Evolution of variants of yeast site-specific recombinase Flp that utilize native genomic sequences as recombination target sites

Swetha Bolusani, Chien-Hui Ma[1], Andrew Paek[1], Jay H. Konieczka[1,2], Makkuni Jayaram[1,*] and Yuri Voziyanov*

School of Biological Sciences/IfM, 911 Hergot Avenue, Louisiana Tech University, Ruston, LA 71272, USA, [1]Molecular Genetics and Microbiology, University of Texas, Austin, 1 University Station A5000, Austin, TX 78712-0162, USA and [2]University of Arizona, Life Sciences North 458, 1501 N. Campbell Avenue, Tucson, AZ 85724, USA

ABSTRACT

As a tool in directed genome manipulations, site-specific recombination is a double-edged sword. Exquisite specificity, while highly desirable, makes it imperative that the target site be first inserted at the desired genomic locale before it can be manipulated. We describe a combination of computational and experimental strategies, based on the tyrosine recombinase Flp and its target site *FRT*, to overcome this impediment. We document the systematic evolution of Flp variants that can utilize, in a bacterial assay, two sites from the human interleukin 10 gene, IL10, as recombination substrates. Recombination competence on an end target site is acquired via chimeric sites containing mixed sequences from *FRT* and the genomic locus. This is the first time that a tyrosine site-specific recombinase has been coaxed successfully to perform DNA exchange within naturally occurring sequences derived from a foreign genomic context. We demonstrate the ability of an Flp variant to mediate integration of a reporter cassette in *Escherichia coli* via recombination at one of the IL10-derived sites.

INTRODUCTION

Site-specific recombinases of the tyrosine family mediate precise DNA rearrangements between a pair of their respective cognate target sites [for a review see Refs (1–5)]. Among more than a hundred known members of this family (6), extensive biochemical and structural information is available for a small subset—Int proteins from phages λ and HP1, XerC/XerD from *Escherichia coli*, Cre from phage P1 and Flp from the *Saccharomyces cerevisiae* plasmid 2 micron circle. Tyrosine recombinases, whose name derives from the active site nucleophile utilized for DNA cleavage, follow the same reaction chemistry as type IB topoisomerases. These recombinases display conserved active site residues and similar 3D folds in their catalytic domains (7–11), features that they share with type I eukaryotic topoisomerases (12,13). At least some of the tyrosine recombinases are able to execute DNA exchange in a variety of cell types tested (14), certifying their potential utility in performing directed genome rearrangements: deletions, integrations, inversions and translocations (15–17). Development of genetic engineering tools has focused primarily on Cre and Flp, with their relatively simple substrates and reaction requirements, and to a lesser extent on λ Int (18–20). A major hindrance, though, is the need to introduce beforehand the native target sites of these recombinases at the genetic loci being targeted for manipulations. Overcoming this requirement by evolving recombinase variants that act on pre-existing target-like sites in a genome will signify an important advance in biotechnology, including gene therapy.

The related structural configurations of the exquisitely sequence-specific recombinases and the relatively sequence-indifferent topoisomerases suggest the possibility of manipulating the former to recognize novel target sites without sacrificing their recombination competence. Directed evolution strategies applied to Cre and Flp proteins (21–26) have produced recombinase variants with relaxed or altered target specificities. Members of the former class recombine both the native and novel targets with comparable efficiencies; those of the latter strongly favor the novel targets. In general, amino acids that directly interact with bases in the target site provide the major determinants of specificity. At the same time, amino acids that interact with the base-contacting residues, as well as those located at the monomer–monomer interfaces and/or in proximity to the catalytic residues, also

contribute to the modulation of target specificity (21–23,25,26). Furthermore, as revealed by structural studies on Cre variants, plasticity in DNA and protein conformations, combined with water-mediated interactions, can result in novel modes of macromolecular recognition (27).

The initial set of evolved novel specificities was limited to a single substitution or a small number of substitutions within the recombinase-binding element of the target site. Furthermore, in order to maintain symmetry, the changes were placed in each of the two binding elements that flank the strand exchange region or spacer. Despite their value as models for understanding macromolecular recognition during recombination, these sites are quite inadequate in representing target-like sites within a genome. The probability of occurrence, at a genomic locale of interest, of a dyad symmetry formed by 11–14 bp sequence strings bracketing a 6–8 bp spacer is extremely low. As elementary mimics for binding site asymmetry, we incorporated a distinct specificity into each half-site of the native Flp recombination target (*FRT*) sequence. Such a hybrid site is recombined efficiently either by a single Flp variant that is relaxed with respect to the two specificities or by a combination of two variants, each of which satisfies one of the two specificities (26).

We describe here Flp variants, obtained through a stepwise molecular evolution regimen, with recombination competence for two *FRT*-like (*FL*) sequence segments that occur naturally within the human IL10 gene (*FL-IL10A or FL-IL10B*). Such *FL* sites, with roughly the same degree of relatedness to *FRT*, can be identified in a genome approximately once every 10 kb. The recombination efficiency of *FL-IL10A* or *FL-IL10B* is increased significantly when its recombination partner is *FRT* or a hybrid site with one Flp-binding element from *FRT* and the other from the *FL* site. The present work establishes the feasibility of mediating locale-specific DNA rearrangements on a genome-wide scale.

MATERIALS AND METHODS

The computer program 'TargetFinder' for searching *FRT*-like genomic sequences

TargetFinder accepts a sequence file as input, and provides an output file listing the position and scores of sub-sequences similar to *FRT*. The scoring algorithm works by first identifying a potential 8 bp spacer within a 34 bp segment, then comparing the 13 bp (probable 'Flp-binding elements') flanking it on either side to the corresponding *FRT* sequence. A potential spacer is defined as an 8 bp string that begins with 5′T and ends with A3′, and has a G–C content <50%. This constraint on G–C content ensures sufficient flexibility of the spacer during the strand exchange steps corresponding to Holliday Junction formation and resolution. Under this context, the first and seventh positions of the binding element are weighted most (score = 4), followed by positions 2–6 (score = 2) and then by positions 8–13 (score = 1). The final score for a genomic site is estimated by multiplying its summed score by 100 and dividing by the summed score for *FRT*. The ranking system may be modified, as desired, by awarding merit points for consecutive matches

with *FRT* positions and for sequences that show dyad symmetry with respect to the spacer.

In vivo recombination assays: negative and positive screens

All *in vivo* recombination assays were performed in the *E.coli* strain DH10B from Invitrogen (*F, mcrA, (mrr-hsdRMS-mcrBC), φ80dlacZM15, lacX74, deoR, recA1, endA1, ara139, D(ara, leu)7697, galU, galK, λ⁻, rpsL, nupG*). The assays utilized either negative or positive screening systems. The 'negative screening system' is based on the excision and loss of a *lacZα* reporter cassette and the resulting change in colony color from blue to white on X-gal indicator plates (24). The 'positive screening system' is based on the excision of the transcription terminator (from the ribosomal RNA operon *rrnB* of *E.coli*) located between a promoter and the *lacZα* gene, causing white to blue change in colony color in presence of X-gal.

The reporters for the positive screen were constructed as follows. First, we introduced an XhoI site between the 3′ end of the *rrnB* region and the promoter of the *bla*-gene of plasmid pBAD24 (28). In the resulting plasmid, a promoterless *lacZα* cassette from pUC18 obtained by PCR amplification (in the *lacZα* cassette, all restriction sites in the multiple cloning sites region except EcoRI site were deleted) was inserted between HindIII and XhoI sites, thus deleting the *rrnB* region. This plasmid was further modified by cloning the promoter for *lacZα*, prepared as a PCR-amplified fragment, between NheI and NcoI sites. This plasmid derivative called pXLPB served as a template for constructing all the positive reporters used in this study. To create a positive reporter for a given pair of recombination target sites, the *rrnB* region from pBAD24 was PCR amplified using oligonucleotide primers harboring the appropriate sequences and then cloned between SphI and HindIII sites of pXLPB.

