAAGTACCCATTGAAGGGTA
ACAAGAACTTGGACTTCACTGACTTCTGTGAAGTTGTTAGATTGATGGAAAACAAGTCTCACTTGACTAAGGAAGG
TTTGGACCAAATCAAGAAGATCAGAAACAGAATGAACACTAACAGAAAG

### Protein sequence:

STLESKLNPSYISGFVDGEGSFMLTIIKDNKYKLGWRVVCRFVISLHKKDLSLLNKIKEFFDVGNVFLMTKDSAQYRVES
LKGLDLIINHFDKYPLITKKQADYKLFKMAHNLIKNKSHLTKEGLLELVAIKAVINNGLNNDLSIAFPGINTILRPDT**SGT**I
LNPFWLSGFVDAEGCFSVVVFKSKTSKLGEAVKLSFILTQSNRDEYLIKSLIEYLGCGNTSLDPRGTIDFKVTNFSSIKDII
VPFFIKYPLKGNKNLDFTDFCEVVRLMENKSHLTKEGLDQIKKIRNRMNTNRK

**I-GzeI with SGT Linker** (SGT Linker bold with underline)

**DNA sequence**

GCTAGCTCTTTGGAACAATCTTCTTTGCCACCAAAGTTGGACCCATCTTACGTTACTGGTTTCACTGACGGTGAAGG
TTCTTTCATCTTGACTATCATCAAGGACAACAAGTACAAGTTGGGTTGGAGAGTTGCATGCAGATTCGTTATCTCTT
TGCACAAGAAGGACTTGTCTTTGTTGAACTCTTTGAAGAACTTCTTCAACACTGGTTCTGTTTTCTTGATGGGTAAG
GGCGCCGCTCAATACAGAGTTGAATCTTTGACTGGTTTGTCTATCATCATCAACCACTTCGACAGATACCCATTGA
ACACTAAGAAGCAAGCTGACTACATGTTGTTCAAGTTGGCTTACAACTTGATCATCAACAAGTCTCACTTGACTGA
AAAGGGTTTGTCTGAACTAGTTTCTTTGAAGGCTGTTATGAACAACGGTTTGAAGGACGAATTGAAGATCGCTTAC
CCAAACATCACTCCAGTTTTGAGGCCTGAAATC**TCGGGTACC**AACATCGATCCATTGTGGTTGGCTGGTTTCACTG
ACGCTGAAGGTTGTTTCTCTGTTGTTGTTTTCAAGTCTAAGACTTCTAAGATCGGTGAAGCTGTTAAGTTGTCTTTC
ATCATCACTCAATCTGTTAGAGACGAATTTTTAATTAAGTCTTTGATCGAATACTTGGGTTGTGGTTACACTTCTTT
GGACGGTAGAGGTGCTATCGACTTCAAGGTTTCTGACTTCTCTTCTCTTAAGAACATCATCATCCCATTCTACGACA
AGTACTACATCCACGGTAACAAGTCTTTGGACTTCAAGGACTTCTCTCGTGTTGTTACTTTGATGGAAAACAAGAA
GCACTTGACTAAGCAAGGTTTGGACGAAATCAAGAAGATCAGAAACGCTATGAACACTAACAGA

**Protein:**

ASSLEQSSLPPKLDPSYVTGFTDGEGSFILTIIKDNKYKLGWRVACRFVISLHKKDLSLLNSLKNFFNTGSVFLMGKGAA
QYRVESLTGLSIIINHFDRYPLNTKKQADYMLFKLAYNLIINKSHLTEKGLSELVSLKAVMNNGLKDELKIAYPNITPVLR
PEI**SGT**NIDPLWLAGFTDAEGCFSVVVFKSKTSKIGEAVKLSFIITQSVRDEFLIKSLIEYLGCGYTSLDGRGAIDFKVSDFS
SLKNIIIPFYDKYYIHGNKSLDFKDFSRVVTLMENKKHLTKQGLDEIKKIRNAMNTNR

