Evolution of Programmable Zinc Finger-recombinases with Activity in Human Cells

Russell M. Gordley, Justin D. Smith, Torbjörn Gräslund and Carlos F. Barbas III *

Departments of Molecular Biology, Chemistry and The Skaggs Institute for Chemical Biology, BCC 550, The Scripps Research Institute, La Jolla, CA 92037, USA

Site-specific recombinases are important tools for genomic engineering in many living systems. Applications of recombinases are, however, constrained by the DNA targeting endemic of the recombinase used. A tremendous range of recombinase applications can be envisioned if the targeting of recombinase specificity can be made readily programmable. To address this problem we sought to generate zinc finger-recombinase fusion proteins (RecZFs) capable of site-specific function in a diversity of genetic contexts. Our first RecZF, Tn3Ch15X2, recombined substrates derived from the native Tn3 resolvase recombination site. Substrate Linked Protein Evolution (SLiPE) was used to optimize the catalytic domains of the enzymes Hin, Gin, and Tn3 for resolution between non-homologous sites. One of the evolved clones, GinL7C7, catalyzed efficient, site-specific recombination in a variety of sequence contexts. When introduced into human cells by retroviral transduction, GinL7C7 excised a 1.4 kb EGFP cassette out of the genome, diminishing fluorescence in \( \sim 17\% \) of transduced cells. Following this template of rational design and directed evolution, RecZFs may eventually mediate gene therapies, facilitate the genetic manipulation of model organisms and cells, and mature into powerful new tools for molecular biology and medicine.

Keywords: substrate linked protein evolution; recombinase; resolvase; zinc finger; transposase

Introduction

Site-specific recombination systems, including Cre-lox, Flp-FRT, and \( \phi \)C31-att, enable researchers to manipulate chromosomal DNA with high-fidelity in vitro and in vivo. Once recombination sites have been introduced into endogenous loci by homologous recombination, site-specific recombinases (SSRs) may catalyze excision, inversion, or integration. This two-step methodology for genome modification has revolutionized forward and reverse genetics. When SSRs are selectively activated in vivo, the resulting conditional knockouts can reveal a gene’s function with exquisite spatial and temporal specificity. What might be the implications for this technology if recombinases could be modified at will to target any desired sequence?

In contrast to the commonly used tyrosine recombinases (Cre and Flp) and large serine integrases (\( \phi \)C31), members of the resolvase/invertase family of serine recombinases are modular in both form and function. Once dimers have bound at the sites of recombination, every subsequent step, including tetramerization, strand cleavage, exchange, and ligation, is mediated solely by the catalytic domain. This modularity makes it possible to retarget recombination by replacing the endogenous DNA binding domain, albeit with some impact on efficiency.

Serine recombinases bind to their cognate crossover sites as head-to-head dimers. In nature, these dimers and adjacent, DNA-bound cofactor proteins assemble into large multimeric synapses and topological constraints ensure selective product formation. Dec-
ades of mechanistic studies, structural characterizations, and analysis of functional chimeras have revealed elegant mechanisms of recombination. A catalytic tetramer forms between two crossover sites and mediates the coordinated cleavage of all four DNA strands by serine nucleophilic attack, covalently linking each strand to a separate monomer. The ensuing intermediate contains a large, planar, hydrophobic surface that divides the tetramer, enabling the 180° rotation required for strand exchange. The recombination reaction is completed when the four free DNA 3' hydroxyls attack the serine esters, generating new phosphodiester bonds.

Mutants of several invertase/resolvase serine recombinases have been found that do not require accessory factors or orthogonal binding sites for their function. Minimal recombination sites for these variants consist of just two inversely repeated DNA sequences that are recognized by the DNA binding domains. We reasoned that if the endogenous DNA binding domains of hyperactive serine recombinases were replaced with designed polydactyl zinc finger proteins, site-specific recombination could be targeted or “reprogrammed” to any desired sequence.

Polydactyl zinc finger proteins bind with high affinity and specificity to DNA. From Cys2-His2 zinc finger motifs, our laboratory has engineered modular zinc finger domains that bind specifically to most GNN, ANN, CNN, and some TNN DNA triplet sequences. DNA binding domains specific for 6 bp to ∼18 bp DNA sites are readily constructed by manual design or by using an automated web-based server. Chimeric proteins containing novel DNA binding domains of this type have effectively directed transcriptional activation and repression, DNA cleavage, and genetic integration. Our first RecZF, similar to Z-resolvases concomitantly assembled by Stark and co-workers, efficiently recombined hybrid recombination sites: two inverted zinc finger binding sites flanking a 20 bp central spacer region. This simple fusion, however, produced chimeras with inherent sequence bias that confined its activity to sites closely related to the sequence recognized by the parent recombinase. Here, we demonstrate the use of substrate linked protein evolution (SLiPE) to engineer a RecZF with desired sequence specificity and demonstrate that the evolved protein is readily reprogrammed with other zinc finger domains to create novel recombinases that function both in bacteria and in human cells. We anticipate that this combination of rational design and directed evolution will ultimately permit site-specific recombination in any endogenous context.