Integration assays *in vivo* in *E.coli*

Two plasmids were constructed for the integration assays in *E.coli*. To construct the first, we introduced an XhoI site between the 3′ end of the *rrnB* region and the promoter of the non-functional *bla*-gene in plasmid pBAD33 (28). The plasmid thus obtained was modified by inserting the PCR-amplified promotorless *lacZα* cassette (∼550 bp) from pUC18 between NsiI and XhoI sites. The forward primer in the amplification step contained the sequence for the desired recombination target site ([*hFRT*]-[*hIL10A*]; Figure 6A). In the resulting plasmid pLac, the *lacZα* gene, being promoterless cannot be expressed. To construct the partner plasmid in integration, an RFP cassette (∼900 bp) lacking a promoter was PCR amplified from p33Rdm11 (24), and was inserted into pXLPB (described above) between SphI and XhoI sites to replace *lacZα*. The forward primer used to amplify the RFP gene contained the *FL-IL10A* site. The resultant construct pRFP can express RFP from the *lac* promoter.

Site-specific integration of pLac into pRFP was assayed by expressing the Flp variant FV7 from the inducible arabinose promoter as described previously (24). Successful integration events were scored by estimating the frequency of blue colonies on X-gal indicator plates. The expected recombination

reaction was verified by digesting plasmids isolated from blue and white (red under UV) colonies with XhoI.

Purification of Flp and Flp variants: *in vitro* DNA binding and recombination reactions

The Flp protein and its variants were purified to ~70% purity as described in (26), and used in DNA binding and recombination assays.

The *in vitro* recombination assays were performed and analyzed as described previously (26). Reactions contained ~2, 4 or 8 pmol of Flp or a Flp variant per pmol of binding element of the recombination substrate.

The substrates for DNA binding, ~200 bp long, were obtained by PCR amplification from plasmids harboring the target sites. The binding reactions were done at 30°C for 30 min in a modified recombination buffer that omits PEG and reduces the final glycerol concentration from 5 to 2.5%. The molar ratio of Flp or a Flp variant to binding element ranged from ~2 to 6. The binding reactions were fractionated by electrophoresis in 1.4% agarose gels at 4°C under a voltage of ~8 V/cm. DNA bands bound or unbound by protein were visualized under UV after ethidium bromide staining. Band intensities were quantified using Bio-Rad Bio-Analyst software.

DNA mutagenesis and shuffling

Selected amino acid positions in Flp were randomized by a PCR-based mutagenesis scheme, using primers containing the requisite degenerate positions (24). Usually, Flp variant libraries were prepared by simultaneously randomizing two residues. Error-prone mutagenesis and DNA shuffling were done as described earlier (29).

Mutagenic PCR was performed using *Taq* polymerase in a buffer that contained 0.25 mM $MnCl_2$, 0.5 mM $MgCl_2$, 10 mM Tris–HCl (pH 8.3), 50 mM KCl, 0.2 mM dNTPs, 100 nM of each primer and 100 ng of DNA template. After amplification, the DNA product was purified using PCR Clean Up Kit (Qiagen), digested with SacI and SphI, and ligated to pBAD33 digested with the same enzymes.

The procedure for DNA shuffling was as follows. DNA fragments encompassing Flp variant genes of interest were amplified using *Taq* polymerase under non-mutagenic conditions. A pool of the PCR product, containing equimolar amounts of each gene, was fragmented to a size range of 50–200 bp by digestion with DNase I (30) [50 mM Tris–HCl (pH 7.4) and 10 mM $MnCl_2$; 5 min on ice]. After purification of the DNA fragments using nucleotide removal kit (Qiagen), they were shuffled and reassembled in a PCR using *Pfu* polymerase, without the addition of exogenous primers. The reconstruction of DNA fragments of the expected size was verified by sampling the reaction mixture by agarose gel electrophoresis. This DNA then served as template for amplification in a second PCR using a pair of synthetic primers designed to anneal just outside the Flp (variant) coding region. Depending on the purpose at hand, amplification was done under non-mutagenic or mutagenic conditions. The latter condition introduced on an average two mutations within the Flp coding region (1.3 kb). The amplified DNA was digested with SacI and SphI, and cloned into the Flp expression vector as described previously (24).

Other methods

The Flp–DNA structure was analyzed using Swiss PDB Viewer 3.7 (31). Standard procedures for plasmid isolation, bacterial, restriction enzyme digestion and so on were as described in Sambrook and Russell (32).

RESULTS

Logic of progressive evolution of Flp recombination specificity to native genomic sites: screening Flp variants with low recombination activity

The directed evolution strategies used in this study emulate those of Voziyanov *et al.* (25) and are based on the following rationale. The first step is to identify, using a computational method, *FRT*-like (*FL*) sites in a genome that could potentially be targeted for recombination by Flp (see below). Since *FL* sites deviate significantly in their sequence from *FRT*, and are unlikely to be symmetric in their potential 'Flp-binding elements', two possible solutions are suggested by previous work (26). A separate Flp variant may be evolved to satisfy each binding element specificity, and the two variants may be coexpressed to mediate recombination. Alternatively, a relaxed specificity variant may be evolved that can mediate recombination on its own. The latter strategy was followed here. Rather than attempting to macroevolve the desired specificity in a single giant step, we followed a more gradual evolutionary path guided by the 'mixed' specificities contained within chimeric substrates constructed from *FRT* and an *FL* sequence of interest. They are referred to as *FLRT* sites (see below). The fraction of *FRT* positions replaced by *FL* sequences in *FLRT* sites increased from early to late stages of evolution.

The exchange of sequence blocks between *FRT* and *FL* in the *FLRT* substrates was generally guided by features of Flp–DNA interactions inferred from footprinting assays (33,34) and observed in crystals of the Flp–Holliday junction complex (11). The N-terminal domain of Flp interacts with the cleavage-proximal bp 1–5 (Figure 1A) of the binding element, with a direct contact between Lys-82 and G-1 through the major grove (11). The C-terminal domain interacts with DNA from the opposite side through the central major groove and the flanking minor grooves, making a direct contact between Arg-281 and G-7 (Figure 1B). Lys-285 is either within hydrogen bonding (or van der Waals) distance of T-9 or is in contact with N7 of G-10, depending on whether the Flp monomer is in its catalytically active or inactive state. A number of additional contacts by both N- and C-terminal domains are possible through water-mediated networks.

