Structure-guided reprogramming of serine recombinase DNA sequence specificity

Thomas Gaj, Andrew C. Mercer, Charles A. Gersbach[2], Russell M. Gordley, and Carlos F. Barbas III[1]

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

Edited by James A. Wells, University of California, San Francisco, CA, and approved November 23, 2010 (received for review September 22, 2010)

Routine manipulation of cellular genomes is contingent upon the development of proteins and enzymes with programmable DNA sequence specificity. Here we describe the structure-guided reprogramming of the DNA sequence specificity of the invertase Gin from bacteriophage Mu and Tn3 resolvase from *Escherichia coli*. Structure-guided and comparative sequence analyses were used to predict a network of amino acid residues that mediate resolvase and invertase DNA sequence specificity. Using saturation mutagenesis and iterative rounds of positive antibiotic selection, we identified extensively redesigned and highly convergent resolvase and invertase populations in the context of engineered zinc-finger recombinase (ZFR) fusion proteins. Reprogrammed variants selectively catalyzed recombination of nonnative DNA sequences >10,000-fold more effectively than their parental enzymes. Alanine-scanning mutagenesis revealed the molecular basis of resolvase and invertase DNA sequence specificity. When used as rationally designed ZFR heterodimers, the reprogrammed enzyme variants site-specifically modified unnatural and asymmetric DNA sequences. Early studies on the directed evolution of serine recombinase DNA sequence specificity produced enzymes with relaxed substrate specificity as a result of randomly incorporated mutations. In the current study, we focused our mutagenesis exclusively on DNA determinants, leading to redesigned enzymes that remained highly specific and directed transgene integration into the human genome with >80% accuracy. These results demonstrate that unique resolvase and invertase derivatives can be developed to site-specifically modify the human genome in the context of zinc-finger recombinase fusion proteins.

gene targeting | protein engineering | site-specific recombination | zinc-finger recombinase

Site-specific recombinases are essential for a variety of diverse biological processes, including the integration and excision of viral genomes, the transposition of mobile genetic elements, and the regulation of gene expression (1). Recently, site-specific recombinases have emerged as powerful tools for advanced genome engineering (2, 3). The exquisite sequence specificities of recombination systems such as Cre/*lox*, FLP/*FRT*, and φC31/*att* allow researchers to accurately modify genetic information for a variety of applications (4–6). However, DNA sequence constraints imposed by site-specific recombinases make routine modification of cellular genomes contingent on the presence of artificially introduced recognition sequences. As a result, a number of attempts have been made to circumvent or reprogram the strict DNA sequence specificities observed in these systems (7–9). Despite these efforts, engineered site-specific recombinase variants often exhibit considerably relaxed DNA sequence specificities (8, 10–14), a detrimental byproduct that often results in adverse off-target chromosomal modification (15–17). Thus, there is significant interest in the development of generalized protein engineering strategies capable of comprehensively redesigning DNA sequence specificity for genome engineering and therapeutic applications.

In recent years, our group and others have demonstrated that engineered zinc-finger recombinase (ZFR) fusion proteins are versatile alternatives to existing site-specific recombinase technologies (13, 18–20). ZFRs are artificial DNA-modifying enzymes generated by fusing site-specific DNA-binding zinc-finger proteins (ZFPs) to highly selective serine resolvase or invertase catalytic domains (21). A ZFR target site consists of two zinc-finger protein binding-sites (ZFBS) flanking a 20-bp core sequence recognized by the serine resolvase or invertase catalytic domain. Unlike zinc-finger nucleases, which are built from the sequence-independent catalytic domain of the *Fok*I restriction endonuclease (22), ZFR sequence specificity is the cooperative product of modular site-specific DNA recognition (23–28) and sequence-dependent catalysis (20). By taking advantage of cooperative sequence specificity, ZFRs are able to mediate the excision of transgenetic elements from the human genome and to catalyze plasmid integration into a genomic locus with exceptional efficiency and specificity (19, 20).

These successes, however, have been dependent on the presence of ZFR target sites that contain 20-bp core sequences derived from the native serine resolvase or invertase recombination sites (19, 20). To address this limitation, we previously used enzyme libraries consisting of random amino acid substitutions to evolve recombinase variants to react with nonnative DNA sequences (13, 19). The resulting collection of broadly dispersed mutations, however, contributed to relaxed DNA sequence specificity. Engineered enzyme variants were subsequently unable to discriminate between nonhomologous recognition sequences (13). To our knowledge, no generalized strategy has thus far been developed that is capable of comprehensively redesigning serine recombinase DNA sequence specificity for targeted genome modification.

Here we describe the structure-guided reprogramming of the DNA sequence specificity of the invertase Gin from bacteriophage Mu and Tn3 resolvase from *Escherichia coli* in the context of engineered ZFR fusion proteins. Structure-guided and comparative sequence analyses were used in tandem to predict the network of amino acid residues that mediate serine resolvase and invertase DNA sequence specificity. Saturation mutagenesis and iterative rounds of positive antibiotic selection were used to identify comprehensively redesigned and highly convergent enzyme populations with desired sequence specificities. Reprogrammed enzymes selectively catalyzed recombination against nonnative recognition sequences with efficiencies and selectivities comparable to their parental enzymes. Alanine-scanning mutagenesis revealed the amino acid residues integral to DNA sequence specificity. Additionally, we demonstrate that redesigned resolvase and invertase variants can be used as rationally

Author contributions: T.G., A.C.M., C.A.G., R.M.G., and C.F.B. designed research; T.G., A.C.M., C.A.G., and R.M.G. performed research; T.G., A.C.M., C.A.G., R.M.G., and C.F.B. analyzed data; and T.G., A.C.M., C.A.G., R.M.G., and C.F.B. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

[1]To whom correspondence should be addressed. E-mail: carlos@scripps.edu.