Results

Design of zinc finger-recombinases (RecZFs)

Before constructing a RecZF, we modeled the recombinase/zinc finger chimera using INSIGHTII by overlaying the crystal structures of the zinc finger protein Zif268 and the GammaDelta resolvase (Figure 1). This analysis allowed us to rationally design the linker used to fuse the zinc finger protein to the catalytic domains of Tn3 resolvase and Hin and Gin invertases. We elected to truncate each recombinase near the C-terminal end of its native flexible linker (residue 145 of Tn3, 143 of Hin, 142 of Gin) and add an additional three residues (SGS) before the start of the canonical zinc finger. The first zinc finger protein selected for fusion, a didactyl protein hereinafter referred to as X2, was designed to bind the DNA nucleotides GAGGAG. Fusion of the X2 zinc finger protein with a hyperactive Tn3 resolvase catalytic domain (with mutations G70S, D102Y, E124Q) produced RecZF Tn3Ch15X2.

Zinc finger-recombinase fusion proteins (RecZFs) bind and function at hybrid recombination sites. These sites are composed of two zinc finger domain binding sites (in inverted repeat) flanking a central spacer region (∼20 bp) (Figure 2(a)). The X2.20tn3c site, for example, is GAGGAGTGATAATTATAATATTTCGCTCTCAT, where each binding site for the X2 zinc finger protein (GAGGAG) is underlined. The intervening spacer region, 20tn3c, is the central 20 base-pairs of site I within the res recombination site of the native Tn3 transposon. Additional spacers include point mutants of 20tn3c, and 20ginc, adapted from the central 20 bp of the gix recombination site (Table 1). Three plasmid-based PCR assays were developed to detect resolution, inversion, and integration catalyzed by RecZFs (Figure 2(b)). When the gene encoding Tn3Ch15X2 was electroporated into Escherichia coli on plasmids bearing X2.20tn3c, significant levels of all three catalytic activities were observed (Figure 2(c)). By contrast, chimeras assembled from hyperactive Hin (H107Y) and Gin (H106Y) invertase catalytic domains (HinX2 and GinX2 and...
GinX2, respectively) demonstrated only weak invertase activity in the same assays with X2.20tn3c (data not shown).

The optimal distance between zinc finger binding sites (ZFBSs) was determined by evaluating Tn3Ch15X2 activity on a panel of substrates. Each RecZF site was composed of two inverted zinc finger binding sites, separated by 18, 20, or 22 bp spacers (Table 1; Figure 2(a) and (b)). This range of spacer distances, initially inferred from the computer model, was examined using resolution assays on a series of substrate plasmids bearing two recombination sites of varied sizes (Figure 2(d)). The final step of the resolution assay was a PCR reaction in which substrate and product fragments were simultaneously amplified (Figure 2(b)). Therefore, the relative intensity of each band on an agarose gel is proportional to the rate of RecZF catalyzed resolution. Comparison of such qualitative rates suggested that whereas Tn3Ch15X2 tolerated 18 and 22 bp spacer arrangements, 20 bp was optimal for the reaction. Our first RecZF can be compared directly against Z-R L6, the best zinc finger-recombinase characterized by Stark and co-workers.3 These two enzymes resolve the substrate X2.20tn3c-G-X2.20tn3c (two X2.20tn3c sites flanking a gene encoding GFPuv) at similar rates (resolution assay, data not shown) and prefer similar spacer lengths (20 bp versus 22 bp).3 The slight difference may be due to different linker lengths: five amino acid residues in our Tn3Ch15X2 resolvase and 14 amino acid residues in Z-R L6.