Depending on the non-*FRT* base pair positions within a chimeric-binding element, selected amino acids, as suggested by the DNA–protein contacts in the Flp structure, were randomized to trigger specificity modulation. After screening this first generation Flp variant libraries for the desired recombination specificity, the subset of positive individual variants was subjected to coupled mutagenesis and gene shuffling, and further screening [(25,26); see also Materials and Methods]. Winners from this test were processed through additional rounds of mutagenic shuffling and screening to identify the emergence of recombination competence on
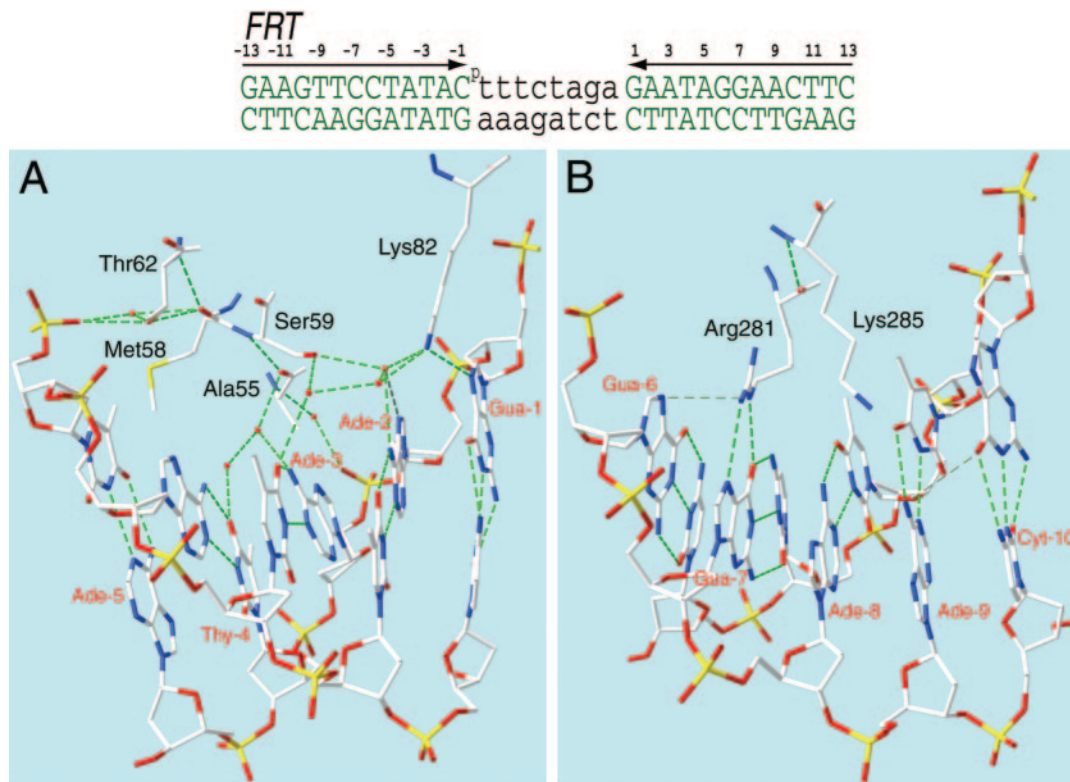
Figure 1. Contacts between Flp and its target site, *FRT*. In the *FRT* sequence, the 13 bp Flp-binding elements flanking the 8 bp spacer are indicated by the horizontal arrows. The individual bp within a binding element are shown in green upper case letters, and numbered in ascending order from the position nearest to the spacer. The spacer sequence is written in black lower case letters, and the scissile phosphates are indicated by 'p'. (**A** and **B**) A subset of the contacts between the Flp protein and *FRT* relevant to the directed evolution of new target specificity are displayed (11).

chimeric sites that were systematically enriched for the genomic sequences, and finally on the genomic site *per se*.

To cope with the suspected low recombination efficiencies at genomic sites, we devised a more sensitive genetic screen than the standard *lacZα* deletion assay employed previously (24). In the new 'positive' system (Materials and Methods), recombination removes a transcription terminator sequence to turn on expression of the *lacZα* reporter. Excision of a small fraction of the terminator copies (theoretically just one copy) would suffice to display a colony as blue on X-gal indicator plates.

Identifying *FRT*-like sites in a genome

To identify *FRT*-like (*FL*) sequences within a genome, we developed the search program 'TargetFinder'. This program divides each of the entire set of 34 bp DNA segments (the size of *FRT*) into two 13 bp elements flanking an 8 bp spacer (strand exchange region), and compares it base pair to base pair with *FRT*. TargetFinder imposes the constraint that the first and last base pair of the spacer are T-A, as in *FRT*, and the overall GC content of the spacer is <50%. Although Flp makes only two phosphate contacts and no direct base-specific contacts within the spacer, a GC rich spacer does not support efficient recombination (35). Each *FL* sequence is assigned a score using a weighted system that grades not all individual positions of the binding element equally (Materials and Methods). Whereas *FL* sequences with high scores (>50) are seen approximately once every 10^4 bp in a genome,

low-scoring *FL* sequences (~30) occur roughly once every 200 bp. Two *FL* sequences from the human IL10 gene (Figure 2B), *FL-IL10A* and *FL-IL10B*, have scores of 60 and 54, respectively, when normalized to the *FRT* score of 100 (Figure 2A). They are the targets for the directed evolution described in this study. Chimeric *FLRT* sites constructed by mixing *FL* and *FRT* sequences (Figure 2C) have scores intermediate between those of *FL* and *FRT*.

Relaxed specificity Flp variants that espouse *FLRT* chimeric sites as recombination targets

A proof-of-principle test of our evolutionary design utilized two *FLRT* sites derived from *FL-IL10B* and *FRT* (Figure 2C). The choice of *FL-IL10B* over *A* was based on the presence of the critical G–C base pair at position 7 (36) in both binding elements of the former, but only one of the latter. However, the two sites are also related. They harbor the important G–C base pair at position 1 in both their binding elements. In addition, they share five identities in one of their binding elements. As such, we anticipated a finite probability that the evolutionary path to *FL-IL10B* might also lead to *FL-IL10A*, either through common relaxed specificity variants or perhaps through independent variants. In *FLRTB(1-8g)*, positions 1–8 were genomic (*g*) sequences derived from *FL-IL10B*, and positions 9–13 were from *FRT*; in *FLRTB(1-10g)*, the genomic sequence extended to position 10 (Figure 2C). Since the contribution towards recombination activity from the remaining three positions
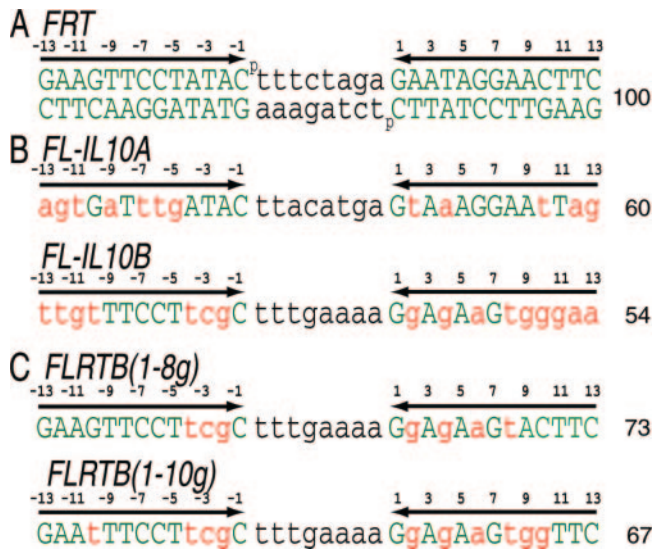
Figure 2. *FRT*, *FRT*-like (*FL*) genomic sites and *FLRT* chimeric sites. (**A**) The *FRT* site is represented as in Figure 1. (**B**) The *FRT*-like (*FL*) sequences from the human IL10 gene used for this study were *FL-IL10A* (GenBank accession no. X78437, positions 1154–1187) and *FL-IL10B* (GenBank accession no. AF418271, positions 1264–1297). The positions within the *FL*-'binding elements' that differ from *FRT* are denoted by lower case red letters. TargetFinder scores for these sites, normalized to a score of 100 for *FRT*, are shown at the right. (**C**) In the chimeric sites between *FRT* and *FL-IL10B*, the spacer as well as positions 1–8 (*FLRTB(1-8g)*; score 73) or positions 1–10 (*FLRTB(1-10g)*; score 67) were derived from IL10B.

(11–13) was minimal (37–39), they were left alone as in *FRT*. The 1-8*g* sequence differed from *FRT* at three positions (2–4) in one binding element of *FLRTB(1-8g)*, but at four positions (2, 4, 6 and 8) in the other. The corresponding differences between *FLRTB(1-10g)* and *FRT* were four (positions 2–4 and 10) and six (positions 2, 4, 6 and 8–10), respectively.