[2]Present address: Department of Biomedical Engineering, Duke University, Durham, NC 27710.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1014214108/-/DCSupplemental.

designed ZFR heterodimers to catalyze recombination against unnatural and asymmetric DNA sequences. Finally, we show that ZFRs composed of the redesigned resolvase and invertase catalytic domains effectively catalyze transgene integration into the human genome with >80% accuracy.

Results and Discussion

Structure-Guided Mutagenesis of Variable Arm Region Amino Acid Residues. To identify the amino acid residues critical to resolvase/invertase DNA sequence specificity, we examined the available crystal structures of the prototypical serine recombinase γδ resolvase (29, 30). Structural analysis of the γδ resolvase dimer in complex with its respective DNA substrate revealed that catalytic domain DNA sequence specificity is mediated by the resolvase arm region (Fig. 1*A*). The interactions between the minor groove of substrate DNA and the γδ resolvase arm region residues Arg 125, Thr 126, Gly 129, Arg 130, Phe 140, Gly 141, and Arg 142 provide the structural and functional foundation for DNA association (31, 32). Comparative sequence analysis of functionally divergent resolvases and invertases and their cognate DNA binding sites suggests that variable arm region positions may also provide the basis for catalytic domain sequence specificity (Fig. 1*B*). Notably, resolvase and invertase variants that recognize nonhomologous DNA sequences have amino acid sequence variation at the positions equivalent to γδ resolvase Arg 130 and Phe 140. These observations, in combination with those originally made by Yang and Steitz following the crystallographic refinement of the γδ resolvase dimer (29), led us to identify a network of divergent amino acids hypothesized to confer resolvase/invertase DNA sequence specificity. The amino acid residues postulated to mediate target site recognition are confined to the resolvase/invertase arm region and consist of the residues equivalent to γδ resolvase Asn 127, Arg 130, Met 134, Phe 140, Lys 143, and Arg 144 (Fig. 1*A*). These amino acid residues were chosen on the basis of their proximity to substrate DNA (<6 Å) and their variability among representative resolvase/invertase family members.

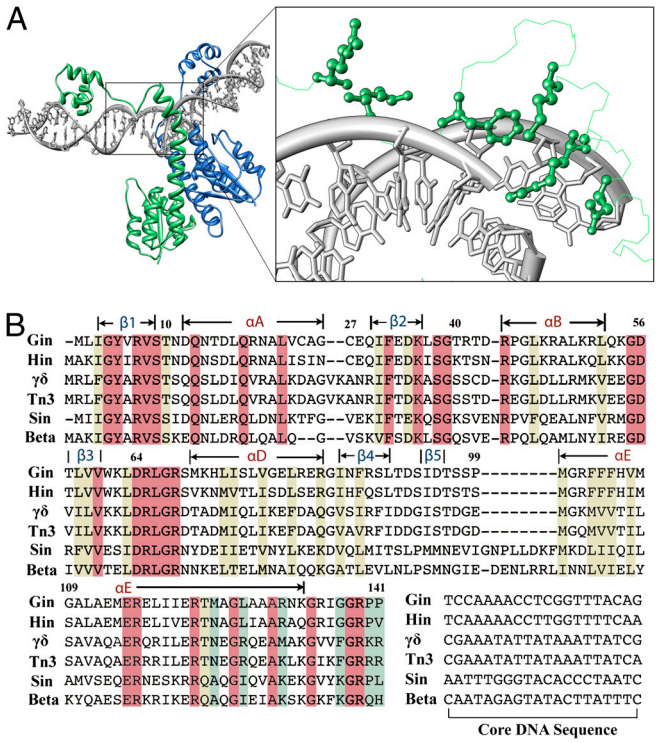

Fig. 1. *Structure-guided analysis of serine resolvase and invertase DNA sequence specificity.* (*A*) γδ resolvase dimer (green and blue) in complex with substrate DNA (gray). The variable arm region residues targeted for saturation mutagenesis, Asn 127, Arg 130, Met 134, Phe 140, Lys 143, and Arg 144, are represented as green balls and sticks (inset, with DNA) (PDB ID: 1GDT). (*B*) Multiple sequence alignment of the representative resolvase/invertase family members—Gin, Hin, γδ, Tn3, Sin, and Beta—and alignments of DNA sequences recognized by these enzymes. The secondary structures observed in the γδ resolvase crystal structure are denoted above the multiple sequence alignment. Conserved residues are highlighted red and conservative amino acid substitutions are highlighted yellow. Arm region amino acid residues targeted for saturation mutagenesis are highlighted green.

To experimentally explore our hypothesis in the context of ZFRs, we attempted to interconvert the DNA sequence specificities of the invertase Gin from bacteriophage Mu and the *E. coli*-derived Tn3 resolvase. These two functionally divergent enzymes share only 39% amino acid sequence identity and their respective DNA target sequences, *gix* and *resI*, share 30% sequence identity (Fig. 1*B*). We introduced biased diversity into the arm regions of the accessory factor-independent mutants of Gin invertase (H106Y) and Tn3 resolvase (G70S, D102Y, and E124Q) by site-directed saturation mutagenesis (33, 34). The amino acid frequencies for each randomized position are available in Table S1. Invertase and resolvase libraries were fused to an unmodified copy of the ZFP H1 to generate ZFR libraries (theoretical library size, 3×10^7 variants). These ZFR libraries were cloned into selection vectors containing the appropriate ZFR target sites (Table 1), and electroporated into *E. coli* TOP 10F′. ZFR libraries were routinely composed of $>5 \times 10^6$ transformants.