The spacer sequence tolerance of Tn3Ch15X2 was also assessed using comparative resolution assays (Figure 2(d)). In this case, one of the two recombination sites contained mutations within 20tn3c, the native Tn3 recombination site sequence (Table 1). Previous studies had evaluated the tolerance of serine recombinases for mutations in this region3,7,38 and had revealed that mutations 2, 3, 4, 7, 8, and 9 bp away from the center of the crossover site (positions depicted in Table 1) are well tolerated, whereas alteration at positions 1, 5, 6, and 10 dramatically inhibits the function of the native Hin and Gamma-Delta recombinases. Mutation at position 1 prevents the efficient ligation of product sites.2 Cytosine and guanosine substitutions at positions 5 or 6 interrupt a critical interaction between the minor groove and a conserved arginine in the recombinase linker (142, Tn3; 140, Hin; 139, Gin). Specificity at position 10 is provided by the endogenous helix-turn-helix DNA binding domains. Tn3Ch15X2 was highly active on substrates with point mutations at positions 4, 6 and 7 (X2.20tc4, X2.20tc6, and X2.20tc7), slightly inhibited by mutation at position 5 (X2.20tc5), and blocked by mutation at position 1 (X2.20tlc). When we assayed a radically different spacer sequence, however, it became clear that simultaneous mutations at
Evolution of RecZFs with optimized substrate restrictive sequence bias of its parent recombinase. Finger protein afforded a chimera that inherited the fusion of a hyperactive catalytic domain with a zinc finger protein. This experiment suggested that the straightforward recombinase gene and substrate on the same plasmid (Figure 3(a)). In this way, each resolution product band, containing a pool of mutants enriched in proportion to their resolvase ability. Because our substrate contained non-homologous sites, we were unable to design a primer, G20S3, to selectively anneal to resolution products at the hybrid recombination site. PCR with flanking primers,31 G20S3 conferred two advantages: dramatic enhancement of product amplification and selective enrichment for high fidelity, sequence-specific recombination.

Table 1. RecZF recombination site components

<table>
<thead>
<tr>
<th>Spacer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>20tn3c</td>
<td>TGATATTTAATATTCTGG</td>
</tr>
<tr>
<td>20tL/tR</td>
<td>TGATATTTAATATTCTGG</td>
</tr>
<tr>
<td>20tR/tR</td>
<td>TGATATTTAATATTCTGG</td>
</tr>
<tr>
<td>18tn3c</td>
<td>GATAATTATATTTG</td>
</tr>
<tr>
<td>22tn3c</td>
<td>CTTGATATTTAATATTCTGG</td>
</tr>
<tr>
<td>20tc5</td>
<td>TGATATTTAATATTCTGG</td>
</tr>
<tr>
<td>20tc6</td>
<td>TGATATTTAATATTCTGG</td>
</tr>
<tr>
<td>20tc7</td>
<td>TGATATTTAATATTCTGG</td>
</tr>
<tr>
<td>20ginc</td>
<td>TTCGATATTTAATATTCTGG</td>
</tr>
<tr>
<td>20(tL/tL)</td>
<td>CCCAGGCGTG3</td>
</tr>
<tr>
<td>20(tR/tR)</td>
<td>CCCAGGCGTG3</td>
</tr>
<tr>
<td>20ginc</td>
<td>X2</td>
</tr>
<tr>
<td>20(tc5)</td>
<td>CGGAGGCGTG3</td>
</tr>
<tr>
<td>20(tc6)</td>
<td>CGGAGGCGTG3</td>
</tr>
<tr>
<td>20(tc7)</td>
<td>CGGAGGCGTG3</td>
</tr>
<tr>
<td>20ginc</td>
<td>X2</td>
</tr>
<tr>
<td>20(tc5)</td>
<td>CGGAGGCGTG3</td>
</tr>
<tr>
<td>20(tc6)</td>
<td>CGGAGGCGTG3</td>
</tr>
<tr>
<td>20(tc7)</td>
<td>CGGAGGCGTG3</td>
</tr>
<tr>
<td>20ginc</td>
<td>X2</td>
</tr>
<tr>
<td>20(tc5)</td>
<td>CGGAGGCGTG3</td>
</tr>
<tr>
<td>20(tc6)</td>
<td>CGGAGGCGTG3</td>
</tr>
<tr>
<td>20(tc7)</td>
<td>CGGAGGCGTG3</td>
</tr>
</tbody>
</table>