Flp variants with recombination competence on *FLRTB(1-8g)* and *FRT*

In the Flp structure, Ala-55 and Ser-59 participate in a network of water-mediated contacts with the cleavage-proximal region of the Flp-binding element (Figure 1A) (11). In addition, Met-58 is in close proximity to Ala-55 and could potentially be involved in contacting DNA or in positioning water molecules that facilitate DNA contact. In previous work (26), we noted the presence of A55H and S59G among the amino acid changes, collectively referred to as sup3, that strongly suppress the recombination defect of a Flp variant Flp(Y60S) on *FRT*. Tyr-60, which is conserved among Flp related yeast tyrosine recombinases, makes a cleavage-proximal contact with a spacer phosphate on the non-cleaved strand opposite the scissile phosphate (11). When transplanted into the wild-type background, position 60 being held as tyrosine, the sup3 group of 10 mutations (listed in Figure 3A) confers relaxed specificity towards native (*FRT*) and two mutated target sites (*mFRT*s) (26). Since one of the binding elements of *FLRTB(1-8g)* differs from that of *FRT* only at three consecutive cleavage-proximal base pairs (2–4), we suspected that individual or combinatorial changes at amino acids 55, 58 and 59, in conjunction with other accessory mutations, will likely accommodate

FLRTB(1-8g) specificity. Two of the four differences present in the second binding element of *FLRTB(1-8g)* lie within the cleavage-proximal bp 1–5 (positions 2 and 4), and two are more distal (positions 6 and 8). In principle, a relaxed specificity Flp variant may recombine *FLRTB(1-8g)* by being indifferent to the asymmetry in the binding elements; alternatively, binding cooperativity between two protein monomers may override the asymmetry.

Because of the proclivity of sup3 mutations to relax Flp specificity, pairwise randomization of residues 55, 58 and 59 were carried out in Flp^{sup3} (26). Of the three separate libraries so obtained, the one with randomized 55–58 pair yielded several variants that were able to recombine *FLRTB(1-8g)*. After randomizing amino acid 59 in this sub-library, its members, together with the wild-type *FLP* gene, were carried through mutagenic shuffling and re-screened by the above test. Four positive variants FV1-4 were identified thus; their mutation profiles are tabulated in Figure 3A. The accuracy of the recombination events performed by them was verified by SphI digestion of plasmids isolated from white and blue colonies obtained in the screen (Figure 3A, right). One of the variants, FV3, was also characterized by *in vitro* binding (Figure 3B) and recombination assays (Figure 3C). As was suspected, these variants were relaxed in their specificity towards *FLRTB(1-8g)* and *FRT*.

Consistent with their origin, FV4 retained 9 out of the 10 mutations that constitute the sup3 group and FV1-3 retained 8 (Figure 3A). The lack of retention was at the randomized positions 55 and 59. All four variants harbored A55S rather than A55H of Flp^{sup3}, whereas FV1-3 additionally contained a switch from S59G of Flp^{sup3} to S59N. FV1-4 had also acquired M58V (a randomized position) and S130P (acquired during mutagenic shuffling). There was a single additional mutation in FV2 (F182S), three such mutations in FV1 (R17C; S193G; A349T) and FV3 (M44V; S234I; T312A), and four in FV4 (M44V; V280I; A349T; T373A). Although the structural basis for the common specificity of FV1-4 on *FLRTB(1-8g)* is unknown, the locations of a subset of the mutations they harbor (S193G in FV1, F182S in FV2, S234I and T312A in FV3 and A349T in FV4) are well suited for potentially altering phosphate and base contacts or interactions involving some of the catalytic residues.

Two Flp variants that utilize *FLRTB(1-10g)* and *FL-IL10A* as recombination targets

Since the non-identities between *FLRTB(1-10g)* and *FRT* include position 10 in one binding element and positions 9 and 10 in the other, a *FLP* library randomized at the codon for Lys-285 (which is within hydrogen bonding distance of T-9 or G-10) was shuffled, under mutagenic conditions, with genes of variants that recombine *FLRTB(1-8g)*. Those variants that showed recombination on *FLRTB(1-10g)* in the positive screen were reshuffled in efforts to improve recombination efficiency. Among the final crop of variants (Figure 4A), FV5 and FV6 gave low to modest recombination (3 and 15% blue colonies, respectively); FV7 and FV8 were significantly more efficient (both ∼80% blue colonies). Not surprisingly, FV7 and FV8 were active on *FRT* as well, yielding nearly 100% recombination (blue colonies)
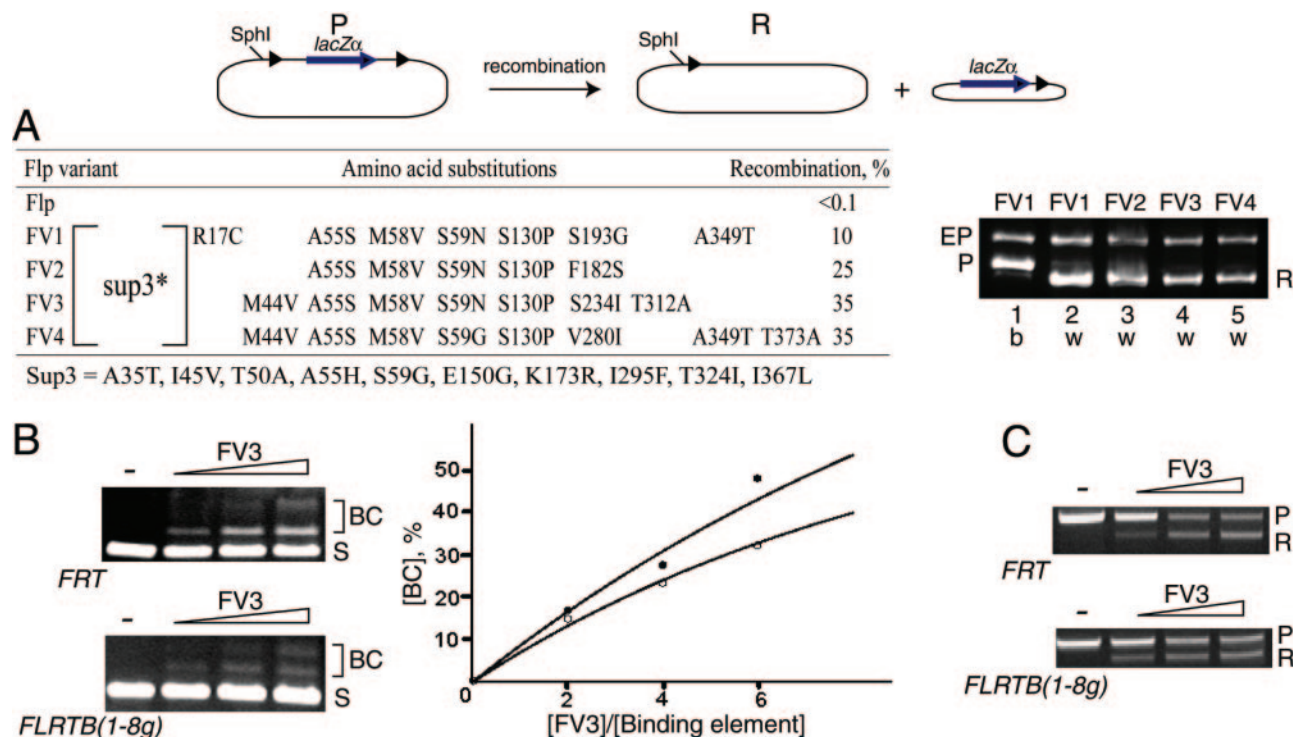
Figure 3. Evolution of Flp variants that show recombination competence on *FLRTB(1-8g)*. The deletion reaction that converts the reporter plasmid P to plasmid R, and forms the basis of the colony color assay, is schematically depicted at the top. (**A**) The template for directed evolution was a variant Flp^{sup3} described previously (26) that showed relaxed specificity towards *FRT* and three altered target sites derived from it. The set of mutations present in Flp^{sup3} are listed (26). The subset of the sup3 mutations that were retained in the Flp variants that recombine *FLRTB(1-8g)* is denoted by sup3* (A35T, I45V, T50A, E150G, K173R, I295F, T324I, I367L). The recombination activity of a variant in the negative system (deletion of *lacZα*) was expressed as the percentage of white colonies on X-gal plates. Plasmids isolated from randomly picked white colonies (w) were digested with SphI to verify authentic recombination. Representative results in lanes 2–5 correspond to FV1–FV4, respectively. Lane 1 shows the digestion result for a blue colony (b) from the FV1 assay. P refers to the parent plasmid, and R to the larger of the two deletion products. The top band (EP) corresponds to the Flp (or Flp variant) expression plasmid. (**B**) The molar ratios of protein to binding element in the *in vitro* binding reactions of FV3 to *FRT* and *FLRTB(1-8g)* were ∼2, 4 and 6 (lanes 2–4). The bound complexes BC contain one FV3 monomer (lower band) or two monomers (upper band) per substrate molecule. The plot shows the levels of binding of FV3 to the two targets. Closed circle, *FRT*; open circle, *FLRTB(1-8g)*. (**C**) *In vitro* recombination activity of FV3 was assayed on *FRT* and *FLRTB(1-8g)* using 2, 4 and 8 monomers of FV3 per binding element (lanes 2–4). P and R stand for parental and recombined plasmids. The small deletion circle that runs towards the bottom of the gel during electrophoresis is not shown.

