

Expanding the zinc-finger recombinase repertoire: directed evolution and mutational analysis of serine recombinase specificity determinants

Shannon J. Sirk^{1,2,3}, Thomas Gaj^{1,2,3}, Andreas Jonsson^{1,2,3}, Andrew C. Mercer^{1,2,3} and Carlos F. Barbas III^{1,2,3,*}

¹The Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, CA 92037, USA,

²Department of Chemistry, The Scripps Research Institute, La Jolla, CA 92037, USA and ³Department of Cell and Molecular Biology, The Scripps Research Institute, La Jolla, CA 92037, USA

Received November 11, 2013; Revised December 13, 2013; Accepted December 17, 2013

ABSTRACT

The serine recombinases are a diverse family of modular enzymes that promote high-fidelity DNA rearrangements between specific target sites. Replacement of their native DNA-binding domains with custom-designed Cys₂-His₂ zinc-finger proteins results in the creation of engineered zinc-finger recombinases (ZFRs) capable of achieving targeted genetic modifications. The flexibility afforded by zinc-finger domains enables the design of hybrid recombinases that recognize a wide variety of potential target sites; however, this technology remains constrained by the strict recognition specificities imposed by the ZFR catalytic domains. In particular, the ability to fully reprogram serine recombinase catalytic specificity has been impeded by conserved base requirements within each recombinase target site and an incomplete understanding of the factors governing DNA recognition. Here we describe an approach to complement the targeting capacity of ZFRs. Using directed evolution, we isolated mutants of the β and Sin recombinases that specifically recognize target sites previously outside the scope of ZFRs. Additionally, we developed a genetic screen to determine the specific base requirements for site-specific recombination and showed that specificity profiling enables the discovery of unique genomic ZFR substrates. Finally, we conducted an extensive and family-wide mutational analysis of the serine recombinase DNA-binding arm region and uncovered

a diverse network of residues that confer target specificity. These results demonstrate that the ZFR repertoire is extensible and highlights the potential of ZFRs as a class of flexible tools for targeted genome engineering.

INTRODUCTION

In recent years, the ability to introduce highly efficient genetic modifications has become an accessible reality in the laboratory (1). Advances in genome engineering are transforming basic biological research and biotechnology by allowing researchers to induce custom alterations into virtually any cell type or organism. Site-specific endonucleases such as ZFNs (2–4), TALENs (5,6) and CRISPR/Cas (7,8) have emerged as powerful and broadly applicable tools for this process. Nevertheless, customizable nucleases are limited by numerous factors including potentially mutagenic off-target effects (9,10) and reliance on the host cell machinery to induce specific genetic modifications (11). Site-specific recombinases (SSRs), such as Cre and FLP, are an alternative class of DNA-modifying tools capable of performing site-specific integration, cassette exchange and chromosomal deletions (12). The utility of many site-specific recombination systems, however, has been hampered by the strict recognition specificities of SSRs for their natural DNA targets, a byproduct of the essential roles they have evolved to perform (13). As a result, application of these enzymes has been limited to cells or organisms that contain rare pre-existing pseudo-recognition sites (14,15) or target sites that have been pre-introduced through time-consuming and labor-intensive procedures. In order for SSRs to achieve the level of convenience and practical utility

*To whom correspondence should be addressed. Tel: +1 858 784 9098; Fax: +1 858 784 2583; Email: carlos@scripps.edu
Present addresses:

Andreas Jonsson, Royal Institute of Technology, Stockholm, Sweden.
Andrew C. Mercer, MedImmune, Gaithersburg, MD, USA.

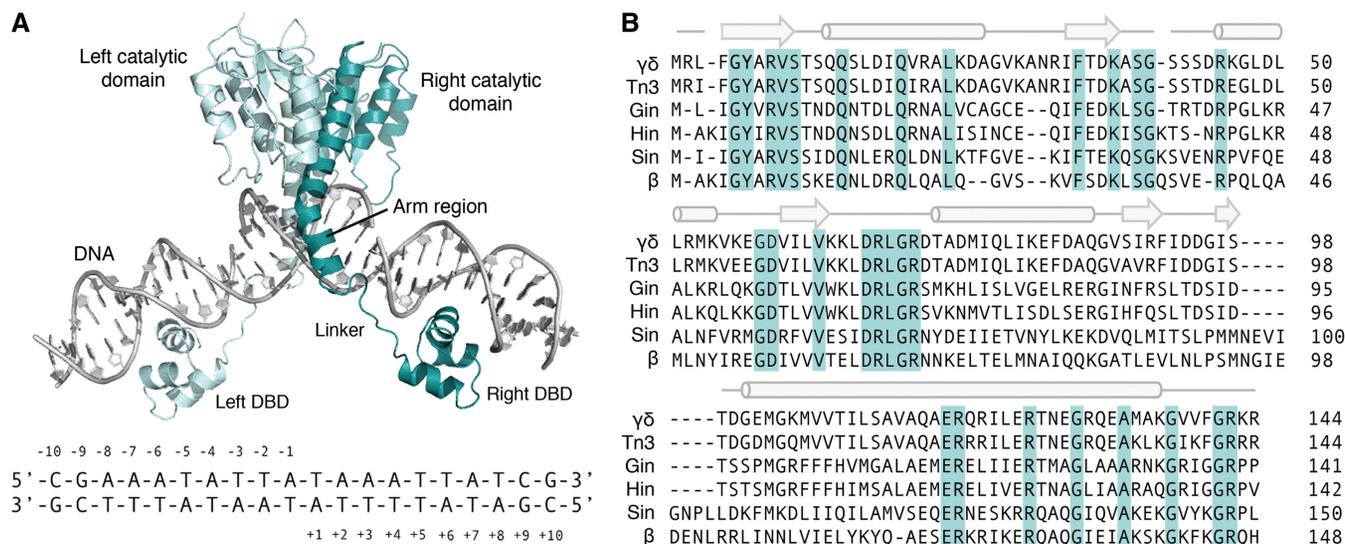


Figure 1. Overview of the small serine recombinases. (A) (Top) Crystal structure of the $\gamma\delta$ resolvase dimer bound to target DNA (PDB ID: 1GDT) (20). ‘Left’ and ‘right’ recombinase monomers are colored light and dark teal, respectively. DBD indicates native DNA-binding domain. Linker and arm region are labeled for the ‘right’ recombinase monomer only. (Bottom) Core sequence recognized by the $\gamma\delta$ resolvase catalytic domain. Base positions are indicated. (B) Sequence alignment of six of the most comprehensively characterized serine recombinase catalytic domains. Conserved residues are highlighted light teal. The α -helical and β -sheet secondary structural elements are denoted above the alignment as cylinders and arrows, respectively.

afforded by targeted nucleases, new and adaptable methods for the design of variants with flexible recombination specificities must be developed (16).

Engineered zinc-finger recombinases (ZFRs) represent a potential solution to this limitation (17,18). ZFRs are composed of custom-designed Cys₂-His₂ zinc-finger domains fused to catalytic domains derived from the resolvase/invertase family of serine recombinases (e.g. $\gamma\delta$ and Tn3 resolvases, Gin and Hin invertases) (19) (Figure 1A). ZFRs recombine hybrid target sites that consist of two inverted zinc-finger binding sites flanking a central 20-bp core sequence recognized by the recombinase catalytic domain (21,22). In nature, unique topological and spatial constraints are imposed onto these enzymes through the presence of multiple binding sites or accessory factor proteins that ensure the specificity of the recombination reaction (11,19). By using various selection strategies, ‘hyperactivated’ recombinase mutants have been identified that allow for unrestricted recombination between minimal recognition sequences (23–27). Because zinc-finger domains can be assembled to recognize a wide variety of unique sequences (28–36), fusion of these hyperactivated catalytic domains with custom zinc-finger proteins allows design of hybrid recombinases with broad targeting capabilities (37,38). Yet ZFR targeting remains constrained by sequence restrictions imposed by the recombinase catalytic domain, which requires the presence of a complementary 20-bp core sequence. To address this limitation, we recently reported the directed evolution of an extended collection of Gin recombinase catalytic domains capable of recognizing an estimated $>10^7$ unique 20-bp core sites (39). These efforts were based on mutagenesis of the C-terminal DNA-binding arm (40), a region of the recombinase that extends from the central E

helix and mediates sequence selectivity through specific interactions with the DNA minor groove (Figure 1A). However, the scope of this technology remains limited because of the presence of conserved amino acid determinants that prohibit complete reprogramming of recombinase catalytic specificity. In particular, our laboratory (39) and others (41–43) have shown that catalytic domains derived from the Gin and Hin recombinases have strict recognition specificity at base positions 6, 5 and 4; only a single A to T substitution at one of these positions is tolerated per half-site. Preliminary specificity profiling of the Tn3 resolvase has revealed similar base requirements at the equivalent half-site positions.