Reprogrammed Variants Selectively Catalyze Recombination Against Nonnative DNA Sequences. Active ZFR variants were selected by enzyme-mediated reconstitution of the gene encoding TEM-1 β-lactamase, a procedure we recently developed to provide quantitative analysis of ZFR activity (13) (Fig. 2*A*). To accomplish this procedure, a GFPuv cassette flanked by external ZFR target sites was inserted into the gene encoding β-lactamase. This strategic insertion results in the aberrant translation of β-lactamase and renders the *E. coli* host cell susceptible to carbenicillin, an ampicillin analog. Expression of an active ZFR from the substrate-containing selection vectors leads to recombination at the ZFR target sites. ZFR-mediated recombination results in the excision of the GFPuv transgene and restoration of the β-lactamase coding sequence. This modification establishes host-cell resistance to carbenicillin and enables the isolation of active, substrate-linked ZFR variants. A round of selection therefore consists of : (*i*) ZFR library transformation and subsequent ZFR-mediated modification of substrate plasmid, (*ii*) purification of a mixture of unmodified and modified plasmid, (*iii*) retransformation of unmodified and modified plasmid mixtures and (*iv*) selective overnight amplification of modified plasmid following inoculation with carbenicillin. ZFR variants can be isolated by restriction digestion and inserted into an unmodified substrate vector for additional rounds of selection. Importantly, ZFR-mediated recombination can be quantitatively assessed following retransformation of modified and unmodified plasmid mixtures by measuring the fraction of carbenicillin-resistant transformants (13).

Following two rounds of selection, the activity of the Gin invertase- and Tn3 resolvase-based ZFR libraries increased several orders of magnitude against the target nonnative *resI* and *gix* core sequences, respectively (Fig. S1). As the selection strin-

Table 1. ZFR target site composition

ZFR target site	ZFR target site DNA sequence
H1.20G	GGA GGC GTG TCCAAAACCATGGTTTACAG CAC GCC TCC
P2.20G	GCA GTG GCG TCCAAAACCATGGTTTACAG CGC CAC TGC
H1.20T	GGA GGC GTG CGAAATATTATAAATTATCA CAC GCC TCC
H1/P2.T/G	GGA GGC GTG CGAAATATTATGGTTTACAG CGC CAC TGC
C4.20G	GCG GGA GGC GTG TCCAAAACCTCGGTTTACAG CAC GCC TCC CGC
C4.20T	GCG GGA GGC GTG CGAAATATTATAAATTATCA CAC GCC TCC CGC

DNA sequence recognized by ZFR catalytic domain is underlined.

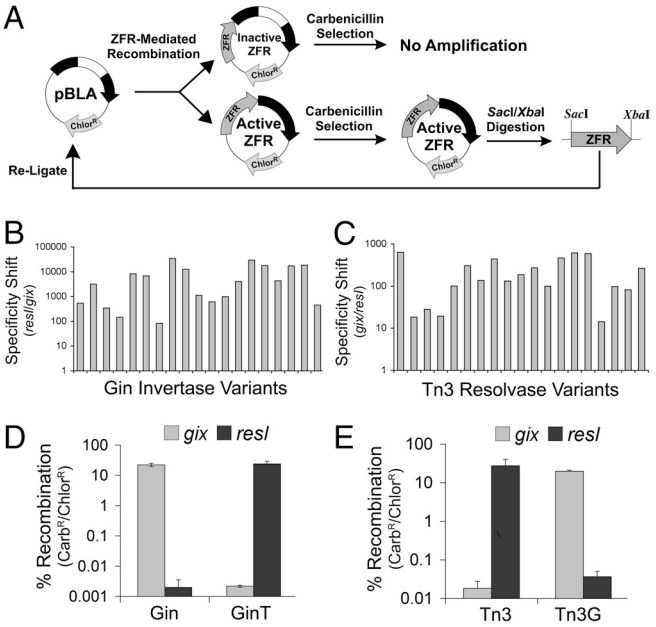

Fig. 2. *Serine resolvase and invertase DNA sequence specificity can be reprogrammed for recombination against nonnative DNA sequences.* (*A*) Schematic representation of the selection strategy used for isolating unique ZFR variants. GFPuv flanked by ZFR target sites (white) was inserted into the gene encoding TEM-1 β-lactamase (black). ZFR-mediated modification resulted in restoration of the β-lactamase gene. Active ZFR variants were enriched by carbenicillin selection and isolated following restriction digestion. (*B*–*C*) Substrate specificity shift analysis of isolated (*B*) Gin invertase and (*C*) Tn3 resolvase variants. Resolvase and invertase derivatives were isolated for analysis following four rounds of positive antibiotic selection. The recombination efficiency of an isolated enzyme variant against the nonnative target core sequence and the native core sequence was determined by measuring the fraction of carbenicillin-resistant transformants following enzyme-mediated reconstitution of the gene encoding TEM-1 β-lactamase. The specificity shift of an enzyme variant was calculated as the quotient of nonnative target core sequence recombination efficiency divided by native core sequence recombination efficiency. (*D*–*E*) Recombination efficiency of parental and reprogrammed (*D*) Gin invertase and (*E*) Tn3 resolvase variants against *gix* (gray) and *resI* (black) DNA sequences. Specificity shift and recombination efficiency are presented on a logarithmic scale. Error bars indicate the standard deviation of three independent replicates.

gency was increased, the activity profile of each ZFR population approached that of the native enzyme-substrate pair. Subsequent sequence analysis of enriched ZFR variants revealed altered resolvase/invertase arm region motifs (Table 2). Striking convergence was observed at the amino acid positions homologous to γδ resolvase Arg 130 and Phe 140. Approximately 90% of sequenced Gin variants contained the mutation L127R, whereas >90% of sequenced Tn3 variants contained either a R130L or R130I substitution. Similarly, >95% of sequenced Gin variants harbored either a G137F or G137M mutation, whereas >66% of analyzed Tn3 variants contained the substitution F140L. Minimal sequence convergence was observed at the amino acid positions corresponding to γδ resolvase Met 134, Arg 143, and Arg 144. We believe amino acid sequence convergence may reflect the functional contribution of each arm region residue to catalysis. The amino acid sequences of these variants are provided in Table S2.