in the positive system and ∼5 and 10% recombination (white colonies), respectively, in the negative system (data not shown). When tested against the genomic sites, both variants showed respectable recombination activity on *FL-IL10A* (∼60% blue colonies) (Figure 4B). The activity on *FL-IL10B*, though detectable, was quite weak (∼0.1%). Thus, by adapting Flp stepwise to chimeric sites, its recombination specificity can be sufficiently relaxed to render a true genomic sequence an acceptable target. There was only one amino acid difference, at position 193, between FV7 and FV8 (Figure 4A). The very similar activity patterns of the two proteins on the different substrates tested is therefore not surprising.

In contrast to FV7 and FV8, wild-type Flp yielded virtually no recombination or extremely poor recombination when a genomic site was offered as the substrate (<0.001% with *FL-IL10B* and ∼0.1% with *FL-IL10A*; Figure 4B). The recombination activity of Flp^{sup3} was nearly undetectable on *FL-IL10B* (<0.001%), and quite weak on *FL-IL10A* (1%) (Figure 4B). This result is consistent with the previous observation that Flp^{sup3} is partially relaxed in its specificity (26).

The predicted recombination events, converting the reporter plasmid P to the shortened form R, were checked by SphI digestion of plasmid DNA isolated from a random set of blue colonies (Figure 4A and B). The band running just above P in Figure 4A and B was consistent with either a recombination event from a dimer form of the parent plasmid or the insertion of the excised circle into a parent monomer. This inference was verified by digestion with SphI plus HindIII, whose target sequences bracket the DNA segment harboring the *rrnB* terminator and the flanking recombination sites. The SphI–HindIII fragment carrying a second inserted copy of the *rrnB* terminator would be ∼1000 bp long. Indeed, a band of this size (X) was detected in digests of the plasmid mixture corresponding to lanes 3 and 4 of Figure 4A (Figure 4C, lanes 3 and 4).

Recombination efficiency of genomic sites can be enhanced by cooperative effects of *FRT* specificity

Because of the ability of FV7 and FV8 to utilize *FRT* with superior efficiency than the genomic sites, the recombination potential of *FL-IL10A*, and in particular that of *FL-IL10B*, could be improved by providing either one of these an *FRT*
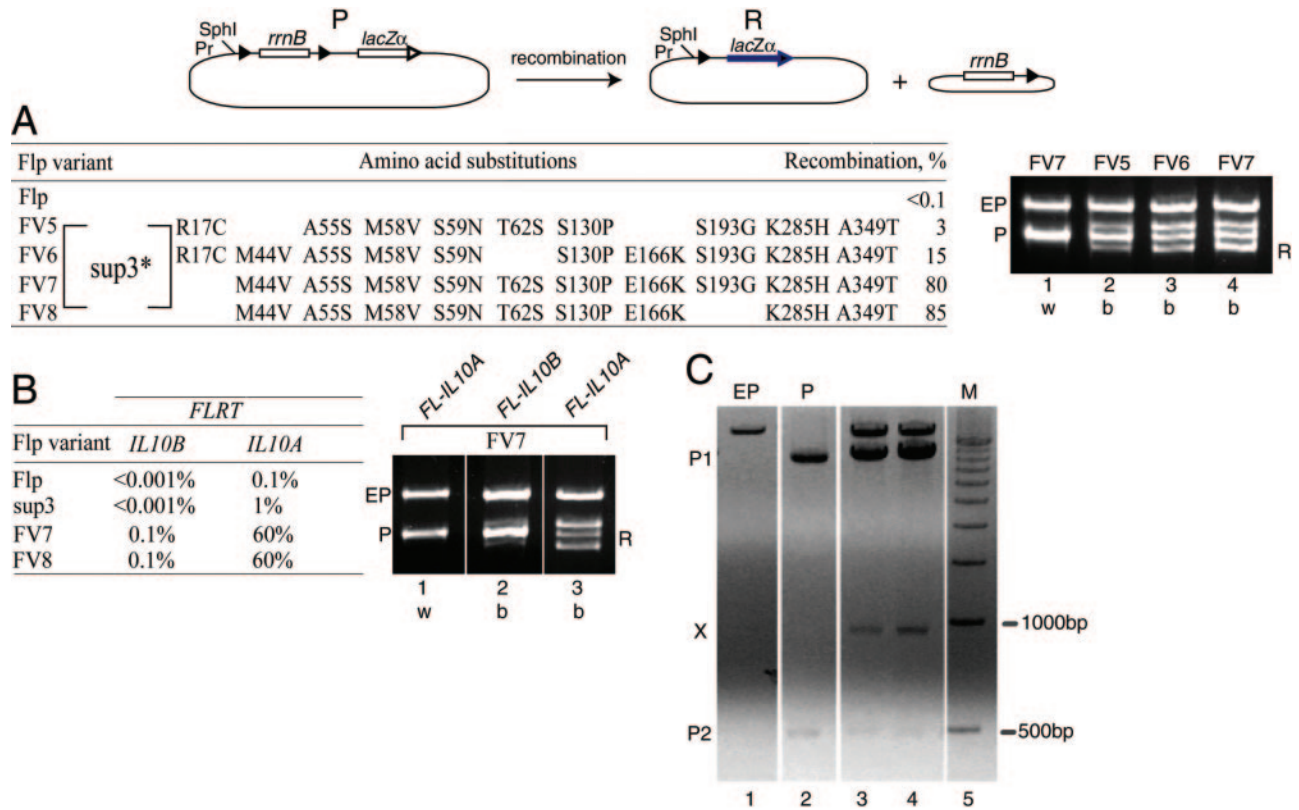
Figure 4. Evolved variants of Flp with recombination activity on *FLRTB(1-10g)*, *FL-IL10A* and *Fl-IL10B*. The recombination event in the parent plasmid P that excises the *rrnB* transcription terminator to cause *lacZ* expression from plasmid R is schematically shown at the top. (**A**) The mutations, in addition to the sup3* set (see legend to Figure 3), present in Flp variants that act on *FLRTB(1-10g)* are tabulated. Recombination assays were done using the positive system (expression of *lacZα*), and the activity of each variant was expressed as the percent of blue colonies on X-gal plates. Plasmids isolated from a random set of blue colonies (b) were digested with SphI to verify the predicted deletion event. Typical results for FV5–FV7 are shown in lanes 2–4, respectively. The result for a white colony (w) obtained with FV7 is shown in lane 1. P and R refer to the parental plasmid and the larger of the two deletion products, respectively. The band at the top denotes the expression plasmid (EP). The extra band above P in lanes 2–4 was either the product of the excision reaction from a plasmid dimer or resulted from insertion of the excised *rrnB* circle into a parental monomer plasmid. (**B**) Recombination efficiencies of Flp, Flp[sup3], FV7 and FV8 on the genomic sites *FL-IL10A* and *FL-IL10B*, as well as the authenticity of recombination, were assayed as described under (A). Plasmids isolated from blue (b) or white (w) colonies were digested with SphI to verify the predicted deletion event. Only the results for FV7 are shown. The reporter substrates are indicated above the respective lanes. Recombination on FL-IL10B was quite weak but detectable. (**C**) Plasmid mixtures corresponding to lanes 3 and 4 of (A) were digested with SphI plus HindIII (lanes 3 and 4), and fractionated with similarly digested Flp expression plasmid (EP) (lane 1) and the parent reporter plasmid (P). The bands marked P1 and P2 refer to the backbone of the recombined reporter and the *rrnB* cassette, respectively. The size of the band marked 'X' agrees with it containing two copies of *rrnB* and three copies of the recombination target site. This is the expected configuration for an intermolecular recombination event that inserts a copy of the excised *rrnB* circle into the parent reporter plasmid. The lane containing the DNA size ladder is marked 'M'.