By using catalytic domains with distinct targeting profiles (19), new ZFRs with extended targeting capabilities could be created (Figure 1B). The β and Sin recombinases are two members of the resolvase/invertase family that recognize core sequences with increased GC content at positions 6, 5 and 4 (Figure 1A and Table 1). The Sin recombinase, originally isolated from the *Staphylococcus aureus* multiresistance plasmid p19789 (44), differs from the Tn3/ $\gamma\delta$ class of recombinases in two major ways: first, it requires the presence of a non-specific DNA-binding protein (e.g. Hbsu) (45), and, second, it coordinates recombination between two 86-bp *resH* sites that contain two binding sites (45) rather than three. The β recombinase, isolated from the *Streptococcus pyogenes* plasmid PSM19035 (46), also requires a host-encoded accessory factor protein (e.g. Hbsu, HU or eukaryotic HMG1 proteins) (47) and recognizes a 90-bp target sequence, *six*, with only two binding sites (48). Here, we report the directed evolution of new, activated β and Sin recombinases with diverse recognition capabilities that significantly expand the targeting

Table 1. The prototypical serine recombinases and their incorporation into ZFRs

Recombinase	Organism	Native function	Target site	Core sequence	Activating mutation(s)	Used as ZFR
$\gamma\delta$	<i>E. coli</i>	Resolvase	<i>res</i>	CGAA ATA TT <u>AT</u> AA ATT ATCG	D102Y, E124Q	N/A
Tn3	<i>E. coli</i>	Resolvase	<i>res</i>	CGAA ATA TT <u>AT</u> AA ATT ATCG	G70S, D102Y, E124Q	Ref. 17, 18, 21, 37, 40, 54
Gin	<i>E. coli</i>	Invertase	<i>gix</i>	CTGT AAA CC <u>GA</u> GG TTT TGGA	H106Y	Ref. 18, 37, 38, 39, 40, 49
Hin	Phage Mu	Invertase	<i>hix</i>	TCCT AAA CC <u>AT</u> GG TTT AGGA	H107Y	Ref. 18
β	<i>S. pyogenes</i>	Resolvase/invertase	<i>six</i>	CAAT AGA GT <u>AT</u> AC TTA TTTC	N95D	Present work
Sin	<i>S. aureus</i>	Resolvase	<i>resH</i>	AATT TGG GT <u>AC</u> AC CCT AATC	Q87R, Q115R	Present work

Dinucleotide cores (e.g. crossover regions) are underlined. Core sequence half-site positions 10-7, 6-4, 3-2, and the dinucleotide core are separated by spaces.

capacity of ZFRs. Additionally, we explore the specificity determinants of the resolvase/invertase family of SSRs and identify critical residues that could be altered to enable the design of recombinases with expanded targeting capabilities.

MATERIALS AND METHODS

Plasmid construction

Split gene reassembly plasmids were constructed as previously described (49). Briefly, GFPuv (Clontech, Mountain View, CA, USA) was polymerase chain reaction (PCR)-amplified with the primers 5'-GFP-ZFR-XbaI-Fwd (5'-TT AATTAAGAGTCTAGAGGAGGCGTGcaatagatagatactatttcCACGCCTCCAGATCTAGGAGGAATTTAAAA TGAG-3') and 3'-GFP-ZFR-HindIII-Rev (5'-ACTGACCTAGAGAAGCTTGGAGGCGTGgaaataagtatactattgCACGCCTCCCTGCAGTTATTTGTACAGTTCA TC-3'), where 'ZFR' corresponds to the specific 20-bp core sequences noted throughout this study (sequence recognized by β is in lowercase). PCR products were cloned into the SpeI and HindIII sites of the split gene reassembly vector. The genes for the β and Sin catalytic domains were custom-synthesized (Blue Heron, Bothell, WA, USA) and fused to the H1 zinc-finger protein (18) by overlap PCR (Supplementary Table S1). ZFR libraries based on these catalytic domains were constructed by error-prone PCR as previously described (18,50). Ala mutants for the Gin, Tn3 and β catalytic domains were generated by mutagenic overlap PCR as described (40). ZFR PCR products were cloned into the SacI and XbaI sites of the split gene reassembly vector, and library sizes were determined to be $\sim 5 \times 10^7$. DNA sequencing indicated ~ 3 amino acid substitutions per ZFR catalytic domain. All oligonucleotides were obtained from IDT (Coralville, IA, USA).

Recombination assays and selections

Recombination assays and selections were performed by split gene reassembly as described (39,40,49).

Substrate specificity profiling

GFPuv was PCR-amplified with the primers 5'-GFP-mutantZFR-XbaI-Fwd (5'-TTAATTAAGAGTCTAGA GGAGGCGTGnnnnnnnnnatactatttcCACGCCTCCAG

ATCTAGGAGGAATTTAAAATGAG-3') and 3'-GFP-wtZFR-HindIII-Rev (5'-ACTGACCTAGAGAAGCTT GGAGGCGTGgaaataagtatactattgCACGCCTCCCTG CAGTTATTTGTACAGTTTCATC-3'), where 5'-GFP-mutantZFR-XbaI-Fwd contained randomized base substitutions at the 10-7, 6-4 or 3 and 2 base positions within the 'left' 10-bp half-site of the 20B or 20S core site, and 3'-GFP-wtZFR-HindIII-Rev contained either the wild-type 20B or 20S core site (sequence recognized by β is in lowercase). PCR products were digested with XbaI and HindIII and ligated into split gene reassembly vectors that contained ZFRs with the β -N95D or Sin-Q87R/Q115R catalytic domains. Vectors were used to transform *Escherichia coli* TOP10F (Life Technologies), and cells were incubated in super broth (SB) medium with 30 μ g/ml chloramphenicol. After 6 or 16 h, cells were plated on solid lysogeny broth (LB) media with 30 μ g/ml chloramphenicol or 30 μ g/ml chloramphenicol and 100 μ g/ml carbenicillin, an ampicillin analog. Recombination frequency was calculated as the number of colonies on chloramphenicol/carbenicillin plates divided by the number of colonies on chloramphenicol-only plates. Colony numbers were determined by automated counting using the GelDoc XR Imaging System (Bio-Rad, Hercules, CA, USA). Individual chloramphenicol/carbenicillin-resistant colonies were analyzed by direct sequencing (Eton Biosciences, San Diego, CA, USA).

RESULTS

Selection of enhanced β and Sin recombinase variants

To incorporate β and Sin into the ZFR architecture, we used directed evolution to select for mutations that promoted unrestricted recombination between minimal 20-bp core sequences derived from site I of the native *six* and *resH* recombination sites (hereafter referred to as 20B and 20S, respectively) (Table 1). Similar selection strategies have previously enabled the identification of hyperactivating mutations for several serine recombinases including Gin and Hin (23), Tn3 and $\gamma\delta$ (24) and Sin (25,26). We note that the serine recombinases promote recombination between pseudo-symmetric 20-bp core sequences that consist of two inverted 10-bp half-site regions (Figure 1A). We used error-prone PCR to introduce ~ 3 amino acid mutations into each catalytic domain and fused each library to an unmodified copy of the H1