Substrate specificity profiling revealed that the majority of the selected variants were reprogrammed to recognize the desired sites (Fig. 2 *B* and *C*). Critically, relaxation of DNA sequence specificity was not observed amongst individually analyzed enzymes. The two most selective Gin and Tn3 variants, GinT and Tn3G, exhibited a $>10^4$ and $\sim 10^3$ shift in substrate specificity, respectively, and catalyzed the recombination of their nonnative core sequences with efficiencies and specificities rivaling those of the parent enzymes (Fig. 2 *D* and *E*). For variants with desired specificities, an analysis of arm region composition vs. core sequence reactivity revealed strict amino acid conservation exclusively at the positions homologous to γδ resolvase Arg 130 and Phe 140 (Table S3).

Alanine-Scanning Mutagenesis Reveals the Molecular Basis of Enzyme Orthogonality. To thoroughly investigate the individual contribution of each targeted arm region amino acid residue to catalysis, alanine-scanning mutagenesis was performed on the native and reprogrammed enzymes (35) (Fig. 3). We observed $>10^5$-fold and $>10^6$-fold reduction in recombination efficiency against the *gix* core sequence following Ala mutagenesis at amino acid positions Gin Leu 127 and Tn3G Ile 130, respectively (Fig. 3*A*). Similarly, introduction of Ala at the equivalent amino acid positions Tn3 Phe 140 or GinT Met 137 led to a $>10^4$-fold reduction in recombination against the *resI* core sequence, whereas a >600-fold reduction in efficiency was observed against the *resI* sequence following mutagenesis of the highly convergent arm region residues Tn3 Arg 130 and GinT Arg 127 (Fig. 3*B*). Ala substitutions at other arm region residues had a negligible impact on recombination efficiency (Fig. 3). The arm region residues shown to have the largest influence on recombination efficiency (the equivalent Gin L127/Tn3G I130, GinT R127/Tn3 R130 and GinT Met 137/Tn3 F140) also exhibited the highest degree of sequence convergence (Table 2). These results support the idea originally formulated by Yang and Steitz that a limited, yet highly conserved network of amino acid residues mediates catalytic domain DNA sequence specificity (29).

ZFR Heterodimers Selectively Catalyze Recombination of Asymmetric DNA Sequences. The arm region amino acid residues functionally implicated in mediating substrate recognition are spatially iso-

Table 2. Amino acid mutation frequencies of reprogrammed invertase and resolvase populations

Gin Met 124	**Gin Leu 127**	**Gin Arg 130**	**Gin Gly 137**	**Gin Pro 140**	**Gin Pro 141**
Asn 22.5%	Arg 93.5%	Arg 41.9%	Phe 61.3%	Pro 70.9%	Pro 35.5%
Ala 22.5%	Lys 6.5%	Lys 19.4%	Met 35.5%	Lys 16.1%	Arg 25.8%
Ser 19.4%		Ile 19.4%	Asn 3.2%	Ile 6.5%	Lys 19.4%
Met 9.7%		Leu 6.5%		Arg 3.2%	His 9.7%
Other 12.8%		Met 6.5%		Ser 3.2%	Other 9.6%
		Other 12.8%			
Tn3 Asn 127	**Tn3 Arg 130**	**Tn3 Lys 134**	**Tn3 Phe 140**	**Tn3 Arg 143**	**Tn3 Arg 144**
Met 24.4%	Leu 62.2%	Arg 53.3%	Leu 66.7%	Pro 66.7%	Arg 48.9%
Val 15.6%	Ile 31.1%	Lys 17.8%	Trp 20.0%	Arg 24.4%	Ser 15.6%
Gln 13.3%	Met 6.6%	Gln 8.9%	Phe 4.4%	Lys 8.9%	Thr 8.9%
Asn 8.9%		Thr 6.7%	Ser 2.2%		Lys 8.9%
His 8.9%		Met 4.5%	Cys 2.2%		Pro 6.7%
Other 33.8%		Other 8.9%	Other 4.4%		Other 20.0%

>30 clones were sequenced for each enzyme population. Amino acid residues targeted for mutagenesis are indicated in bold.

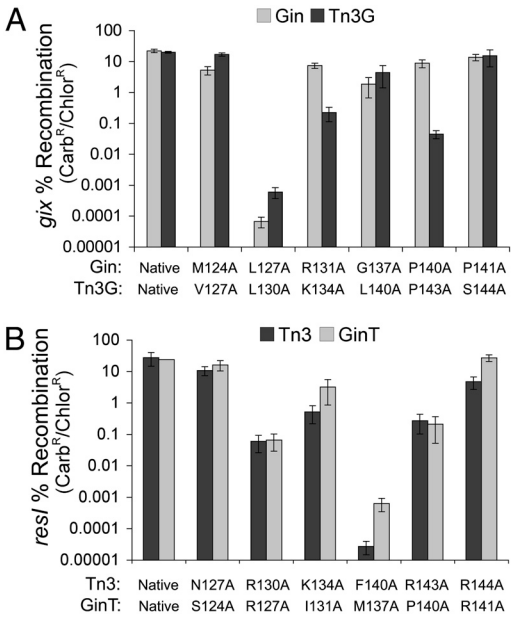

Fig. 3. *Functional analysis of the resolvase/invertase amino acid residues implicated in mediating DNA sequence specificity.* Recombination efficiency of alanine-substituted parental (Gin and Tn3) and reprogrammed (GinT and Tn3G) resolvase and invertase variants against (*A*) *gix* and (*B*) *resI* DNA sequences. Recombination efficiency is presented on a logarithmic scale. Error bars indicate the standard deviation of three independent replicates.