site with a matching spacer as the recombination partner (Figure 5A and B). Similarly, FV7- or FV8-mediated recombination between *FL-IL10B* and a hybrid site [h*FRT*]-[h*FL-IL10B*], containing one binding element from *FL-IL10B* and the other from *FRT* (see the *IL10B2* reporter configuration in Figure 5B), was much more efficient than that between two copies of *FL-IL10B*. The 'h' in the hybrid site refers to a half-site. In contrast, wild-type Flp yielded no recombination or poor recombination when a genomic site was partnered with *FRT* or with a hybrid site. Similarly, the recombination activity of Flp[sup3] was nearly undetectable (<0.001%) when one of the recombination partners of the reporter cassette was *FL-IL10B* (the *IL10B2* or *IL10B3* cassettes in Figure 5B). However, it showed significantly improved recombination (20%) when *FL-IL10A* was paired with *FRT* (the *IL10A3* cassette in Figure 5B). This result is not surprising, given that Flp[sup3] recombines *FRT* efficiently, and is partially relaxed in its specificity (26).

The FV7 variant recombination activity was detected *in vitro* in a deletion reaction using *FRT* and *FL-IL10A* as target sites, as in the substrate configuration *IL10A3* in Figure 5B (Figure 5C).

Site-specific integration of a reporter plasmid at the *FL-IL10A* site

Based on the 'hybrid vigor' of sites containing one *FL*-binding element and one *FRT*-binding element (Figure 5), we successfully performed a gene targeting experiment, as diagrammed in Figure 6A. One of the plasmid substrates for recombination pLac contained a promoterless *lacZα* cassette, whereas the second one pRFP contained the RFP (red fluorescent protein) gene expressed from the upstream Lac promoter. As a result, colonies harboring them were white (or faintly red) under visible light, but distinctly red under UV light. Plasmids pLac and pRFP were compatible
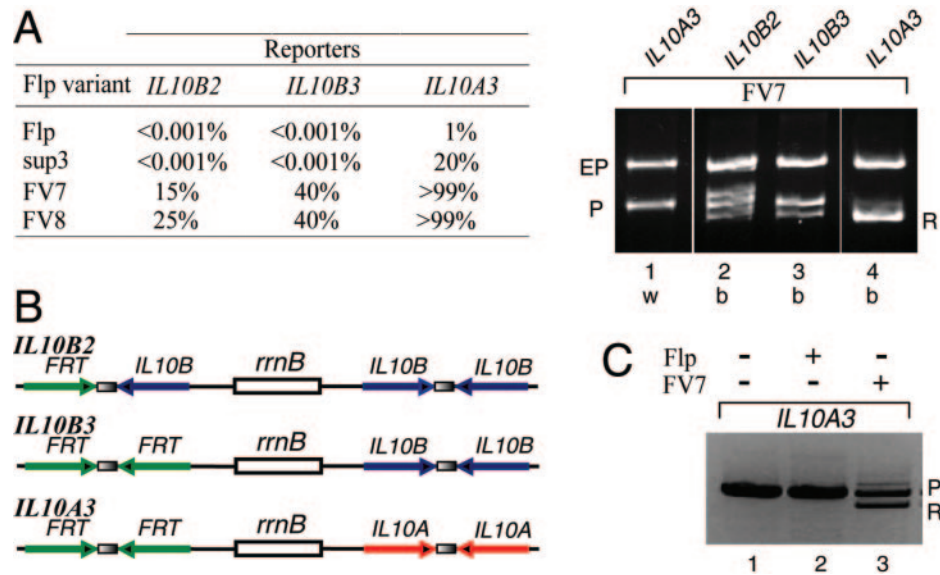
Figure 5. Recombination between a genomic site and a partner *FRT* site or a hybrid site containing one Flp-binding element from *FRT*. (**A** and **B**) Recombination was assayed in the positive screen using the indicated proteins and the schematically diagrammed reporter constructs. The recombinase-binding elements are represented by colored arrows. Recombination events for FV7 were ascertained by SphI digestion of plasmids from blue colonies (b). Lanes 2–4 represent *IL10B2*, *IL10B3* and *IL10A3* reporter cassettes, respectively. Lane 1 shows the control digestion of plasmids from a white colony obtained in the assay with the *IL10A3* reporter. (**C**) The *in vitro* recombination reactions were carried out using the reporter plasmid carrying the *IL10A3* recombination cassette depicted in (B). The reactions in lanes 2 and 3 contained ∼8 monomers of Flp and FV7, respectively, per binding element. P and R refer to the parent plasmid and the larger of the two deletion products. The smaller deletion circle, which runs towards the bottom of the gel, is not shown. The band just above P in lane 3 is due to an intermolecular recombination event between the smaller deletion circle and P.
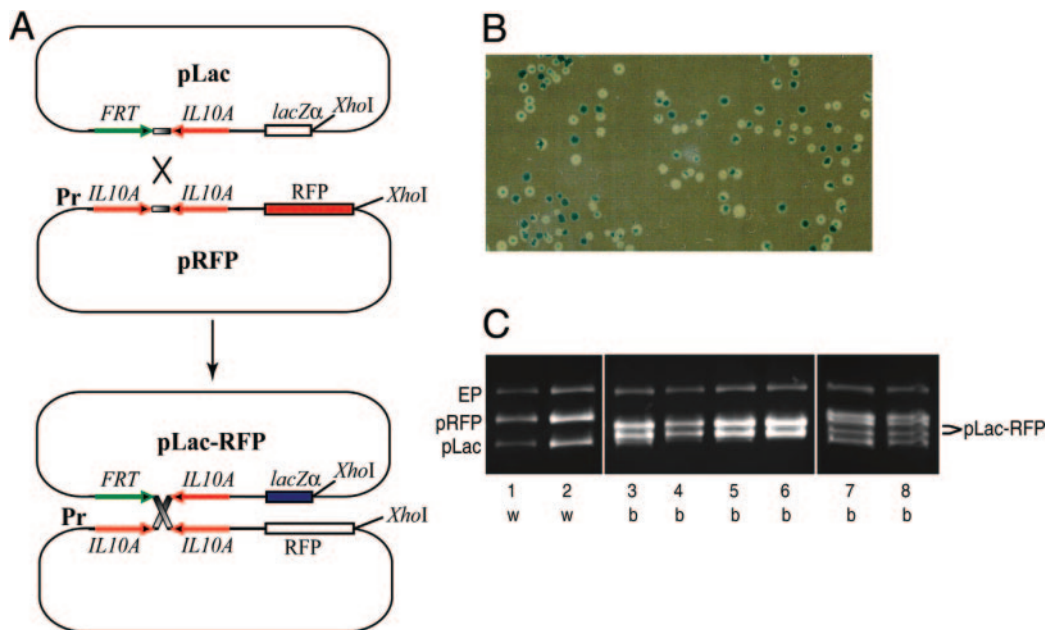
Figure 6. Site-specific integration at the *IL10A* site promoted by FV7. (**A**) The expected site-specific recombination event between pLac and pRFP plasmids is outlined. In the recombinant fusion plasmid pLac-RFP, *lacZα* would be positioned downstream of the Lac promoter (Pr). (**B**) Upon induction of FV7, expression of *lacZα* was observed in ∼45% of the colonies (blue on X-gal plates). (**C**) Digestion of plasmids from randomly picked white (w) and blue (b) colonies with XhoI confirmed the integration event.

with each other, as the former contained the ColE1 replication origin and the latter the p15a origin (24). The [h*FRT*]– [h*FL-IL10A*] hybrid site was placed upstream of the *lacZα* gene and the native *FL-IL10A* site upstream of the RFP gene. Thus, the *lacZα*-containing plasmid mimicked transfected foreign DNA, whereas the RFP-containing plasmid approximated genomic DNA. The FV7 expression plasmid, harboring the R1 replication origin, was introduced into the

experimental strain by transformation (24). Following P_{BAD} promoter-driven induction of F7V, ~45% of the colonies were blue, indicating expression of *lacZα* by acquisition of the Lac promoter via integration (Figure 6B). The authenticity of the recombination event was confirmed by XhoI digestion of plasmids isolated from randomly picked white and blue colonies (Figure 6C). Among the white colonies, only the parental reporter plasmids were detected (Figure 6C, lanes 1 and 2). Among blue colonies, the majority contained the fusion plasmid alone (Figure 6C, lanes 3–6), whereas a smaller number contained the parent plasmids as well as the fusion plasmid (Figure 6C, lanes 7–8).