ZF Binding site

<table>
<thead>
<tr>
<th>X2</th>
<th>Y2</th>
<th>Z2</th>
<th>H3</th>
<th>P3</th>
</tr>
</thead>
<tbody>
<tr>
<td>GAG</td>
<td>GAG</td>
<td>GAG</td>
<td>GAG</td>
<td>GAG</td>
</tr>
<tr>
<td>GAG</td>
<td>GAG</td>
<td>GAG</td>
<td>GAG</td>
<td>GAG</td>
</tr>
<tr>
<td>TAG</td>
<td>GAG</td>
<td>GAG</td>
<td>GAG</td>
<td>GAG</td>
</tr>
<tr>
<td>GGC</td>
<td>GTG</td>
<td>GGC</td>
<td>GTG</td>
<td>GGC</td>
</tr>
<tr>
<td>GGC</td>
<td>GTG</td>
<td>GGC</td>
<td>GTG</td>
<td>GGC</td>
</tr>
</tbody>
</table>

a Bold signifies a mutation of the 20tn3c spacer sequence.
b Product of inversion between spacers 20tn3c and 20ginc.
c Product of resolution between spacers 20ginc and 20tn3c.
d Product of inversion between spacers 20ginc and 20tn3c.

multiple positions were not well tolerated. Resolution was inefficient with a substrate that contained a spacer sequence (20ginc) derived from the native Gin site that differed from 20tn3c at 12 of 20 positions. This experiment suggested that the straightforward fusion of a hyperactive catalytic domain with a zinc finger protein afforded a chimera that inherited the restrictive sequence bias of its parent recombinase.

Evolution of RecZF with optimized substrate specificity

With the aim of generating tools for endogenous genome recombination, we sought to eliminate RecZF spacer sequence bias. RecZFs were enriched by SLiPE31 for proteins with the ability to efficiently recombine two non-homologous spacer sequences, 20tn3c and 20ginc. The SLiPE approach united recombinase gene and substrate on the same plasmid such that each resolvase, expressed in E. coli, is provided with the opportunity to modify its parent plasmid (Figure 3(a)). In this way, each resolution product encoded an active recombinase and was physically distinguishable from substrate plasmids. Subsequent PCR amplification generated a unique product band, containing a pool of mutants enriched in proportion to their resolvase ability. Because our substrate contained non-homologous sites, we were able to design a primer, G20S3, to selectively anneal to resolution products at the hybrid recombination site X2.20(tL/tR) (Table 1). Compared to selection PCR with flanking primers,31 G20S3 conferred two advantages: dramatic enhancement of product amplification and selective enrichment for high fidelity, sequence-specific recombination.

Hin, Gin, and Tn3 catalytic domains were amplified by error-prone PCR and were subsequently fused to (error-free) zinc finger protein G to generate three libraries of mutant RecZFs: HinL1X2, GinL1X2, and Tn3L1X2. After three rounds of SLiPE, RecZFs within each selected pool were recombined using the DNA shuffling method first described by Stemmer.39 Five additional rounds of selection were sufficient to enrich for functional X2.20ginc-G-X2.20tn3c resolvases (Figure 3(b)). Eight highly active clones were identified from each pool by colony PCR screens and resolution assays. Mutations present in 50% or more of these clones (Figure 3(c) and (d)) are grouped into four regions: the active site loop that includes the catalytic serine residue (112V, D13G; where all numbers correspond to Tn3 equivalent positions), the long E-helix and proximal side of the core sub-domain (K65R, G70S, M73V, 180M, T108A), the solvent exposed sub-domain surface (K53E), and the flexible linker (K151M). D13G was the most frequent mutation observed in Hin and Gin catalytic domains. This substitution may destabilize the adjacent A-helix, thereby increasing the flexibility of this region. The rate enhancement evolved in the mutant enzymes might arise from more optimal positioning of catalytic active site residues for strand cleavage. The same effect might be achieved with mutations that alter the relative positions of the E-helix (residues 103–137) and core sub-domain (1–102). Several mutations selected in Gin (M70V and T96A) and Tn3 (180M and V108A) domains occurred at residues that mediate this intramolecular interaction. Substitutions around the catalytic serine and helical interface may afford the flexibility required to form the tetrameric synapic intermediate on a suboptimal spacer. One of the most active catalytic domains, GinL7C7 (D12G, N14S, N20D, K50E, M70V, 194V, Y109H, M114V, K148M; where Y109H is a reversion to wild-type Gin and K148M is a linker mutation), was selected for further characterization.