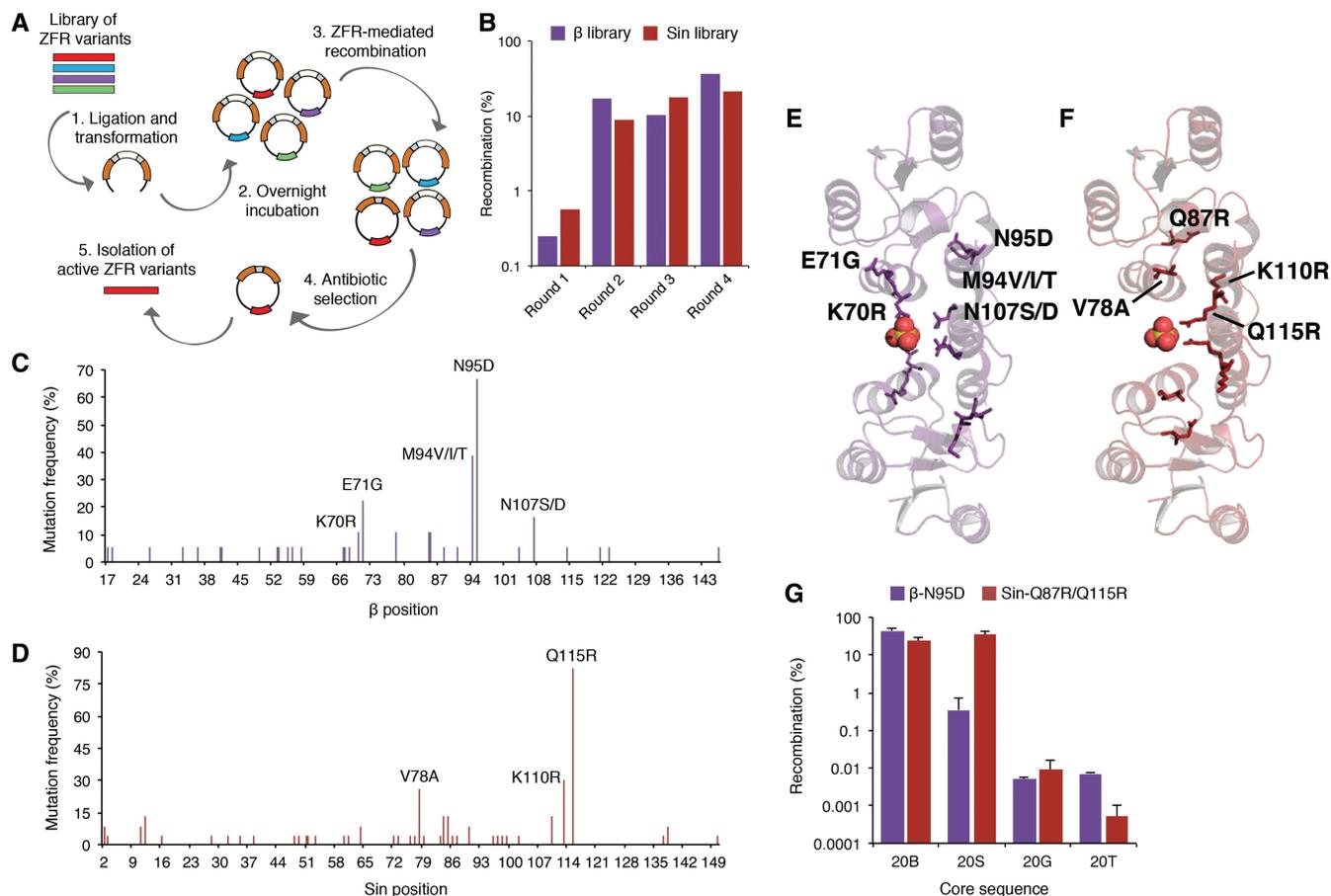


Figure 2. Directed evolution of enhanced β and Sin catalytic domains. (A) Schematic representation illustrating the split gene reassembly selection strategy. ZFR variants are shown in various colors; β -lactamase gene is in orange and GFPuv gene is in white. (B) Selection of β and Sin variants that recombine minimal core sites from the *six* and *resH* recombination sites, respectively. (C, D) Frequency and position of the mutations that activate the (C) β and (D) Sin catalytic domains. Highly recurrent mutations are indicated. (E, F) Crystal structure of the activated Sin-Q115R tetramer; view of dimer interface from above the N-terminus of the E helix (PDB ID: 3PKZ) (51). Highly recurrent (E) β and (F) Sin mutations shown as sticks and mapped onto the rotated Sin dimer, residues labeled on upper monomer only. Sulfate ion shown as spheres. (G) Recombination activity of β -N95D and Sin-Q87R/Q115R on the 20B, 20S, 20G and 20T core sequences. Recombination was determined by split gene reassembly. Error bars indicate standard deviation ($n = 3$).

zinc-finger protein (18), which recognizes the sequence 5'-GGAGGCGTG-3'. All 'wild-type' Sin mutants in this work contain the fixed substitution I100T, which was previously shown to enhance Sin-mediated recombination (26) but had negligible activating effect in our system. We selected active ZFR mutants by split gene reassembly (49), a method that links recombinase activity with cell survival in the presence of carbenicillin (Figure 2A). After only two rounds of selection, we observed a >1000-fold increase in recombination for each ZFR library (Figure 2B). After the fourth round of selection, we sequenced ~30 clones for each recombinase and observed a diverse collection of mutations for both ZFR libraries (Supplementary Table S2 and S3). We identified 32 distinct substitutions for β and 44 unique substitutions for Sin. Among sequenced β clones, ~66% contained the substitution N95D; >22% contained E71G or M94V; and >11% contained K70R, M94T or N107S (Figure 2C). For Sin, ~82% of all clones harbored Q115R; ~26% contained V78A; and >13% contained I113T, K84R, D85G or K110R (Figure 2D). Among these, Rowland *et al.*

previously identified V78A, D85G, K110R and Q115R (25). Notably, the majority of the selected mutations clustered within or near the central E helix and recombinase dimer interface (Figure 2E and F). The location of these substitutions is similar to those mutations previously shown to enhance Gin-, Hin- (23) and Tn3-mediated recombination (24), indicating that a conserved mechanism for activation might involve stabilization of the recombinase synaptic tetramer (52).

Selected β and Sin variants recombine DNA with high efficiency

To determine the extent to which the selected mutations promote recombination, we used split gene reassembly to evaluate the activity of individual ZFRs composed of various β and Sin catalytic domains on the 20B and 20S core sequences, as well as on core sites derived from the native Gin and Tn3 recombination sites (hereafter referred to as 20G and 20T, respectively) (Table 1). We found that each selected β and Sin mutant recombined its intended

Table 2. Recombination by selected β and Sin catalytic domains

Recombinase	Mutations	Core sequence				
		20B	20S	20G	20T	
β	None	--		--	--	
	M94V	+++		--	--	
	N95D	++++		--	--	
	M94T, R104H	+		--	--	
	M94V, N107S	++++		--	--	
	V58A, N95D	+		--	--	
	M94T, N95D	+++		--	--	
	E71G, M94V, N95D	++		--	--	
	N68S, E71G, V88A, N95D	+++		--	--	
	K33R, N49S, E71G, N95D	+++		--	--	
	R18P, R41P, D55G, R67G, E71G, M94I, N107K, Y114N	+++		--	--	
	Sin	None		--	--	--
		I2V		++	--	--
		Q115R		+	--	--
Q87R, Q115R			++++	--	--	
T32A, N97D, Q115R			+	--	--	
I11V, D12N, V78A, Q115R			++++	--	--	
I11V, D12N, V78A, Q115R, L150P			+	--	--	
T77I, D85G, K110R, I133V, V138I			+	--	--	
I11T, V61A, D85G, K110R, I113T, Q115R			++	--	--	
I64A, V78A, K84R, I90V, I113T, Q115R, Q137R			++++	--	--	
V53A, E76G, E83G, D85G, V99A, N102S, I113S, Q115R			++++	--	--	

Symbols indicate recombination efficiency. +++++, >35% recombination; +++, 20–35%; ++, 6–19%; +, 1–5%; –, <0.1%; --, <0.01%. The limit of detection of recombination by split gene reassembly is $\sim 10^{-5}\%$. All Sin variants are derived from the I100T background strain.

DNA target >1000-fold more efficiently than the corresponding wild-type enzymes (Table 2). One β and four Sin clones demonstrated slightly relaxed specificity on 20T, while no variants effectively recombined the 20G target (Table 2). The β and Sin mutants that showed the strictest recognition specificity for their intended DNA targets, β -N95D and Sin-Q87R/Q115R, also harbored the most prevalent mutation from each library (Figure 2G). Intriguingly, in the case of Sin, a single auxiliary substitution (Q87R) was selected for only in the presence of Q115R. Cross-comparative specificity analysis between these two clones revealed that Sin-Q87R/Q115R, but not β -N95D, recombined the 20S and 20B target sites with comparable efficiencies (Figure 2G), indicating that β -N95D exhibits more stringent recognition specificity than Sin-Q87R/Q115R.

Specificity profile of the β recombinase

To develop a more detailed understanding of the factors underlying serine recombinase substrate recognition, we evaluated the specificity profiles of the β and Sin catalytic domains. To accomplish this, we adapted our split gene reassembly selection method to identify the specific bases tolerated by each recombinase at every position within their respective 10-bp half-site regions (Figure 3A). Previous studies with the Gin recombinase revealed a pseudo-modular recognition pattern within each 10-bp half-site; recognition was segmented into four discrete regions (e.g. non-specific base recognition at positions 10, 9, 8 and 7; strict specificity at positions 6, 5 and 4; specific recognition of positions 3 and 2; and non-specific recognition at the dinucleotide core) (39). Based

on these findings, we constructed a series of mutant 20B and 20S substrate libraries that contained fully randomized base combinations within three of the four half-site sub-domains (i.e. positions 10-7, 6-4 and 3 and 2) (Figure 3B). To ensure efficient recombination, we elected not to introduce substitutions within the central dinucleotide core (i.e. the region in which crossover takes place between compatible 2-bp overhangs). Therefore, to maximize the effectiveness of our selection system, we introduced mutations only within a single 10-bp half-site region (Figure 3B). This approach facilitated straightforward retrieval by DNA sequencing of all tolerated/recombined core sites.