**I-SscMI with SGT Linker** (SGT Linker bold with underline)

**DNA sequence:**

TTCTACTCTACTTCTAACATCTACGTTAACAAGAACATCAACCCATGGTTCTTGACTGGTTTCATCGACGGTGAAGG
TTGTTTCAGAATCTCTTTGACTAAGGTTACTAGAGCTATCGGTTGGAGAGTTCAATTGTTCTTCCAAATCAACTTGC
ACAAGAAGGACATCGCTTTGTTGGAAGACATCAGAGACTACTTCGGTGTTGGTATGATCCACAAGTCTGGTACTAA
CTTGGTTCAATACAGAATCCAAACTTTCGACGAATTGTCTATCTTGATCAACCACTTGAACGACTACCCATTGGTTT
CTCAAAAGAAGTGGGACTTCGAATTGTTCAAGCAAGCTCACGAATTGGTTAAGATGAACGAACACTTGAACAAGG
AAGGTATCTTGAAGATCGTTTCTTTGAAGGCTTCTTTGAACCTAGGTTTGTCTGAAGCTTTGAAGTTGGCTTTCCCA
AACGTTAAGAACGCTACTAAGTTGACGTCTTCT**TCGGGTACC**ATCCCAGACCCACACTGGTTCTCTGGTTTCACGT
CTGCTGAAGGTTGTTTCATGGTTGGTATCGCTAAGTCTAAGGAATCTACTACTGGTTACCAAGTTTACTTGTCTTTC
ATCGTTACTCAACACGTTAGAGACGAATTGTTGCTTAAGTGTTTGATCGACTACTTCTCTTGTGGTAGATTGGCTAG
AAAGAGAGACGTTTACGAATACCAAGTTTCTAAGTTCTCTGACGTTGAAAAGTTCATCGATTTCTTCGACAAGTAC
CCAATCTTGGGTGAAAAGTCTAAGGACTACTTGGACTTCAGAACTGTTTCTGAAATCATGAGATCTAAGGACCACT
TGACTGAAGTTGGTGTTGCTAAGGTTAGAATCATCAAGCAAGGTATGAACAGAGGTAGA

**Protein:**

FYSTSNIYVNKNINPWFLTGFIDGEGCFRISLTKVTRAIGWRVQLFFQINLHKKDIALLEDIRDYFGVGMIHKSGTNLVQY
RIQTFDELSILINHLNDYPLVSQKKWDFELFKQAHELVKMNEHLNKEGILKIVSLKASLNLGLSEALKLAFPNVKNATKL
TSS**SGT**IPDPHWFSGFTSAEGCFMVGIAKSKESTTGYQVYLSFIVTQHVRDELLLKCLIDYFSCGRLARKRDVYEYQVSK
FSDVEKFIDFFDKYPILGEKSKDYLDFRTVSEIMRSKDHLTEVGVAKVRIIKQGMNRGR

**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 alpha helix across the missing portion of the I-OnuI structure. [102] 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 (A.R. Lambert, unpublished results). Calculation of domain interface properties and energetics were performed using Rosetta. (28)

**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.

**Yeast growth, transformation, and plasmid recovery**

Saccaromyces cerevisiae strain EBY100 was transformed using the lithium-acetate protocol described by Gietz and Schiestl.[102] 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 hours, to a density of 90-150 million cells/mL. Cells were then induced in SC + 2% galactose for 2-3 hours at 30°C, followed by 12-18 hours 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 (2009).[(42)] Briefly, expression was measured by incubating 0.25-0.5 x $10^6$ induced yeast cells per sample in 100 μL yeast staining buffer (YSB) [10 mM HEPES, 10mM NaCl, 180 mM KCl, 5 mM $CaCl_2$, 0.1% galactose, 0.2% BSA, pH 7.5], containing biotin-conjugated anti-Myc antibody. Cells were incubated for 1-2 hours at 4°C, washed with an excess of buffer, and then counter-stained with streptavidin-allophycocyanin (APC) for 1 hour at 4°C. Binding activity of surface-expressed LHEs was determined by incubating 0.5-50 nM fluorophore-labeled ds-oligo with approximately 2-5x$10^5$ cells/sample in 100 μL yeast staining buffer (yielding an estimated