Characterization of RecZF specificity and substrate tolerance

Rounds of selection on particular spacer sequences (20tn3c and 20ginc) might have given rise to RecZFs with a novel substrate bias, and perhaps even fostered the ability to recombine those sequences in the absence of flanking zinc finger binding sites. To explore the latter possibility, we fused two new zinc finger proteins to catalytic domain GinL7C7. These novel zinc finger proteins, H3 and P3, target 5′ GGAGGCGTG3′ and 5′ GCAGTGCCG3′, respectively. The resulting RecZFs, are GinL7C7H3 and GinL7C7P3. Substrates in which the H3 and P3 target sequences replaced 5′ GGAGGAG3′ (H3.20ginc-G-H3.20tn3c, P3.20ginc-G-P3.20tn3c) were prepared (Table 1). RecZFs were cloned into each of the substrates and assayed for resolution and inversion (Figure 4(a)). Recombination occurred only in samples in which the binding site and zinc finger protein were matched. This result suggested that RecZF
function is restricted to loci flanked by cognate zinc finger binding sites and that the zinc finger binding domains can be readily exchanged without diminishing the activity of the enzymes. To further establish RecZF zinc finger mediated specificity, we constructed substrate variants in which a single nucleotide in the binding site had been altered. In these cases, the four copies of 5′GAGGAG3′ were replaced with 5′CAGGAG3′ (Y2.20ginc-G-Y2.20tn3) and 5′TAGGAG3′ (Z2.20ginc-G-Z2.20tn3c) (Table 1).

Study of GinL7C7X2 demonstrated that it resolves these two derivatives at a significantly lower rate than its cognate substrate (X2.20ginc-G-X2.20tn3c) (Figure 4(b)).

In order to rapidly characterize the spacer sequence bias of the GinL7C7 catalytic domain, substrate libraries were prepared in which 5 bp spacer regions were randomized. Our four libraries contained mutations at positions 2–6 or 6–10 (Table 1), within sites X2.20tn3c or X2.20ginc. GinL7C7X2 was cloned into each library and assayed for inversion. After purifying inversion PCR products, we sequenced the aggregate population of functional recombination sites (Figure 5(a)). The resulting chromatograms suggested a very broad substrate tolerance, particularly within the 5 bp adjacent to the zinc finger binding site (Figure 5(b)). This outcome shows that RecZFs can be successfully targeted to sequences unrelated to a parental recombination site.

RecZF recombination in the human genome

Our ultimate goal is to design RecZFs that catalyze targeted and site-specific recombination at any desired site in any genome. In order to evaluate recombination by our RecZFs in human cells, we inserted a reporter cassette that encodes enhanced green fluorescent protein (EGFP) driven by a CMV promoter and flanked by recombination sites H3.20ginc and H3.20tn3c into 293 embryonic kidney cells using the Invitrogen Flp-In system.
Figure 4. Characterization of the dependence of recombinase sequence specific activity on zinc finger domain specificity. (a) Resolution and inversion assays of GinL7C7H3 (H3) and GinL7C7P3 (P3), on substrates H3.20ginc-G-H3.20tn3c (H3) and P3.20ginc-G-P3.20tn3c (P3). Successful resolution increases the intensity of the product band (1.1 kb) relative to the substrate band (1.9 kb). Inversion generates a unique product band (1.4 kb). (b) Resolution assays of GinL7C7X2 on substrates X2.ginc-G-X2.20tn3c (X2), Y2.20ginc-G-Y2.20tn3c (Y2) and Z2.ginc-G-Z2.20tn3c (Z2). Successful resolution increases the intensity of the product band (1.0 kb) relative to the substrate band (1.8 kb).

(Feature 6(a)). Because only one copy of the cassette is present in each cell,40 site-specific resolution will generate an EGFP knockout and recombinase activity will be directly proportional to the percentage of cells with diminished fluorescence. We cloned GinL7C7H3 and GinL7C7P3 into the pBabe-Puromycin expression vector (pBP).41 The two constructs and the empty vector were introduced into the reporter cell line by retroviral transduction, and enriched by puromycin selection. When transduced cells reached confluence, nine days post-transduction, they were subjected to fluorescence-activated cell sorting (FACS) analysis (Figure 6(b) and (c)) and genomic PCR (Figure 6(d)). Both assays were in agreement: RecZFs catalyzed genomic recombination efficiently and with zinc finger-mediated specificity. GinL7C7H3 generated the expected product PCR band (~200 bp) and lowered EGFP fluorescence in 17.0(±0.8)% of transduced cells. By contrast, GinL7C7P3 and the empty vector (pBP) generated neither product band nor significant numbers of non-fluorescent cells (1.7(±0.2)% and 2.3(±0.4)% respectively). We verified the PCR result by sequencing the ~200 bp band, confirming that it was the expected product of site-specific resolution. In an effort to more tightly link genotype and phenotype, we isolated populations of EGFP− and EGFP+ cells transduced with pBP-GinL7C7H3. Subsequent genomic PCR analysis substantiated the use of FACS as a measure of site-specific excision (Figure 6(d)).