DISCUSSION

In this study, we have demonstrated the evolution of two closely related Flp variants that mediate recombination at a native site derived from the human *IL10* gene. A second site within *IL10* serves as a weaker substrate. The choice of these sites as potential recombination targets is based on their organizational and sequence similarities to *FRT*, as reported by the search program TargetFinder. The evolved variants retain *FRT* specificity, which is stronger than its acquired specificity. As discussed below, this dual specificity offers an important technical advantage in targeted integration or replacement of foreign or resident sequences, respectively, in a genomic context.

The integrase from *Streptomyces* phage φC31, a serine family site-specific recombinase, has been found to be useful for genetic manipulations in mammalian cells. Normally, the enzyme mediates integrative recombination between phage and bacterial attachment sites, *attP* and *attB*, respectively. However, it can also promote recombination between a plasmid harboring *attB* and certain native sequences in mammalian genomes that resemble *attP*, referred to as pseudo *attP* sites (40–42). By directed evolution, it has been possible to generate integrase variants that preferentially target one such pseudo *attP* site present on human chromosome 8 (43). There is a significant difference between this φC31 integrase variant and the Flp variant FV7 obtained by us in their specificity attributes. What has been accomplished with the former is to obtain ~2.5-fold higher preference and recombination efficiency for a site that the wild-type integrase also recognizes as a target for recombination. In contrast, the genomic sequence FL-IL10A that FV7 utilizes as recombination substrate is not a target, or an extremely poor target for wild-type Flp; it is discriminated against by two to three orders of magnitude relative to FRT.

An alternative approach to applying a recombinase enzyme for genome engineering is to attach to it a 'tunable' specificity module in order to direct it to a particular genomic target site. Following this logic, Akopian *et al.* (44) fused the DNA recognition domain from the mouse transcription factor Zif268 to a relaxed specificity variant of Tn3 resolvase. They demonstrated that this hybrid recombinase can act at a locale that contains the cognate DNA sequence for the transcription factor. A related strategy for promoting directed gene replacement is to stimulate homologous recombination at desired genomic locales by introducing double-strand DNA breaks with the help of zinc finger nucleases (45).

In principle, the ability to construct designer recombinases either by directed evolution or by incorporation of specificity domains will significantly expand their role in biotechnological applications. Together with nucleases that serve as site-specific stimulators of homologous recombination, the novel specificity recombinases will provide the capacity of manipulating large genomes at multiple locales within a single cell.

The logic of evolving novel recombination potential through relaxation of specificity

Since genomic *FL* sites have multiple differences from *FRT* in their 'Flp-binding' elements and almost certainly lack binding site symmetry, evolution of unique specificity to such a site would be tremendously challenging. More plausible is the relaxation of Flp specificity that permits recombination of an *FL* site, albeit at a low level, without accompanying loss of native specificity. Based on this logic, the template for directed evolution is the gene for Flp^{sup3} (26), already relaxed towards *FRT* and two of its mutant derivatives (*mFRT*s), each containing a symmetric single base pair change in the binding elements. The evolutionary steps, dictated by the features of Flp–DNA interactions, involve (i) targeted randomization of specific residues, (ii) iterative shuffling of gene libraries under mutagenic conditions, and (iii) low and high sensitivity screens for recombination events. The adaptive drive is provided by chimeric sites that diverge in steps from *FRT* towards *FL*.

Although evolution was directed to *IL10B* specificity, it led to better recombination efficiency on *IL10A*. This outcome is consistent with (i) the sequence relatedness between the two *FL* sites in one of the binding elements, (ii) cooperative interactions between Flp monomers during recombination, and (iii) relaxation being the source of the new specificity. The potential to generate specificity to more than one closely related genomic *FL* sites is a bonus of the evolutionary design.

Advantages of relaxed specificity in genetic engineering

Absent the capability to design or evolve a unique specificity recombinase variant with recombination competence on a native genomic sequence, a relaxed specificity variant has finite advantages in accomplishing directed genetic rearrangements. As is evident from the present findings, the *FRT* specificity preserved in FV7 can be exploited in the form of a hybrid site or as a partner site to enhance the recombination potential at *FL* sites such as *FL-IL10A* and *FL-IL10B*. This cooperativity assisted augmentation of FV7 permits successful DNA integration by recombination at *FL* sites in a bacterial host. Indeed, targeted integration and replacement would be among the most prominent applications of a designer recombinase enzyme. The probability that relaxed specificity will give rise to degenerate recombination events and undesirable rearrangements is essentially nil. Since successful recombination requires absolute homology between the 8 bp spacers of recombination partners, the design of the spacer sequence can impose exquisite target selectivity even when the recombinase specificity encompasses more than one genomic *FL* site. Consistent with this idea, the several recombination events that we have analyzed in the *E.coli* system have not revealed aberrant reactions.

Future prospects

Using the Flp recombinase as a model system, we have demonstrated here the feasibility of obtaining recombination competence on native genomic sequences by coupling a structure-based evolutionary strategy with mutagenic DNA shuffling. In principle, a similar strategy can be utilized to generate a library of variants whose recombination specificities span a large number of genomic target sites. It is possible that not every one of these sites will work in its natural context due to potential masking effects of chromatin organization. Even so, being able to direct integration events at multiple locales within a large genome would be an important advance in bioengineering.