100 pM enzyme concentration, assuming $10^4$-$10^5$ molecules per yeast surface), supplemented with 5mM calcium. Yeast were incubated for 2 hours 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: 2.5-5 x 10^5 cells were stained with biotinylated anti-HA antibody (Covenance) 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 $MgCl_2$ (for catalytic activity) or $CaCl_2$ (for binding without cleavage). These samples were incubated at 37°C for 15 min – 1 hour, 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, 2009.[115] Briefly, 5-10 million induced yeast were incubated in 50 µL YSB, as described above (approximating 15-30 nM enzyme), supplemented with 5 mM $MgCl_2$ or $CaCl_2$, 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, 2011.[115] Target sites for each enzyme were inserted into the TLR construct using standard molecular biology techniques. Lentivirus was produced as described previously.(43) Briefly, HEK293T cells were

transiently co-transfectled with 6ug CVL-backbone TLR plasmids, 1.5ug pMD2G envelope plasmid (VSV-G) and 3ug psPAX2 for viral packaging. Cells were incubated in 10mL DMEM without Phenol Red supplemented with 3-4% FBS and glutamine. 48 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 $0.2x10^6$ HEK293T cells with 0.5, 1, and 2 μL of their respective unconcentrated reporter lentivirus. Three days after transduction, cells with integrated reporters were selected by treatment with 1 μg/mL puromycin for 5 days. The cultures with the lowest number of surviving cells (those initially receiving 0.5 μL 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, $0.1x10^6$ HEK293T cells were seeded in a 24-well plate 24 hours prior to transfection. Cells were transiently transfected with 0.5 μg of HE-expression construct, with the addition of 0.5ug 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 hours after transfection and analyzed on a BD LSRII™ for BFP, mCherry, and GFP fluorescence. $0.1x10^6$ 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 1M 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.[(2)]

**Circular dichroism melting curves**

Circular dichroism (CD) thermal denaturation experiments were preformed at 10 μM protein concentration in 150 mM NaCl, 50 mM phosphate buffer. Measurements were made using a JASCO J-815 CD spectrometer with a Peltier thermostat. Circular dichroism 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 sec. Denaturation was performed over a 25 to 96 °C temperature range. The melting temperature was determined using JASCO software. Percent folded protein was determined using the formula $(X_{obs} - X_u)/(X_n - X_u)*100\%$, 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.

## Funding

This work was supported by National Institutes of Health [RO1CA133832, RL1 GM133833, 5RL1GM84433-04 and U19AI096111].

# References


1. Friedmann T, Roblin R. Gene therapy for human genetic disease? *Science*. 1972;175(4025):949–955.

2. Rogers S. Gene therapy: a potentially invaluable aid to medicine and mankind. *Res. Commun. Chem. Pathol. Pharmacol.* 1971;2(4):587–600.

3. Fox MS, Littlefield JW. Reservations concerning gene therapy. *Science*. 1971;173(3993):195.

4. Zanjani ED, Anderson WF. Prospects for in utero human gene therapy. *Science*. 1999;285(5436):2084–2088.

5. Anon. Trial watch: novel HIV gene therapy enters Phase I trial. *Nat Rev Drug Discov*. 2009;8(4):267.

6. Symonds GP, Johnstone HA, Millington ML, et al. The use of cell-delivered gene therapy for the treatment of HIV/AIDS. *Immunol. Res.* 2010;48(1-3):84–98.

7. Rossi JJ, June CH, Kohn DB. Genetic therapies against HIV. *Nat Biotechnol*. 2007;25(12):1444–54.

8. Gibbs JB. Mechanism-based target identification and drug discovery in cancer research. *Science*. 2000;287(5460):1969–1973.

9. Fox JL. US authorities uphold suspension of SCID gene therapy. *Nat. Biotechnol.* 2003;21(3):217.

10. Kohn DB, Sadelain M, Glorioso JC. Occurrence of leukaemia following gene therapy of X-linked SCID. *Nat. Rev. Cancer*. 2003;3(7):477–488.

11. Williams DA, Baum C. Medicine. Gene therapy--new challenges ahead. *Science*. 2003;302(5644):400–401.

12. Hacein-Bey-Abina S, Von Kalle C, Schmidt M, et al. LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science*. 2003;302(5644):415–419.

13. Hacein-Bey-Abina S, Garrigue A, Wang GP, et al. Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. *J. Clin. Invest.* 2008;118(9):3132–3142.