We evaluated the ability of β -N95D and Sin-Q87R/Q115R to recombine DNA substrate libraries at two time points: 6h and 16h. After 6 or 16h of incubation in liquid culture, we subjected cells harboring the library members to antibiotic selection on LB agar plates, followed by sequencing of individual transformants to (i) ensure that recombination had occurred and (ii) identify tolerated recombination substrates. As anticipated, we observed that ZFRs that were allowed to recombine DNA for only 6h demonstrated greater recognition stringency than those allowed to react for 16h (Figure 3C and D). Previous work by our laboratory indicated that 6-h incubation is sufficient to allow for high levels of ZFR-mediated recombination to occur (49). We sequenced 30 clones for each β -N95D substrate library and 10 clones for each Sin-Q87R/Q115R substrate library. Both enzymes, regardless of incubation time, yielded outputs that converged on the sequence motif GT at positions 3 and 2 (Figure 3E; data for β -N95D shown only). We also observed strong convergence toward G at position 4 for

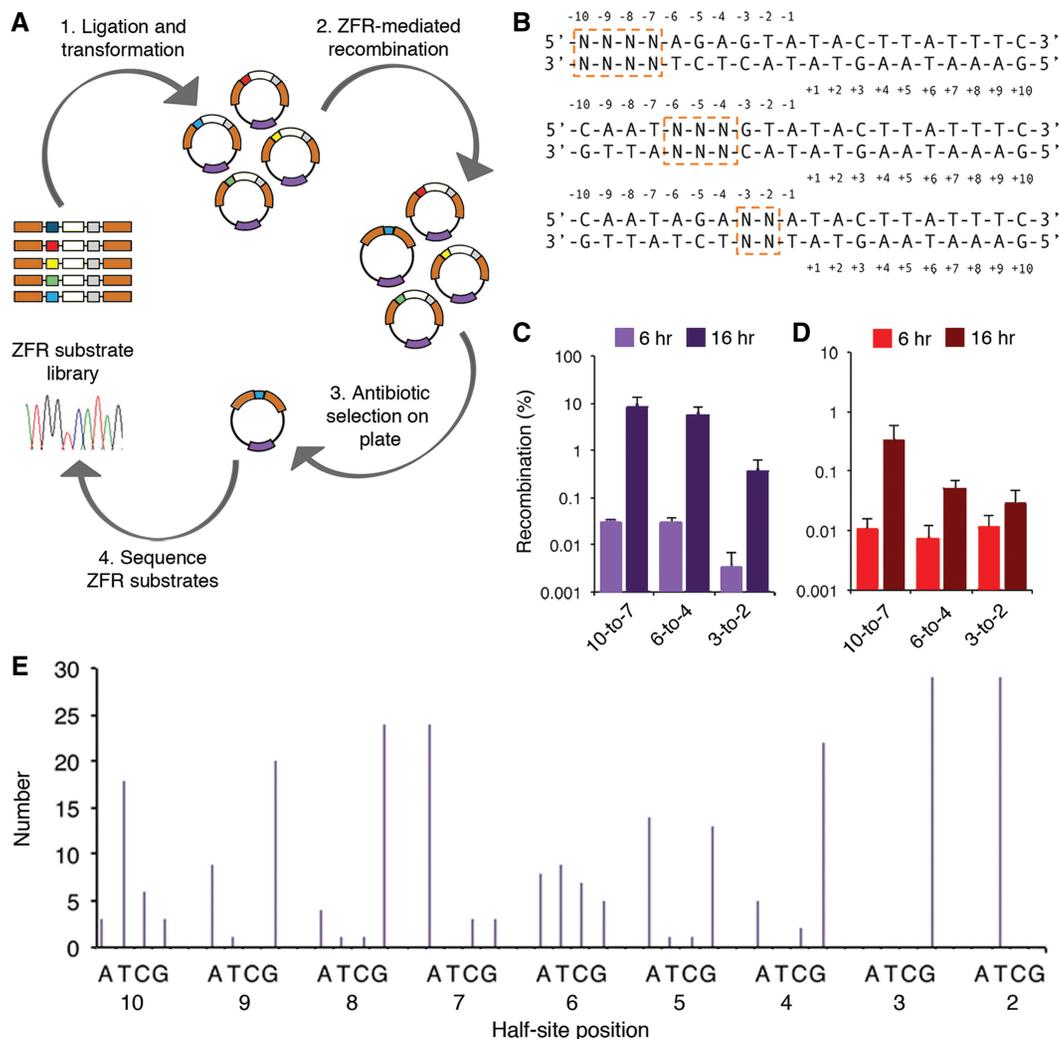


Figure 3. Specificity of the β -N95D catalytic domain. (A) Schematic representation illustrating the genetic screen used to profile recombinase specificity. Recombinase substrate library shown in various colors; ZFR gene is in purple, β -lactamase gene is in orange and GFPuv gene is in white. (B) Randomization strategy used for specificity profiling. Randomized bases are boxed. Note that only ‘left’ half-site of the upstream ZFR target site contained base substitutions. (C and D) Recombination by (C) β -N95D and (D) Sin-Q87R/Q115R for each 20B and 20S core site library, respectively, at 6 and 16 h. (E) Number of selected base sequences (out of 30) at each position within the 20B half-site. Thirty clones were sequenced from each 6-h library output. Recombination was determined by split gene reassembly. Error bars indicate standard deviation ($n = 3$).

both catalytic domains. Although no strong consensus was observed after 16 h at positions 10, 9, 8, 7, 6 or 5 for β -N95D, substrates incubated for 6 h showed sequence convergence at 7 of 9 positions (exclusion of thymine at position 7 is an artifact of the system, as its presence allows for the introduction of stop codons within the 20-bp core) (Figure 3E; data for β -N95D shown only). In particular, β -N95D recognition at position 6 was fully degenerate, and the enzyme had a strong preference for A or G at position 5. To a lesser degree, we also observed variability at positions 9 (A or G) and 10 (T >> C > A = T). Interestingly, the consensus sequence derived for β -N95D shares only 60% sequence identity with the native core sequence recognized by the wild-type β recombinase (Table 3); however, both target sites maintain a strong preference for A and G bases.

Although we constructed the Sin substrate libraries in a manner identical to those for β , we were unable to

conclusively determine its specificity profile due to repeated selection of a single, potentially artifactual consensus sequence (data not shown). Based on these findings, we focused subsequent studies on the β -N95D catalytic domain.

β -N95D-mediated recombination of core sequences from the human genome

To determine the accuracy and utility of the β -N95D specificity profile, we next investigated whether ZFRs that contained β -N95D could recombine pre-determined 20-bp core sites from the human genome. Based on our earlier findings, we derived a degenerate 20-bp core site and identified 17 recombination sites from several therapeutically relevant human genes including *breakpoint cluster region (BCR)*, *achromatopsia/cyclic-nucleotide gated ion channels 2 (CNGA3)* and *3 (CNGB3)*, *factor VIII*, *factor IX* and *retinal pigment epithelium (RPE65)*

Table 3. β -Mediated recombination of core sequences derived from the human genome