Discussion

Technologies currently used for targeted endogenous genome modification are largely based on either homologous recombination (HR) or site-specific recombination.1 Although HR can be readily targeted to any genetic sequence, the frequency of recombination is very low. Because HR relies on endogenous DNA repair machinery, the frequency of recombination is also cell-type dependent and proportional to the degree of homology between substrates. SSRs, by contrast, catalyze recombination between unrelated substrates, in any cellular environment. Applications of site-specific recombination, however, are circumscribed by the sequence specificity of known recombinases. Although much effort has been invested in improving the efficiency of HR (by the introduction of double-strand breaks,29 triplex forming oligonucleotides,42 or adeno-associated virus43) and in altering SSR substrate preference,31,44 these limitations continue to preclude many applications of genome engineering and gene therapy.

In many ways, RecZFs combine the best elements of HR and SSRs: efficient, targeted recombination of unrelated substrates in any cell type. Our studies suggest that it is now feasible to attempt resolution, inversion, or integration at, or between, genomic loci of interest. Note that Gin and Hin inverter catalytic domains were readily evolved to catalyze resolution by applying selection pressure with SLiPE. This gain in functionality, however, may appear to be offset by a coincident loss of control. HR generates stable products in the desired location and orientation, while serine integrases like dC31 achieve the same end with unidirectional recombination. Simple RecZF reactions lack such precision because they are mediated by hyperactive catalytic domains insensitive to substrate orientation and topology. We envision a variety of strategies to control RecZF-mediated recombination. The orientation of serine recombination is guided by the 2 bp overhang, at the center of the crossover site.2,45 An overhang that is not its own reversed compliment (unlike AT, in spacers 20G and 20T) should guide RecZF reactions in the same way; sites in direct repeat would allow resolution, whereas inverse repeats would allow inversion. Although it is possible that a unidirectional RecZF system may be created in the future, more immediate degrees of control may be gained by adapting strategies previously developed for Cre and Flp, including sub-optimal half-sites46,47 and recombination mediated cassette exchange (RMCE).38 In addition to these techniques, stable integration might be achieved via RecZF targeted transposition.

The novel functionality of our RecZF would eventually allow current SSR methodologies to be employed in any genetic context. The deliberate disruption of particular genes is an obvious use for

Evolution of Programmable Zinc Finger-recombinases
RecZFs, and in this role they could facilitate reverse genetics in a variety of species in which HR is inefficient. Non-specific disruption or activation could be mediated by RecZFs in a manner comparable to transposases. RecZFs could also be used to manipulate the genomes of model organisms, thereby generating useful disease models in a manner analogous to HR49 and SSRs.50

RecZFs may ultimately be used for therapeutic "genome surgery" to correct genetic defects and deliver life-enhancing genes. The small size of RecZF genes (∼800 bp) would allow a single vector53 to express the four different chimeras required for endogenous excision or integration between non-homologous sites. Similar gene therapies are marred by health risks, which often outweigh their therapeutic benefit. Retroviral integrases can deliver genes with high efficiency, but non-specific integration can activate oncogenes. The specificity of HR makes it a good candidate for gene correction, but the associated DNA damage response may diminish the viability of treated cells.64 SSRs do not trigger a DNA damage response and would seem excellent vectors for therapeutic genes. Indeed, the site-specific integrase φC31 can target pseudo-sites in the mouse and human genomes44,55,56 enabling successful in vivo treatment of murine disease models for junctional epidermolysis bullosa,57 Duchenne muscular dystrophy,58 and hereditary tyrosinemia type I.59 Unfortunately φC31 shows significant levels of toxicity and inter-chromosomal recombination in human cells.56,60

We expect that if toxicity within the RecZFs system becomes an issue with particular proteins, it can be mitigated by the careful choice of zinc finger domains. Although the tridactyl proteins in this study bind only 9 bp, hexadactyl zinc finger proteins that bind to 18 bp can target a single site in the human genome. The specificity of these proteins has been demonstrated in vitro61; hexadactyl ZFs mediate regulation of single genes in human cells and whole plants.62,63 As we have demonstrated here, a single mismatch between the zinc finger domain and its target sequence substantially disables the activity of our RecZFs (Figure 4(b)). The ability to rapidly tune activity through DNA binding domain modification is a unique feature of these recombinases. Other issues, including intermediate disassociation,64 pseudo-site presence, and half-site activity,65 must be addressed as RecZFs are evaluated for therapeutic use. However, in light of the
remarkable functional plasticity evinced in this study, RecZFs are promising tools that should facilitate a level of genomic modification heretofore inaccessible and may empower both the study of gene function and new therapies.