14. Howe SJ, Mansour MR, Schwarzwaelder K, et al. Insertional mutagenesis combined with acquired somatic mutations causes leukemogenesis following gene therapy of SCID-X1 patients. *J. Clin. Invest.* 2008;118(9):3143–3150.

15. Kaiser J. Gene therapy. Seeking the cause of induced leukemias in X-SCID trial. *Science*. 2003;299(5606):495.

16. Kaiser J. Gene therapy. Panel urges limits on X-SCID trials. *Science*. 2005;307(5715):1544–1545.

17. Buckley RH. Gene therapy for SCID--a complication after remarkable progress. *Lancet*. 2002;360(9341):1185–1186.

18. Anon. Gene therapy|[mdash]|a loss of innocence. *Nature Medicine*. 2000;6(1):1.

19. Hollon T. Researchers and regulators reflect on first gene therapy death. *Nature Medicine*. 2000;6(1):6.

20. Friedmann T. A brief history of gene therapy. *Nat. Genet.* 1992;2(2):93–98.

21. Blaese RM, Culver KW, Miller AD, et al. T lymphocyte-directed gene therapy for ADA-SCID: initial trial results after 4 years. *Science*. 1995;270(5235):475–480.

22. Sponzilli I, Notarangelo LD. Severe combined immunodeficiency (SCID): from molecular basis to clinical management. *Acta Biomed*. 2011;82(1):5–13.

23. Lenarsky C, Parkman R. Bone marrow transplantation for the treatment of immune deficiency states. *Bone Marrow Transplant.* 1990;6(6):361–369.

24. Rappeport JM, O'Reilly RJ, Kapoor N, Parkman R. Hematopoietic stem cell transplantation for severe combined immune deficiency or what the children have taught us. *Immunol Allergy Clin North Am*. 2010;30(1):17–30.

25. Cavazzana-Calvo M, Hacein-Bey S, de Saint Basile G, et al. Gene therapy of human severe combined immunodeficiency (SCID)-X1 disease. *Science*. 2000;288(5466):669–672.

26. Gaspar HB, Parsley KL, Howe S, et al. Gene therapy of X-linked severe combined immunodeficiency by use of a pseudotyped gammaretroviral vector. *Lancet*. 2004;364(9452):2181–2187.

27. Hacein-Bey-Abina S, Von Kalle C, Schmidt M, et al. LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science*. 2003;302(5644):415–419.

28. Stein S, Ott MG, Schultze-Strasser S, et al. Genomic instability and myelodysplasia with monosomy 7 consequent to EVI1 activation after gene therapy for chronic granulomatous disease. *Nat. Med.* 2010;16(2):198–204.

29. Sheridan C. Gene therapy finds its niche. *Nature Biotechnology*. 2011;29(2):121–128.

30. Bank A, Dorazio R, Leboulch P. A phase I/II clinical trial of beta-globin gene therapy for beta-thalassemia. *Ann. N. Y. Acad. Sci.* 2005;1054:308–316.

31. Cavazzana-Calvo M, Payen E, Negre O, et al. Transfusion independence and HMGA2 activation after gene therapy of human β-thalassaemia. *Nature*. 2010;467(7313):318–322.

32. Cartier N, Hacein-Bey-Abina S, Bartholomae CC, et al. Hematopoietic stem cell gene therapy with a lentiviral vector in X-linked adrenoleukodystrophy. *Science*. 2009;326(5954):818–823.

33. Humeau LM, Binder GK, Lu X, et al. Efficient lentiviral vector-mediated control of HIV-1 replication in CD4 lymphocytes from diverse HIV+ infected patients grouped according to CD4 count and viral load. *Mol. Ther.* 2004;9(6):902–913.

34. Modlich U, Navarro S, Zychlinski D, et al. Insertional transformation of hematopoietic cells by self-inactivating lentiviral and gammaretroviral vectors. *Mol. Ther.* 2009;17(11):1919–1928.

35. Titeux M, Pendaries V, Zanta-Boussif MA, et al. SIN retroviral vectors expressing COL7A1 under human promoters for ex vivo gene therapy of recessive dystrophic epidermolysis bullosa. *Mol. Ther.* 2010;18(8):1509–1518.