lated from the resolvase dimer interface (Fig. 4*A*). This observation suggests that enzyme variants with unique sequence specificities may be able to form heterotetrameric synaptosomes and catalyze recombination against asymmetric DNA sequences. To explore this possibility, we challenged pairs of ZFRs with asymmetric DNA substrates with the expectation that catalysis would be dependent on structurally compatible and sequence-specific ZFR heterodimer pairs. We began by designing an asymmetric core sequence comprised of DNA elements derived from the native *resI* and *gix* DNA sequence. The artificial core sequence (*resI*/*gix*) was then flanked with orthogonal H1 and P2 ZFBS to generate the asymmetric ZFR target site H1/P2.T/G (Table 1). Two of these target sites were then introduced into the gene encoding TEM-1 β-lactamase in positions flanking the GFPuv cassette. As previously described, ZFR-mediated modification of the inserted cassette results in host-cell resistance to carbenicillin. Highly selective ZFRs comprised of the native and reprogrammed Tn3 resolvase and Gin invertase catalytic domains were fused to the ZFPs H1 and P2 (Table S4).

We anticipated that ZFR-mediated modification of asymmetric DNA sequences would be dependent on both compatible ZFR architectures and the presence of ZFR heterodimers (ZFR_L and ZFR_R) with complementary DNA binding and catalytic domain sequence specificities (Fig. 4*B*). Therefore, recombination at the H1/P2.T/G target site was expected to occur exclusively in the presence of the structurally compatible and target site-complementary heterodimers $GinT_{H1}$ :: Gin_{P2} and $Tn3_{H1}$ :: $Tn3G_{P2}$. Indeed, the $GinT_{H1}$ :: Gin_{P2} ZFR heterodimer catalyzed recombination against asymmetric DNA sequences with efficiencies comparable to the native enzyme-substrate pair (Fig. 4*C*). A >10,000-fold increase in recombination efficiency was observed for $GinT_{H1}$ :: Gin_{P2} relative to a series of either noncognate or incompatible enzyme pairs. Similarly, the heterodimer $Tn3_{H1}$ :: $Tn3G_{P2}$ catalyzed recombination >1,000-fold more efficiently than the analyzed structurally incompatible heterodimers (Fig. 4*D*). Furthermore, ZFR monomers with only half site-complementary sequence specificity were unable to catalyze recombination against asymmetric DNA sequences (Fig. 4 *C–D*). These results support the hypothesis that rationally designed ZFR heterodimers endowed with reprogrammed DNA sequence specificity can be used to efficiently and selectively target unnatural and asymmetric DNA sequences.

Engineered ZFRs Accurately Direct Plasmid Integration into the Human Genome. A major goal of ZFR-based technologies is targeted transgene integration. To gauge the potential of the engineered ZFRs to accurately and efficiently direct plasmid integration into the human genome, we used our previously described enhanced green fluorescence protein (EGFP)-based reporter system (20) (Fig. 5*A*). Briefly, a ZFR target site (Table 1) was inserted upstream of a promoterless EGFP transgene and stably integrated into a single genomic location in human embryonic kidney (HEK) 293 cells. A donor plasmid harboring a ZFR target site adjacent to a cytomegalovirus (CMV) promoter and a constitutively expressed puromycin resistance gene was cotransfected with a ZFR expression vector. With this system, ZFR-mediated integration into the cell genome results in puromycin resistance. Site-specific integration yields two phenotypically distinct products: (*i*) a forward integration product with enhanced cellular fluorescence or (*ii*) a reverse integration product with diminished cellular

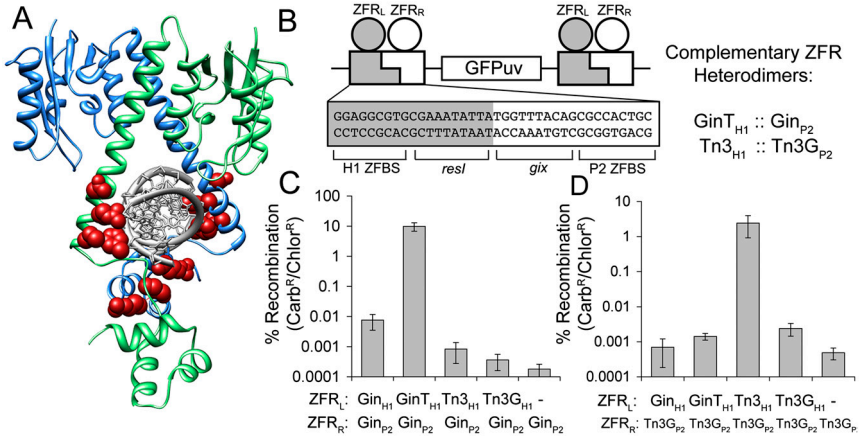

Fig. 4. *ZFR heterodimers catalyze recombination against unnatural and asymmetric DNA sequences.* (*A*) γδ resolvase dimer (green and blue) in complex with substrate DNA (gray). Arm region amino acid residues implicated in mediating substrate recognition (red spheres) are sequestered from the resolvase dimer interface (PDB ID: 1GDT). (*B*) Schematic representation of ZFR-mediated modification of asymmetric DNA sequences. The H1/P2.T/G DNA sequence is depicted. (*C–D*) Recombination efficiencies of (*C*) Gin invertase-based and (*D*) Tn3 resolvase-based ZFR heterodimers against the asymmetric ZFR target site H1/P2.T/G. "-" indicates no ZFR_L added. Recombination efficiency is presented on a logarithmic scale. Error bars indicate the standard deviation of three independent replicates.