ACKNOWLEDGEMENTS

REFERENCES

1. Azaro,M.A. and Landy,A. (2002) λ Integrase and λ Int family. In Craig,N.L., Craigie,R., Gellert,M. and Lambowitz,A.M. (eds), *Mobile DNA II*. ASM Press, Washington, DC, pp. 118–148.
2. Barre,F.-X. and Sherratt,D.J. (2002) Xer site-specific recombination: promoting chromosome segregation. In Craig,N.L., Craigie,R., Gellert,M. and Lambowitz,A.M. (eds), *Mobile DNA II*. ASM Press, Washington, DC, pp. 1149–161.
3. Jayaram,M., Grainge,I. and Tribble,G. (2002) Site-specific recombination by the Flp protein of *Saccharomyces cerevisiae*. In Craig,N.L., Craigie,R., Gellert,M. and Lambowitz,A.M. (eds), *Mobile DNA II*. ASM Press, Washington, DC, pp. 192–218.
4. Rice,P.A. (2002) Theme and variation in tyrosine recombinases: structure of a Flp-DNA complex. In Craig,N.L., Craigie,R., Gellert,M. and Lambowitz,A.M. (eds), *Mobile DNA II*. ASM Press, Washington, DC, pp. 219–229.
5. Van Duyne,G.D. (2002) A structural view of tyrosine recombinase site-specific recombination. In Craig,N.L., Craigie,R., Gellert,M. and Lambowitz,A.M. (eds), *Mobile DNA II*. ASM Press, Washington, DC, pp. 93–117.
6. Nunes-Duby,S.E., Kwon,H.J., Tirumalai,R.S., Ellenberger,T. and Landy,A. (1998) Similarities and differences among 105 members of the Int family of site-specific recombinases. *Nucleic Acids Res.*, **26**, 391–406.
7. Guo,F., Gopaul,D.N. and van Duyne,G.D. (1997) Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse. *Nature*, **389**, 40–46.
8. Hickman,A.B., Waninger,S., Scocca,J.J. and Dyda,F. (1997) Molecular organization in site-specific recombination: the catalytic domain of bacteriophage HP1 integrase at 2.7 A resolution. *Cell*, **89**, 227–237.
9. Kwon,H.J., Tirumalai,R., Landy,A. and Ellenberger,T. (1997) Flexibility in DNA recombination: structure of the lambda integrase catalytic core. *Science*, **276**, 126–131.
10. Subramanya,H.S., Arciszewska,L.K., Baker,R.A., Bird,L.E., Sherratt,D.J. and Wigley,D.B. (1997) Crystal structure of the site-specific recombinase, XerD. *EMBO J.*, **16**, 5178–5187.
11. Chen,Y., Narendra,U., Iype,L.E., Cox,M.M. and Rice,P.A. (2000) Crystal structure of a Flp recombinase–Holliday junction complex: assembly of an active oligomer by helix swapping. *Mol. Cell*, **6**, 885–897.
12. Stewart,L., Redinbo,M.R., Qiu,X., Hol,W.G. and Champoux,J.J. (1998) A model for the mechanism of human topoisomerase I. *Science*, **279**, 1534–1541.
13. Cheng,C., Kussie,P., Pavletich,N. and Shuman,S. (1998) Conservation of structure and mechanism between eukaryotic topoisomerase I and site-specific recombinases. *Cell*, **92**, 841–850.
14. Kilby,N.J., Snaith,M.R. and Murray,J.A. (1993) Site-specific recombinases: tools for genome engineering. *Trends Genet.*, **9**, 413–421.
15. Rodriguez,C.I., Buchholz,F., Galloway,J., Sequerra,R., Kasper,J., Ayala,R., Stewart,A.F. and Dymecki,S.M. (2000) High-efficiency deleter mice show that FLPe is an alternative to Cre-loxP. *Nature Genet.*, **25**, 139–140.
16. Mills,A.A. and Bradley,A. (2001) From mouse to man: generating megabase chromosome rearrangements. *Trends Genet.*, **17**, 331–339.
17. Glaser,S., Anastassiadis,K. and Stewart,A.F. (2005) Current issues in mouse genome engineering. *Nature Genet.*, **37**, 1187–1193.
18. Christ,N. and Droge,P. (2002) Genetic manipulation of mouse embryonic stem cells by mutant lambda integrase. *Genesis*, **32**, 203–208.
19. Corona,T., Bao,Q., Christ,N., Schwartz,T., Li,J. and Droge,P. (2003) Activation of site-specific DNA integration in human cells by a single chain integration host factor. *Nucleic Acids Res.*, **31**, 5140–5148.
20. Tan,S.M. and Droge,P. (2005) Comparative analysis of sequence-specific DNA recombination systems in human embryonic stem cells. *Stem Cells*, **23**, 868–873.
21. Buchholz,F. and Stewart,A.F. (2001) Alteration of Cre recombinase site specificity by substrate-linked protein evolution. *Nat. Biotechnol.*, **19**, 1047–1052.
22. Santoro,S.W. and Schultz,P.G. (2002) Directed evolution of the site specificity of Cre recombinase. *Proc. Natl Acad. Sci. USA*, **99**, 4185–4190.
23. Rufer,A.W. and Sauer,B. (2002) Non-contact positions impose site selectivity on Cre recombinase. *Nucleic Acids Res.*, **30**, 2764–2771.
24. Voziyanov,Y., Stewart,A.F. and 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.
25. Voziyanov,Y., Konieczka,J.H., Stewart,A.F. and Jayaram,M. (2003) Stepwise manipulation of DNA specificity in Flp recombinase: progressively adapting Flp to individual and combinatorial mutations in its target site. *J. Mol. Biol.*, **326**, 65–76.
26. Konieczka,J.H., Paek,A., Jayaram,M. and Voziyanov,Y. (2004) Recombination of hybrid target sites by binary combinations of Flp variants: mutations that foster inter-protomer collaboration and enlarge substrate tolerance. *J. Mol. Biol.*, **339**, 365–378.
27. Baldwin,E.P., Martin,S.S., Abel,J., Gelato,K.A., Kim,H., Schultz,P.G. and Santoro,S.W. (2003) A specificity switch in selected Cre recombinase variants is mediated by macromolecular plasticity and water. *Chem. Biol.*, **10**, 1085–1094.
28. Guzman,L.M., Belin,D., Carson,M.J. and Beckwith,J. (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.*, **177**, 4121–4130.
29. Buchholz,F., Angrand,P.O. and Stewart,A.F. (1998) Improved properties of FLP recombinase evolved by cycling mutagenesis. *Nat. Biotechnol.*, **16**, 657–662.
30. Lorimer,I.A. and Pastan,I. (1995) Random recombination of antibody single chain Fv sequences after fragmentation with DNaseI in the presence of Mn2+. *Nucleic Acids Res.*, **23**, 3067–3068.
31. Guex,N., Diemand,A. and Peitsch,M.C. (1999) Protein modelling for all. *Trends Biochem. Sci.*, **24**, 364–367.
32. Sambrook,J.F. and Russell,D.W. (2001) *Molecular Cloning: A Laboratory Manual, 3rd edn.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
33. Bruckner,R.C. and Cox,M.M. (1986) Specific contacts between the FLP protein of the yeast 2-micron plasmid and its recombination site. *J. Biol. Chem.*, **261**, 11798–11807.
34. Panigrahi,G.B. and Sadowski,P.D. (1994) Interaction of the NH2- and COOH-terminal domains of the FLP recombinase with the FLP recognition target sequence. *J. Biol. Chem.*, **269**, 10940–10945.
35. Umlauf,S.W. and Cox,M.M. (1988) The functional significance of DNA sequence structure in a site-specific genetic recombination reaction. *EMBO J.*, **6**, 1845–1852.

36. Senecoff,J.F.P., Rossmeissl,P.J. and Cox,M.M. (1988) DNA recognition by FLP recombinase of the yeast 2 micron plasmid: a mutational analysis of the FLP binding site. *J. Mol. Biol.*, **201**, 405–421.

37. Gronostajski,R.M. and Sadowski,P.D. (1985) Determination of DNA sequences essential for FLP-mediated recombination by a novel method. *J. Biol. Chem.*, **260**, 12320–12327.

38. Jayaram,M. (1985) Two-micrometer circle site-specific recombination: the minimal substrate and the possible role of flanking sequences. *Proc. Natl Acad. Sci. USA*, **82**, 5875–5879.

39. Senecoff,J.F., Bruckner,R.C. and Cox,M.M. (1985) The FLP recombinase of the yeast 2-micron plasmid: characterization of its recombination site. *Proc. Natl Acad. Sci. USA*, **82**, 7270–7274.

40. Groth,A.C., Olivares,E.C., Thyagarajan,B. and Calos,M.P. (2000) A phage integrase directs efficient site-specific integration in human cells. *Proc. Natl Acad. Sci. USA*, **97**, 5995–6000.

41. Thyagarajan,B., Olivares,E.C., Hollis,R.P., Ginsburg,D.S. and Calos,M.P. (2001) Site-specific genomic integration in mammalian cells mediated by phage C31 integrase. *Mol. Cell. Biol.*, **21**, 3926–3934.

42. 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 phiC31 integrase in the human genome. *J. Mol. Biol.*, **357**, 28–48.

43. Sclimenti,C.R., Thyagarajan,B. and Calos,M.P. (2001) Directed evolution of a recombinase for improved genomic integration at a native human sequence. *Nucleic Acids Res.*, **29**, 5044–5051.

44. Akopian,A., He,J., Boocock,M.R. and Stark,W.M. (2003) Chimeric recombinases with designed DNA sequence recognition. *Proc. Natl Acad. Sci. USA*, **100**, 8688–8691.

45. Porteus,M.H. and Carroll,D. (2005) Gene targeting using zinc finger nucleases. *Nat. Biotechnol.*, **8**, 967–973.