36. Zhou S, Mody D, DeRavin SS, et al. A self-inactivating lentiviral vector for SCID-X1 gene therapy that does not activate LMO2 expression in human T cells. *Blood*. 2010;116(6):900–908.

37. Smih F, Rouet P, Romanienko PJ, Jasin M. Double-strand breaks at the target locus stimulate gene targeting in embryonic stem cells. *Nucleic Acids Res*. 1995;23(24):5012–9.

38. Certo MT, Ryu BY, Annis JE, et al. Tracking genome engineering outcome at individual DNA breakpoints. *Nat. Methods*. 2011;8(8):671–676.

39. Branzei D, Foiani M. Regulation of DNA repair throughout the cell cycle. *Nat. Rev. Mol. Cell Biol*. 2008;9(4):297–308.

40. Ciccia A, Elledge SJ. The DNA damage response: making it safe to play with knives. *Mol. Cell*. 2010;40(2):179–204.

41. Karran P. DNA double strand break repair in mammalian cells. *Curr. Opin. Genet. Dev*. 2000;10(2):144–150.

42. McCammon JM, Amacher SL. Using zinc finger nucleases for efficient and heritable gene disruption in zebrafish. *Methods Mol. Biol.* 2010;649:281–298.

43. Doyon Y, McCammon JM, Miller JC, et al. Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases. *Nat. Biotechnol.* 2008;26(6):702–708.

44. Bibikova M, Beumer K, Trautman JK, Carroll D. Enhancing gene targeting with designed zinc finger nucleases. *Science*. 2003;300(5620):764.

45. Bibikova M, Carroll D, Segal DJ, et al. Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. *Mol Cell Biol*. 2001;21(1):289–97.

46. Bronson SK, Plaehn EG, Kluckman KD, et al. Single-copy transgenic mice with chosen-site integration. *Proc. Natl. Acad. Sci. U.S.A.* 1996;93(17):9067–9072.

47. Cai CQ, Doyon Y, Ainley WM, et al. Targeted transgene integration in plant cells using designed zinc finger nucleases. *Plant Mol. Biol.* 2009;69(6):699–709.

48. Pessach IM, Notarangelo LD. Gene therapy for primary immunodeficiencies: looking ahead, toward gene correction. *J. Allergy Clin. Immunol.* 2011;127(6):1344–1350.

49. Cohen-Tannoudji M, Robine S, Choulika A, et al. I-SceI-induced gene replacement at a natural locus in embryonic stem cells. *Mol Cell Biol*. 1998;18(3):1444–8.

50. Gouble A, Smith J, Bruneau S, et al. Efficient in toto targeted recombination in mouse liver by meganuclease-induced double-strand break. *J Gene Med*. 2006;8(5):616–22.

51. Heyer W-D, Ehmsen KT, Liu J. Regulation of homologous recombination in eukaryotes. *Annu. Rev. Genet*. 2010;44:113–139.

52. Kandavelou K, Ramalingam S, London V, et al. Targeted manipulation of mammalian genomes using designed zinc finger nucleases. *Biochem. Biophys. Res. Commun*. 2009;388(1):56–61.

53. Gupta A, Meng X, Zhu LJ, Lawson ND, Wolfe SA. Zinc finger protein-dependent and -independent contributions to the in vivo off-target activity of zinc finger nucleases. *Nucleic Acids Res.* 2011;39(1):381–392.

54. Cornu TI, Thibodeau-Beganny S, Guhl E, et al. DNA-binding Specificity Is a Major Determinant of the Activity and Toxicity of Zinc-finger Nucleases. *Mol Ther*. 2007.

55. Radecke S, Radecke F, Cathomen T, Schwarz K. Zinc-finger nuclease-induced gene repair with oligodeoxynucleotides: wanted and unwanted target locus modifications. *Mol. Ther*. 2010;18(4):743–753.

56. Porteus MH, Baltimore D. Chimeric nucleases stimulate gene targeting in human cells. *Science*. 2003;300(5620):763.

57. Porteus MH, Carroll D. Gene targeting using zinc finger nucleases. *Nat. Biotechnol.* 2005;23(8):967–973.

58. Ramalingam S, Kandavelou K, Rajenderan R, Chandrasegaran S. Creating designed zinc-finger nucleases with minimal cytotoxicity. *J. Mol. Biol.* 2011;405(3):630–641.