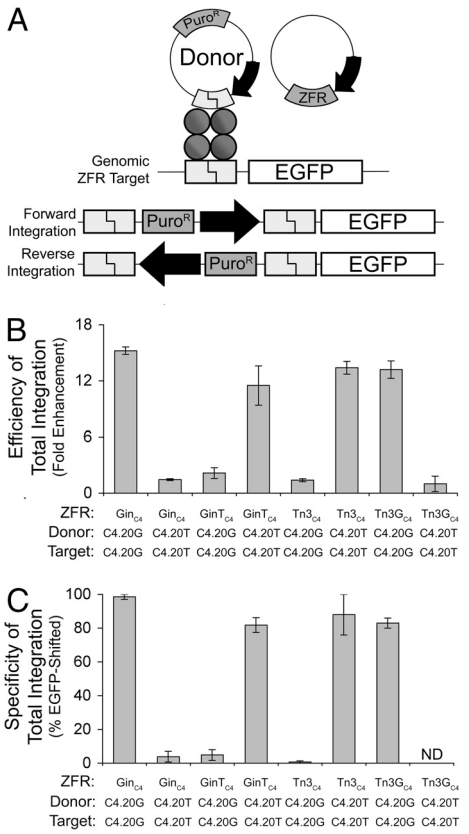

Fig. 5. *Engineered ZFRs accurately and efficiently direct plasmid integration into the human genome.* (*A*) Schematic representation of ZFR-mediated plasmid integration into the human genome. Donor plasmid containing ZFR target-site, CMV promoter (black arrow), and constitutively expressed puromycin-resistance gene were cotransfected with ZFR expression vector into human cells that contain a single genomic ZFR target site. ZFR-mediated integration resulted in the formation of puromycin-resistant cells and two phenotypically distinct integration products. (*B*) Efficiency of ZFR-mediated plasmid integration into the human genome. Following cotransfection of ZFR expression vector and donor plasmid, puromycin selection was used to assess total integration efficiency. Efficiency of ZFR-mediated integration is represented as fold-enhancement relative to donor plasmid only. (*C*) Specificity of ZFR-mediated plasmid integration into the human genome. Following puromycin selection and clonal expansion, specificity of ZFR-mediated integration was determined by flow cytometry. The efficiency and specificity of ZFR-mediated integration was analyzed with various ZFR (Gin_{C4}, $Tn3_{C4}$, $GinT_{C4}$, and $Tn3G_{C4}$), donor (C4.20G, C4.20T) and genomic target (C4.20G, C4.20T) combinations. Error bars indicate the standard deviation of three independent replicates. "ND" indicates not detectable.

fluorescence (Fig. 5*A*). Thus, the efficiency of ZFR-mediated plasmid integration can be assessed through antibiotic selection whereas specificity can be determined through a combination of flow cytometry and genomic PCR analysis.

Reprogrammed ZFR variants catalyzed plasmid integration into the human genome as efficiently as enzymes derived from native resolvase and invertase (Fig. 5*B*). Relative to transfection with donor plasmid alone, $GinT_{C4}$ and $Tn3G_{C4}$ enhanced the efficiency of stable transgene integration by 11.5 ± 2.1 and 13.2 ± 0.93-fold, respectively, in cells harboring the cognate ZFR target sites C4.20T and C4.20G (Table 1). Importantly, $GinT_{C4}$ and $Tn3G_{C4}$ did not enhance the efficiency of plasmid integration in the presence of noncomplementary ZFR target sites (2.16 ± 0.58 and 1 ± 0.816-fold, respectively), suggesting strict sequence specificity (Fig. 5*B*). Moreover, as evidenced by measurement of cellular fluorescence, ZFRs comprised of the reprogrammed domains were able to accurately mediate plasmid integration into the human genome (Fig. 5*C*). Following puromycin selection and clonal expansion, flow cytometry analysis revealed the overall specificity of transgene integration for $GinT_{C4}$ and $Tn3G_{C4}$ to be $81.8 \pm 4.41\%$ and $83 \pm 2.84\%$, respectively. The specificity of total integration for $GinT_{C4}$ and $Tn3G_{C4}$ against noncognate ZFR target sites was observed to be $4.84 \pm 3.2\%$ and 0%, respectively, further demonstrating that reprogrammed variants do not exhibit deleteriously broadened substrate specificity profiles. Critically, the specificity of total integration of each reprogrammed enzyme compared favorably to the native resolvase/invertase domains (Fig. 5*C*), indicating the experimental feasibility of successfully directing transgene integration into the human genome using enzymes with comprehensively redesigned DNA sequence specificities.