Materials and Methods

Construction of RecZF substrates

Each substrate plasmid contained a recombination cassette (e.g., X2.20tn3c-G-X2.20tn3c; Figure 2(a)), composed of two RecZF recombination sites (Table 1) flanking a GFPUV gene (Clontech). Cassettes were assembled by amplifying the GFPUV gene with primers that each encoded a particular RecZF site. The PCR product was cloned (XbaI, HindIII) into pBSS, a variant of pBluescriptII SK(−) (pB; Stratagene) in which the 1.2 kb SS stuffer66 is inserted between the SacI and XbaI restriction enzyme sites.

Construction of RecZF genes

The Tn3 resolvase catalytic domain was PCR amplified from the plasmid pWL625 (ATCC) in two segments, so as to introduce hyperactivating mutations D102Y and E124Q. These fragments were combined, and subsequently fused to the zinc finger protein X2 (amplified from pRTBV2-HS2#11) by overlap PCR. The completed Tn3X2 gene was digested with SacI and XbaI and ligated into similarly digested pBSS-X2.20tn3c-G-X2.20tn3c. After screening colonies by resolution assay, a hyperactive single clone, Tn3Ch15X2, was selected for further work. In addition to the hyperactivating mutations D102Y and E124Q characterized by Arnold and co-workers,11 Tn3Ch15X2 also contained the novel mutation S70G.

The Gin invertase catalytic domain was amplified by PCR from the genome of bacteriophage Mu (ATCC) in two segments, so as to introduce the hyperactivating mutation H106Y. These fragments were combined, and subsequently fused to the zinc finger protein X2 by overlap PCR. The completed GinX2 gene, containing the hyperactivating mutation H106Y characterized by Klippel and co-workers,10 was digested with SacI and XbaI and ligated into similarly digested pBSS-X2.20tn3c-G-X2.20tn3c.

The Hin invertase catalytic domain was amplified by PCR from the genome of Salmonella enterica (ATCC) in three segments, so as to remove the SacI site and introduce the hyperactivating mutation H107Y. These fragments were combined, and subsequently fused to the zinc finger protein X2 by overlap PCR. The completed HinX2 gene, containing the hyperactivating mutation H107Y characterized by Merickel and co-workers,36 was digested with SacI and XbaI and ligated into similarly digested pBSS-X2.20tn3c-G-X2.20tn3c.

Catalytic domains selected in each round of evolution were amplified by PCR and fused to an error-free copy of zinc finger protein X2 by overlap PCR. These new pools of RecZF genes were digested with SacI and XbaI and ligated into similarly digested pBSS-X2.20tn3c for the next round of selection.

The tridactyl RecZF GinL7C7H3 and GinL7C7P3 were constructed by fusing the GinL7C7 catalytic domain to either the H3 zinc finger protein (amplified from pMal-HLTR3-HS1#4) or the P3 zinc finger protein (amplified from pMal-PBS-(s)HS2-J2). When cloning GinL7C7H3 and GinL7C7P3 for transduction and expression in human cells, the assembled RecZF genes were digested with BglII and EcoRI and ligated between between BamHI and EcoRI in pBabe-Puromycin41 to create pBP-GinL7C7H3 and pBP-GinL7C7P3.

Assay of plasmid recombination

RecZFs, ligated behind the lac promoter on substrate plasmids, were electroporated into E. coli cells. On both solid and liquid media, these cells were allowed to grow overnight at 37 °C. Colonies and purified plasmid (miniprep; QIAGEN) were substrates for PCR assays of
resolution, inversion, and integration (Figure 2(b) and (c)). In each case, product formation correlated with the appearance of a unique band as visualized on an agarose gel. The resolution assay (Figure 1(b); primers pUC18-Prim1 (TGGCCGCGCTCTTCTTGATTAC) and pUC18-Prim2 (AATGTTATAGCTCCTACCTAGT)) amplified plasmid fragments from both substrate (1814 bp) and product (1039 bp) in proportion to the relative abundance of each. Inversion (primers pUC18-Prim2 and I-GFP-Mid-Prim2 (GGAAAAGTGCCACCTAG)) and integration (primers pUC18-Prim1 and pACYC184-Prim3 (GTTACCTCGGTTCAAGACTCGGTG)) were verified by the appearance of a single band (1263 bp and 370 bp, respectively); only product plasmids contained complementary primer binding sites (PBS; Figure 2(b)). Detection of RecZF catalyzed integration required a second, non-homologous, plasmid. For this purpose, the X2.20tn3c site was introduced into pACYC184 (New England Biolabs) between XbaI and HindIII to create pA-X2.20tn3c. The resolution product pb-X2.20tn3c-Tn3Ch152 was co-transformed with pA-X2.20tn3c and the two compatible plasmids were co-maintained under carbenicillin and chloramphenicol selection. Integrative products were detected when primers annealing to each plasmid backbone (pUC18-Prim1 and pACYC184-Prim3) were able to complement each other, generating a 371 bp PCR product. Co-transformation with unmodified pACYC184, verified the absence of non-specific integration (Int(-); Figure 2(c)). All PCR assays were carried out using 30 ng of plasmid DNA and a program of 1 cycle of 5 min at 94 °C; 30 cycles of 30 s at 94 °C, 30 s at 55 °C, 90 s at 72 °C; and a final cycle of 7 min at 72 °C.