Target site	Gene	Core sequence	Recombination (%)
β wild-type		CAAT AGA GT AT AC TTA TTTC	44 \pm 9
β consensus		TRGA NRG GT AT AC CYN TCYA	ND
β degenerate		NNNV NNR GT WW AC YNN BNNN	ND
β -AT 1	CNGB3	<u>AAAC</u> AGA GT AT AC <u>CCT</u> <u>CATT</u>	92 \pm 63
β -AT 2	CNGB3	<u>GTTA</u> <u>CTA</u> GT AT AC TTA <u>TCCT</u>	37 \pm 34
β -AT 3	Factor VIII	<u>GTGC</u> <u>CAG</u> GT AT AC TGT <u>GTTA</u>	36 \pm 9
β -AT 4	CNGB3	<u>TAAC</u> <u>TCA</u> GT AT AC TFG <u>GGGG</u>	32 \pm 10
β -AT 5	Factor VIII	<u>TGGG</u> <u>GGA</u> GT AT AC <u>CTT</u> <u>TTTC</u>	27 \pm 4
β -AT 6	BCR	<u>ACTC</u> <u>TTG</u> GT AT AC TGT <u>TCTG</u>	3.2 \pm 1
β -TT 1	RPE	<u>CAGG</u> TGA GT TT AC <u>CAA</u> <u>TCTG</u>	0.07 \pm 0.008
β -TT 2	CNGA3	<u>AAAA</u> <u>GAG</u> GT TT AC TCA <u>GCTC</u>	0.03 \pm 0.02
β -TT 3	CNGB3	<u>CTAA</u> <u>TTA</u> GT TT AC <u>CAG</u> <u>TGAA</u>	0.03 \pm 0.01
β -TT 4	CNGB3	<u>GTCA</u> <u>CCA</u> GT TT AC <u>CAT</u> <u>TTCT</u>	0.02 \pm 0.01
β -TT 5	CNGB3	<u>TTAA</u> <u>ATA</u> GT TT AC TGG <u>TGCA</u>	0.01 \pm 0.003
β -AA 1	CNGB3	<u>GAGA</u> <u>GTG</u> GT AA AC TAC <u>CTGC</u>	0.01 \pm 0.004
β -AA 2	Factor IX	<u>GAGA</u> <u>CAG</u> GT AA AC TAC <u>GCAG</u>	<0.01
β -TA 1	Factor VIII	<u>TGAA</u> <u>GTG</u> GT TA AC TAT <u>GCAA</u>	<0.01
β -TA 2	CNGB3	<u>TTTC</u> <u>TAA</u> GT TA AC TTT <u>TTAC</u>	<0.01
β -TA 3	Factor VIII	<u>CCTC</u> <u>TAG</u> GT TA AC CAA <u>TTTG</u>	<0.01
β -TA 4	CNGB3	<u>GGCC</u> <u>CTG</u> GT TA AC TCT <u>CCTC</u>	<0.01

Dinucleotide core composition is denoted in target site (e.g. β -AT 1 contains an AT dinucleotide core). Positions 10-7, 6-4, 3-2, and the dinucleotide core are separated by spaces. Base mismatches between genomic and wild-type core sequences are underlined. Recombination was measured by split gene reassembly. ND indicates not determined. Error values indicate standard deviation ($n = 3$). Abbreviations for nucleotide substitutions are as follows: N = A, T, C, or G; V = A, C, or G; B = T, C, or G; R = A or G; Y = T or C; W = A or T.

genes (GRCh37.p10 primary reference assembly) (Table 3). The dinucleotide core sequence within our degenerate recombination site was restricted to WW base combinations (i.e. AA, AT, TA or TT). This restriction was based on previous findings indicating that conservative base substitutions are tolerated by the Gin and Tn3 recombinases at the dinucleotide core (39). The selected recombination sites displayed varying degrees of sequence similarity to the wild-type core sequence (Table 3). We flanked each 20-bp genomic core site with zinc-finger binding sites for the H1 zinc-finger protein (18) and evaluated recombination by split gene reassembly (49). We found that β -N95D effectively recombined 6 of 17 (~35%) target sites, with recombination efficiencies from 3 to 95% (Table 3). Surprisingly, we found that only sites containing an AT dinucleotide core were recombined, indicating that β -N95D, unlike the Gin and Tn3 recombinases, might exhibit strict dinucleotide core specificity. The highest level of recombination was observed on target β -AT 1, which shared the greatest degree of sequence similarity with both the wild-type and consensus recombination sites.

Mutational analysis of the serine recombinase arm region

While substrate specificity profiling provides insight into the base requirements for site-specific recombination, the relative importance of the specific amino acid residues that mediate these interactions remains largely unknown. Crystal structures of the $\gamma\delta$ resolvase dimer in complex with its target DNA have revealed that extensive protein-DNA contacts between the C-terminal arm region and the DNA minor groove dictate recombinase recognition (20) (Figure 1A); however, a more detailed

understanding of these interactions is required to facilitate reprogramming of recombinase specificity toward diverse target sites. To better understand the factors that confer target specificity, we used alanine-scanning mutagenesis (53) to investigate the role of each arm region residue in recombination. We took a family-wide approach, targeting the DNA-binding arms of three functionally distinct and hyperactivated recombinases: Tn3-G70S/D102Y/E124Q (Arg120 through Arg143), Gin-H106Y (Glu117 through Pro141) and β -N95D (Ile125 through His147). We introduced Ala substitutions into every arm position for each catalytic domain and fused each mutant to the H1 zinc-finger protein. Native Ala residues were substituted with Gly. We evaluated the ability of each variant to recombine its intended 20-bp core sequence by split gene reassembly.

For each recombinase, we identified a network of 7–10 residues indispensable for catalysis (i.e. a >100-fold reduction in recombination was observed on Ala or Gly substitution) (Figure 4). This network consisted of both evolutionarily conserved residues, presumably important for non-specific DNA-binding, and variable residues that likely contribute to specific target recognition. Conserved residues, numbered according to Tn3, are Ile122, Arg125, Thr126, Gly129, Lys134, Gly137 and Gly141. Among non-conserved residues essential for recombination, we observed substantial positional variation. These residues are Arg130, Ala133, Ile138 and Phe140 for Tn3 (Figure 4A); Glu117, Ile119 and Leu 127 for Gin (Figure 4B); and Ile125, Lys137 and Phe142 for β (Figure 4C). Mapping of the residues critical for Tn3-mediated recombination onto the crystal structure of the closely related $\gamma\delta$ resolvase dimer revealed that each essential position is either directly in contact with