Conclusions

We demonstrate that targeted mutagenesis of critical serine resolvase/invertase arm region amino acid residues can be used to extensively reprogram DNA sequence specificity for catalysis. Redesigned enzyme variants can effectively catalyze recombination of nonnative DNA sequences and do not exhibit relaxed DNA sequence specificity. In addition, we show that, as a result of the unique structural attributes inherent to the serine resolvase/invertase family, ZFR heterodimers can be readily generated to selectively target unnatural and asymmetric DNA sequences. This strategy eliminates the requirement for symmetrical DNA target sites, a condition that would have dramatically limited the practical utility of this technology. ZFRs comprised of the reprogrammed resolvase and invertase variants can efficiently catalyze transgene integration into the human genome with >80% accuracy. The cooperative sequence specificity afforded by the combination of modular, site-specific DNA recognition and programmable, sequence-dependent catalysis should expand the targeting capacity of ZFRs and enable the previously unattainable discrimination of highly homologous regions of the human genome. We believe the results described in this study validate the hypothesis that ZFR sequence specificity is reprogrammable and that ZFRs with unique DNA sequence specificity can be developed to accurately modify the human genome.

Materials and Methods

Construction of Zinc-Finger Recombinase Libraries and Target Site Vectors. Construction of ZFR libraries and target site vectors is described in *SI Text*.

Selection of Reprogrammed Zinc-Finger Recombinase Variants. ZFR libraries were digested with *Xba*I and *Sac*I and ligated into 1 μg of *Xba*I/*Sac*I-digested selection vectors pBCBlaH1.20T and pBCBlaH1.20G. Ligation products were purified by ethanol precipitation and transformed into *E. coli* TOP 10F' (Invitrogen) by electroportation. Following 1 h recovery in 2 mL super optimal broth with catabolite repression (SOC), cells were plated on LB Agar with chloramphenicol to determine transformation efficiency. Transformants were then inoculated into 100 mL Super Broth (SB) medium and 30 μg/mL chloramphenicol. ZFR-expressing cells were cultured for 16 h at 37 °C. Library sizes were routinely found to be $>5 \times 10^6$ colony forming units per μg. Ten individual colonies from each ZFR library were sequenced to assess library quality. The following day, plasmid was harvested by miniprep (Invitrogen) and 1 μg was retransformed into *E. coli* TOP 10F′. Transformed cells harboring a mixture of modified and unmodified ZFR-linked substrate vector were inoculated into 100 mL SB selection medium containing 30 μg/mL chloramphenicol and 100 μg/mL carbenicillin. Substrate vector plasmids containing active ZFR variants were selectively enriched in liquid culture and purified by Maxi-Prep (Invitrogen). The activity of transformed ZFR populations was measured on solid media by calculating the number of ZFR-modified carbenicillin-resistant transformants relative to the number of chloramphenicol-resistant transformants ($Carb^R/Chlor^R$). The activity of individual ZFR variants was calculated in an identical manner. Enriched ZFR populations were recovered by restriction digestion with *Sac*I and *Xba*I and subsequently ligated into unmodified pBCBlaH1.20T and pBCBlaH1.20G for further rounds of selection.

ZFR-Mediated Modification of Asymmetric DNA Sequences. ZFR-mediated modification of asymmetric DNA sequences is described in *SI Text*.

ZFR-Directed Plasmid Integration Assay. To generate ZFR expression vectors, the reprogrammed Gin invertase and Tn3 resolvase catalytic domains were amplified and digested with *Bgl*II and *Bam*HI and subsequently cloned into the previously described expression vector pcDNAC4 (20). The pBabe-puromycin based donor plasmids pBSS-CMV-C.20G and pBSS-CMV-C.20T as well as the EGFP containing target reporter cell lines, 293-C.20G and 293-C.20T, were constructed as previously described (20). Target cells (293-C.20G and 293-C.20T) were seeded onto polylysine-coated 6-well plates at a density of 5×10^5 cells per well and maintained in DMEM containing 10% (vol/vol) FBS, 150 μg/mL hygromycin and 1% penicillin/streptomycin (Gibco/BRL, Invitrogen). After a 24 h incubation, reporter cells were cotransfected with 2 μg ZFR expression vector (pcDNA Gin C4, GinT C4, Tn3 C4, or Tn3G C4) and 200 ng donor plasmid (pBSS-CMV-C.20G or pBSS-CMV-C.20T) using Lipofectamine 2000 (Invitrogen) under conditions specified by the manufacturer. At 48 h posttransfection, cells were split evenly into 6-well plates containing puromycin and hygromycin and allowed to grow for 14 d before crystal violet staining. A 1:100 dilution of each sample was also plated in hygromycin as a control for total colony forming units.

The efficiency of ZFR-directed plasmid integration was determined by colony counting following staining with crystal violet solution 12 d postinoculation with puromycin (2 μg/mL). The specificity of total integration was determined as previously described (20). Briefly, 18 d posttransfection, 50,000 cells were analyzed by flow cytometry (FACScan Dual Laser Flow Cytometer, BD Biosciences) to measure the percentage of EGFP-shifted cells. Individual colonies were isolated with cloning cylinders (Corning) and seeded onto 6-well plates with media containing hygromycin (150 μg/mL) and puromycin (2 μg/mL). EGFP expression by clonal populations was determined by flow cytometry. Genomic DNA of clonal populations was isolated with the QIAmp DNA MiniKit (QAIGEN) and PCR was performed to confirm integration at the target locus.

ACKNOWLEDGMENTS. This study was supported by the National Institutes of Health (NIH) Grants GM065059, R21CA126664, F32CA125910 and The Skaggs Institute for Chemical Biology.