59. Miller JC, Tan S, Qiao G, et al. A TALE nuclease architecture for efficient genome editing. *Nat. Biotechnol.* 2011;29(2):143–148.

60. Bobis-Wozowicz S, Osiak A, Rahman SH, Cathomen T. Targeted genome editing in pluripotent stem cells using zinc-finger nucleases. *Methods*. 2011;53(4):339–346.

61. Mahfouz MM, Li L, Shamimuzzaman M, et al. De novo-engineered transcription activator-like effector (TALE) hybrid nuclease with novel DNA binding specificity creates double-strand breaks. *Proc. Natl. Acad. Sci. U.S.A.* 2011;108(6):2623–2628.

62. Stoddard BL. Homing endonuclease structure and function. *Q Rev Biophys*. 2005;38(1):49–95.

63. Paques F, Duchateau P. Meganucleases and DNA double-strand break-induced recombination: perspectives for gene therapy. *Curr Gene Ther*. 2007;7(1):49–66.

64. Arnould S, Chames P, Perez C, et al. Engineering of large numbers of highly specific homing endonucleases that induce recombination on novel DNA targets. *J Mol Biol*. 2006;355(3):443–58.

65. Heath PJ, Stephens KM, Monnat RJ, Stoddard BL. The structure of I-CreI, a group I intron-encoded homing endonuclease. *Nat Struct Biol*. 1997;4(6):468–76.

66. Takeuchi R, Lambert AR, Mak AN-S, et al. Tapping natural reservoirs of homing endonucleases for targeted gene modification. *Proc. Natl. Acad. Sci. U.S.A.* 2011;108(32):13077–13082.

67. Duan X, Gimble FS, Quiocho FA. Crystal structure of PI-SceI, a homing endonuclease with protein splicing activity. *Cell*. 1997;89(4):555–64.

68. Gimble FS. Invasion of a multitude of genetic niches by mobile endonuclease genes. *FEMS Microbiol Lett*. 2000;185(2):99–107.

69. Jurica MS, Stoddard BL. Homing endonucleases: structure, function and evolution. *Cell Mol Life Sci*. 1999;55(10):1304–26.

70. Thermes V, Grabher C, Ristoratore F, et al. I-SceI meganuclease mediates highly efficient transgenesis in fish. *Mech Dev*. 2002;118(1-2):91–8.

71. Arnould S, Perez C, Cabaniols JP, et al. Engineered I-CreI Derivatives Cleaving Sequences from the Human XPC Gene can Induce Highly Efficient Gene Correction in Mammalian Cells. *J Mol Biol*. 2007;371(1):49–65.

72. Gao H, Smith J, Yang M, et al. Heritable targeted mutagenesis in maize using a designed endonuclease. *Plant J*. 2010;61(1):176–187.

73. Windbichler N, Papathanos PA, Catteruccia F, et al. Homing endonuclease mediated gene targeting in Anopheles gambiae cells and embryos. *Nucleic Acids Res*. 2007;35(17):5922–5933.

74. Urnov FD, Rebar EJ, Holmes MC, Zhang HS, Gregory PD. Genome editing with engineered zinc finger nucleases. *Nat. Rev. Genet*. 2010;11(9):636–646.

75. Sethuraman J, Majer A, Friedrich NC, Edgell DR, Hausner G. 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.* 2009;26(10):2299–2315.

76. Ashworth J, Havranek JJ, Duarte CM, et al. Computational redesign of endonuclease DNA binding and cleavage specificity. *Nature*. 2006;441(7093):656–9.

77. Ashworth J, Taylor GK, Havranek JJ, et al. Computational reprogramming of homing endonuclease specificity at multiple adjacent base pairs. *Nucleic Acids Res*. 2010;38(16):5601 – 5608.

78. Volna P, Jarjour J, Baxter S, et al. Flow cytometric analysis of DNA binding and cleavage by cell surface-displayed homing endonucleases. *Nucleic Acids Res*. 2007;35(8):2748–58.

79. Doyon JB, Pattanayak V, Meyer CB, Liu DR. Directed evolution and substrate specificity profile of homing endonuclease I-SceI. *J Am Chem Soc*. 2006;128(7):2477–84.