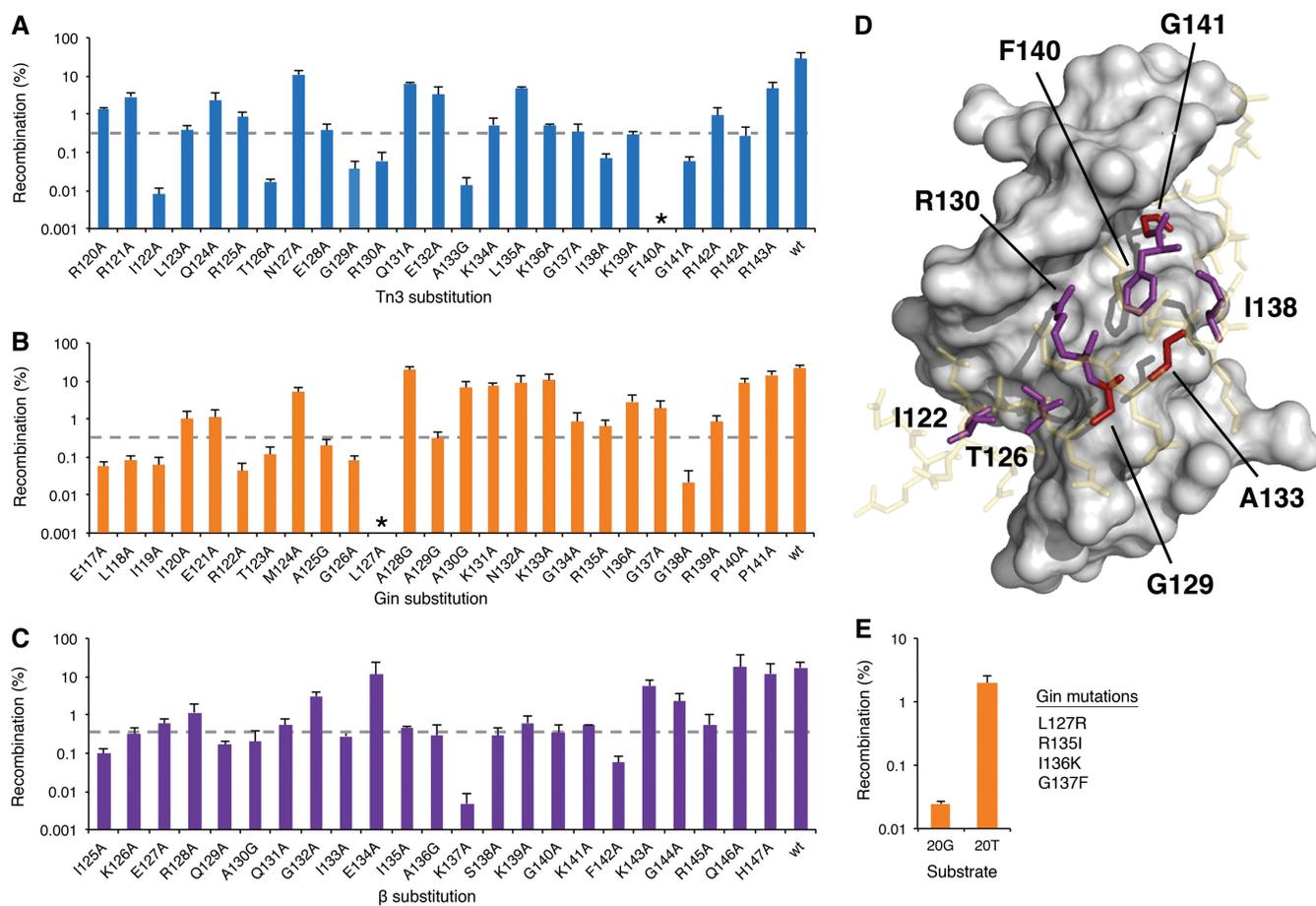


Figure 4. Alanine-scanning mutagenesis of the serine recombinase arm region. (A–C) Recombination activity of mutant (A) Tn3, (B) Gin and (C) β catalytic domains on their native and minimal DNA targets. Asterisk indicates $<0.0001\%$ recombination. Dotted lines indicate threshold below which mutants were considered non-functional. (D) Crystal structure of the $\gamma\delta$ resolvase arm region (sticks) in contact with substrate DNA (gray surface). Conserved and variable residues important for recombination are shown in red and purple, respectively. Inert residues are shown in yellow (PDB ID: 1GDT) (20). (E) Recombination by a Gin chimera substituted with residues predicted to impart specificity onto the 20T core site. Recombination was determined by split gene reassembly. Error bars indicate standard deviation ($n = 3$).

or located in proximity to target DNA (Figure 4D). In particular, this analysis suggests that evolutionarily conserved residues associate with the DNA phosphate backbone, whereas those residues implicated in specific DNA recognition pack into the interior of the DNA minor groove. To directly test whether variable arm residues coordinate recombinase specificity, we attempted to switch the catalytic specificity of the Gin recombinase to that of Tn3. For this, we introduced the four variable residues speculated to be essential for specific recognition by Tn3 (Arg130, Ile138, Lys139 and Phe140) into the analogous locations within the Gin catalytic domain (L127R, R135I, I136K and G137F). We evaluated recombination by this chimera on both the 20T and 20G ZFR target sites. Remarkably, this Gin mutant displayed switched specificity, demonstrating a >1000 -fold preference for the Tn3 target over the Gin target (Figure 4E). Unlike an earlier Gin mutant that was evolved to recognize 20T and contained the substitutions M124S, L127R, R131I, G137M and P141R (40), this chimera was generated entirely by rational methods, guided by experimental data, and used distinct residues not previously suspected to contribute to catalytic specificity.

DISCUSSION

Here we describe the directed evolution of novel β and Sin variants that freely recombine minimal 20-bp core sites derived from the *six* and *resH* site I recombination target sequences. Two selected variants, β -N95D and Sin-Q87R/Q115R, recombined their intended DNA targets with high efficiency and specificity. These results support the use of selection by split gene reassembly for the rapid identification of hyperactivating mutations for the serine recombinases. Cross-comparative analysis revealed that β -N95D would likely be superior to Sin-Q87R/Q115R for targeting applications based on its ability to discriminate the closely related 20S target site. This finding, as well as our inability to profile the specificity of Sin-Q87R/Q115R, suggests that Sin may not be an ideal enzyme for incorporation into ZFRs intended for highly specific genome engineering. However, it may be a promising candidate for applications that require ‘generalist’ catalytic activity (18,49,54) such as non-specific gene transfer.

Selection-based screening revealed the complete specificity profile of β -N95D and led to the derivation of a

consensus target site, which enabled the identification of pseudo-recognition sites present in the human genome. These findings indicate that β -N95D can recombine therapeutically relevant DNA targets and that substrate specificity profiling is an effective tool for identifying ZFR recombination sites within the human genome. Toward a more complete understanding of the mechanisms governing serine recombinase recognition specificity, we also performed an extensive family-wide mutational analysis of the serine recombinase DNA-binding arm region, which indicated that recombinase catalytic specificity is the product of a network of evolutionarily conserved and variable arm positions.

To our knowledge, β -N95D is the first β recombinase variant capable of catalyzing unrestricted recombination between minimal crossover sites derived from the native *six* target site. Rowland *et al.* previously identified 36 substitutions that bypass or disrupt the functions of the Sin regulatory tetramer, enabling Sin-mediated recombination on various *resH*-derived crossover sites (25). In particular, four mutations—T77I, V78A, K110R and Q115R—promoted recombination to near completion. Further, Q115R facilitated recombination in the presence of the inhibitory mutation R54E, suggesting that Q115R might be the most strongly activating mutation of the group. The increased range of sensitivity afforded by split gene reassembly allowed us to more accurately rank the varying levels of activation achieved by each substitution, revealing the following hierarchy: Q115R > V78A > K110R >> T77I. Although highly active on its intended DNA target, we also found that Sin-Q115R displayed low levels of non-specific recombination on several non-cognate core sites, including those derived from the Gin and β target sequences. Profiling of our activated Sin recombinase population revealed that the Sin-Q87R/Q115R double mutant displayed similarly high levels of activity with reduced levels of non-specific recombination. Q87R had not been identified in previous screens, indicating that its selection might be contextually dependent on the presence of Q115R.

The majority of the activating mutations selected in this study lie within the E helix, and more specifically within the recombinase dimer interface. Similarly, the hyperactivating mutations for the Hin (H107Y), Gin (H106Y), Tn3 (G70S, D102Y and E124Q) and $\gamma\delta$ (D102Y) recombinases are also located in this region, indicating a conserved mode of action for these enzymes (52). Keenholtz *et al.* recently solved the crystal structure of the Sin-Q115R catalytic domain in the absence of substrate DNA (51). These studies revealed that the activated Sin tetramer is stabilized by two Arg115 residues present on adjacent recombinase monomers, suggesting that activating mutations facilitate recombination either by destabilizing the dimeric configuration or by stabilizing the tetrameric conformation. Favorable stacking interactions between arginine residues have been proposed as a possible mechanism for tetrameric stabilization (51), although this does not explain the role of the Lys or Cys residues also observed at this position among hyperactivated Sin variants (25). Detailed examination of the crystal structure also revealed the presence of a

negatively charged sulfate ion bound by the Arg115 residues from one subunit of each rotating dimer (51). This sulfate ion is in proximity to nearly all of the selected β and Sin activating mutations. When docked with the substrate DNA from the DNA-bound $\gamma\delta$ resolvase dimer structure (20), the sulfate ion mapped nearly perfectly onto the scissile phosphate of the substrate DNA (51), indicating that activating mutations, such as Sin Q115R, may function by stabilizing the active tetramer geometry, leading to a persistent 'ready' conformation that facilitates elevated levels of catalysis. Recent work has indicated that activating Sin mutations promote DNA cleavage rather than stabilize the cleaved DNA product and that Sin catalytic activity is tightly coupled with its oligomerization state (55). The absence of structural information for β makes it difficult to assess the exact role that N95D has on β -mediated recombination. However, mapping of this mutation onto the Sin-Q115R crystal structure reveals that N95D is in proximity to the sulfate ion, indicating that this substitution might promote recombination through a mechanism similar to Sin-Q115R.

Based on the degenerate β -N95D recombination site used to identify pseudo-target sites, we estimate that β could recombine $\sim 6.7 \times 10^7$ unique 20-bp core sequences in the context of our ZFR platform. This complements our previous work with the Gin recombinase catalytic domain (39), which was estimated to recombine $\sim 3.77 \times 10^7$ distinct core sites. Combined with our archive of >45 pre-selected zinc-finger domains (28,30,31), we estimate that ZFRs based on β -N95D could be generated to recognize nearly 20000 unique 44-bp DNA sequences with even greater targeting capacity anticipated through the future development of β -N95D TAL effector recombinases (56). Notably, our studies revealed that only pseudo-recognition sites that contained an AT dinucleotide core could be effectively recombined by β -N95D. While this could indicate a potential limitation with regard to controlling the directionality of integration, additional studies are required to investigate this unique feature and develop a complete understanding of substrate recognition by β -N95D. In particular, more comprehensive methods for specificity profiling that consider the context-dependency of base substitutions will improve our knowledge of ZFR target recognition and should lead to the design of ZFRs with enhanced targeting capabilities. We previously showed that Gin recombinase catalytic specificity could be re-engineered toward a broad collection of unnatural core sites (39). Similar studies are required to determine whether β -N95D catalytic specificity is also re-programmable. Currently, the ability of β -N95D to recombine core sites beyond the scope of existing ZFRs suggests the possibility of targeting a subset of previously inaccessible genomic target sites. Several previous studies have demonstrated that β is capable of catalyzing site-specific recombination in mammalian cells with pre-introduced copies of the native *six* target sites (57–60). Further studies are required to determine whether ZFRs based on β -N95D are capable of catalyzing targeted integration into endogenous genomic loci.