1. Grindley ND, Whiteson KL, Rice PA (2006) Mechanisms of site-specific recombination. *Annu Rev Biochem* 75:567–605.
2. Sorrell DA, Kolb AF (2005) Targeted modification of mammalian genomes. *Biotechnol Adv* 23:431–469.
3. Akopian A, Marshall Stark W (2005) Site-specific DNA recombinases as instruments for genomic surgery. *Adv Genet* 55:1–23.
4. Bischof J, Basler K (2008) Recombinases and their use in gene activation, gene inactivation, and transgenesis. *Methods in Molecular Biology* 420:175–195.
5. Ginsburg DS, Calos MP (2005) Site-specific integration with PhiC31 integrase for prolonged expression of therapeutic genes. *Adv Genet* 54:179–187.
6. Wilson TJ, Kola I (2001) The LoxP/CRE system and genome modification. *Methods in Molecular Biology* 158:83–94.
7. Buchholz F, Angrand PO, Stewart AF (1998) Improved properties of FLP recombinase evolved by cycling mutagenesis. *Nat Biotechnol* 16:657–662.
8. Buchholz F, Stewart AF (2001) Alteration of Cre recombinase site specificity by substrate-linked protein evolution. *Nat Biotechnol* 19:1047–1052.
9. Sarkar I, Hauber I, Hauber J, Buchholz F (2007) HIV-1 proviral DNA excision using an evolved recombinase. *Science* 316:1912–1915.
10. Voziyanov Y, Stewart AF, Jayaram M (2002) A dual reporter screening system identifies the amino acid at position 82 in Flp site-specific recombinase as a determinant for target specificity. *Nucleic Acids Res* 30:1656–1663.
11. Bolusani S, et al. (2006) Evolution of variants of yeast site-specific recombinase Flp that utilize native genomic sequences as recombination target sites. *Nucleic Acids Res* 34:5259–5269.
12. Dorgai L, Yagil E, Weisberg RA (1995) Identifying determinants of recombination specificity: construction and characterization of mutant bacteriophage integrases. *J Mol Biol* 252:178–188.
13. Gersbach CA, Gaj T, Gordley RM, Barbas CF, 3rd (2010) Directed evolution of recombinase specificity by split gene reassembly. *Nucleic Acids Res* 38:4198–4206.
14. Konieczka JH, Paek A, Jayaram M, Voziyanov Y (2004) Recombination of hybrid target sites by binary combinations of Flp variants: mutations that foster interprotomer collaboration and enlarge substrate tolerance. *J Mol Biol* 339:365–378.
15. Hacein-Bey-Abina S, et al. (2003) LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science* 302:415–419.
16. Schmidt EE, et al. (2000) Illegitimate Cre-dependent chromosome rearrangements in transgenic mouse spermatids. *Proc Natl Acad Sci USA* 97:13702–13707.
17. Loonstra A, et al. (2001) Growth inhibition and DNA damage induced by Cre recombinase in mammalian cells. *Proc Natl Acad Sci USA* 98:9209–9214.
18. Akopian A, He J, Boocock MR, Stark WM (2003) Chimeric recombinases with designed DNA sequence recognition. *Proc Natl Acad Sci USA* 100:8688–8691.
19. Gordley RM, Smith JD, Graslund T, Barbas CF, 3rd (2007) Evolution of programmable zinc finger-recombinases with activity in human cells. *J Mol Biol* 367:802–813.
20. Gordley RM, Gersbach CA, Barbas CF, 3rd (2009) Synthesis of programmable integrases. *Proc Natl Acad Sci USA* 106:5053–5058.
21. Smith MC, Thorpe HM (2002) Diversity in the serine recombinases. *Mol Microbiol* 44:299–307.
22. Carroll D (2008) Progress and prospects: zinc-finger nucleases as gene therapy agents. *Gene Ther* 15:1463–1468.
23. Segal DJ, Dreier B, Beerli RR, Barbas CF, 3rd (1999) Toward controlling gene expression at will: selection and design of zinc finger domains recognizing each of the 5′-GNN-3′ DNA target sequences. *Proc Natl Acad Sci USA* 96:2758–2763.
24. Dreier B, et al. (2005) Development of zinc finger domains for recognition of the 5′-CNN-3′ family DNA sequences and their use in the construction of artificial transcription factors. *J Biol Chem* 280:35588–35597.
25. Dreier B, et al. (2001) Development of zinc finger domains for recognition of the 5′-ANN-3′ family of DNA sequences and their use in the construction of artificial transcription factors. *J Biol Chem* 276:29466–29478.
26. Gonzalez B, et al. (2010) Modular system for the construction of zinc-finger libraries and proteins. *Nat Protoc* 5:791–810.
27. Segal DJ, et al. (2003) Evaluation of a modular strategy for the construction of novel polydactyl zinc finger DNA-binding proteins. *Biochemistry* 42:2137–2148.
28. Maeder ML, et al. (2008) Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification. *Mol Cell* 31:294–301.
29. Yang W, Steitz TA (1995) Crystal structure of the site-specific recombinase gamma delta resolvase complexed with a 34 bp cleavage site. *Cell* 82:193–207.
30. Li W, et al. (2005) Structure of a synaptic gamma delta resolvase tetramer covalently linked to two cleaved DNAs. *Science* 309:1210–1215.
31. Hughes RE, et al. (1990) Cooperativity mutants of the gamma delta resolvase identify an essential interdimer interaction. *Cell* 63:1331–1338.
32. Sluka JP, et al. (1990) Importance of minor-groove contacts for recognition of DNA by the binding domain of Hin recombinase. *Biochemistry* 29:6551–6561.
33. Klippel A, Cloppenborg K, Kahmann R (1988) Isolation and characterization of unusual gin mutants. *Embo J* 7:3983–3989.
34. Arnold PH, et al. (1999) Mutants of Tn3 resolvase which do not require accessory binding sites for recombination activity. *Embo J* 18:1407–1414.
35. Cunningham BC, Wells JA (1989) High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. *Science* 244:1081–1085.