80. Gruen M, Chang K, Serbanescu I, Liu DR. An in vivo selection system for homing endonuclease activity. *Nucleic Acids Res*. 2002;30(7):e29.

81. Rosen LE, Morrison HA, Masri S, et al. Homing endonuclease I-CreI derivatives with novel DNA target specificities. *Nucleic Acids Res*. 2006;34(17):4791–4800.

82. Seligman LM, Chisholm KM, Chevalier BS, et al. Mutations altering the cleavage specificity of a homing endonuclease. *Nucleic Acids Res*. 2002;30(17):3870–9.

83. Gimble FS, Moure CM, Posey KL. Assessing the plasticity of DNA target site recognition of the PI-SceI homing endonuclease using a bacterial two-hybrid selection system. *J Mol Biol*. 2003;334(5):993–1008.

84. Chames P, Epinat JC, Guillier S, et al. In vivo selection of engineered homing endonucleases using double-strand break induced homologous recombination. *Nucleic Acids Res*. 2005;33(20):e178.

85. Sussman D, Chadsey M, Fauce S, et al. Isolation and characterization of new homing endonuclease specificities at individual target site positions. *J. Mol. Biol.* 2004;342(1):31–41.

86. Jarjour J, West-Foyle H, Certo MT, et al. High-resolution profiling of homing endonuclease binding and catalytic specificity using yeast surface display. *Nucleic Acids Res*. 2009;37(20):6871–6880.

87. Chou W-C, Liao K-W, Lo Y-C, et al. Expression of chimeric monomer and dimer proteins on the plasma membrane of mammalian cells. *Biotechnology and Bioengineering*. 1999;65(2):160–169.

88. Liao K-W, Chou W-C, Lo Y-C, Roffler SR. Design of transgenes for efficient expression of active chimeric proteins on mammalian cells. *Biotechnology and Bioengineering*. 2001;73(4):313–323.

89. Mohler WA, Blau HM. Membrane-bound neomycin phosphotransferase confers drug-resistance in mammalian cells: A marker for high-efficiency targeting of genes encoding secreted and cell-surface proteins. *Somatic Cell and Molecular Genetics*. 1994;20(3):153–162.

90. Bolduc JM, Spiegel PC, Chatterjee P, et al. Structural and Biochemical Analyses of DNA and RNA Binding by a Bifunctional Homing Endonuclease and Group I Intron Splicing Factor. *Genes Dev.* 2003;17(23):2875–2888.

91. Chevalier BS, Kortemme T, Chadsey MS, et al. Design, activity, and structure of a highly specific artificial endonuclease. *Mol Cell*. 2002;10(4):895–905.

92. Hirsch C, Gauss R, Horn SC, Neuber O, Sommer T. The ubiquitylation machinery of the endoplasmic reticulum. *Nature*. 2009;458(7237):453–460.

93. Vembar SS, Brodsky JL. One step at a time: endoplasmic reticulum-associated degradation. *Nat. Rev. Mol. Cell Biol.* 2008;9(12):944–957.

94. Shusta EV, Holler PD, Kieke MC, Kranz DM, Wittrup KD. Directed evolution of a stable scaffold for T-cell receptor engineering. *Nat. Biotechnol.* 2000;18(7):754–759.

95. Chevalier BS, Monnat RJ, Stoddard BL. The homing endonuclease I-CreI uses three metals, one of which is shared between the two active sites. *Nat Struct Biol*. 2001;8(4):312–6.

96. Chevalier B, Sussman D, Otis C, et al. Metal-dependent DNA cleavage mechanism of the I-CreI LAGLIDADG homing endonuclease. *Biochemistry*. 2004;43(44):14015–26.

97. Gai SA, Wittrup KD. Yeast surface display for protein engineering and characterization. *Curr. Opin. Struct. Biol.* 2007;17(4):467–473.

98. Eastberg JH, McConnell Smith A, Zhao L, et al. Thermodynamics of DNA target site recognition by homing endonucleases. *Nucleic Acids Res.* 2007;35(21):7209–7221.

99. Geese WJ, Kwon YK, Wen X, Waring RB. In vitro analysis of the relationship between endonuclease and maturase activities in the bi-functional group I intron-encoded protein, I-AniI. *European Journal of Biochemistry*. 2003;270(7):1543–1554.