Finally, alanine-scanning mutagenesis of the entire arm regions of the Gin, Tn3 and β catalytic domains revealed the identities of residues essential for recombination. Past mutational studies have focused on identifying residues that directly participate in catalysis (61,62). Our analysis instead was aimed at locating key residues that confer target specificity. We suspect that highly conserved residues, such as Ile 122, Arg 125, Thr 126, Gly 129, Lys 134, Gly 137 and Gly 141 (numbered according to Tn3), likely contribute critical, but largely non-specific, interactions with DNA, whereas variable positions mediate recombinase specificity. The information gathered from this study allowed us to accurately predict a set of four Gin arm residues that, on substitution with the corresponding Tn3 residues, resulted in a chimeric Gin recombinase with specificity switched from that of Gin to Tn3. Together, these data indicate that a network of variable arm residues coordinates resolvase/invertase target recognition and that mutagenesis of these essential residues is an effective approach for altering recombinase catalytic specificity. These data suggest that future directed evolution studies should be focused on these positions, as they may be responsible for the emergence of the novel recognition features among these enzymes. Intriguingly, our data also suggest that recombinase specificity might be more evolutionarily flexible at core positions 3 and 2, as structural analysis indicates that the majority of the variable positions implicated in mediating specificity are positioned near these bases. In contrast, those residues that contact positions 6, 5 and 4 were observed to be highly conserved Gly, Arg or Tyr residues, indicating that these positions may not be as amenable to redesign. The ability of β to recognize GC content at these positions, compared with the AT-restricted recognition of Gin and Tn3, expands the ZFR toolbox and presents the opportunity for extended evolution and development of the platform. Future studies aimed at investigating cooperative effects among arm residues and their relationship to those residues that participate directly in catalysis should guide the design of smarter libraries and facilitate the generation of new recombinases with extended specificity. In particular, by combining our knowledge of the degeneracy of substrate sequence tolerance with our ability to intelligently target the appropriate specificity determinants for directed evolution, we should be able to generate highly efficient and specific ZFRs for virtually any genomic target, including safe-harbor sites and specific disease loci.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank D. Ekiert, M. Santa-Marta and R. Gordley for contributing to preliminary studies, and S. Carpenter, I. Ahmed and R. Tingle for technical assistance. Molecular graphics were generated using PyMol (<http://pymol.org>).

FUNDING

The Skaggs Institute for Chemical Biology and the National Institutes of Health (NIH) [R21CA126664]; NIH Pioneer Award [DP1CA174426 to C.F.B.]; Swedish Research Council Research [623-2009-7281 to A.J.]. Funding for open access charge: NIH Pioneer Award [DP1CA174426].

Conflict of interest statement. None declared.

REFERENCES

- Gaj, T., Gersbach, C.A. and Barbas, C.F. III (2013) ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol.*, **31**, 397–405.
- Bibikova, M., Beumer, K., Trautman, J.K. and Carroll, D. (2003) Enhancing gene targeting with designed zinc finger nucleases. *Science*, **300**, 764.
- Perez, E.E., Wang, J., Miller, J.C., Jouvenot, Y., Kim, K.A., Liu, O., Wang, N., Lee, G., Bartsevich, V.V., Lee, Y.L. *et al.* (2008) Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. *Nat. Biotechnol.*, **26**, 808–816.
- Gaj, T., Guo, J., Kato, Y., Sirk, S.J. and Barbas, C.F. III (2012) Targeted gene knockout by direct delivery of zinc-finger nuclease proteins. *Nat. Methods*, **9**, 805–807.
- Miller, J.C., Tan, S., Qiao, G., Barlow, K.A., Wang, J., Xia, D.F., Meng, X., Paschon, D.E., Leung, E., Hinkley, S.J. *et al.* (2011) A TALE nuclease architecture for efficient genome editing. *Nat. Biotechnol.*, **29**, 143–148.
- Cermak, T., Doyle, E.L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., Baller, J.A., Somia, N.V., Bogdanove, A.J. and Voytas, D.F. (2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res.*, **39**, e82.
- Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A. *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science*, **339**, 819–823.
- Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E. and Church, G.M. (2013) RNA-guided human genome engineering via Cas9. *Science*, **339**, 823–826.
- Gabriel, R., Lombardo, A., Arens, A., Miller, J.C., Genovese, P., Kaeppl, C., Nowrouzi, A., Bartholomae, C.C., Wang, J., Friedmann, G. *et al.* (2011) An unbiased genome-wide analysis of zinc-finger nuclease specificity. *Nat. Biotechnol.*, **29**, 816–823.
- Fu, Y., Foden, J.A., Khayter, C., Maeder, M.L., Reyon, D., Joung, J.K. and Sander, J.D. (2013) High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.*, **31**, 822–826.
- Sancar, A., Lindsey-Boltz, L.A., Unsal-Kacmaz, K. and Linn, S. (2004) Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. *Annu. Rev. Biochem.*, **73**, 39–85.
- Branda, C.S. and Dymecki, S.M. (2004) Talking about a revolution: The impact of site-specific recombinases on genetic analyses in mice. *Dev. Cell*, **6**, 7–28.
- Grindley, N.D., Whiteson, K.L. and Rice, P.A. (2006) Mechanisms of site-specific recombination. *Annu. Rev. Biochem.*, **75**, 567–605.
- Thyagarajan, B., Guimaraes, M.J., Groth, A.C. and Calos, M.P. (2000) Mammalian genomes contain active recombinase recognition sites. *Gene*, **244**, 47–54.
- Chalberg, T.W., Portlock, J.L., Olivares, E.C., Thyagarajan, B., Kirby, P.J., Hillman, R.T., Hoelters, J. and Calos, M.P. (2006) Integration specificity of phage ϕ C31 integrase in the human genome. *J. Mol. Biol.*, **357**, 28–48.
- Gaj, T., Sirk, S.J. and Barbas, C.F. III (2014) Expanding the scope of site-specific recombinases for genetic and metabolic engineering. *Biotechnol. Bioeng.*, **111**, 1–15.
- Akopian, A., He, J., Boockvar, M.R. and Stark, W.M. (2003) Chimeric recombinases with designed DNA sequence recognition. *Proc. Natl Acad. Sci. USA*, **100**, 8688–8691.