Directed evolution

Libraries of RecZF mutants were created by error-prone PCR by the method of Zaccaro and co-workers. Amplification of the hyperactive Hin, Gin, and Tn3 catalytic domains performed in the presence of the dNTP analogues, dTTP (12.5 μM) and d-o-4-dTTP (12.5 μM), generated templates with randomly placed nucleotide analogs. Subsequent overlap PCR fused each catalytic domain (containing an average of 3.2 amino acid changes) to an error-free copy of zinc finger protein X2. These RecZF libraries were subsequently digested with SacI and XbaI and ligated into similarly digested pBSS-X2.20ginc-G-X2.20tn3c for the first round of functional selection. Plasmids were electroporated into ~10^8 E. coli cells, allowed to grow overnight at 37 °C in liquid culture, and isolated by miniprep. This purified plasmid was the substrate for a selection PCR (with primers pUC18-Prim1 and pACYC184-Prim3) and one outside the 3′ recombination site (pUC18-Prim1), generated a product band inside the GFPUV gene (I-GFP-Mid-Prim1 (CAGCACGCGTGCTTGTAGTTCC)) and one outside the 3′ recombination site (pUC18-Prim1), generated a product band containing only functional members of the substrate library.

Substrate tolerance assay

Substrate libraries, X2.15t(5N)/5N, X2.11t(5N)/4t, X2. (15g/5N), and X2.(11g/5N/4 g), were generated using primers in which 5 bp regions of the 3′ spacer had been randomized such that each library was a derivative of substrates X2.20ginc-G-X2.20tn3c and X2.20tn3c-G-X2.20ginc. Cloning, performed in the manner described, afforded libraries whose average number of molecules (~10^7) far exceeded the number of possible sequences (1024). Sequence randomization was confirmed by sequencing the aggregate population with pUC18 Prim1. GinL7C7X2 was ligated into each substrate library and transformed E. coli were incubated at 37 °C overnight in selective liquid media (5 μl carbenicillin, 2 ml SOC media, 3 ml SB media). Substrate and product plasmids in the aggregate culture were isolated by miniprep. Inversion PCR with the purified plasmid, performed with one primer inside the GFPUV gene (I-GFP-Mid-Prim1 (CAGCACGCGTGCTTGTAGTTCC)) and one outside the 3′ recombinase site (pUC18-Prim1), generated a product band containing only functional members of the substrate library.

RecZF site-specific genomic recombination

The CMV promoter and EGFP gene (Clontech), flanked by recombination sites H3.20ginc and H3.20tn3c, were inserted between BglII and HindIII in pcDNAS/FRT (Invitrogen). Co-transfection of the CMV-EGFP substrate plasmid and Flp expression plasmid (pOG44, Invitrogen) allowed site-specific integration into the single FLP recombinase target (FRT) site present in the Flp-In™/293 cell line (Invitrogen). A single colony from this isogenic, hygromycin-resistant population was isolated, characterized by FACS, and used as the substrate cell line (SubC) in all subsequent experiments. Cells were maintained in DMEM containing 10% fetal bovine serum and antibiotics (Gibco/ BRL, Invitrogen).

Recombination experiments. Genomic PCR of aggregate cell populations and of sorted samples were carried out using 400 ng and 100 ng of genomic DNA, respectively, and a program of 1 cycle of 5 min at 94 °C; 35 cycles of 30 s at 94 °C, 30 s at 55.7 °C, 30 s at 72 °C; and a final cycle of 7 min at 72 °C.

Acknowledgements

We thank Adrian Huber for the initial gene cloning and Jing Guo, Acilino Vieira, Lauren Schwimmer,
and the Barbas laboratory for discussions and technical advise.

References

Evolution of Programmable Zinc Finger-recombinases


*Edited by J. Karn*

(Received 16 November 2006; received in revised form 22 December 2006; accepted 4 January 2007)  
Available online 12 January 2007