100. Ulge UY, Baker DA, Monnat RJ Jr. Comprehensive computational design of mCreI homing endonuclease cleavage specificity for genome engineering. *Nucleic Acids Res.* 2011;39(10):4330–4339.

101. Murphy PM, Bolduc JM, Gallaher JL, Stoddard BL, Baker D. Alteration of enzyme specificity by computational loop remodeling and design. *Proc. Natl. Acad. Sci. U.S.A.* 2009;106(23):9215–9220.

102. Gietz RD, Schiestl RH. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat Protoc*. 2007;2(1):31–34.

103. Gietz RD, Schiestl RH. Frozen competent yeast cells that can be transformed with high efficiency using the LiAc/SS carrier DNA/PEG method. *Nat Protoc*. 2007;2(1):1–4.

104. Grizot S, Epinat J-C, Thomas S, et al. Generation of redesigned homing endonucleases comprising DNA-binding domains derived from two different scaffolds. *Nucleic Acids Res*. 2010;38(6):2006–2018.

105. Silva GH, Belfort M. Analysis of the LAGLIDADG interface of the monomeric homing endonuclease I-DmoI. *Nucleic Acids Res*. 2004;32(10):3156–68.

106. Epinat JC, Arnould S, Chames P, et al. A novel engineered meganuclease induces homologous recombination in yeast and mammalian cells. *Nucleic Acids Res*. 2003;31(11):2952–62.

107. Leaver-Fay A, Tyka M, Lewis SM, et al. ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Meth. Enzymol.* 2011;487:545–574.

108. Rohl CA, Strauss CEM, Misura KMS, Baker D. Protein structure prediction using Rosetta. *Meth. Enzymol.* 2004;383:66–93.

109. Das R, Baker D. Macromolecular Modeling with Rosetta. *Annual Review of Biochemistry*. 2008;77(1):363–382.

110. Li H, Pellenz S, Ulge U, Stoddard BL, Monnat RJ. Generation of single-chain LAGLIDADG homing endonucleases from native homodimeric precursor proteins. *Nucleic Acids Res*. 2009;37(5):1650–1662.

111. Silva GH, Belfort M, Wende W, Pingoud A. From monomeric to homodimeric endonucleases and back: engineering novel specificity of LAGLIDADG enzymes. *J. Mol. Biol.* 2006;361(4):744–754.

112. Fajardo-Sanchez E, Stricher F, Pâques F, Isalan M, Serrano L. Computer design of obligate heterodimer meganucleases allows efficient cutting of custom DNA sequences. *Nucleic Acids Research*. 2008;36(7):2163 –2173.

113. Takeuchi R, Lambert AR, Mak AN-S, et al. Tapping natural reservoirs of homing endonucleases for targeted gene modification. *Proc. Natl. Acad. Sci. U.S.A.* 2011;108(32):13077–13082.

114. Volná P, Jarjour J, Baxter S, et al. Flow cytometric analysis of DNA binding and cleavage by cell surface-displayed homing endonucleases. *Nucleic Acids Res*. 2007;35(8):2748–2758.

115. Jarjour J, West-Foyle H, Certo MT, et al. High-resolution profiling of homing endonuclease binding and catalytic specificity using yeast surface display. *Nucleic Acids Res*. 2009;37(20):6871–6880.

116. Sather BD, Ryu BY, Stirling BV, et al. Development of B-lineage predominant lentiviral vectors for use in genetic therapies for B cell disorders. *Mol. Ther.* 2011;19(3):515–525.

# Vita

Sarah Baxter was born in Hartford, CT in 1982. She graduated from Loomis Chaffee High School in 2001, and received her Bachelor of Arts from Yale University in 2005. Her undergraduate research was in the laboratory of Dr. Michael Cappello, studying Ancylostoma ceylanicum. After graduation she worked with Dr. Philip Rosenthal's lab at UCSF, studying Plasmodium falciparum. In 2006, Sarah was admitted to the Medical Scientist Training Program at the University of Washington. She completed her Doctor of Philosphy in Immunology in 2012 for her thesis work on protein engineering, and completed her Doctor of Medicine in 2014.