18. Gordley, R.M., Smith, J.D., Graslund, T. and Barbas, C.F. III (2007) Evolution of programmable zinc finger-recombinases with activity in human cells. *J. Mol. Biol.*, **367**, 802–813.
19. Smith, M.C. and Thorpe, H.M. (2002) Diversity in the serine recombinases. *Mol. Microbiol.*, **44**, 299–307.
20. Yang, W. and Steitz, T.A. (1995) Crystal structure of the site-specific recombinase gamma delta resolvase complexed with a 34 bp cleavage site. *Cell*, **82**, 193–207.
21. Prorocic, M.M., Wenlong, D., Olorunniji, F.J., Akopian, A., Schloetel, J.G., Hannigan, A., McPherson, A.L. and Stark, W.M. (2011) Zinc-finger recombinase activities *in vitro*. *Nucleic Acids Res.*, **39**, 9316–9328.
22. Nomura, W., Masuda, A., Ohba, K., Urabe, A., Ito, N., Ryo, A., Yamamoto, N. and Tamamura, H. (2012) Effects of DNA binding of the zinc finger and linkers for domain fusion on the catalytic activity of sequence-specific chimeric recombinases determined by a facile fluorescent system. *Biochemistry*, **51**, 1510–1517.
23. Klippel, A., Cloppenborg, K. and Kahmann, R. (1988) Isolation and characterization of unusual gin mutants. *EMBO J.*, **7**, 3983–3989.
24. Arnold, P.H., Blake, D.G., Grindley, N.D., Boocock, M.R. and Stark, W.M. (1999) Mutants of Tn3 resolvase which do not require accessory binding sites for recombination activity. *EMBO J.*, **18**, 1407–1414.
25. Rowland, S.J., Boocock, M.R., McPherson, A.L., Mouw, K.W., Rice, P.A. and Stark, W.M. (2009) Regulatory mutations in Sin recombinase support a structure-based model of the synaptosome. *Mol. Microbiol.*, **74**, 282–298.
26. Rowland, S.J., Boocock, M.R. and Stark, W.M. (2005) Regulation of Sin recombinase by accessory proteins. *Mol. Microbiol.*, **56**, 371–382.
27. Haffter, P. and Bickle, T.A. (1988) Enhancer-independent mutants of the Cin recombinase have a relaxed topological specificity. *EMBO J.*, **7**, 3991–3996.
28. Segal, D.J., Dreier, B., Beerli, R.R. and Barbas, C.F. III (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.
29. Beerli, R.R., Segal, D.J., Dreier, B. and Barbas, C.F. III (1998) Toward controlling gene expression at will: specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. *Proc. Natl Acad. Sci. USA*, **95**, 14628–14633.
30. Dreier, B., Beerli, R.R., Segal, D.J., Flippin, J.D. and Barbas, C.F. III (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.
31. Dreier, B., Fuller, R.P., Segal, D.J., Lund, C.V., Blancafort, P., Huber, A., Kokscha, B. and Barbas, C.F. III (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.
32. Maeder, M.L., Thibodeau-Beganny, S., Osiak, A., Wright, D.A., Anthony, R.M., Eichinger, M., Jiang, T., Foley, J.E., Winfrey, R.J., Townsend, J.A. *et al.* (2008) Rapid “open-source” engineering of customized zinc-finger nucleases for highly efficient gene modification. *Mol. Cell*, **31**, 294–301.
33. Sander, J.D., Dahlborg, E.J., Goodwin, M.J., Cade, L., Zhang, F., Cifuentes, D., Curtin, S.J., Blackburn, J.S., Thibodeau-Beganny, S., Qi, Y. *et al.* (2011) Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). *Nat. Methods*, **8**, 67–69.
34. Doyon, Y., McCammon, J.M., Miller, J.C., Faraji, F., Ngo, C., Katibah, G.E., Amora, R., Hocking, T.D., Zhang, L., Rebar, E.J. *et al.* (2008) Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases. *Nat. Biotechnol.*, **26**, 702–708.
35. Bhakta, M.S., Henry, I.M., Ousterout, D.G., Das, K.T., Lockwood, S.H., Meckler, J.F., Wallen, M.C., Zykovich, A., Yu, Y., Leo, H. *et al.* (2013) Highly active zinc-finger nucleases by extended modular assembly. *Genome Res.*, **23**, 530–538.
36. Kim, H.J., Lee, H.J., Kim, H., Cho, S.W. and Kim, J.S. (2009) Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly. *Genome Res.*, **19**, 1279–1288.
37. Gordley, R.M., Gersbach, C.A. and Barbas, C.F. III (2009) Synthesis of programmable integrases. *Proc. Natl Acad. Sci. USA*, **106**, 5053–5058.
38. Gersbach, C.A., Gaj, T., Gordley, R.M., Mercer, A.C. and Barbas, C.F. III (2011) Targeted plasmid integration into the human genome by an engineered zinc-finger recombinase. *Nucleic Acids Res.*, **39**, 7868–7878.
39. Gaj, T., Mercer, A.C., Sirk, S.J., Smith, H.L. and Barbas, C.F. III (2013) A comprehensive approach to zinc-finger recombinase customization enables genomic targeting in human cells. *Nucleic Acids Res.*, **41**, 3937–3946.
40. Gaj, T., Mercer, A.C., Gersbach, C.A., Gordley, R.M. and Barbas, C.F. III (2011) Structure-guided reprogramming of serine recombinase DNA sequence specificity. *Proc. Natl Acad. Sci. USA*, **108**, 498–503.
41. Hughes, K.T., Youderian, P. and Simon, M.I. (1988) Phase variation in Salmonella: analysis of Hin recombinase and hix recombination site interaction *in vivo*. *Genes Dev.*, **2**, 937–948.
42. Glasgow, A.C., Bruist, M.F. and Simon, M.I. (1989) DNA-binding properties of the Hin recombinase. *J. Biol. Chem.*, **264**, 10072–10082.
43. Hughes, K.T., Gaines, P.C., Karlinsey, J.E., Vinayak, R. and Simon, M.I. (1992) Sequence-specific interaction of the Salmonella Hin recombinase in both major and minor grooves of DNA. *EMBO J.*, **11**, 2695–2705.
44. Rowland, S.J. and Dyke, K.G. (1989) Characterization of the staphylococcal beta-lactamase transposon Tn552. *EMBO J.*, **8**, 2761–2773.
45. Rowland, S.J., Stark, W.M. and Boocock, M.R. (2002) Sin recombinase from *Staphylococcus aureus*: synaptic complex architecture and transposon targeting. *Mol. Microbiol.*, **44**, 607–619.
46. Rojo, F. and Alonso, J.C. (1994) A novel site-specific recombinase encoded by the *Streptococcus pyogenes* plasmid pSM19035. *J. Mol. Biol.*, **238**, 159–172.
47. Alonso, J.C., Weise, F. and Rojo, F. (1995) The *Bacillus subtilis* histone-like protein Hbsu is required for DNA resolution and DNA inversion mediated by the beta recombinase of plasmid pSM19035. *J. Biol. Chem.*, **270**, 2938–2945.
48. Canosa, I., Rojo, F. and Alonso, J.C. (1996) Site-specific recombination by the beta protein from the streptococcal plasmid pSM19035: minimal recombination sequences and crossing over site. *Nucleic Acids Res.*, **24**, 2712–2717.
49. Gersbach, C.A., Gaj, T., Gordley, R.M. and Barbas, C.F. III (2010) Directed evolution of recombinase specificity by split gene reassembly. *Nucleic Acids Res.*, **38**, 4198–4206.
50. Guo, J., Gaj, T. and Barbas, C.F. III (2010) Directed evolution of an enhanced and highly efficient FokI cleavage domain for zinc finger nucleases. *J. Mol. Biol.*, **400**, 96–107.
51. Keenholz, R.A., Rowland, S.J., Boocock, M.R., Stark, W.M. and Rice, P.A. (2011) Structural basis for catalytic activation of a serine recombinase. *Structure*, **19**, 799–809.
52. Burke, M.E., Arnold, P.H., He, J., Wenwieser, S.V., Rowland, S.J., Boocock, M.R. and Stark, W.M. (2004) Activating mutations of Tn3 resolvase marking interfaces important in recombination catalysis and its regulation. *Mol. Microbiol.*, **51**, 937–948.
53. Cunningham, B.C. and Wells, J.A. (1989) High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. *Science*, **244**, 1081–1085.
54. Proudfoot, C., McPherson, A.L., Kolb, A.F. and Stark, W.M. (2011) Zinc finger recombinases with adaptable DNA sequence specificity. *PLoS One*, **6**, e19537.
55. Mouw, K.W., Steiner, A.M., Ghirlando, R., Li, N.S., Rowland, S.J., Boocock, M.R., Stark, W.M., Piccirilli, J.A. and Rice, P.A. (2010) Sin resolvase catalytic activity and oligomerization state are tightly coupled. *J. Mol. Biol.*, **404**, 16–33.
56. Mercer, A.C., Gaj, T., Fuller, R.P. and Barbas, C.F. III (2012) Chimeric TALE recombinases with programmable DNA sequence specificity. *Nucleic Acids Res.*, **40**, 11163–11172.
57. Diaz, V., Rojo, F., Martinez, A.C., Alonso, J.C. and Bernad, A. (1999) The prokaryotic beta-recombinase catalyzes site-specific recombination in mammalian cells. *J. Biol. Chem.*, **274**, 6634–6640.

58. Diaz, V., Servert, P., Prieto, I., Gonzalez, M.A., Martinez, A.C., Alonso, J.C. and Bernad, A. (2001) New insights into host factor requirements for prokaryotic beta-recombinase-mediated reactions in mammalian cells. *J. Biol. Chem.*, **276**, 16257–16264.
59. Servert, P., Garcia-Castro, J., Diaz, V., Lucas, D., Gonzalez, M.A., Martinez, A.C. and Bernad, A. (2006) Inducible model for beta-six-mediated site-specific recombination in mammalian cells. *Nucleic Acids Res.*, **34**, e1.
60. Servert, P., Diaz, V., Lucas, D., de la Cueva, T., Rodriguez, M., Garcia-Castro, J., Alonso, J., Martinez, A.C., Gonzalez, M. and Bernad, A. (2008) *In vivo* site-specific recombination using the beta-rec/six system. *Biotechniques*, **45**, 69–78.
61. Olorunniji, F.J. and Stark, W.M. (2009) The catalytic residues of Tn3 resolvase. *Nucleic Acids Res.*, **37**, 7590–7602.
62. Keenholtz, R.A., Mouw, K.W., Boocock, M.R., Li, N.S., Piccirilli, J.A. and Rice, P.A. (2013) Arginine as a general acid catalyst in serine recombinase-mediated DNA cleavage. *J. Biol. Chem.*, **288**, 29206–29214.